Faculté de Philosophie, Arts et Lettres
Institut Langage et Communication

# Agent-defocussing structures in Spanish, English and French

## A contrastive corpus-based account of the interaction between different passive-like structures

Émeline PIERRE

Thèse soutenue en vue de l'obtention du grade

de Docteur en Langues et Lettres

**Membres du jury**

Prof. Barbara De Cock (Université catholique de Louvain), promotrice
Prof. Anne Catherine Simon (Université catholique de Louvain), présidente
Prof. Marie-Aude Lefer (Université catholique de Louvain)
Prof. Bettina Kluge (Stiftung Universität Hildesheim)
Prof. Elena Martínez Caro (Universidad Complutense de Madrid),
Prof. Victoria Vázquez Rozas (Universidade de Santiago de Compostela)

Louvain-la-Neuve

17 Septembre 2021

PIERRE, Émeline. 2021. *Agent-defocussing structures in Spanish, English and French: A contrastive corpus-based account of the interaction between different passive-like structures*. Dissertation doctorale, Université catholique de Louvain.

# Table of contents

List of Tables ..... vi
List of Figures ..... vii
List of Graphs ..... viii
List of abbreviations ..... xii
Acknowledgments ..... xiv
1. Introduction ..... 1
1.1. Delimitation of the object of research ..... 3
1.2. Justification of the study, objectives and research questions ..... 6
1.3. Preview of the thesis ..... 8
2. Core features of passive-like structures ..... 11
2.1. Periphrastic passives and *get*-passives ..... 11
2.1.1. Brief account of voices ..... 11
2.1.2. Canonical passives ..... 13
2.1.3. *Get*-passives ..... 20
2.2. *Se*-constructions ..... 24
2.2.1. Spanish *se*-constructions ..... 24
2.2.2. French *se*-constructions ..... 30
2.3. R-impersonals ..... 31
2.3.1. Non-anaphoric third person plurals ..... 31
2.3.2. Man-impersonal *on* ..... 34
2.4. Numeral-based constructions: *uno* and *one* ..... 39
2.4.1. Construction with *uno* ..... 39
2.4.2. Construction with *one* ..... 41
2.4.3. Comparison with the man-impersonal *on* ..... 42
3. Comparing and contrasting corpus data ..... 45
3.1. Contrastive linguistics and its multiple facets ..... 45
3.2. Crosslinguistic studies and impersonal constructions ..... 47
3.2.1. English and other languages ..... 48
3.2.2. Spanish and other languages ..... 50
3.2.3. French and other languages ..... 54
4. Defining the theoretical framework ..... 57
4.1. Distance between speech and writing ..... 57

4.1.1. Clarifying key terms ..... 57
4.1.2. Issue of primacy ..... 58
4.1.3. Spoken and written, divide or continuum? ..... 59
4.1.4. Distinctions between spoken and written productions ..... 60
4.2. Register, a variational construct ..... 62
4.2.1. Terminology ..... 62
4.2.2. Typical features of *register* ..... 63
4.2.3. Contrasting different terminological approaches ..... 65
4.3. Transitivity, a multifactorial approach ..... 68
4.3.1. Exploring transitivity and Hopper & Thompson's approach ..... 68
4.3.2. Criticism of Hopper & Thomspon's work (1980) and other contributions ..... 71
5. Data and methodology ..... 73
5.1. Data description ..... 73
5.1.1. Written data ..... 74
5.1.2. Oral data ..... 76
5.1.3. Features of the final datasets and extraction method ..... 79
5.1.4. Overview of the final datasets ..... 80
5.1.5. Extraction process ..... 82
5.2. Annotation protocol ..... 83
5.2.1. Verb-related parameters ..... 84
5.2.2. Participant-related parameters ..... 94
5.2.3. Context-related parameters ..... 99
5.3. Interrater agreement ..... 108
5.4. Statistical tests ..... 109
5.4.1. Multiple Correspondence Analyses ..... 109
5.4.2. Chi-square tests ..... 110
6. General overview of the data ..... 113
6.1. Overall distribution ..... 113
6.2. Type/token ratio ..... 119
6.3. Exploring statistical results ..... 123
7. Passive-like structures in French ..... 127
7.1. Panoramic overview ..... 128
7.2. Verb-related parameters ..... 132
7.2.1. Tense ..... 132
7.2.2. Aspect ..... 135

7.2.3. Verb mode ..... 138
7.2.4. Type of verb process ..... 141
7.2.5. Kinesis ..... 148
7.2.6. Realis/irrealis dichotomy ..... 150
7.2.7. Affirmation ..... 152
7.2.8. Other parameters ..... 154
7.2.9. Summary ..... 154
7.3. Participant-related parameters ..... 154
7.3.1. Number of participants ..... 155
7.3.2. Agent ..... 158
7.3.3. Patient ..... 163
7.3.4. Individuation and affectedness of the patient ..... 169
7.3.5. Summary ..... 172
7.4. Context-related parameters ..... 173
7.4.1. Presence of a spatiotemporal restriction ..... 173
7.4.2. Presence of a prepositional phrase ..... 176
7.4.3. Other parameters ..... 179
7.4.4. Summary ..... 180
7.5. Portraying the French system ..... 181
8. Passive-like structures in Spanish ..... 187
8.1. Panoramic overview ..... 188
8.2. Verb-related parameters ..... 191
8.2.1. Tense ..... 191
8.2.2. Aspect ..... 194
8.2.3. Verb mode ..... 198
8.2.4. Type of verb process ..... 200
8.2.5. Kinesis ..... 208
8.2.6. Realis/irrealis dichotomy ..... 210
8.2.7. Affirmation ..... 211
8.2.8. Other parameters ..... 212
8.2.9. Summary ..... 213
8.3. Participant-related parameters ..... 213
8.3.1. Number of participants ..... 213
8.3.2. Agent ..... 216
8.3.3. Patient ..... 219

8.3.4. Individuation and affectedness of the patient ..... 223
8.3.5. Summary ..... 226
8.4. Context-related parameters ..... 226
8.4.1. Presence of a spatiotemporal restriction ..... 227
8.4.2. Presence of a prepositional phrase ..... 228
8.4.3. Presence of another passive-like structure ..... 233
8.4.4. Other parameters ..... 235
8.4.5. Summary ..... 236
8.5. Numeral-based *uno* ..... 237
8.6. Portraying the Spanish system ..... 238
9. Passive-like structures in English ..... 241
9.1. Panoramic overview ..... 242
9.2. Verb-related parameters ..... 245
9.2.1. Tense ..... 246
9.2.2. Aspect ..... 248
9.2.3. Verb mode ..... 250
9.2.4. Type of verb process ..... 252
9.2.5. Kinesis ..... 257
9.2.6. Realis/irrealis dichotomy ..... 259
9.2.7. Affirmation ..... 261
9.2.8. Other parameters ..... 261
9.2.9. Summary ..... 262
9.3. Participant-related parameters ..... 263
9.3.1. Number of participants ..... 263
9.3.2. Agent ..... 265
9.3.3. Patient ..... 267
9.3.4. Individuation of the patient ..... 269
9.3.5. Affectedness of the patient ..... 271
9.3.6. Summary ..... 272
9.4. Context-related parameters ..... 272
9.4.1. Presence of a spatiotemporal restriction ..... 272
9.4.2. Presence of a prepositional phrase ..... 274
9.4.3. Other parameters ..... 276
9.5. *Get*-passive, an alternative to periphrastic passive? ..... 277
9.5.1. Verb-related parameters ..... 278

9.5.2. Participant-related parameters ..... 280
9.5.3. Context-related parameters ..... 280
9.6. Numeral-based *one* ..... 282
9.7. Portraying the English system ..... 283
10. A contrastive view of the data ..... 289
10.1. Contrasting passive-like structures in French, Spanish and English ..... 289
10.1.1. The periphrastic passive ..... 289
10.1.2. The non-anaphoric third person plural ..... 293
10.1.3. Se-constructions in French and in Spanish ..... 296
10.1.4. Se-constructions and periphrastic passives, complementary or overlapping structures? 297
10.1.5. Man-impersonal *on*, standing alone? ..... 300
10.2. Mode and register, key factors in contrastive studies ..... 304
10.2.1. Structures and variation ..... 304
10.2.2. Parameters and variation ..... 309
11. Conclusion ..... 315
References ..... 329
Appendices ..... 347
Appendix 1: FSGraph created in Unitex for data extraction ..... 347
Appendix 2: Code used for the language, the passive-like structure and the mode/register ..... 353
Appendix 3: Example of script used to obtain kappa score in Rstudio ..... 354
Appendix 4: Interrater agreement results ..... 357
Appendix 5: MCA graphs excluding the man-impersonal *on* ..... 358
Appendix 6: Chi-square tests used in Chapter 10 ..... 359

# List of Tables

Table 1. Overview of the passive-like structures included in the study (EN= English, FR= French, and SP = Spanish) ..... 2
Table 2. Classification of the pronoun on from Fløttum et al. (2007) ..... 35
Table 3. Types of corpus-based cross-linguistic comparisons (Granger 2010) ..... 46
Table 4. Features of formal and informal registers (Briz 2010) ..... 67
Table 5. Overview of the different subcomponents in the PROCEP corpus ..... 78
Table 6. Overview of the features of the corpora ..... 81
Table 7. Classification of the parameter verb mode ..... 93
Table 8. Types of processes (1) ..... 94
Table 9. Types of processes (2) ..... 94
Table 10. Coding system for the parameter 'person reference' ..... 100
Table 11. Coding system in the three languages for prepositional phrases ..... 106
Table 12. Examples of kappa scores ..... 108
Table 13. General distribution of the data by structure and language ..... 115
Table 14. Overall frequencies for se-constructions in French and Spanish ..... 117
Table 15. Overall frequencies for the man-impersonal construction ..... 117
Table 16. Overall frequencies for the periphrastic passive in Spanish, English and French ..... 118
Table 17. Overall frequencies for the non-anaphoric third person plural in Spanish, French and English ..... 118
Table 18. Type token ratios per language ..... 120
Table 19. Top three most frequent verbs for the non-anaphoric third person plural, the periphrastic passive, the se-construction and the man-impersonal on ..... 122
Table 20. Raw and relative frequencies per ten thousand words (pttw) in French data ..... 127
Table 21. Classification of overt agents in French periphrastic passives ..... 159
Table 22. Distribution of overt agentsin periphrastic passives by mode and register ..... 162
Table 23. Classification of patients in French passive-like structures ..... 165
Table 24. Overview of the distinctive features of French passive-like structures ..... 182
Table 25. Raw and relative frequencies per ten thousand words (pttw) in Spanish data ..... 187
Table 26. Classification of animate and inanimate agents in Spanish periphrastic passives ..... 217
Table 27. Classification of animate and inanimate patients in Spanish passive-like structures ..... 220
Table 28. Overview of the distinctive features of Spanish passive-like structures ..... 238
Table 29. Raw and relative frequencies per ten thousand words (pttw) in English data ..... 241
Table 30. Classification of animate and inanimate agents in English ..... 265
Table 31. Classification of animate and inanimate patients in English passive-like structures ..... 268
Table 32. Overview of the distinctive features of English passive-like structures ..... 284
Table 33. Absolute and relative frequencies of the periphrastic passive in English, French and Spanish ..... 289
Table 34. Absolute and relative frequencies of the non-anaphoric third person plural in English, French and Spanish ..... 293
Table 35. Absolute and relative frequencies of se-construction in French and Spanish ..... 296
Table 36. Absolute and relative frequencies for the non-anaphoric third person plural, the man-impersonal on, the periphrastic passive and the se-construction ..... 300

Table 37. Distribution of non-anaphoric third person plurals and periphrastic passives in English ..... 305
Table 38. Code used for each language ..... 353
Table 39. Code used for each passive-like structure ..... 353
Table 40. Code used for each register/mode ..... 353
Table 41. Results interrater agreement (kappa score and % of agreement) ..... 357

# List of Figures

Figure 1. Transitivity scale established by Hopper & Thompson (1980) ..... 69
Figure 2. Hopper & Thompson's transitivity scale modified by Malchukov (2006) ..... 71
Figure 3. Example of a code allowing the identification of the text files containing the words vegetariano and vegetarianismo in the Spanish Wikipedia Corpus ..... 75
Figure 4. Example of an csv. file allowing the identification of food habits in the Spanish Wikipedia Corpus ..... 75
Figure 5. Number of words per language through the different registers and modes ..... 80
Figure 6. Example of graph used to extract the periphrastic passive in English ..... 82
Figure 7. Classification dynamic vs stative verbs (Quirk et al. 1985) ..... 87
Figure 8. Classification of participants with their corresponding code in parentheses ..... 97
Figure 9. Classification scheme of the nineteen parameters under study ..... 107
Figure 10. Distribution of languages, excluding the man-impersonal *on* ..... 123
Figure 11. Distribution of languages, including the man-impersonal *on* ..... 123
Figure 12. Distribution of the nineteen parameters ..... 124
Figure 13. Factor map for mode and register ..... 126
Figure 14. Distribution of parameters in French passive-like structures ..... 128
Figure 15. Factor map for mode and register in French passive-like structures ..... 129
Figure 16. Distribution of the four passive-like structures in French ..... 130
Figure 17. Correlation(s) in French data ..... 131
Figure 18. Distribution of parameters in Spanish data ..... 188
Figure 19. Distribution of the four passive-like structures in Spanish ..... 189
Figure 20. Factor map for mode and register in Spanish ..... 190
Figure 21. Correlation(s) in Spanish data ..... 190
Figure 22. Distribution of parameters in English data ..... 242
Figure 23. Factor map for mode and register in Spanish ..... 243
Figure 24. Distribution of the four passive-like structures in English ..... 244
Figure 25. Correlation(s) in English data ..... 245
Figure 26. Summary of the specific characteristics of the periphrastic passives in English, French and Spanish ..... 290
Figure 27. Summary of the specific characteristics of the non-anaphoric third person plurals in English, French and Spanish ..... 294
Figure 28.Summary of the specific characteristics of the se-construction in French and Spanish ..... 296
Figure 29. Summary of the specific characteristics of the se-construction and the periphrastic passive in Spanish ..... 298

Figure 30. Summary of the shared characteristics between on versus the periphrastic passive, the non-anaphoric third person plural and the se-construction ..... 301
Figure 31. Influence of mode and register on three sets of passive-like structures ..... 309
Figure 32. Parameters most prone to variation according to mode and register ..... 313
Figure 33. Summary of the distinctive features for French passive-like constructions (in blue verb-related parameters, in violet participant-related parameters, and in burgundy context-related parameters). ..... 317
Figure 34. Summary of the distinctive features for Spanish passive-like constructions (in blue verb-related parameters, in violet participant-related parameters, and in burgundy context-related parameters) ..... 319
Figure 35. Summary of the distinctive features for English passive-like constructions (in blue verb-related parameters and in violet participant-related parameters) ..... 321
Figure 36. Graph used to extract the non-anaphoric third person plural in French ..... 347
Figure 37. Graph used to extract the periphrastic passive in French ..... 347
Figure 38. Graph used to extract the man-impersonal *on* ..... 348
Figure 39. Graph used to extract the se-construction in French ..... 348
Figure 40. Graph used to extract the get-passive ..... 349
Figure 41. Graph used to extract the periphrastic passive in English ..... 349
Figure 42. Graph used to extract the non-anaphoric third person plural in English ..... 350
Figure 43. Graph used to extract the numeral-based *one* ..... 350
Figure 44. Graph used to extract the periphrastic passive in Spanish ..... 351
Figure 45. Graph used to extract the se-construction in Spanish ..... 351
Figure 46. Graph used to extract the non-anaphoric third person plural in Spanish ..... 352
Figure 47. Graph used to extract the numeral-based *uno* ..... 352
Figure 48. Distribution of parameters in French without *on* ..... 358
Figure 49. Factor map for mode and register in French without on ..... 358

## List of Graphs

Graph 1. Overview of the final dataset composition ..... 73
Graph 2. Type/token ratio per language and structure ..... 120
Graph 3. Type token ratio per register and mode ..... 121
Graph 4. Distribution of tenses in French passive-like structures ..... 132
Graph 5. Distribution of present, past and future tenses in French periphrastic passives ..... 133
Graph 6. Distribution of tenses in the French man-impersonal on ..... 134
Graph 7. Distribution of perfective and imperfective forms in French passive-like structures (parameter aspect) ..... 135
Graph 8. Distribution of imperfective and perfective forms in French periphrastic passives 136
Graph 9. Distribution of imperfective and perfective forms in French man-impersonals *on* 137
Graph 10. Distribution of verb modes in French passive-like structures ..... 138
Graph 11. Distribution of verb modes in French periphrastic passives ..... 139
Graph 12. Distribution of verb modes in French man-impersonals *on* ..... 140
Graph 13. Type of verb process in French passive-like structures ..... 141
Graph 14. Distribution of verb processes in French man-impersonals ..... 146
Graph 15. Distribution of verb processes in French periphrastic passives ..... 147

Graph 16. Distribution of dynamic and stative verbs in French passive-like structures (parameter kinesis) ..... 148
Graph 17. Distribution of dynamic and stative verbs in French periphrastic passives ..... 149
Graph 18. Distribution of dynamic and stative verbs in French man-impersonals *on* ..... 149
Graph 19. Distribution of realis and irrealis clauses in French passive-like structures ..... 150
Graph 20. Distribution of realis and irrealis clauses in French periphrastic passives ..... 151
Graph 21. Distribution of affirmative and negative clauses in French passive-like structures ..... 152
Graph 22. Affirmation in French periphrastic passives ..... 153
Graph 23. Affirmation in French man-impersonals *on* ..... 153
Graph 24. Distribution of one- and two-participant clauses in French passive-like structures ..... 155
Graph 25. Distribution one- and two-participant clauses in French man-impersonals ..... 157
Graph 26. Presence of a patient in French passive-like structures ..... 163
Graph 27. Distribution of animate and inanimate patients in French passive-like structures 164
Graph 28. Distribution of patients in the French periphrastic passive ..... 167
Graph 29. Distribution of patients in French man-impersonals *on* ..... 168
Graph 30. Distribution of individuated and non-individuated patients in French passive-like structures ..... 170
Graph 31. Distribution of individuated and non-individuated patients in French periphrastic passives ..... 171
Graph 32. Distribution of individuated and non-individuated patients in French man-impersonals on ..... 171
Graph 33. Presence of a spatiotemporal restriction in French passive-like structures ..... 174
Graph 34. Distribution of spatiotemporal restrictions in French periphrastic passives ..... 175
Graph 35. Distribution of spaciotemporal restrictions in French man-impersonals ..... 175
Graph 36. Presence of prepositional phrases in French passive-like structures ..... 176
Graph 37. Distribution of prepositional phrases in French passive-like structures ..... 177
Graph 38. Distribution of prepositional phrases in French periphrastic passives ..... 178
Graph 39.Distribution of prepositional phrases in French man-impersonals ..... 179
Graph 40. Distribution of present, past and future tenses in Spanish passive-like structures 191
Graph 41. Distribution of tenses in Spanish periphrastic passives ..... 193
Graph 42. Distribution of tenses in Spanish se-constructions ..... 194
Graph 43. Distribution of imperfective and perfective forms in Spanish passive-like structures (parameter aspect) ..... 195
Graph 44. Distribution of imperfective and perfective forms in the Spanish se-constructions ..... 196
Graph 46. Distribution of imperfective and perfective forms in Spanish non-anaphoric third person plurals ..... 197
Graph 45. Distribution of imperfective and perfective forms in Spanish periphrastic passives ..... 197
Graph 47. Distribution of verb modes in Spanish passive-like structures ..... 198
Graph 48. Distribution of verb modes in Spanish periphrastic passives ..... 199
Graph 49. Distribution of verb modes in Spanish se-constructions ..... 200
Graph 50. Distribution of verb processes in Spanish passive-like structures ..... 201
Graph 51. Distribution of verb processes in Spanish periphrastic passives ..... 205
Graph 52. Distribution of verb processes in Spanish se-constructions ..... 206

Graph 53. Distribution of verb processes in Spanish non-anaphoric third person plurals ..... 207
Graph 54. Distribution of dynamic and stative verbs in Spanish passive-like structures ..... 208
Graph 55. Distribution of dynamic and stative verbs Spanish se-constructions ..... 208
Graph 56. Distribution of dynamic and stative verbs in Spanish non-anaphoric third person plurals ..... 209
Graph 57. Distribution of realis and irrealis clauses in Spanish periphrastic passives ..... 210
Graph 58. Distribution of realis and irrealis clauses in Spanish se-constructions ..... 210
Graph 59. Distribution of affirmative and negative clauses in Spanish passive-like structures ..... 211
Graph 60. Distribution of one- and two-participant clauses in Spanish passive-like structures ..... 214
Graph 61. Distribution of one- and two-participant clauses in Spanish se-constructions..... 215
Graph 62. Distribution of one- and two-participant clauses in Spanish non-anaphoric third person plurals ..... 216
Graph 63. Animate versus inanimate patients in Spanish passive-like structures ..... 220
Graph 64. Distribution of patients in Spanish periphrastic passives ..... 222
Graph 65. Distribution of patients in Spanish non-anaphoric third person plurals ..... 222
Graph 66. Distribution of individuated and non-individuated patients in Spanish passive-like structures ..... 223
Graph 67. Individuated and non individuated patients in Spanish periphrastic passives ..... 224
Graph 68. Individuated and non individuated patients in Spanish se-constructions ..... 225
Graph 69. Individuated and non individuated patients in Spanish non-anaphoric third person plurals ..... 225
Graph 70. Use of spatiotemporal restrictions in Spanish ..... 227
Graph 71. Presence of a spatiotemporal restriction in Spanish non-anaphoric third person plurals ..... 228
Graph 72. Use of prepositional phrases in Spanish passive-like structures ..... 229
Graph 73. Distribution prepositional phrases in Spanish passive-like structures ..... 230
Graph 74. Distribution of prepositional phrases in Spanish periphrastic passives ..... 232
Graph 75. Distribution of prepositional phrases in Spanish se-constructions..... 232
Graph 76. Presence of another passive-like structure with Spanish passive-like structures .. 233
Graph 77. Presence of another passive-like structure with Spanish periphrastic passives .... 234
Graph 78. Presence of another passive-like structure with Spanish se-constructions..... 235
Graph 79. Distribution of present, past and future tenses in English passive-like structures 246
Graph 80. Distribution of tenses in English periprhastic passives ..... 247
Graph 81. Distribution of tenses in English non-anaphoric third person plurals ..... 247
Graph 82. Distribution of imperfective and perfective forms in English passive-like structures (parameter aspect) ..... 248
Graph 83. Distribution of imperfective and perfective forms in English periphrastic passives ..... 249
Graph 84. Distribution of verb modes in English passive-like structures..... 250
Graph 85. Distribution of verb modes in English periphrastic passives ..... 251
Graph 86. Distribution of verb processes in English passive-like structures..... 252
Graph 87. Distribution of verb processes in English periphrastic passives ..... 255
Graph 88. Distribution of verb processes in English non-anaphoric third person plurals ..... 256
Graph 89. Distribution of dynamic and stative verbs in English passive-like structures (parameter kinesis) ..... 257

Graph 90. Distribution of dynamic and stative verbs in English periphrastic passives ..... 258
Graph 91. Distribution of dynamic and stative verbs in English non-anaphoric third person plurals ..... 259
Graph 92. Distribution of (ir)realis clauses in English passive-like structures ..... 260
Graph 93. Distribution of realis and irrealis clauses in English periphrastic passives ..... 260
Graph 94. Distribution of affirmative and negative clauses in English passive-like structures ..... 261
Graph 95. Distribution of one- and two-participant clauses in English passive-like structures ..... 263
Graph 96. Distribution of one- and two-participant clauses in English non-anaphoric third person plurals ..... 264
Graph 97. Distribution of animate versus inanimate patients in English passive-like structures ..... 267
Graph 98. Distribution of patients in English periphrastic passives ..... 268
Graph 99. Distribution of (non-)individuated patient in English passive-like structures ..... 270
Graph 100. Distribution of (non-)individuated patient in English periphrastic passives ..... 270
Graph 101. Distribution of spatiotemporal restrictions in English periphrastic passives ..... 273
Graph 102. Presence of a prepositional phrase in English passive-like structures ..... 274
Graph 103. Distribution of prepositional phrases in English passive-like structures ..... 274
Graph 104. Distribution prepositional phrases in English periphrastic passives ..... 276
Graph 105. Distribution of types of verb processes in the get-passive ..... 279

# List of abbreviations

CL: contrastive linguistics

EN: English

FML: formal

FR: French

INF: informal

MCA: multiple correspondence Analysis

ON: man-impersonal *on*

Pttw: per ten thousand words

PP: periphrastic passive

SE: se-construction

SP: Spanish

TC: tertium comparationis

3PP: non-anaphoric third person plural

# Acknowledgments

First and foremost, I am immensely grateful to my supervisor, Dr. Barbara De Cock for the opportunity she gave me five years ago when she offered me to start a doctoral journey by her side. I would like to thank you for your invaluable supervision, constant support, help and understanding through the whole process. It was a genuine pleasure and privilege to work with you.

I am also deeply grateful to the members of my supervisory panel Dr. Marie-Aude Lefer and Dr. Anne Catherine Simon for their advice and assistance at several stages of my research project.

I would like to express my sincere gratitude to the members of my jury: Dr. Victoria Vázquez Rozas, Dr. Elena Martínez Caro and Dr. Bettina Kluge. I am extremely grateful to Dr. Victoria Vázquez Rozas and to Dr. Elena Martínez Caro who welcomed me in their respective university during my research stays in the Universidade de Santiago de Compostela in 2018 and in the Universidad Complutense de Madrid in 2020.

I am also particularly indebted to my funding organisms: the Fonds spéciaux de recherche (FSR) and the Fund for Scientific Research (F.R.S. – FNRS) which gave me the opportunity to work in the best conditions possible.

Heartfelt thanks to my colleagues. I am very much thankful to the two persons whom I shared the office with. Bénédicte, you introduced me to the world of PhD students. It was a pleasure sharing the c.466 with you ("On ferme la porte?"). Aurélie, I mainly shared the office 'virtually' with you but I have no words to describe how important and supportive you were during this whole PhD journey. I would also like to thank all the linguists from the third and fourth floors in the faculty, who, for some, are now much more than colleagues. I had the best working environment I could ever have wished for. My special thanks also go to Thomas François for his help to realise the interrater agreement and to Alain Guillet and the Support en "Méthodologie et Calcul statistique" from UCLouvain for their advice on the statistical part of this doctoral dissertation. I am also greatly thankful to all the proofreaders who reviewed parts of this manuscript and helped me to track down the errors (Alysson, Aurélie, Floriane, Mélanie, Pauline and Miguel).

A PhD journey is not only a professional adventure. I am extremely grateful to my family and friends. To my parents for their daily unflagging and loving support as well as their precious help, especially during the last weeks of my thesis. To my brother and his wife, for cheering me up with kind messages (and for IT support). And my last (but not least!) round of acknowledgments goes to all the people I had the opportunity to become friends with during my academic path. Through them I could escape from the sphere of work and share more relaxed moments throughout these years.

# 1. Introduction

One could wonder what the following sentences, extracted from various personal conversations with colleagues, friends or family members, have in common.

- *Queridas Aurélie et Emeline: Tengo buena noticia: tendremos un miembro más en otoño ;-)* ***Se aprobó*** *la solicitud de FSR con [Nombre]*[1].
- *Tous leurs congés* ***ont été annulés***[2].
- *If they don't pay, they will* ***get arrested****.*
- *Hotels will open on 19.5 so maybe* ***they will open*** *borders too, without quarantine.*
- *Tu sais comment* ***on*** *fait ce genre de graphiques?*[3]
- *Pues sí, pero* ***uno*** *no valora la importancia de la salud mental*[4].

At first sight, they may appear as bearing little resemblance to one another. However, they share an important feature, namely that their agent is either absent or vague. They are then some illustrations of the phenomena which lie at the heart of the present thesis, namely agent-defocussing strategies. These mechanisms diminish the salience of the agent and confer language production with varying degrees of impersonality. Language has developed a wide range of syntactic devices to defocus the agent or, to put it differently, to provide the discourse with an impersonal, "faceless character" (Achard 2015: ix). Impersonal constructions allow to portray the agent in a particular way by reducing its prominence (Langacker 2011).

The present thesis explores six of these constructions in English, French and Spanish: the periphrastic passive (available in the three languages), the English *get*-passive, the non-anaphoric third person plural (also used in the three languages), the *se*-construction (available in French and in Spanish), the French man-impersonal *on* and the numeral-based *one* in English and its Spanish counterpart *uno.* At first sight, the structures under study in the investigation are of two different natures: (i) similar with regard to the surface structure. This is the case for periphrastic passives (available in the three languages), the non-anaphoric third person plural, even though the Spanish non-anaphoric third person plural does not have an explicit pronoun since Spanish is a pro-drop language, the numeral-based pronouns in English and Spanish and the *se*-construction in French and Spanish; (ii) or standing alone structures, i.e., lacking a formal equivalent in the other languages as is the case for the *get*-passive in English and the *man*-impersonal *on* in French (see Table 1 for a summary). Certain structures that also function as agent-defocussing mechanisms have not been included in this work. The rationale behind the selection will be elaborated upon later on, notably in Chapter 2.

[1] 'Dear Aurélie and Emeline: I have good news: we will have one more member in autumn ;-) The FSR application of [Name] has been approved.'
[2] 'All their holidays have been cancelled.'
[3] 'Do you know how to make these kinds of graphs?'
[4] 'Yes, but one does not value the importance of mental health.'

| Passive-like structures | Languages | | |
|---|---|---|---|
| | EN | FR | SP |
| **Periphrastic passive** | + | + | + |
| ***Get*-passive** | + | - | - |
| **Passive construction based on a reflexive pronoun** | - | + | + |
| **Non-anaphoric third person plural** | + | + | + |
| **Man-impersonal** | - | + | - |
| **Numeral-based indefinite pronoun** | + | - | + |

*Table 1. Overview of the passive-like structures included in the study (EN= English, FR= French, and SP = Spanish)*

The general concept of impersonalisation has, however, proven to be fairly difficult to delineate since it can acquire a different meaning according to the linguistic tradition, the language or the focus adopted by the researcher. Moreover, the term can be used to cover a large series of grammatical phenomena, which are, in some cases, not directly related (see Creissels 2008 or Malchukov & Ogawa 2011). The consequent heterogeneity of this notion has led some linguists to consider impersonality as a near-vacuous term, which requires the specification of the particular meaning assigned to it in each given context (Fernández 2013).

In the literature, impersonality usually receives two readings (Siewierska 2008): a structural one, associated with the complete absence of a canonical subject, and a functional one, linked to agent-defocussing. These two readings correspond respectively to a syntactic approach of impersonality (lack of grammatical subject), also labelled subject-based approach, and a semantic one (lack of human agentivity or reduced referentiality of the human agent). Within the syntactic approach, Siewierska (2008) defines the grammatical or canonical subject as "one realized by a verbal argument which is fully referential and manifests the morpho-syntactic properties of subjects in a language" (2008:2). Considering this definition, impersonality viewed from the syntactic perspective can include constructions with four options regarding the subject (Siewierska 2008: 2-6): (i) a not fully referential subject, such as the non-anaphoric third plural or the man-impersonal; (ii) a subject deprived of the canonical subject properties, for instance constructions based on a reflexive marker in Romance languages; (iii) a subject acting as place filler with no semantic or referential power, such as expletive subjects in some meteorological constructions; (iv) a total absence of subject.

The second approach to impersonality involves a semantic perspective on the matter, where impersonal constructions allow to defocus the agent by diminishing its prominence (Siewierska 2008, Langacker 2011). According to Siewierska (2008:7), the defocussing process can be performed through: (i) the non- or under-elaboration of the agent; (ii) the demotion of agent from its prototypical argument slot; (iii) both processes. As pointed out by the author, there is, however, a considerable degree of overlap between the syntactic and the semantic perspectives. Indeed, she highlights that impersonal structures which have a semantically under- or non-elaborated agent correspond to those with no fully referential subjects whereas constructions with a demoted subject considerably match the other three types of constructions described in the syntactic approach. Agentless passives, which fall into the fourth category of the syntactic-based approach, are considered as undergoing a process of both demotion and non-elaboration

and cover thus the third possibility of constructions in the semantic perspective. To these two approaches Siewierska adds that a third type of characterization, from a morphological perspective, is also frequently applied to impersonality and corresponds to the invariant marking of the verb for person (Siewierska 2008:2). Given its higher degree of inclusiveness (Malchukov & Ogawa 2011:20), the semantic approach presented by Siewierska (2008), namely the agent-defocussing nature of the mechanism, will be adopted as the guiding element in the selection of the objects of study to be compared.
Some of these subtypes of impersonality are linked to passivisation. Indeed, the canonical periphrastic passive, which is traditionally seen as the prototypical device to serve passivisation, appears as one of the central mechanisms which spring to mind when it comes to expressing impersonal meanings since agent-defocussing is one of its principal functions (Givón 1979). Nevertheless, as already clear from the definition of impersonality offered by Siewierska (2008), the canonical passive is not the only construction allowing to defocus the agent and to express impersonal content. The passive voice and its relation to impersonality will be further detailed in Chapter 2.

In the present investigation, not only the canonical periphrastic passive but also a selection of other constructions serving impersonality will be surveyed and compared. Throughout this work, these structures will be referred to as *passive-like structures*. Nevertheless, in the theoretical chapters, the term *impersonal construction* will also be employed to respect the terminological decision of other researchers. Although the domain of impersonal constructions is heterogeneous, Malchukov & Ogawa (2011) point out that it has a certain structure and a series of relationships can be established between impersonal constructions. In order to select the constructions studied in the present work and examine the interaction between them, it is crucial to establish a common ground against which the comparison can be drawn.

## 1.1. Delimitation of the object of research

Prior to the selection of the object of study, and to any comparative work and contrastive analysis in general, it is crucial to establish a *tertium comparationis* (henceforth TC). Indeed, depending on the framework of reference, two items may appear as either different or similar (Jaszczolt 2003). In the introductory chapter of his influential work on functional contrastive analysis, Chesterman (1998) uses the following riddle from Carroll's masterpiece, *Alice in Wonderland*: *Why is a raven like a writing desk?* to illustrate the notion of *tertium comparationis*. At first sight, a raven has little in common with a writing desk. However, the author proves the contrary[5], highlighting the importance in creating and selecting an adequate common background of sameness when conducting comparative analyses.

The main objective in the use of such a TC is to form a language-neutral platform of comparison that serves as referent point when comparing different elements. The shared common ground allows to compare like with like (James 1980, Chesterman 1998, Connor & Moreno 2005,

[5] The author, referring to Gardner (1970), suggests that a raven and a writing desk share four kinds of likeness: purely formal, homonymic, semantic and functional (Chesterman 1998:5).

Ebeling & Ebeling 2020) and, through the use of delimiting features, focus on potential (dis)similarities across languages. Using such a starting point is all the more important since the present study includes three languages and compares six different items which are not all available in the three languages. Such methodological process allows to draw constant comparisons against a neutral basis and not directly against the different languages, which may result in a bias towards one or the other language (Dupont 2019:9). The decisions of (dis)similarities are thus made in reference to the *tertium comparationis* (Chesterman 1998:29). Usually, a contrastive work adopts three distinct methodological steps (James 1980, Krzeszowski 1990): (i) a description of the item, each of the studied languages being described separately; (ii) based on these descriptions, the creation of the TC which would further lead to the comparison (iii). Although it is largely acknowledged that the use of a neutral framework of reference consists in an essential requirement of any contrastive work (García 2002, Connor & Moreno 2005, Lefer 2009, Ebeling & Ebeling 2020), researchers tend to neglect the explicit mention of the selected common platform of reference (Krzeszowski 1990:15, Lefer & Cartoni 2011:90).

Traditionally, a *tertium comparationis* is either based on formal criteria, i.e., based on surface structure, or semantic criteria, depending essentially on the translation (Chesterman 1998, Kurteš 2009). Formal TC is, most of the time, untenable in syntactic contrastive studies (Krzeszowski 1990:16). As indicated by the author, a formal comparison of the *present perfect tense* in English with the *passé composé* in French would fail to capture all the facets of the phenomena since, even though their surface structure appear as similar, they do not match semantically (Krzeszowski 1990). However, Chesterman (1998) points out that both formal and semantic TC are somewhat related since contrastive analyses based on the form assume a potential semantic relation between the systems. Similarly, the author observes that, when using a comparative common ground based on meaning, researchers in translational studies tend to leave out data presenting substantially different structures (Chesterman 1998:30). As will be developed later on, by selecting a series of formally different impersonal structures, the present investigation intends to overcome this limitation. In addition to these two types of TC, several authors suggest that, to adequately fit the object of study, the frame of reference can be designed at different levels of analysis. This observation results in the possibility of a series of *tertia comparationis* which Chesterman (1998) summarises as follows: statistical equivalence, translation equivalence, system equivalence, semantico-syntactic equivalence, rule equivalence, substantive equivalence and pragmatic equivalence. Given the variety of TC available, Connor & Moreno (2005:5) further suggest that all levels of research, including the material selection process to design comparable corpora, could benefit from the use of this methodological approach. They thus extend the notion of TC, usually used on microlinguistic levels, to macrolinguistic ones.

In establishing and using a common platform of comparison, the linguist faces two limitations. The first one concerns the issue of circularity (Krzeszowski 1990, Chesterman 1998), originating from an inevitable risk of confusion between the initial reference framework and the consequent result of the comparison. Krzeszowski describes this risk of circularity as follows (1990: 20):

> "We compare in order to see what is similar and what is different in the compared materials; we can only compare items which are in some respect similar, but we cannot use similarity as an independent criterion in deciding how to match items for comparison since similarity (or difference) is to *result* from the comparison and not to *motivate* it".

In an attempt to avoid this circularity, Chesterman (1998:58) suggests to view the assumption of similarity as the initial point of the comparison and the *tertium comparationis* as the consequent result of the comparative work. Zhu highlights that, even though this approach has received considerable attention, it remains a controversial proposal (2017:35). In line with Chesterman (1998), Ebeling & Ebeling (2020:99) suggest that any comparative work should be based on a perceived equivalence between grammatical items. Within this perspective, the objectives of the study are thus viewed as evaluating this equivalence.

A second limitation is, as pointed out by Chesterman (1998: 37), the subjectiveness of similarity judgements, which can vary according to researchers and the particular aims of their study. Similarly, TC largely vary since they are established following the specific needs of each comparative work. *Tertium comparationis*, despite its crucial importance in setting the bases of contrastive analysis, displays thus some limitations when it comes to offering a precise definition of the concept.

Given the above, it can safely be assumed that the establishment and the choice of the *tertium comparationis* is of crucial importance and heavily orientates the aim and consequent results of the research.

Due to the colossal number of impersonal mechanisms available in human language and the lack of general consensus on the notion of impersonality, this work adopts Creissels' methodological approach (2008) to impersonals, which focuses on a definite set of constructions serving impersonal purposes (presented in Table 1) and does not intend to embrace impersonality from a general perspective.

The recommendation to resort to a functional framework of reference (Conner & Moreno 2005) is followed in the present investigation since it includes various objects of study, displaying different surface structures and possible meanings. In an attempt to avoid the issue of circularity, the starting point of this analysis is the assumption of the following functional similarity between the selected structures: defocussing or demoting the agent of a verbal clause through under-elaboration, non-elaboration or both processes. Two brief comments on the selection should be added. First, given the ample scope of this study (three languages and, as will be explained later on, two modes and two registers) all agent-defocussing mechanisms could unfortunately not be included. It is, among others, the case of the generic use of the second person singular, already extensively analysed[6] and often discussed alongside impersonal strategies as those included in this study. Despite its agent-defocussing potential, the generic second person plural possesses a stronger deictic basis than the strategies included in this study, with a consequent referential ambiguity (Kluge 2016). In addition, the generic *you* has less

[6] For a more detailed account of this phenomenon, see, among others, Barrajón López 2005, Kluge 2010, 2012, 2016, Hugo Rojas 2011, Cela & Candea 2013 and Posio 2016.

frequently been presented as a possible alternative to the passive than some other constructions based on personal pronoun such as the non-anaphoric third person plural (Weiner & Labov 1973). A second comment has to be made regarding the two passive-like constructions based on a numeral, i.e., *uno* and *one.* These structures can fulfil other syntactic functions than the subject one, appearing in these cases without any verbal clause. These low frequency constructions are included in the descriptive analysis only for information purposes. As we will see, given their object nature, some of the parameters developed in the taxonomy used in this investigation do not apply.

## 1.2. Justification of the study, objectives and research questions

Prior research on passive-like structures does exist. Nevertheless, studies usually focus on one language (Monge 1954, Gelabert-Desnoyer 2008, Guirado 2011, Fernández 2012, Devís Márquez 2017, Aijón Oliva 2020, De Cock 2020, to cite only a few), contrast two or more languages but with a focus on one structure (e.g., Espinoza 1997, McEnery & Xiao 2006, Coussé & van der Auwera 2012 or Giardano 2015) or adopt a typological perspective (Siewierska 1984, Rousseau 2000, Sansò 2006, Creissels 2008, Siewierska & Papastathi 2011, Gast & van der Auwera 2013). When adopting a contrastive perspective, in most cases, the object of research is rather restricted or the studies are either conducted on small-sized corpora or lack empirical bases. To our knowledge, a deep understanding of passive-like structures built from a multifactorial contrastive and corpus-based approach is thus lacking.

In the present dissertation, I strive to show how a series of passive-like structures interact across registers and modes by establishing a comprehensive panorama of their syntactico-pragmatic features. This research represents an added value for the field of contrastive linguistics since it adopts a multidimensional perspective which includes six structures, three languages, nineteen parameters, and two registers and modes. In addition to these multiple variables, findings are explored and strengthened through advanced statistical models, often absent in previous studies on the topic though indispensable in contrastive research.

The present thesis intends thus to fill in this gap of a fine-grained usage-based contrastive panorama of six passive-like mechanisms in English, French and Spanish: the canonical periphrastic passive, the *get*-passive, the *se*-construction, the non-anaphoric third person plural, the man-impersonal *on* and the numeral-based constructions *one* and *uno.* Variation is the constant thread running through this work. I develop the theme by investigating the following three areas: variation across syntactico-pragmatic features, across languages, and across modes and registers.

**Variation across specific features**

- What are the main distinctive grammatical features of the different passive-like structures under study? Do these structures appear in different linguistic environments?

The first research topic concerns the different specific syntactico-pragmatic behaviours displayed by each construction involved. Through a twofold perspective analysis, I intend to

highlight the convergent and divergent features of the selected passive-like structures. The first axis of the study focuses on the verb and the participants included in the clause as well as their respective characteristics. The departure point of this analysis is largely based on Hopper & Thompson's work (1980) on transitivity. The second axis explores the linguistic context in which the passive-like construction appears. Several scholars have already emphasized the importance of the linguistic context in the interpretation and selection process of impersonal constructions (Landragin & Tanguy 2014, Gjesdal 2008, Yao 2016, Boutet 1986, Fløttum 2007 to cite only a few). A special interest will be devoted to determining which element(s) of the context play a key role in these processes.

**Variation across languages**

- How do language specificities impact the selection of the passive-like constructions?

The languages selected for the present investigation pertain to the Indo-European family of languages. Nevertheless, it is an undeniable fact that these three languages will display differences, each language possessing its own grammatical specificities and mechanisms. What is of particular interest here is to bring to light where the potential (dis)similarities lie and examine whether the origin of the language, i.e. their belonging to either the Germanic or Romance group of languages, possibly impacts the interpretation and the selection process. The findings stemming from this analysis will have to be cautiously interpreted since a restricted number of languages for each group is studied in the present investigation. They can, however, serve as a preliminary basis for further research. Three points are to be considered: (i) the frequency of use of the structures, (ii) as not all the structures are available in the three languages, I will intend to determine whether some formally different constructions may have a similar role and grammatical characteristics. I will devote particular attention to the man-impersonal *on* construction since, among the three languages investigated in the present study, such mechanism is only available in French. Scholars have pointed out that several impersonal mechanisms such as the non-anaphoric third person plural, the *se*-construction or the *get*-passive are available as possible alternative to man-impersonals (Yao 2016, Siewierska 2011). I will examine, from both a syntactico-semantic and a pragmatic perspective, to which extent these structures can be considered as viable equivalents. (iii) I will also focus on the structures available in the three languages (that is, the periphrastic passive and the non-anaphoric third person plural) or in two out of the three languages (the *se*-construction and *uno*/*one)*. It is hypothesised that apparently similar constructions may display slightly different grammatical properties and may be accompanied by different contextual elements.

**Variation across mode and register**

- To which extent do mode and register affect the selection of a passive-like construction?

A third part will be devoted to mode and register specificities. Within these terms, I include oral versus written modes and formal versus informal registers. Teich et al. argue that "particular situational settings have linguistic correlates at the level of lexico-grammar" (Teich et al. 2016:1669). I assume that the use of a combination of data including different degrees of formality and modes of production allows a more detailed interpretation of the studied phenomena. Two working hypotheses can be formulated in this regard. It has been widely

proved that, overall, different discursive characteristics can be found both between oral and written modes as well as between formal and informal registers (Biber 1995, Koch and Oesterreicher 2001). I postulate then that the six passive-like structures possess different grammatical properties and occur with different contextual elements according to the register/mode they are used in. A second hypothesis can be formulated concerning their productivity. I assume that, in addition to a difference in grammatical properties and contextual elements, some constructions are more specific to certain registers and modes. I will then examine whether the impact of the register and mode is similar across the three languages, an aspect long neglected in contrastive studies.

To sum up, through a bottom-up approach, which means I will conduct the analysis starting from my corpus data, I intend to create a solid synthesis by investigating three types of variations, namely variation across specific features, languages and registers and modes. This will allow to draw a comprehensive portrait of the use of passive-like structures and their behaviour according the languages and the context in which they are used. These three axes of research will be studied in light of each other to obtain the most fine-grained panorama as possible. Thus, the project intends to offer a clearer insight into the functioning of a selection of passive-like structures and may contribute to a better understanding of the differences between the languages under study and the impact of register/mode.

## 1.3. Preview of the thesis

This thesis includes nine chapters besides the introduction and the conclusion: three theoretical chapters, one methodological chapter and five analytical chapters. The next chapter (Chapter 2) describes the core features of the six passive-like structures included in the present work in the following order: the periphrastic passive, the *get*-passive, the *se*-construction, the non-anaphoric third person plural, the man-impersonal *on* and the numeral-based *uno* and *one.*

Chapter 3 presents first the framework within which this investigation is rooted, namely contrastive linguistics practices (3.1). In a second step, it offers a review of the existent literature on crosslinguistic studies exploring impersonal constructions (3.2). This section is subdivided in three parts, corresponding to works either related to English, French or Spanish.

The last theoretical chapter (Chapter 4) discusses the key notions of this work. Sections (4.1) and (4.2) defines mode and register, two concepts that have been fairly neglected so far in large corpus-based contrastive studies. Section (4.3) presents the notion of transitivity, which lies at the heart of the annotation protocol.

Chapter 5, the methodological chapter, offers an overview of the data sets, explains how they were compiled and details the annotation protocol adopted in this investigation. It also provides information on the inter-annotation agreement performed to test the reliability of the annotation scheme as well as information on the statistical tests included to strengthen the results yielded by the present work.

Chapter 6, the first analytical chapter, suggests a general descriptive overview of the data of the thesis and is subdivided in three parts: the first part presents the overall distribution of the data, the second part provides information on the type/token ratio of the passive-like structures and the third part explores the overall statistical results.

Chapters 7, 8 and 9 are dedicated respectively to the analysis of the French, the Spanish, and the English passive-like structures included in the study. They are all organized in five distinct parts: first a general overview of the data distribution is presented through the graphs provided by Multiple Correspondences Analysis tests, the next three sections detail the behaviour of verb-related, participant-related and context-related parameters. The final section intends to capture the picture of the interaction of the different passive-like structures in each language.

Chapter 10 compares the results found in the three language-specific analytical chapters. It first contrasts the findings obtained regarding the passive-like constructions and second it focuses on the influence mode and register exercise on the distribution of the data.

The thesis closes with the main conclusions and contributions of this work and opens future avenues for further research on the topic.

# 2. Core features of passive-like structures

Chapter 2 sketches the main morpho-syntactic and functional characteristics of the six types of structures analysed in this study.

The first section (Section 2.1) will be devoted to the canonical passive as it is a widely acknowledged device used for agent defocussing, attested in the three languages. I will add a brief description of the *get*-passive, a structure closely related to the canonical passive but available in English only. Then, I will turn to the passive construction based on the reflexive pronoun *se*, available in Spanish and French (Section 2.2). This construction is considered as a frequent candidate competing with the periphrastic passive, especially in Spanish. The next section (Section 2.3) will summarise the main characteristics of R-impersonal constructions (Siewierska 2011), i.e., the non-anaphoric third personal plural and the French man-impersonal. The last part will describe the numeral-based indefinite pronouns *uno* in Spanish and *one* in English (Section 2.4).

Note that some of these structures have already been largely described (for example the periphrastic passive or the *se*-construction in Spanish). Considering that the present study falls within a contrastive approach, the main objective is not to provide an exhausting formal description of each structure but rather highlight the salient syntactico-semantic properties of the structures and compare their respective behaviour in order to establish a fine-grained panorama of their mutual interaction. Therefore, this section aims at briefly describing the constructions under study from a grammatical perspective and suggests a summary of the leading investigations already carried out in the field, without providing many detailed explanations as it would largely go beyond the scope of my investigation.

## 2.1. Periphrastic passives and *get*-passives

Passivization can be considered as a quasi-universal phenomenon since typological studies indicate its presence in all the main language families (Siewierska 1984). Therefore, I will try to provide, as far as possible, a description of the phenomenon that would fit the three languages. When relevant, potential differences will be highlighted.

### 2.1.1. Brief account of voices

Before getting at the heart of the matter, that is the description of the canonical or periphrastic passive, a brief account of the notion of voice will be provided since the periphrastic passive is directly related to this central notion. Le Bellec (2009) reports on the difficulty to reach a consensus on an exact definition of voice, which is intrinsically linked to the notion of diathesis and sometimes considered as synonymous. Acknowledging this hotly debated issue, I will adopt the following, though reductive, view of the phenomenon which considers *voice* as a

grammatical category related to the form of the verb. Within this approach, the notion of voice can be defined as the "property of a verb which makes clear whether the subject of the verb performs the action or receives the action described by the verb (Pence & Emery 1963:42)". Thus, analysing the voice provides insights into the relation between the predicate and its potential arguments. The concept of diathesis would enlarge the scope of the relation predicate/argument(s) through the inclusion of a communicative perspective on the matter (Le Bellec 2009).

Adopting one of the traditional views on the topic, a tripartite division of the concept can be made: active (considered as the unmarked option), middle or pronominal and passive voice (viewed as marked). The system to represent the concept of voice differs according to the language. While the form and use of active and passive voices in English, Spanish and French converge on numerous points, it is not the case for what is commonly termed the middle voice. While Spanish and French possess a clear indicator of the middle voice, namely the clitic *se*, English is deprived of such device. Middle readings in English are rather rendered through an intransitive use of verbs which are normally considered as transitive such as in (1a) and (2), where the grammatical subject of these structures is the recipient/beneficiary of the action. Regarding the presence of an explicit agent, English rejects its presence, unlike Spanish and French where, though rare, its expression is possible in a restricted number of situations. Example (1b) would be the potential equivalent middle voice in Spanish. When used intransitively, these verbs are usually accompanied by adverbials or prepositional phrases (Carter & McCarthy 2006), as illustrated by the adverbs *easily* and *quickly* in examples (1b) and (2).

(1) *a. This screen cleans easily.*
*b. Esta pantalla se limpia fácilmente.* (Examples Barta 2014:6)

(2) *Fish cooks quickly.*

(Carter & McCarthy 2006 : 2006)

Quirk et al. (1985) report that these grammatical features lead some scholars to only distinguish two voices in the English language and rather categorize the middle phenomenon as either active or passive. A similar trend to consider middle constructions as an expression of the passive voice is sometimes observed in Spanish and French, though largely rejected among scholars (Wonder 1990, Arce et al. 1994, Maldonado 1992, Quesada 1998, Gómez Torrego 2015 on Spanish; Merk 1988, Wilmet 1998, Le Bellec 2009 on French; Frias-Conde 2006 on Romance languages). Regarding the frequency, unlike French and Spanish, the middle voice in English is far less common. This may partly be attributed to the fact that numerous pronominal verbs in Romance languages correspond to either active or passive options in English (Grady 1969, Vinay & Darbelnet 1995 on French). I will come back to the expression of the middle voice when discussing the *se*-constructions in French and Spanish and for now focus my attention on the passive.

### 2.1.2. Canonical passives

Passivization is traditionally defined as "the process by which a non-agent is promoted into the role of 'main topic' of the sentence (Givón 1979:117). Its primary function is thus agent defocussing. As pointed out by several scholars, there is no general consensus over a clear definition of what constitutes a passive (Granger 1983, Siewierska 1984). Granger (1983:73) listed a series of examples traditionally considered as passives, some of them reported in (3-8):

(3) The dog was killed by his own master.
(4) My thesis is finished.
(5) He was much surprised at the news.
(6) My dog got killed in a road accident.
(7) The village seemed quite deserted.
(8) He felt disappointed.

Since there are nearly as many definitions of the passive as researchers interested in the phenomenon, throughout this study, the term *periphrastic passive* will be used to refer to a construction fulfilling syntactic, lexical and morphological requirements (Jalenques 2016), which are briefly discussed below.

(9) Une clôture électrique entoure la propriété.
X Y
(10) La propriété est entourée par une clôture électrique.
Y X

(Examples from Jalenques 2016 :2)

First, the second argument of the verb (Y in (9)), traditionally the direct object, in the active equivalent, is promoted to the syntactical subject position (Y in (10)) in the passive construction. The expression of the first argument (Y) is usually not compulsory. A periphrastic passive does, however, not merely derive from an active equivalent (Gaatone 1998, Jalenques 2106). This is especially true since some periphrastic passives do not have active counterparts at all. Also, in some cases, the formal active candidate of a passive construction does not convey the exact same semantic meaning (e.g., Lamiroy 1993 and Gaatone 1999). Consider (11) and (12) from Gaatone (1998: 19):

(11) *Maya n'a pas invité beaucoup d'enfants.*
'Maya did not invite many children.'
(12) *Beaucoup d'enfants n'ont pas été invités par Maya.*
'Many children were not invited by Maya.'

In the first example, it can be imagined that within the context of a specific activity Maya only invited a few children to take part in it. Conversely, it can be inferred from the second example that only a few children among a much larger group were invited by Maya.
Passivization is thus a much more complex process than a simple structural modification of an active option through the reversal of two arguments of the verb.

Second, as stated above, the periphrastic construction involves a finite or non-finite form of the auxiliary *be* in English, *ser* in Spanish and *être* in French. Note that in Spanish, past participles can combine either with the copular *ser* or *estar*. The latter is much more restricted than *ser* and usually associated with resultative meanings (Marín 2004, Bosque 2014). This has led scholars to question to which extent *estar* with a past participle can be considered to be a periphrastic passive (Contí 2004). Therefore, this study focuses on *ser*-passives. The status of the *be* auxiliary may appear problematic as it can be considered as a mere grammatical indicator of the passive, an auxiliary verb or a main verb. The status of *be*, thus deprived from any lexical meaning, is the most widely accepted view (Siewierska 1984, Langacker & Munro 1975).

Third, the particle *be* must be followed by a past participle which concords in genre and number with the subject. The majority of these past participles denote dynamic actions (Granger 1983:190 for English) or, in a smaller proportion, are classified as cognitive, perception, reaction or locative verbs. As will be explained in the following section, the nature of the participle sometimes appears ambiguous.

In line with the suggestion of various researchers (e.g., Comrie 1977 and Keenan 1975), I consider the phenomenon of passivization as allowing patient promotion and agent demotion to operate independently from one another. Within the term of periphrastic passive, I therefore include both personal and impersonal passives such as illustrated in (13) and (14).

(13) *Une enquête a été menée par les autorités.*
'An investigation was conducted by the authorities.'

(14) *Il a été procédé à une enquête.*
'An investigation was conducted.'

(Helland 2002:9)

(15) *It was hoped that it would rain.*

(Müller & Ørsnes 2013:142)

(16) *It is said that figs are better for us than bananas.*

While personal passives include both moves (patient promotion and agent demotion), impersonal passives only involve agent demotion. As illustrated in (13), which exemplifies a personal passive, the grammatical subject *par les autorités* ('by the authorities'), that is, the agent, is demoted and the grammatical object *une enquête* ('an investigation'), that is, the patient, is promoted in topic position. Conversely, the impersonal passives in (14-16), only imply one move. In (14-16) there is no promotional move of a noun phrase from comment to topic position. The presence of one unique move in impersonal passives is possible through the use of expletives pronouns (*it* in English and *il* in French). It has to be mentioned that the meaning conveyed by the French and the English impersonal passive as in (14) is usually realized through an impersonal *se*-construction in Spanish (Sharp 2005: 117).

### Distinction between adjectival and verbal past participles

When selecting and classifying the occurrences of periphrastic passives, the linguist will be confronted to the issue of verbal versus adjectival nature of the past participle. To select the data, I followed Granger's methodology (1983) and, within periphrastic passive consisting in be+Ved, I only retained the forms with corresponding infinitive, which Svartvik labels 'verbal base noncompound' (1966:23). Nevertheless, in some contexts, distinguishing verbal from non-verbal participles becomes an arduous task. Several grammarians (Granger 1983, Lamiroy 1993, Desclés & Guentchéva 1993, Creissels 2000, Carlier 2002) have tackled the issue of the nature of the past participle. A summary of the main points of their approaches is provided hereunder in order to establish sound theoretical bases to select the occurrences of the periphrastic passive analysed in this study.

Desclés & Guentchéva (1993) differentiate a passive construction from an adjectival construction thanks to the presence of defocused agent (1993: 79). They suggest that the passive predicate contains an agent while the adjectival predicate does not. In Example (17) someone performed the action of opening the door while Example (18) describes a state and there is, thus, no underlying agent.

(17) *La porte a été ouverte (par quelqu'un).*
'The door was opened (by someone).'
(18) *La porte est ouverte (et non pas fermée).*
'The door is open (and not closed).'

(Examples Desclés & Guentchéva 1993 :79)

Creissels adopts a slightly different view as he argues that passive and adjectival predicates should not be opposed as the former displays characteristics of the latter (2000:134). He suggests that the construction *être + participe passé* 'be+past participle' can take on either a passive or a resultative meaning. The term resultative corresponds to what Desclés & Guentchéva (1993) named adjectival.
Creissels suggests that the construction *être + participe passé* is considered as passive only "si on peut mettre la séquence $N_{sujet}$ + être + $V_{pp}$ en correspondance avec une séquence $X_{sujet}$ + V + $N_{objet}$ (…) (2000 :133).[7]" Examples (19) and (20) illustrate Creissels' suggestion.

(19) *Les coupables sont recherchés -> On recherche les coupables.*
'The culprits are looked for > We are looking for the culprits.'
(20) *Ils sont recherchés par la police -> La police les recherche*
'They are looked for by the police > The police are looking for them.'
(21) *Ma maison est construite.*
'My house is built.'

(Examples from Creissels 2000)

(22) *His bills are paid, so that he owes nothing now.*

[7] 'if the sequence $N_{subject}$ + be + $V_{pp}$ can match a sequence $X_{subject}$ + V + $N_{object}$ (Creissel (2000:133).' [free translation]

(23) *His bills are paid regularly every month.*

(Examples from Granger 1983)

In some cases, Creissels shows that the construction can receive either a passive or a resultative meaning. For example, this is the case in (21), where the verb can either be interpreted as resultative where the meaning is "Ma maison est déjà construite, elle a été construite en six mois seulement (2000: 137)" or as passive where the meaning corresponds to "Ma maison est construite par l'entreprise X" (2000:137). The presence of an adjunct such as in (22) and (23) may facilitate the distinction between both types of readings, hence the importance of the context in assigning one or the other meaning (Poutsma 1926, Granger 1983, Creissels 2000).

Carlier (2002), in line with Creissels (2002)'s work, adds that a resultative construction may sometimes receive an agent complement.

(24) *Méfiez-vous! L'électricité est installée par un bricoleur inexpérimenté.*
Beware! Electricity is installed by an inexperienced handyman.

(25) *(…) car il est dansé ici une autre danse [une autre danse est ici dansée].*
because another dance is danced here [another dance is here danced].

(26) *Sur chaque face, il est enregistré trois chansons [trois chansons sont enregistrées sur chaque face]*.
On each side, are recorded three songs [three songs are recorded on each side].

(Examples from Carlier 2002)

She explains that in (24) the action performed by the agent affects the result of the process that is, a possible danger caused by the electrical circuit (2002:54). Carlier (2000) also notes that the periphrastic passive can produce impersonal constructions. This is possible both with a passive and a resultative structure such as exemplified in (25) and (26).

A criterion to select periphrastic passives based solely on a mere dichotomy between the adjectival and passive nature of the participle would be too simplistic. The picture is much more complex and the boundaries between the two are sometimes fuzzy. The selection process used in the present investigation is further based on a tripartite classification elaborated by Granger (1983). She distinguishes three main categories of verbs considered as *be* Ved: passives (P) (the most common category), verbal pseudopassives ($PsP_v$) and adjectival pseudopassives ($PsP_a$). The category of passives, as illustrated in (27), includes those occurrences that possess an active counterpart.

(27) *That attitude **was maintained by** the government in the further nine days of debates in the Lords.*
(28) *Well, that would **be modernized** and labour-saving, that seems to me; that's a big enough house.*
(29) *The ministry is **satisfied** that providing the normal method of gassing technique is followed, then death from hydro-cyanic acid gas is quick and painless.*

(Examples from Granger 1983)

Adjectival pseudopassives (27) do not have the possibility of alternating with an active option, as was the case for the passives and possess similar features to central adjectives. Finally, verbal pseudopassives do not alternate with a semantically active equivalent either nor display adjectival characteristics (28).

Several distinctions can be made within each of the three categories to refine the classification as Granger obtains a final set of seven categories (1983). I will, however, not enter into further detail in the present investigation as it goes far beyond the scope attributed to this work. For a more comprehensive account of the classification and the difference between the types of verbs in canonical passive constructions, see Granger (1983).

I will complete the description of the internal structure of the periphrastic passive and the nature of its verb with a short note on the agent. Though the expression of the agentive force is rare (Weiner & Labov 1983, Granger 1983 on English, Siewierska 1984, Lamiroy 1993 on French), its presence is not impossible (Xiao et al. 2016, Jespersen 1933, Svartvik 1966, Siewierska 1984 to only cite a few). The use of an agent complement depends on grammatical and extralinguistic factors. Passives with an overt agent are referred to as *long passives* and those lacking an agent complement as *short passives* (Desclés & Guentchéva 1993, Biber et al. 1999). Scarcity of explicit agents is an expected phenomenon since one of the primary functions of the passive is agent defocussing. In English, the preponderance of short passives has been observed in both spoken and written registers (Xiao et al. 2016). Lamiroy (1993:62) notes, however, that the presence of the agent complement is compulsory when the verb has a metaphorical meaning such as in (30).

(30) *Plusieurs personnes ont été blessées par les mots de Paul*.
'Several people were hurt by Paul's words.'

Before turning to the description of the functions of the periphrastic passive, I will briefly mention and illustrate some constructions that are not included in this work.

(31) *Ma chienne (Maya) est croisée Husky / Berger allemand.* (Corpus Yahoo FR)
'My female dog (Maya) is cross-breeded Husky/ German shepherd'

(32) *Oui les chiens et autres animaux puent quand ils sont mouillés* (Corpus Yahoo FR) 'Yes dogs and other animals stink when they are wet.'

Since the present study is interested in the potential presence and impact of an agentive force, I will focus on passive constructions with and without agent as well as resultative constructions with an agent complement which allows a passive reading. True adjectival passive constructions will be discarded. Thus, examples (31) and (32), automatically extracted and considered as periphrastic passives by the software Unitex, will not be selected for the purpose of this analysis since, in those cases, the construction 'être + participle passé' has only an adjectival meaning.

(33) *Trahi par ses proches, il quitta le pays*
'Betrayed by his relatives, he left the country.'

(34) *La pierre ôtée, on vit en dedans de la tombe* (V.Hugo)

‘The stone removed, one lives inside the tomb (V. Hugo)’

(Lamiroy 1993 :53)

Truncated passives, that is detached from the auxiliary *be*, such as (33 and 34) will also be discarded. The main reason motivating this methodological choice is the absence of an auxiliary verb in the construction. Deprived of this grammatical element, a series of parameters related to the verb would be irrelevant, with a subsequent impact on the comparability of the data. Furthermore, the absence of auxiliary complicates the extraction process.

Now the phenomenon of the periphrastic passive is delimited, I will intend to summarise its principal functions.

The main function, commonly agreed among scholars, is the promotion of the patient (e.g., Givón 1979, Shibatani 1988, Gaatone 1999, Carter & McCarthy 2006, Raineri 2010) and the subsequent demotion of the agent, that is usually even supressed from the construction. The defocussing process is mainly carried out for the following reasons (Jespersen 1933): (i) the object is unknown or evident from the context, (ii) there is a special interest in hiding it (for instance the need to attenuate the personal implication of the agent), (iii) the patient is of preponderant importance. Siewierska (1984: 223) suggests an additional reason to promote the patient by specifying that, in English, when “a definite patient is higher on the animacy hierarchy than an indefinite agent”, the passive option would most likely be the preferred candidate. Similar findings on the impact of the degree of animacy on the argument structure were also found by Ashby & Bentivoglio (1993:73) in French and Spanish oral data.
Promotional moves also allow to place greater focus on the process rather than on the arguments since they are either suppressed or relegated to a position of second importance (Gaatone 1999, Raineri 2010). This holds particularly true in the case of impersonal periphrastic passive since they contain indirect transitive and intransitive verbs.
Nevertheless, promotional and defocussing processes cannot be considered as the only purpose of periphrastic passive since, in some contexts, even though situated in second position, the presence of an explicit agent is still required and of crucial importance. It can be assumed, thus, that the periphrastic passive fulfils some additional purposes.

The use of periphrastic passive allows to organize the discourse on two points. First, it maintains the given/new distribution of the information (Siewierska 1984, Raineri 2010). Second, it develops textual cohesion by facilitating the connection between the sentences (Jespersen 1924) in creating a syntactic pivot (Dixon 1979). Gaatone (1999:224) explains that the main objective is to maintain the same given support where some anaphoric relations between the different syntactic elements already exist. He provides (35) to illustrate this function of the periphrastic passive, where the subject *wish* of the first clause becomes the patient of the subordinate clause but still occupies the subject slot, that is in topic position. Required information on the agentive force, necessary in this context, is brought by the presence of an agent complement.

(35) *Un vœu est toujours exaucé lorsqu’il est fait par un enfant au moment même où il vient au monde.*
‘A wish is already granted when it is made by a child at the very moment when he/she comes into the world.’

The use of an active option in (35) would have involved a change in the noun phrase occupying the topic position, which would have prejudiced textual cohesion.

A last reason commonly offered to explain the use of periphrastic passive is register sensitivity (Nash 1980, Siewierska 1984, Leech & Svartvik 1995). The passive, compared to the active, is usually considered as a more elaborate mechanisms and appears as a preferred device to meet stylistic needs (e.g., impersonal style) in formal registers. This study will offer more insights into the exact relation between the use of periphrastic passives and registers.

### A brief note on frequency, productivity and genre

Svartvik (1966) assumes that the degree of colloquialism of the production is the key factor impacting the frequency of use of the passive rather than the medium, written or spoken, considered by the author as of secondary importance. It seems commonly accepted that the canonical passive is stylistically marked and serves to fulfil a series of requirements of the formal discourse (Jespersen 1924, Weiner & Labov 1983, Siwierska 1984, Yamamoto 1984). Its subsequent high frequency in formal registers, compared to informal ones, is, therefore, of no surprise. Biber (1988) further specifies that, typically, productions with either an abstract or a technical content display a high number of canonical passives.

Compared with French and Spanish, English is known to largely resort to the periphrastic passives. This high productivity may partly be explained by the structure of the language. Siewierska puts forward two properties that may be at play: "the rigid word order of English, together with the possibility of passivizing not only direct, but also indirect and prepositional objects (…) (1984:230)." These characteristics are also highlighted by Vinay and Darbelnet (1995) that add the high frequency of passive in English is also a matter of perspective on life.

> "Il y a une certaine objectivité anglaise qui se plaît à constater un phénomène sans l'attribuer à une cause précise, ou qui ne mentionne la cause ou l'agent qu'accessoirement. On ne peut s'empêcher d'établir un rapport entre cette construction et la répugnance des Anglo-saxons à formuler tout de suite un jugement ou même une opinion (Vinay & Darbelnet 1995 : 136)."

The existent literature on canonical passives seems to indicate that Spanish, among the three languages selected in this study, resorts the least to this device. Quesada (1998) even refers to the low frequency of periphrastic passives in Spanish spoken language by using the terms "almost non-existent" (Quesada 1998). Some researchers (García Yebra 1990, De Cock 2014) specify that Spanish only resorts to passives when actives cannot meet pragmatico-discursive requirements, for instance for reasons of clarity or rhythm. Between highly frequent in English and almost non-existent in Spanish, French occupies an intermediate position (García Yebra 1990, Gauchola 2012).

A common form frequently mentioned alongside the periphrastic passive is the *get*-passive. The next section is devoted to a brief description and comparison of the two forms since they are believed to, sometimes, fulfil similar functions.

### 2.1.3. *Get*-passives

Besides the canonical passive formed with the *be* auxiliary, English possesses the *get*-passive, a variant termed as the non-canonical passive by some scholars (Alexiadou 2005). However, as it will be demonstrated, even though some of their usages overlap, the *get*-passive cannot be considered as a fully structural variant of the *be*-passive. The exact semantico-syntactic distinction between the *get*-passive and the *be*-passive of passives has not been studied in detail so far (Thompson et al. 2018) but some reflexions on the main distinctive features usually pointed out in the literature are integrated simultaneously to the description of the use of the *get*-passive when relevant.

#### *2.1.3.1. Description of the construction*

The non-canonical passive consists in the verb *get* + a past participle. As the *be*-passive, the *get*-passive, allows a shift of the patient towards the subject position, placing the latter in topic position. Scholars indicate that the use of the non-canonical passive allows to put the focus on the event and/or its outcome rather than focus on the patient (Alexiadou 2005, Carther & McCarthy 1999, Thompson et al. 2018). Thompson et al. even report some "level of patient agency" (2018: 3). The patient in the *get*-passive appears as more affected[8] than in the be-passive and can be charged, to some extent, for the denoted event (Thompson 2018, Honda 2012, Rhülemann 2007). Compared to the *be*-passive, the surface subject of the get-passive can be considered as more actively involved in the event (Hatcher 1949, Lakoff 1971).

Though its apparent basic structure, precisely delimiting the phenomenon of the *get*-passive is a difficult task since the structure may adopt different readings, each one displaying a variable degree of affinity with the *be*-passive construction. Each reading may be placed on a continuum ranging from a more active pole such as (36) and (37) where patients are actively involved in carrying out the action to a full passive one (38) (Svartvik 1966)[9].

(36) *My parents got divorced when I was a junior in high school.*

(37) *The sisters (sc.babysitters) we spoke to simply love a good page of written instructions, full of such small but important details as where the soothers are kept, who needs a night light, and if the teeth get brushed before or after PJs go on.*

[8] The term *affected* may have different acceptations. Within the description of the *get*-passive, I will follow Thompson's view according to which *affected* means that "the patient undergoes some manner of change as a result of the action or event (Thompson 2018:3)."

[9] For an extensive account of the different related forms of *get*-passive see Carter & McCarthy (1999:46).

(38) *He got killed trying to save some other man.*

(Anderwald 2018:4,12)

Coming back to Examples (36) and (37), some *get*-passives, which are called *middle,* have a reciprocal (36) and reflexive (37) meaning. Example (36) can be understand as they divorced from each other. In (37), the participants, here supposedly children, do brush their teeth on their own. The argument structure of these types of *get*-passives differs from a truly passive since middle structures only imply one participant, the patient which is, to some extent, simultaneously the doer and the receiver of action. In this sense, the *get*-passive differs from the *be*-passive since the canonical passive does not accept neither the reflexive nor the reciprocal readings (Arce Arenales et al. 1994, Alexiadou 2005).

Only the most common type of *get*-passives is studied in the present investigation, namely the one displaying the internal structure X *get* V-en (by Z), where a patient (X) undergoes a process expressed through a *get* verb (get V-en), accompanied or not by an overt agent (Z) (Carter &McCarther 1999, Collins 1996) as illustrated in (2).

As far as register preference and productivity of the structure are concerned, two main points can be raised. Several researchers (Leech & Svartvik 1975, Huddleston 1984, Anderwald 2018, Rühlemann 2007) note that the *get*-passive appears to be highly register sensitive as corpus-based studied reveal that the construction is more commonly used in the oral mode and, more specifically in informal situations. Regarding its productivity, the *get*-passive is less frequent than the *be*-passive option (e.g., Quirk et al. 1972, Granger 1983). This may be partly attributed to the fact that it suffers more syntactic restrictions than the canonical passive, such as the impossibility of occuring in a negative clause or in the interrogative form as will be developed in the next section.

#### *2.1.3.2. Specific features of the get-passive*

As Anderwald (2018) points out, thoroughly describing the use and behaviour of the *get*-passive is not such a straightforward task. Some reflexions about its description and behaviour can, however, be summarised as follows.

First, as is the case for the canonical periphrastic passive, the *get*-passive can occur with adjectival and verbal participle. In a series of cases, the participle is listed in reference dictionaries as adjective (e.g., *drunk, tired* (*of*), *entangled* or *rid of*) or verb (e.g., *delivered, paid*, *smashed* or *saved*). These cases do not pose problem and can easily be discarded. However, such a clear distinction is not always possible and the interpretation may sometimes result ambiguous as both usages are possible. Honda (2012: 74) suggests example (39) to illustrate this ambiguity.

(39) *John got involved in the project quickly.*

A first reading understands the construction as follows. An implicit agent, for example, other members of the project, behaves in such a way that John was quickly involved in the project. In this first acceptation the participle *involved* is analysed as a verbal passive participle. A second reading would consider that John becomes involved in the project through his own actions. In this case, *get involved* would be interpreted as an adjectival participle. Fleisher (2006) notes that aspect may facilitate the distinction between adjectival and verbal readings, the presence of a perfective aspect leading towards a verbal passive participle. For the purpose of this study, only full verbal passive participles were selected.

Second, another characteristic of the non-canonical passive is the particular nature of the verb *get.* Its grammatical status differs from a typical auxiliary on three points. The first one is the fact that the non-canonical passive does not accept the negation (Toyota 2008, Honda 2012, Carter & McCarthy 1999). While sentences like (40) are possible with the *be*-passive, the negative option results ungrammatical (41) with *get*-passive and would require the support of *do* to be acceptable.

(40) *He was not caught*
(41) **He got not caught* (Toyota 2008:115)

Another feature of the *get*-passive is that it does not accept the interrogative form. And the last distinction pointed out by Honda (2012) is the impossibility for the *get*-passive to delete the repetition of the second participle in a coordinate phrase. Examine (42) and (43).

(42) *He was caught and so was his friend.*
(43) **He got caught and so got his friend.* (Toyota 2008:115)

Whereas the *be*-passive does not require the repetition of the verb *caught after* the conjunction *and*, a structurally similar construction with the *get*-passive is ungrammatical.
These restrictions lead some scholars to consider *get* as a lexical (Honda 2012) or semi-lexical verb rather than a full auxiliary.

Third, the non-canonical passive has a special relation to the expression of the agent. Researchers (Carter & McCarthy 1999, Medina 2009, Svartvik 1966, Granger 1983) largely agree on the fact that, usually, the *get*-passive does not co-occur with a by-phrase mentioning an explicit agent. On that point, it resembles the *be*-passive since research has shown its preference to occur without prototypical agent. This lack of explicit agent allows to place greater focus on the event and on its consequences on the patient (Carter &McCarthy 1999). Nonetheless, in some cases, an agent might occur. In these cases, they are mainly either non-human (44), indefinite, impersonal or depersonalized (Carter & McCarthy 1999, Granger 1983). Carter & McCarthy (1999) add that where explicit agents are used, they are new information, considered as central for the understanding of the situation.

(44) *She's going to get eaten by the wolf.*
(45) *S1 You get intimidated by?*
*S2 The staff on the labour ward.*

(Carter & McCarthy 1999:51)

In (45), Speaker 1 is explicitly asking Speaker 2 to provide additional elements about the participant(s) involved in the event as he/she considers it as important information to maintain effective communication.

Fourth, unlike the *be*-passive, the *get*-passive is highly semantically marked since it mostly occurs in adversative contexts or in situations considered as problematic from the speaker's perspective (Carter &McCarthy 1999). The structure conveys the speaker's stance towards this adversative situation, in which the consequence of the event denoted can be perceived as fortunate or unfortunate for the patient (Siewierska 1984, Alexiadou 2005, Hatcher 1949, Carter & McCarthy 1999, Stein 1979, Matthews 1993). Chappell (1980) and Carter & McCarthy (1999) stress that the structure allows the expression of the speaker's view rather than the inherent properties of the event itself. Given this ability to express the speaker's attitude towards the involvement of the patient and the event, Carter & McCarthy (1999) qualify the *get*-passive as attitudinal marker.

In direct relation with the stance, *get*-passives typically co-occur with similar types of verbs, mostly with an unfortunate or problematic connotation. In their corpus-based study, Carther & McCarthy (1999) establish a list among which verbs as *get arrested, get killed, get critised* or *get penalised* can be found. However, it has to be noted, as highlighted by Sussex (1982) or Carter & MacCarthy (1999), that the semantic of the verb is not the decisive criterion in the selection of the *get*-passive, since the construction can also appear with non-adversative verbs such as *pay* (46) or *write* (47).

(46) *And she's got a contract. She's actually she didn't get paid for it, her (…).*
(47) *Most things got written up by scribes.*

(Examples from Carter & McCarthy 1999:51)

In (46), Carter and McCarthy (1999) report that the adversative attitude is, in this case, only rendered through the use of the negative particle, hence the importance to consider the co-text when analysing the *get*-passive. In (47), with no additional contextual information to guide the interpretation, the problematic stance usually reported in the literature seems to be absent and the focus rather lays on the central role played by scribes in the past. Carter & McCarthy report (1999:50) that, even though these cases with no obviously fortunate or unfortunate situations do not form the bulk of the occurrences, they do appear in small proportions.

A last property concerns the co-occurrence of the structure with adverbials. Carter & McCarthy (1999) note that very few get-passives are complemented with an adverb or an adverbial phrase. The authors assume that the use of such grammatical element would confer greater attention to the verb and, consequently, diminish to importance of the patient. Several studies (Alexiadou 2005, McIntyre 2005, Carter &McCarthy 1999) report on the restricted types of adverbs that *get*-passives can receive, mostly allowing to intensify or focus on the event and not affecting the whole clause. The category of adverbials used with *get*-passive rarely provides information on the manner, place or time denoted by the event. The specific adverb placement, i.e., excluding the medial position between the *get* and the past participial, constitutes another restriction imposed on the *get*-passive, which distinguish it from the *be*-passive (Toyota 2008, Carter &McCarthy 1999).

Carter &McCarthy (1999:46) summarise the *get*-passive using the following terms "the form is closely related to *be*-passive, with a different focus on agent, event and patient, and with some marking of attitude (…)", to which it can be added that *get*-passives mostly occurred in adversative or problematic circumstances.

Even though the *get*-passive possesses a wide range of restrictions, recent diachronic corpus-based studies have highlighted its increasing use. Since it is commonly referred to as a possible device for expressing passive meanings, it is worth including its study in the present investigation and bring further insights into its possible alternation with the canonical passive.

## 2.2. *Se*-constructions

### 2.2.1. Spanish *se*-constructions

Be it in Spanish or in French, the particle *se* is multifaceted and plays a key role in numerous constructions. In the present investigation, the focus will be put only on non-paradigmatic uses of *se*, which means that reflexive and reciprocal uses of *se* as well as its use as a pronominal and aspectual marker will be disregarded. Similarly, the use of *se* as a variant of *le*, required in some cases by the Spanish language, will not be included. Even though Spanish and French *se* share various characteristics, I will first start with a separate description of the Spanish construction since it is an extremely widely used mechanism and then compare it with the French equivalent, far less frequent.

#### *2.2.1.1. Origin of the structure, status of se and terminology*

Researchers largely agree that the *se*-construction developed from the reflexive morpheme *se* and, with time, various forms of the construction lost the reflexive character and adopted a passive meaning (Cabañas Maya 2006, Calvo Montoro 1983, Clements 2006, Laslop &Diaz 2010). The status of *se* has changed through time. From a fully reflexive marker, it acquired the status of an indicator of passiveness, with no explicit syntactic role in the sentence. The exact status of *se* is, however, questioned by several researchers. While some of them consider it as a mere particle with no syntactical function (Gómez Torrego 1984, Luján 1999), others view it as a marker of the passive (Calvo Montoro 1983, Gili Gaya 1964), of impersonality (Mendikoetxea 1999, Gómez Torrego 1992) or of middle voice (Pedersen 2013). Devís Márquez (2017) is the only scholar to attribute a specific role to the particle since he considers it as the unifying element between the different form of *se*-constructions. According to him, the function of the particle is to confer a different degree of indefiniteness and non-specificity to a clause, whose expression of the subject is impossible (2017).

In Spanish, various authors have already studied the *se*-construction, most of the time assigning it a different label (among others Martín Zorraquino 1979, Calvo Montoro, Gómez Torrego

1992, Molina Redondo 1997, Mendikoetxea 1999, Maldonado 1999, Devís Márquez 2003). No general consensus on terminology has emerged so far. To ease the reading, I will adopt the terms 'passive with *se'* (*pasiva con se*) and 'impersonal with *se'* (*impersonal con se*) following Campos (1989), Mendikoetxea (1999), or Dieck (2016)'s perspective (to only cite a few). It has to be mentioned, however, that some authors (e.g., Quesada 1997 or Molina Redondo 1997) reject the *impersonal* term. As it will be explained, the construction requires a human agent (mostly defocused) and, according to them, the passiveness only emerged from the specific reading made out of a middle voice construction.
In addition to adequate terminology, another issue concerns the way the construction is classified. As a detailed account of the different typologies suggested so far to classify the *se* construction is beyond the scope of this study, a summary of the main ones will be sketched.

#### *2.2.1.2. Classification*

Clearly defining and delimiting the Spanish *se*-constructions with a passive meaning is an arduous task since it can be used in multiple contexts and can undertake various forms. For the sake of this study I will consider that the *se*-construction consists in: SE + active form of 3rnd person + subject/object. The agreement between the subject and the verb is the guiding thread of the classification used in the present investigation. Following this criterion, two main types of constructions can be distinguished, namely the passive with *se* (also termed *pasiva refleja* 'reflexive passive' (Calvo Montoro 1983, Molina Redondo 1997, Gómez Torrego 1992) and the impersonal *se* passive. While the latter does not contain a subject agreeing with the verb, the former does. Let us first consider the passive with *se.*

##### Passive with *se*

Broadly speaking, the passive with *se* can take three different forms according to the nature of the noun phrase: inanimate object (a), indefinite animate object (b) and preposition governing a subordinate clause (c) (Gómez Torrego 1992, Fernández 2007). When the nominal phrase refers to a definite animate object and requires the preposition *a* to be grammatically correct or when there is no object at all, it is classified as impersonal with *se,* and will be described in the next section.

(48) se + transitive verb + inanimate direct object.

*a) Se divulgaron rumores sobre un nuevo encarcelamiento.*
'Rumors were disclosed about a new imprisonment.'
*a') Se dieron los premios a los niños.*
'The prizes were given to the children'
*a") Se vive la vida alegremente.*
'Life is lived happily.'

(49) se + transitive verb + animate direct object, mainly indefinite or generic
*b) Se necesitan ayudantes.*

'Assistants are needed.'

(50) se + transitive verb + subordinate clause
*c) Se cree que se suspenderán las negociones y que habrá nuevas elecciones*
'It is believed that negotiations will be suspended and that there will be new elections' (Mendikoetxea 1999: 1970, 1680)

The passive specificity of these types of *se*-constructions can be explained, from a formal point of view, by the notional object of a transitive verb acting as the grammatical subject of the sentence (Mendikoetxea 1999). A parallel phenomenon can be observed in periphrastic passives.

In Types (a) and (b) a transitive verb agrees with a direct object, either inanimate or animate. In case of animate object, the noun phrase is indefinite or generic (Mendikoetxa 1999, Fernández 2007, Molina Redondo 1997). Mendikoetxa (1999) reports that, when inanimate, there is no restriction regarding the thematic role of the subject (either agent, destinatary, source or experiencer). Ditransitive verbs, that is occurring with two complements, and verbs alternating between a transitive and an intransitive use, can also be used in passives with *se*. These forms are respectively illustrated in (a') and (a''). Some authors (see for example Molina Redondo (1997)) classify Type (b) with the impersonal with *se*.

In Type (c), consisting of *se*+ transitive verb+ subordinate clause, the second clause is considered as the subject. This type of construction is especially found in the journalistic productions (Mendikoetxa 1999). Before turning to the second main category of constructions with *se*, it is worth adding that in passives with *se*, the grammatical subject is placed after the verb and the construction does not suffer aspectual restrictions. Mendikoetxa (1999) considers these two characteristics as reasons explaining the high frequency of occurrence of passives with *se* compared to other types of *se*-constructions.

### Impersonal with se

Impersonal constructions with *se*, also termed *impersonal refleja* 'impersonal reflexive' (Calvo Montoro 1983), are the second main type of *se*-constructions. Gómez Torrego (1992) considers these types of *se*-constructions as impersonal both from a syntactic and a semantic perspective. They are considered as distinct from the passive with *se* since the notional object is also the grammatical object and not the grammatical subject (Mendikoetxea 1999).

(51) a) se + transitive verb + A + animate direct object
*Se recibió con aplausos a los jugadores*
'The players were received with applause'

(52) b) se + intransitive verb
*Se vive bien en España*
'People live well in Spain'

(Examples from Gómez Torrego 1992)

The distinctive criterion is mainly of syntactic rather than semantic nature as impersonals with *se* comply with the agent defocussing process (Molina Redondo 1992). In this form of constructions with *se*, the nominal phrase is preceded by the preposition *a* and does not possess any explicit grammatical subject as was the case for the passive with *se* where the notional object is promoted to grammatical subject position (Molina Redondo 1992, Dieck 2016). The use of the preposition facilitates the distinction of impersonals with *se* from full middle constructions with a reflexive and reciprocal meanings (Molina Redondo 1992:21). Besides its use with transitive verbs, the impersonal with *se* can also occur with intransitive verbs as illustrated in (b). These uses are reported as having ambiguous readings as they can alternate between a passive and an active one (Campos 1989).

Nevertheless, researchers reckon that the limit between passive and impersonal with *se* might, in some cases, be blurred (Maldonado 1992, Mendikoetxea 1999) as all these different forms of *se* share various central features which can be summarised as follows (Quesada 1997, Molina Redondo 1992):

- The verb does not have a true grammatical subject
- The structure possesses an active morphology (and syntax)
- The defocussing process allows to place the focus on the patient
- Though defocused, the verb expresses an event requiring a human agent
- The construction conveys generic, indefinite or collective nuance

Instead of a division between the two types of constructions, Maldonado (1992:277) rather suggests a gradual continuum within which the contrast would lie in the importance given to the thematic subject and the defocused agent.

A largely polemical issue regarding the syntax of the *se*-construction is the one of the agreement between the verb and the notional object such as in (58) where the verb *vende* 'sell' does not agree with the plural object *libros* 'books'.

(53) Se vende libros
(54) Se venden libros

(Mendikoetxea 1999: 1677)

Without entering in too much details, there are two positions. The more traditional and prescriptive view considers the absence of concordance as ungrammatical and rejects it. According to the Real Academia Española, this variant of the construction would have emerged from a combination of the impersonal and the passive with *se*. The supporters of a rather descriptive approach classify the phenomenon as a variant of the passive with *se* (Toscano Mateus 1953, Mendikoetxea 1999). The absence of concordance is attested for a long time and relatively frequent, more specifically in oral productions (Lope Blanch 1983, Gili Gaya 1964, Mendikoetxea 1999). This increasing use is observed both in peninsular and non-peninsular Spanish and is, thus, generally not considered as a regional feature (Toscano Mateus 1953, De Mello 1995). Even though widely attested, the variant with agreement between the verb and the

noun phrase still largely predominates. This study adopts the descriptivist's view of the phenomenon and includes the possible occurrences of *se*-constructions with no agreement within the category of passive with *se.*

### Middle passive

To these two main forms of the *se*-constructions, a third one, though controversial, can be added: the *middle* constructions which can be divided into 'middle passive' (55) and 'middle impersonal passive' (57) (see for example Laslop & Díaz 2010 or Mendikoetxea 1999 on the terminology).

(55) *La gripe se cura con estas pastillas*
'The flu is cured with these pills'
(56) *Esta camisa se lava muy bien*
'This shirt can be easily washed'
(57) *A estos niños se les asusta fácilmente*
'These children can be easily scared'

Middle passives are limited to transitive verbs and the notional object is the grammatical subject. Conversely, in middle impersonal passives, the notional object is the grammatical object. Middle passives share the following four characteristics: (i) they are used with the particle *se*, (ii) they contain a grammatical inanimate subject, (iii) they are constructed with a transitive verb.

While in the passive with *se* and the impersonal with *se,* the expression of the agent was either rare or syntactically blocked, it can be explicit with middle passives and impersonal passives. In Example (55), the noun phrase *estas pastillas* 'these pills' is the expressed agent. A second main difference is the nature of the agent (Molina Redondo 1997). Whereas the passive with *se* and the impersonal with *se* require a human agent, though defocused, middle constructions allow inanimate objects with an agent-like role as illustrated in (55) with *estas pastillas*. Various researchers report that these types of clauses express an intrinsic property of the grammatical subject (Dieck 2016, Mendikoetxea 1999, Molina Redondo 1997). This assumption is, however, questioned by Devís Márquez (2017) since he emphasizes that it is rather the reading adopted than the property of the subject itself that allows to distinguish these structures from the impersonals with *se*.

Middle constructions with *se* can only be used in the imperfect aspect and may, in some cases, require an additional element to activate the passive/impersonal readings (Mendikoetxea 1999:1667). In Example (56) and (57) this role is undertaken by the adverbs *bien* 'good' and *fácilmente* 'easily'.

Since these structures share a series of properties with the structures classified as impersonal (Mendikoetxea 1999), they will be included in the analysis.

In order to maintain this study to a manageable scope, I have considered the aforementioned types of *se*-constructions as a continuum rather than as discrete entities. I have therefore

analysed them as a whole rather than separately. Investigating the different types of *se* would constitute an interesting avenue for further research and complement the results of the present work.

### Presence of an explicit agent and productivity of the construction

The presence of an explicit agent within the *se*-construction is rare (Gómez Torrego 1992, Cabañas Maya 2006, Calvo Montoro 1983, Fernández 2007) and is usually explained by the pragmatic function of the structure, that is defocussing the agentive force and a series of grammatical restrictions. However, *se*-constructions containing an overt agent introduced by a *por*-phrase do exist as illustrated in (58).

(58) Se están dictando cursos por profesoras especializadas
*'Courses are being given by specialized professors.'*

(Example from Rauch 2018:21)

Such a use is attested in several contexts and seems to be more common in the written mode for legal and administrative purposes as well as in newspapers (Martín Zorraquino 1979, Gómez Torrego 1992, Mendikoetxea 1999). Gómez Torrego (1992) mentions a relative reluctancy from the defenders of the norm to consider the use of an agent complement as grammatically correct, more particularly for impersonal constructions with *se.* Mendikoetxea (1999) reports that so far there is no clear agreement among grammarians on the (un)grammaticality of the use of an explicit agent.

It seems largely agreed that *se*-constructions are the most frequent device to defocus the agent in the Spanish language (Calvo Montoro 1983, Quesada 1999, Mendikoetxea 1999, Cabañas Maya 2006, Laslop & Diaz 2010) and supplant other impersonal mechanisms such as *uno* or the periphrastic passive. The high versatility of the structure may be a factor contributing to its particularly productive use. Even though there is a general tendency to link the constructions with *se* to more formal and distancing genres (Jisa &Tolchinsky 2009, De Cock 2020), the strategy is also largely present in informal spoken conversations (Quesada 1999, De Cock 2020). Dieck (2016) points to the lack of solid corpus-based study to empirically describe the preferred modes and registers of the constructions.

### Interpretation of *se*

The use of the construction with *se* can serve several purposes. It can be used in a purpose of generalization (59) where the utterance can be understood as all Spanish people have the habit to go to church on Sundays or of conveying indetermination such as in (60) where a specific indeterminate person killed the immigrant (Devís Márquez 2003:435).

(59) *En España se solía ir a misa los domingos.*
'In Spain they were used to go to church on Sundays'.
(60) *El periódico dice que se ha asesinado a un inmigrante.*
'The newspaper says that an immigrant has been killed'.

(Devís Márquez 2003: 435)

Though classified as an impersonal mechanism, the construction has the possibility of being interpreted as including the speaker, the hearer, or exclude both of them (Blanca & Lourdes 2005). It can also endorse corporative reading, that is including a socially definite group of persons (Suarez 1996). All these strategies may serve different pragmatic purposes such as distancing or politeness (Dieck 2016, De Cock 2020)

### 2.2.2. French *se*-constructions

The second part of this section will be devoted to the description of the *se*-construction in French and will be followed by a comparison with its Spanish counterpart and the canonical passive.

Termed *passif moyen* 'middle passive' (Gaatone1998), *se-moyen '*middle *se' (*Zribi-Hertz 1982) or *passif pronominal* 'pronominal passive' (Lamiroy 1993), the French *se* construction has not always been considered as included in the domain of the passive (Gaatone 1998). However, in some contexts, the *se* construction (61) adopts a similar defocussing function as the periphrastic passive (62) (Gaatone 1998, Zribi-Hertz 1982).

(61) *Le français se parle dans de nombreux pays.*
'French is spoken in numerous countries.'

(62) *Le français est parlé dans de nombreux pays.*
'French is spoken in numerous countries.'

(Gaatone 1998: 241)

In Examples (61) and (62), both mechanisms allow to place the focus on the event in defocussing the agent. The *se*-construction displays the following features:

- *se*-construction reject the presence of an agent complement (Weiner & Labov 1983, Lamiroy 1993, Gaatone 1998, Zribi-Hertz 1982, Lagae 2002)
- defocused agent is necessarily human (Ruwet 1972, Lagae 2002, Zribi-Hertz 1982)
- restricted to transitive verbs (Lagae 2002, Zribi-Hertz 1982)
- the subject of the middle-*se* is usually inanimate (Lagae 2002)
- usually more common with the imperfective aspect (Lagae 2002)

The passive constructions in *se faire*, illustrated in (63a), can have a similar reading to the periphrastic passive (63b) (see Le Bellec 2015). In some contexts, *se faire* competes, therefore, with the canonical passive. However, no occurrence of *se faire* was found in our data.

(63) a. Je me suis fait agresser
'I got attacked'

b. J'ai été agréssé
'I was attacked' (Example from Le Bellec 2015: 3)

Even though the French and the Spanish constructions with *se* share several features, they also differ in various aspects. First, its uses are much more limited in French since it is impossible to distinguish three acceptations as was developed in the previous section for Spanish. Lamiroy (1993:67) distinguishes three major differences between the two languages. She mentions that the French use of *se* seems to appear mainly with inanimate nominal clauses in the third person while with the Spanish *se* human nominal clauses would be accepted. A second fundamental distinction (1993:68) is the limited use of the French *se* to direct transitive verbs while the Spanish *se* also accepts indirect transitive and intransitive verbs. A last point highlighted by Lamiroy (1993:69) is related to the aspectual domain. In most cases, French does not allow the use of *se* with perfect tenses while Spanish does. While Spanish prefers the pronominal construction rather than the periphrastic passive, French seems to favor the man-impersonal *on* (Lamiroy 1993).

The following five main points allow to distinguish the canonical passive from the pronominal passive (Lamiroy 2000: 138): (i) the passive operator, (ii) the aspect, (iii) type of subject, (iv) agent complement and (v) type of verb. The periphrastic passive and the passive construction with *se* are formally and functionally different constructions (Lamiroy 1993, Lagae 2002) but they also share at least the following three features. Passivization is carried out through a 'passive operator' (1993:64) be it a state operator or a reflexive one. She adds that both constructions allow to diminish the number of arguments. And, both in the periphrastic passive and the *se* impersonal construction, an agent is semantically implicated in the process, be it formally expressed or not. It can thus be observed that the periphrastic passive and the *se*-constructions display a number of similarities but also differ on a wide range of features. This contrastive work strives to provide further details on the different distinctive behaviour of these two structures.

## 2.3. R-impersonals

This section describes two structures considered as R-impersonals, namely the non-anaphoric third person plural and the man-impersonal *on*. Siewierska defines the term *R-impersonal*, which applies to human-impersonal pronouns, as "impersonals triggered by a reduction in referentiality" (2011: 57).

### 2.3.1. Non-anaphoric third person plurals

The non-anaphoric third person plural is a widely acknowledged morphosyntactic resource for agent defocussing (Malchukov & Siewierska, 2011; Aijón Oliva, 2020; Siewierska & Papastathi, 2011; Siewierska, 2010; Weiner & Labov, 1983). This type of construction, which is attested cross-linguistically, is considered by Siewierska "to be impersonal under the functional, agent defocussing view of impersonality which associates defocussing of an agent with loss of subject status and/or lack of full referentiality" (2010: 74).

From a formal perspective, this impersonal mechanism is constructed with a verb agreeing with a third person plural subject pronoun. The subject can be overtly realized, as it is the case in

English and French, or realized through pro-dropping as the Spanish language does, since, in pro-drop languages, the formal expression of the clause subject is not necessary, the subject being encoded in the verbal morphology. The presence of an overt pronoun is generally considered as blocking the reduction in referentiality. As demonstrated by Lapidus & Otheguy (2005:157) a nonspecific *ellos* 'they' is possible in some very specific contexts such as in Spanish of Latin America spoken by individuals whose exposure to English is intensive.

Non-anaphoric third person plurals can be distinguished from their anaphoric counterparts by the fact that they lack an overt antecedent in the clause, the antecedent being specific or not. Even though cases of third person plurals with unspecific referents (type *everyone* or *someone* such as in example (64)) can be considered impersonal to some extent, they are excluded from the study. As showed in (64), *they* functions as an anaphor of the pronoun *someone*.

(64) *Harry walked round and round his new room. **Someone** knew he had moved out of his cupboard and **they** seemed to know he hadn't received his first letter.*

(65) *If a student comes, **they**'ll have to wait.*

(Examples from Siewierska & Papastathi 2011: 579)

Cases of epicene third person plurals, which are, as explained by Siewierska & Papastathi (2011), third person plurals used anaphorically to refer to an entity which may be of either gender (65), are also discarded since they cannot be considered as impersonal.

From a semantic perspective, the impersonal construction refers to a non-referential human subject, be it specific or collective within which the speaker and the addressee are typically excluded from[10] (e.g., Creissels 2008, Cabredo Hofherr 2003, Siewierska & Papastathi 2011, Malamud 2013), echoing the third person initial nature of the structure. The exclusion of such categories of participants is a salient semantic property allowing to differentiate the impersonal third person plural from other impersonal strategies.

When the third person plural is used impersonally, the degree of referentiality of the subject pronoun weakens, ranging from specific to universal. As pointed out by some scholars (see for example Siewierska 2011), few typological studies have been designed to intend classifying the interpretation of impersonal third person plurals (Aijón Oliva 2020 on Spanish, Siewierska & Papastathi 2011 crosslinguistically, Cabredo Hofherr 2004 on a sample of European languages, Casielles 1996 on arbitrary subjects). Various researchers (see e.g., Alonso-Ovalle 2000, Malamud 2004, Cabredo Hofherr 2004, Siewierska 2011) suggest to group the different acceptations of this impersonal mechanism under three categories: generic (66), episodic (67) (also termed arbitrary or existential (see e.g., Siewierska 2011), much more anchored in time than the generic uses, and specific (68), "involving some sort of contact of the speaker with the referent of the subject, for example, visual (at the door, on TV), […] audio (on the phone, radio), written (via an e-mail or letter) Siewierska (2011:65). "

(66) *En Italie **ils savent** préparer les pâtes*

[10] Note that, in some specific situations, the inclusion of the speaker may be considered as the decision to include or exclude the speaker is rather a matter of perspective than reference (Malamud 2013:5).

'In Italy they know how to prepare pasta'

(67) *Ils **ont retrouvé** une moto dans la cour*
'They have found a motorbike in the courtyard'

(68) ***Tocan** a la puerta. Quieren entrar.*
'They're knocking at the door. They want to come in.'

(Siewierska 2011:67,73)

As it will be further detailed in the next section, this tripartite typology has also been suggested as a possible mean to classify the French man-impersonal *on*.

The factors conditioning the referentiality can be of varied guise such as noun phrases, verbal phrases, space or time builders (Alonso-Ovalle 2000, Casielles 1993, Malamud 2013 to only mention a few). Siewierska and Papastathi (2011) sum up the different factors and identify five main criteria which guided the classification of the different types of impersonal *they*:

a. "group as opposed to individual identity of the referent"
b. "specificity of the referent"
c. "specific time reference and anchoring in time"
d. "inclusion of the speaker among the range of potential referents"
e. "inferential nature of the denoted event"

Siewierska and Papastathi (2011:581)

Siewierska and Papastathi (2011), elaborating on the typology established by Cabredo Hofherr (2006), further develop this classification and refine it by distinguishing six types of interpretations, from the most specific to the most universal:

a. Specific existential
*They're knocking on the door.*
b. Inferential
*They've been frying chips here.*
c. Vague
*They've found his bike in the back of a barn.*
d. Corporate
*They changed the tax laws last year.*
e. Universal
*In Spain, they eat late.*
f. Use with the speech act verb *say*
*They say he met vampires in the Black Forest.*

(Examples from Siewierska and Papastathi (2011: 581-585)

Universal and corporate uses (respectively (d) and (e)) both refer to a collective, the only difference between the two lying in the more or less restricted extent of the group identified. Regarding universal uses, the identification is usually achieved through the use of locative expressions such as the prepositional phrase *In Spain* in (d). By contrast, corporate interpretations are delimited by the verbal clause and its arguments such as *changed the tax*

*laws* (e). Changing the tax laws falls to a certain group of persons, specialized in that activity. Vague uses of non-anaphoric third person plurals refer to a set of unknown or irrelevant persons. In (c) someone found a bike but the specification of the person's identity is either not relevant or unknown by the speaker at the moment of the speech act. In the case of inferential uses of *they*, "the event in question is not known to have taken place, but is merely inferred from (visual, acoustic, etc.) evidence" (Gast & van der Auwera 2013:140). Example (b) is an illustration of this category, where it is inferred from the smell that a certain activity involving frying food took place. The last category of non-anaphoric third person plurals gathers the specific existential uses of the construction which includes a situationally known and/or specific entity such as in (a) where a specific and physically present referent is involved.

Including such impersonal devices as the non-anaphoric third person plural construction in this study is crucial in establishing a contrastive panorama of passive-like construction since several researchers note that the non-anaphoric third person plural is often used as a substitute or alternatives to the periphrastic passive in spoken registers (see e.g., Weiner & Labov 1973 or Siewierska 2010). Excerpt (69) illustrates such a use of the third person plural, in an oral exchange between two participants.

(69) *Well, there you are, boy. Platform nine — platform ten. Your platform should be somewhere in the middle, but they don't seem to have built it yet, do they?*

(Siewierska 2010:75)

Since they are both impersonal strategies based on a reduction in referentiality, non-anaphoric third person plurals and man-impersonals may frequently function as translational equivalent (see Jisa & Viguié 2005 or Siewierska 2011). Nevertheless, as already briefly mentioned above and as will be further developed, they cannot be considered as equivalent strategies and do not function in the same situational contexts. Siewierska (2011:86) adds that the non-anaphoric third person plural regularly competes with active impersonals based on a reflexive such as the *se*-construction.

As non-anaphoric third person plurals and man-impersonal are often considered as competing constructions, I will now turn to the description of the man-impersonal *on* in French.

### 2.3.2. Man-impersonal *on*

The French pronoun *on* belongs to the so-called man-impersonal constructions, widely attested among a variety of European languages (see for example Siewierska (2011) crosslinguistically, Giacalone Ramat and Sansò (2007) on European languages or Coussé and van der Auwera (2012) on Swedish *man* and Dutch *men* only to cite a few). A typical man-impersonal construction consists of a non-referential human subject realized by a word originating from 'human' or 'man'. The French pronoun *on*, etymologically related to the Latin word *homo* (Siewierska 2011: 60), is known to have a polysemous status (Schapira 2006, Fløttum et al. 2007, Landragin and Tanguy 2014, Yao 2016), even leading Fløttum et al. to consider the mechanism as a semantic paradox (2007:7). Various scholars have already tackled the issue of

classification and suggested different typologies for the interpretations of *on* (Koenig 1999, Fløttum et al. 2007, Cabredo Hofherr 2008, Landragin and Tanguy 2014, Yao 2016).

Working crosslinguistically, some scholars, among which Cabredo Hohherr (2004) and Siewierska (2011), set out the basis of a tripartite classification of man-impersonals: generic, episodic and specific (presented in 2.3.1). Broadly speaking, the generic usage is realized when "the referent of the subject corresponds to anybody or some unspecified or loosely specified group of individuals" (Siewierska 2011:62). With a generic acceptation, man-impersonals can occur in all types of non-assertive, irrealis situations.
When considered as episodic, the constructions "denote partially specified sets of human subjects or even individuals (Siewierska 2011: 59)." Compared with the generic usage, episodic man-impersonals are more frequently anchored in time and/or space, which allows to considerably reduce the potential set of referents included in the construction. Typically, but not exclusively, episodic uses are constructed with a perfective tense (Siewierska 2011:63). Siewierska (2011) further distinguishes vague from corporate episodic uses of man-impersonals. While the referents of the former would be deducible from the linguistic context, the later allows to identify the "group to which the referents of the subject belong but not the identity of the individuals in question" (Siewierska 2011: 64). The author exemplifies the difference between the two types of episodic uses with the following predicates: *raising blinds* and *raising taxes*. While the first noun phrase doesn't bring to mind any specific group of individuals, the second one automatically implies a specific entity, for example the government. Besides generic and episodic uses, man-impersonals can also be used specifically, which allows to interpret the referent of the construction as specific and definite. Specific uses typically refer to a first person be it singular or plural. Nevertheless, as shown by Boutet in (70), the substitution of another referent is not excluded. In (70), the French man-impersonal *on* refers to a third person plural, the presence of the pronoun *me* excluding the speaker from the interpretation.

(70) *On me reproche mon caractère*
'I am criticized for my character'

(Boutet 1986 :34)

Focusing on the French man-impersonal *on*, Fløttum et al. (2007) suggest a slightly different referential typology than the tripartite one previously described and summarise the different acceptations of *on* in two categories: indefinite or personal such as indicated in the summary table (Table 2).

| **Indéfini** | | | **personnel** | |
|---|---|---|---|---|
| **impersonnel** | générique | spécifique | neutre | stylistique |
| ***Il, ce*** | *Tous les hommes, chacun* | *Ils, quelqu'un* | *nous* | *Je, tu, vous, il elle, ils, elles* |

*Table 2. Classification of the pronoun on from Fløttum et al. (2007)*

In its indefinite use *on* can be interpreted as impersonal (71), generic (72) or specific (73).

(71) *On était le 15 octobre*
'It was October 15th.'
(72) *On peut toujours rêver*
'We can always dream.'
(73) *On frappe à la porte*
'Somebody is knocking on the door.'

(Fløttum et al. 2007:30)

When used as indefinite, the pronoun has lost its regular pronominal properties and does not refer to any other discourse element in the context. Koening qualifies this phenomenon 'discourse inertness' (1999:243), which echoes the behaviour of other impersonal mechanisms such as the periphrastic passives. Within this acceptation, *on* is even considered as having undergone a grammaticalization process and transformed into an impersonal marker (Creissel 2008:13). In a similar vein, *on* can appear in fixed expressions such 'on aurait dit' within which the construction lost referentiality (Landragin and Tanguy 2014). In this particular context, *on* represents a source of information or a figure of authority (Borra 2010 and Fløttum et al. 2007). Gaatone notes that indefinite *on* constructions and periphrastic passives frequently have (partially) equivalent meanings and can therefore be used as alternatives (1998:243). This characteristic allows to further strengthen its legitimacy to appear within the group of so-called passive-like structures. Other scholars (e.g., Landragin and Tanguy 2014 or Boutet 1986) even questioned the referential status of *on*, emphasizing its importance in the construction of the discourse and considering it as a marker of cohesion rather than a referential phenomenon.

In its personal use, the referents of *on* are more or less definite and identified or may be identifiable such as in examples (74) and (75) where the construction alternately refers to the speaker and to the addressee.

(74) ***On** fait aller, **on** fait ce que **on** peut (pour je) / Alors, **on** a bien dormi ? **On** est contente (pour tu)*
*'I do what I can (for I) / So, did you sleep well? You're happy (for you)'*

(75) *Tante, **on** va dormir maintenant.*
'Aunt, **we** are going to sleep now.'

(Fløttum et al. 2007: 31)

As seen from above, the man-impersonal *on* may refer to a wide range of human agents. It can substitute one individual (the speaker, the addressee or an outsider specific person), a set of individuals consisting in the speaker and the addressee(s), a group of persons, displaying various degrees of definiteness, within which the speaker is either included or not or be used in fixed expression without referentiality. Fløttum et al. (2007) highlight that, without any clues indicating the contrary, the interpretation includes the speaker. In generic contexts, speaker and

addressee are largely considered as potentially included (Siewierska 2011), even though in some specific cases, these referents may be excluded[11].

As it has been pointed out, *on* can substitute a broad range of referents, leading Siewierska (2011) to consider this feature as a potential factor favouring a higher frequency of occurrence of the phenomenon.

From a morpho-syntactic perspective, *on* is limited to subject position, always followed by a singular verb. Note that in a restricted number of languages such as Icelandic (Siewierska 2011), the man-impersonal construction can appear in another position. In French, the construction is, however, restricted to subject position. While the verb always agrees with the singular subject, the adjectives and past participles potentially co-occuring with the man-impersonal construction do not, however, always agree in the singular form and can, therefore, be used as clues to identify the referents (Borra 2010; Landragin and Tanguy 2014). Identical in appearance to its personal counterpart, the only feature that distinguishes personal constructions from impersonal ones, is the reduction in referentiality in the latter type.

As previously highlighted in this study, the linguistic context in which the French man-impersonal *on* occurs, is of crucial importance in the identification of the potential referents (Landragin and Tanguy 2014, Gjesdal 2008, Yao 2016, Boutet 1986, Fløttum 2007). A series of contextual factors can guide the interpretation and further lead to the inclusion or exclusion of sets of participants:

a. Presence of a pronoun
   *On me reproche mon caractère* (Boutet 1986 :34)
   'I am criticized for my character'
b. agreement of the attribute
   *On est toutes très fatiguées* (Fløttum 2007 :35)
c. Verb tenses (some tenses may trigger a particular interpretation. For instance, the present and past conditional tenses in French, describing irreal events, tend to favour a generic interpretation)
d. Modal verbs tend to favour a generic interpretation over a specific one.
e. Type of clause (Fløttum 2007 :44),
   *Elle semait partout des souvenirs comme on jette des graines en terre.*
   'She was sowing memories everywhere, like throwing seeds in the ground.'
   The comparative subordinate used with the conjunction *comme* in the example tends to favour generic or non specific readings
f. Presence of a specific complement restricting the possibility of intepretation (Fløttum 2007:43)
   *On me dit d'embrasser le bébé*
   'I am told to kiss the baby' (Fløttum 2007:43)
g. Presence of spatio-temporal space builders (Boutet 1986, Siewierska 2011)
   *à Paris on était quand même payés raisonnablement*

[11] For an ample discussion of these cases, see Siewierska (2011: 63).

‘In Paris we were still paid reasonably’ (Boutet 1986: 42)

The use of these grammatical elements in the context are clues that may guide the interpretation towards a certain category of *on* and their role may differ according to the situation.

### Impersonal alternatives in Spanish & English

The French man-impersonal *on* has no form-based equivalent construction in the other languages studied in this investigation. Working in Spanish, Yao (2016) suggest that, depending on the morphosyntactic and pragmatic intentions of the speaker, several mechanisms are available to fulfil the role of man-impersonals such as *uno*, the non-anaphoric third person plural and the *se*-construction. Siewierska (2011) also points out that French *on* is sometimes used as a translation for the *get*-passive such as in excerpt (76) from Siewierska (2011:60), (a) being the original source in English and (b) the equivalent excerpt translated in French. It can be observed that the impersonal nuanced conveyed by the *get*-passive in English is rendered by a *on* in French.

(76) a) *Red card!” said Dean furiously. “In soccer you get shown the red card and you’re out of the game!*
*b)On n’est pas au football, l’interrompit Ron. On ne peut pas renvoyer les joueurs*.
‘We are not at a football game, interrupted Ron. We/one cannot send away the players (here)’.

(Example from Siewierska 2011:60)

Siewierska (2011:61) also underlines that, beside its possible alternative with the *get*-passive, there is a considerable “degree of overlap” in the use of man-impersonals and non-anaphoric third person plurals. As some scholars have shown, these two impersonal strategies can be grouped into three main types (generic, episodic and specific) (cf. Siewierska 2011; Alonso-Ovalle 2000, Malamud 2005; Cabredo Hofherr 2006). The present work intends to determine to which extent these strategies share common features.

### Man-impersonal in modes and genres

It is assumed that the use of *on* is considerably influenced by the mode and the genre of the discourse (e.g., Siewierska 2011, Fløttum et al. 2004, Coveney 2000).

The construction *on* may display a preference regarding the type of verb it appears with. In scientific articles, Fløttum et al. (2004:103) report that the construction is preferentially used with the following category of verbs: discursive (*décrire* ‘to describe’, *illustrer* ‘to illustrate’, *discuter* ‘to discuss’), cognitive (*comprendre* ‘to understand’, *observer* ‘to observe’), communicative, attitudinal (*affirmer* ‘to state’, *objecter* ‘to object’), perceptive (*entendre* ‘to hear’, *voir* ‘to see’) and verbs related to the research area (*analyser* ‘to analyse’, *comparer* ‘to compare’). As far as referentiality is concerned, in scientific articles, the *on* mainly substitutes three sets of referents (Fløttum et al. 2004:105), presented here by frequency of use: (i) the

author and a restricted discursive community, which would correspond to a first person singular or plural and the fellow colleagues of the author(s); (ii) the author(s) and (iii) the author(s) and the reader(s).

The use of *on* in writing is widely attested especially in generic contexts or when it refers to the first person plural in more definite contexts (Siewierska 2011). In academic writing and literature, that is more formal genres, Grevisse (1994) points out that *on* is a stylistic device that may substitute a personal pronoun to express modesty, politeness or criticism. In informal genres, *on* frequently competes with other personal pronouns, especially the first person plural, without endorsing a particular stylistic flavour (Yao 2016). In informal spoken language, it has been observed that the use of the first person plural has decreased over time, being substituted by *on* (Coveney 2000). When in spoken generic contexts, French would rather opt for a generic second person singular than a man-impersonal construction (Fonesca-Greber & Waugh 2003, Siewierska 2011). Unfortunately, the study of this relationship falls outside the scope of the present work.

Man-impersonals are often considered as equivalent forms to numeral-based impersonal strategies. However, as will be explained in the next section, even though in some situations the two types of constructions can be used in similar contexts, each strategy possesses its own specific features.

## 2.4. Numeral-based constructions: *uno* and *one*

In what follows I will sketch the main characteristics of the last impersonal type of device included in this study that is, the one based on a numeral. This option is available in Spanish with *uno* and in English with *one*.

### 2.4.1. Construction with *uno*

*Uno* originates from the Latin indefinite article *unus*, which through the action of time has acquired impersonal meaning, attested since approximately the fifteenth century (Ridruejo 1981, Pozas Loyo 2010). *Uno* can only be considered as a defocussing mechanism on the semantic level since, syntactically, the pronoun occupies, among others, the subject position and always involved a human agentive force. This defocussing process is thus realized by "under-elaboration, e.g., by designating a class (rather than an individual), by establishing a vague reference" (De Cock 2020:2).

When used impersonally, various researchers have already suggested a taxonomy of its different interpretations (Hollaender Jensen 2002, Gelabert-Desnoyer 2008, Rasson 2017), which indicates that the pronoun can display substantial variation in its degree of specificity and can serve a wide range of functions. The use of *uno* can involve three types of participants

with a varying degree of implication: the speaker, the hearer(s) and any external person, absent from the communicative situation.

Several readings can be made and can be placed on a continuum ranging from a specific reading to a generic one. Traditionally, *uno* is associated to the 1st person singular form (Gómez Torrego 1992, Mendikoetxea 1999), thus referring to the speaker (77). Within that acceptation, two common functions of *uno* are to distanciate the speaker from what he/she says or to resort to a greater level of politeness since it allows him/her to express his/her aims and intentions in a less direct style (Dieck 2016).

(77) *Qué bien vives, tío – Es que uno sabe montárselo bien*
'What a good life you live, man - You know how to make a good life'

(78) *Cuando uno mira al sol, se queda como ciego*
'When you look at the sun, you're like a blind man'

(Gómez Torrego 1992: 15)

On the opposite extremity of the continuum, *uno* endorses a generic function and refers to 'everybody', as illustrated in (78) with a general statement applying to each and every one.

Between these two extremities ranging from the most specific to the most generic, *uno* can occupy a series of intermediate positions. Rasson (2017) distinguishes a total of nine categories related either to genericity (full or reduced genericity), specificity (linked to the speaker, the addressee or a third person), intermediated positions (linked to the speaker, the addressee or a third person), and ambiguous reference[12].

To clearly distinguish between the different interpretations of *uno* is a complex task. Since, when used as a 1st person singular, the speaker confers, to some extent, a sense of generality to his/her words which contributes to blur the distinction of the different acceptations of *uno* (Company Company & Pozas Loyo 2009). To decide which reading is the most appropriate, it is worthy to consider the linguistic context, which, most of the time, contains clues to guide the interpretation (Lavandera 1984, Rasson 2017, De Cock 2020).

Even though relatively frequent in subject position, *uno* can also be used as a direct or indirect object and in prepositional phrases illustrated in examples (79) to (81), respectively.

(79) *Cuando uno comienza anda uno dándole mucha lata a los medios, para que los medios lo ayuden a uno [Reforma, 25-01-2006]*
'When one begins (one's career), one is always begging the media for help'

(80) *A uno siempre le gusta descubrir qué secretos, qué costumbres se traen de sus tierras [Casa verde, 308]*
'One always likes to discover what secrets, what habits they bring from their lands'

[12] For an in-depth description of the interpretations of *uno* see Rasson 2017.

(81) *Nieves asintió: no había como la tierra de uno [Casa verde, 180]*
'Nieves nodded: there was nothing like one's land'

(Examples and translations from Pozas Loyo 2010)

A last point that is worth mentioning is the fact that, though rare (De Cock 2014), *uno* possesses a feminine equivalent, *una*. This form of the numeral based construction requires two conditions to apply: (i) the speaker has to be a woman and (ii) the statement in which it occurs has to involve only women (Dieck 2016). Nevertheless, the masculine form *uno* is also frequently used to refer to female referents. For a more detailed understanding of the alternation between *uno* and *una* see for instance De Cock (2014).

#### Frequency, genres and context of use

It is widely accepted that *uno* is a relatively infrequent phenomenon. When occurring, De Cock (2020) found a correlation with the spoken genre. Numerous authors point out its extremely low frequency compared to other impersonal mechanisms, especially the impersonal form of the second person singular (Lorenzo 1989, Blanco Canales 2004, Pozas Loyo 2010, Guirado 2011). *Uno* frequently occurs with pronominal and reflexives verbs (Ridruejo 1981, Pozas Loyo 2010). This pattern of use is, however, expected since the presence of the clitic *se* blocks the use of an impersonal construction with *se* since a double *se* would produce an ungrammatical sentence.

### 2.4.2. Construction with *one*

The English language also possesses an impersonal strategy based on a numeral, that is *one.* Similarly to *uno*, the pronoun *one* can appear in different syntactic positions, that is subject or object complement. It differs from the Spanish construction on the fact that *uno* can take on possessive and reflexive forms (one's, oneself) (Altenberg 2005:98), which is not possible in Spanish.

As explained above, the interpretations of *uno* can vary from generic to specific with a wide range of intermediate positions. *One* can also have different degrees of specificity (Wales 1996, Altenberg 2005) ranging from generic to contextually restricted. However, *one* is generally considered as a "first-person oriented generic pronoun" (Moltmann 2010:440), that is simultaneously containing features of the speaker's perspective and of genericity as illustrated in (82). On the basis of his own experience, that is getting sick after eating mushrooms, John infers that other people than him may also get sick after eating mushrooms and, thus extend his experience to a generic statement.

(82) *John found out that one gets sick when one eats these mushrooms.*

(Moltmann 2010:448)

In line with Moltmann (2010), Malamud explains that "(…) every use of *one* is a generalisation drawn either on the basis of the speaker's own experience (…) or via the speaker taking the perspective of (simulating) each person in a relevant set" (2012:9). In addition to the special link to the first person singular, a particular connection to conversational participants is also reported (Safir 2004, Moltmann 2006), with a tendency to include them in the reference.

The pronoun *one*, though considered as generic, is however restricted in its context of use since it cannot appear in episodical context or with existential meaning such as in (83) (Achard 2010, Fenger 2018).

(83) *One has called for you, but I don't know what it was about*

(Example from Fenger 2017:298)

As was the case for its Spanish equivalent, clearly distinguishing between the different acceptations of *uno* is challenging, hence the importance of the linguistic context.

Concerning genre, Rasson also specifies that even though *uno* and *one* are etymologically similar and share a series of interpretations, *one* has developed a more formal acceptation (2017:40) than its Spanish counterpart and is particular of the written medium. When in informal situations the speaker tends to opt for a personal pronoun (Quirk et al. 1985, Biber et al. 1999, Altenberg 2005) or another impersonal strategy such as the noun *people* or the agentless passive (Altenberg 2005, Svartvik and Sager 1996).

### 2.4.3. Comparison with the man-impersonal *on*

English *one* and Spanish *uno* are often thought of as similar mechanisms to *on* when the latter has an indefinite reference. However, even though they may seem equivalent, they have a distinct origin and function differently. While *on* comes from the Latin *homo*, *uno* and *one* originate from the numeral *unus*. Several differences can also be noted from the syntactic and morphologic points of view. While we have seen that *on* can only be used in subject position, *one/uno* can play several syntactic functions such as subject, direct and indirect object complement or prepositional complement (Rasson 2017). However, the subject position seems to be the most frequent one (Ridruejo 1981, Company and Pozas Loyo 2009, Rasson 2017).

Cabredo Hohherr notes that *on* differs from *uno* and *one* in the sense that these mechanisms based on a numeral do not accept an episodic interpretation (2008:44). Koening (1999) specifies that, in episodic contexts where the French language would preferentially opt for *on*, the impersonal flavor would be achieved through the use of *someone* rather than the construction *one*. Similarly, even though some properties of the man-impersonals and the Spanish *uno* overlap, they cannot be considered as equivalent since the range of uses of the construction based on a numeral is much more limited than for the man-impersonals (Siewierska 2011:60). A similar overlap on some aspects can be observed with man-impersonals and non-anaphoric third person plural.

As this study intends to draw a detailed contrastive portrait of the syntactico-semantic and pragmatic behaviours of various passive-like constructions, it will enable to pinpoint the salient features of each passive-like structures, examine to which extent some structures may be used as alternatives and examine whether the present results concord with previous work on the topic.

# 3. Comparing and contrasting corpus data

Since the objects of study have been described, let us turn to the presentation of the general framework within which this work is rooted, namely contrastive linguistics. The first part of this chapter broadly presents the field of contrastive linguistics (3.1). The second part of the chapter is devoted to a review of existing contrastive and crosslinguistic studies on defocussing strategies (3.2).

## 3.1. Contrastive linguistics and its multiple facets

The present investigation is rooted in contrastive linguistics practices. Broadly speaking, contrastive linguistics aims at "uncovering differences and similarities between two (or more) languages (…) and capturing the distinctive features (…)" (Granger 2003:22). This notion is also frequently referred to as *cross-linguistic studies* (Krzeszowski 1989). Contrastive linguistics can adopt either a theoretical or a practical-oriented focus and bring insightful information in a variety of fields such as foreign language teaching and translation studies, revealing "what is general and what is language-specific" (Gast 2013: 1).

While the contrastive approach to linguistics seems to have gained favour in the last decades, it was not always so. The main reason for the temporary decline of such types of studies in the 1980s and 1990s was the general disappointment and questioning as regards their efficiency in improving foreign language teaching practices and in understanding second language acquisition mechanisms (Granger 2003, Köning 2012). However, this loss of impetus did not impede contrastive linguistics to regain importance, mainly thanks to the following three interrelated factors (Altenberg & Granger 2002: 4): (i) a phenomenon of internationalisation and a subsequent demand for multilingual and cross-cultural competences, (ii) developments within the general field of linguistic and an increasing interest in real-life communication (iii) the rapid development and progress of computer- aided and corpus-based research. The use of corpora, i.e., large collections of language productions, as starting point of contrastive studies has offered new perspectives to contrastive linguistics. Whereas early contrastive studies mainly relied on intuitions, with the subsequent risk to lead to over-generalisations (Granger 2003), contrastive linguistics now enjoys the possibility of using corpus data, which constitute sound empirical and reliable bases for cross-linguistic investigations. From a purely theoretical and intuition-based approach, CL shifted towards a more corpus-informed one, providing a consistent and effective "interface between theory and application" (Gast 2013:1) and leading CL to occupy a dominant position in linguistics (Altenberg & Granger 2002).

Before describing the main different applications of contrastive linguistics it is important to distinguish this discipline from typological linguistics. While the former relies on two or a very limited set of languages (most of the time related languages), the later includes a large variety of languages, usually considered as a representative selection of the world's languages (König 2012:10). Their purposes also differ in that CL aims at precisely charting potential differences

and similarities between a small set of languages whereas typological studies are rather interested in bringing into light language universals and in showing that languages do not vary randomly. As suggested by König, it can be thus assumed that these two disciplines share their synchronic orientation but "differ greatly in their scope, in their granularity and in their explanations" (2012:10).

Contrastive linguistics can rely on different types of resources depending on the objectives and the needs of the study. Granger (2003:19) distinguishes two main types of corpora: (i) those termed *comparable corpora* consisting of original texts in two or more languages, matched by criteria such as the time of composition, the text category, the target audience, etc. (Johansson & Hasselgård 1999) and (ii) those consisting of original texts in one language and their translations into one or more languages, also called *translation corpora.* The comparison can combine three types of corpus-based data, summarised in Table 3 by Granger (2010): texts in original languages, in source languages and in translated languages.

| | Type of comparison | Type of corpus | Discipline |
|---|---|---|---|
| 1. | $OL_x \Leftrightarrow OL_y$ | Multilingual comparable corpus of original texts | CL |
| 2. | $SL_x \Leftrightarrow TL_y$ | Multilingual translation corpus | CL & TS |
| 3. | $SL_x \Leftrightarrow TL_x$ | Monolingual comparable corpus of original and translated texts | TS & CL |
| 4. | $TL_x \Leftrightarrow TL_y$ | Multilingual comparable corpus of translated texts | TS |

OL = original language
SL = source language
TL = translated language

*Table 3. Types of corpus-based cross-linguistic comparisons (Granger 2010)*

The present study is interested in uncovering potential similarities and differences between original productions in English, Spanish and French, i.e., a small set of related languages, with future potential pedagogical applications and inputs into the field of translation. It is thus concerned with the first type of comparison presented by Granger (2010), i.e., multilingual comparable corpus of original texts. *Multilingual* is here understood as "two or more monolingual subcorpora in different languages, each collected according to similar design criteria" (Zanettin 2011:14). Multilingual comparable corpora can be specific that is, including language productions from a restricted domain (e.g., newspaper articles, academic abstracts, job interviews) or more general, including a wider variety of text types and particularly useful for terminological studies (Altenberg & Granger 2002).

Multilingual comparison is considered as the contrastive linguistics 'domain of expertise par excellence' (Granger 2010:4). Using multilingual comparable corpora in CL offers a series of advantages. The first benefit in working with these corpora regards their nature. As they include "original texts in the two or more languages under comparison - i.e., language spontaneously produced by native speakers of those languages" (Granger 2003:19), they allow contrastive analyses to rely on solid and reliable empirical data. These data can further be easily analysed qualitatively and quantitively, which offers new interesting perspectives compared to early

intuition-based contrastive approaches. Another advantage in using multilingual comparable corpora is the wide availability of texts. Granger (2010) notes that this type of corpus is the most easily accessible, particularly in English. As regards less well documented languages with fewer electronic corpora available, recent and considerable developments in computerized competences make it possible to easily design corpora suitable for comparative works involving multilingual original texts.

Although comparable corpora have major advantages, they nevertheless suffer from some drawbacks. The main issue the linguist is confronted with when investigating pairs of original languages is the problem of comparability. This issue is double. First, the researcher has to deal with the appropriate selection of comparable language productions. In addition to the usual formal constraints in corpus design such as size, time, and place of production, the contrastivist has to take into account the fact that some types of texts are culturally dependent and may lack equivalents in one of the languages studied (Granger 2003:22, Halliday 1978). The choice of comparable text types requires thus crucial reflection and cultural awareness prior to any multilingual corpus-based contrastive analysis. As the languages studied in the present research are considered as socio-culturally linked, i.e., "languages whose speech communities overlap in some way" (Gast 2013:1), it is assumed that the distance between the different socio-cultural practices is relatively limited and will not impact the comparability of our data in a significant way. Second, there is the problem of equivalence (Johansson 2007) in the studied phenomenon, or, put differently, how to delimitate the object of study and ensure that likes are compared to likes. This problem can be overcome with the prior establishment of a *tertium comparationis* as detailed in the introduction of the present work.

As explained above, the main issues generally reported in the literature can be quite easily overcome with a minute and consistent work in the selection and preparation of the data, which makes multilingual corpora the most appropriate material to serve the needs of this study. Now the general framework of the study has been set, I will turn to a revision of the main contrastive works carried out in the field of impersonality.

## 3.2. Crosslinguistic studies and impersonal constructions

In most cases, the structures included in the present investigation have already been extensively described from a monolingual and theoretical perspective (see Gómez Torrego 1992, Devís Márquez 2003 or Cabañas Maya 2006 for Spanish; Gaatone 1999, Creissels 2000 or Flottum et al. 2007 for French; Granger 1983, McIntyre 2005, Carter & McCarthy 2006 for English, to only cite a few). This body of literature served as a basis for the description of the structures presented in Chapter 2. I will set these works aside for now and rather proceed to a review of contrastive studies on agent defocussing mechanisms. This section will be divided as follows. The first part (32.1) includes the major contributions in relation to English and possible other languages. The state of the art for Spanish studies is established in (3.2.2). Finally, the section ends with a review of works including French constructions (3.2.3). As they are assumed to be

valid for any language, contributions from typological studies have been included in Chapter 2 alongside the detailed description of each structure examined in the present investigation.

### 3.2.1. English and other languages

Contrastive studies on passive-like structures including English as one of the languages of the pair or of the small set of languages studied abound. It may be assumed that this phenomenon is partly due to the undeniable role of English as a global language. The careful review of the literature reveals that the most frequent combination of language included in contrastive work combined English with another Germanic language (see e.g., Malamud 2012 or Gast 2015 on English and German, Van der Auwera et al. 2012 on English, German and Dutch, Fredriksson 2016 on English and Swedish or Van Olmen 2018 on English, Dutch and Afrikaans). Cross-linguistic studies contrasting English with a genetically more distant language are isolated, that is only a very restricted number of studies has been found with the same language combination (see e.g., Jia & Dou 2002 or Xiao & McEnery 2006 on English and Chinese, Warburton 1975 on English and Greek, Kaiser 2015 on English and Finnish or Kibort 2004 on English and Polish)[13]. Studying languages from the same sub-family, that is Germanic language in this case or, on the contrary, from distant and pertaining to another language family than the Indo-European one (e.g., Finnish or Chinese) may easily highlight their shared properties or reveal the distinctive features of these languages. We are, however, convinced that the existent body of literature on impersonalization will greatly benefit from the language combination included in this study in bringing fine-grained information on English, Spanish and French. Though it is expected that the differences found in our combination of language compared to, for instance, English and Greek, might be less spectacular, the granularity of the present study allows to bring to light subtle details and provide useful information to fields such as translation or teaching requiring precise information to produce appropriate content in the target language. Interestingly, the studies including only Germanic languages lay their focus of investigation principally on the use of human-impersonal pronoun as defocussing mechanisms while studies including more distant languages tend to grant more importance to the canonical passive. Since they seem to have aroused most interest, I will first start with the review of human-impersonal pronouns. In a second phase I will complement these findings with some major contributions related to other agent defocussing mechanisms.

Contrary to some other Germanic languages such as German or Dutch, English does not have a man-impersonal pronoun anymore[14]. Van der Auwera et al. (2012) notes that such pronoun existed in Old English but disappeared in the 15$^{th}$ century. Various researchers attempted thus to discover how present-day English expresses the nuances conveyed by this mechanism and which factors guide the speaker towards one or another equivalent strategy (Van der Auwera et al. 2012, Malamud 2012 or Gast 2015). Gast (2015) reports that several factors have to be considered when determining the adequate cross-linguistic correspondence to express similar

[13] Note that the combination English-Spanish will be discussed in section (3.2.2) devoted to Spanish.

[14] Cheshire notes that the pronoun *man* seems to be re-emerging in some varieties of English, though conveying different values from the earlier one (Cheshire 2013:609) and still being at the early stages of grammaticalisation.

propositional content than man-impersonal pronouns. A first element impacting the decisional process is the highly sensitive character of these pronouns to the context within which they appear. Their uses as universal (84) or existential quantifier (85) would correspond to different impersonal strategies in English.

(84) *Man lebt nur einmal.*
'You/everyone only lives once.'

(85) *Man hat geklopft*
'They knocked at the door.'

(Examples and translations from Gast 2015:7)

In addition to the linguistic context, two other essential parameters impacting the selection process of the adequate English correspondence are the syntactic environment and semantic factors (e.g., modality or veridicality) (Gast 2015). Taking into account these variables, Gast (2015) reports that, in proceedings of the European Parliament, the most common strategies to convey similar meanings to man-impersonals are the passive, the non-anaphoric first person plural and then nominal constructions and infinitives (Gast 2015:9). Van der Auwera et al. (2102) add that man-impersonals can also be partly covered by the generic second person singular and the non-anaphoric third person plural. Gast (2015) notes that the English impersonal pronoun *one*, usually associated with relatively formal registers, does not seem to represent a viable alternative to man-impersonal pronouns as it only appears in a relatively restricted number of cases. Similar findings were reported by Van Olmen & Breed (2018) in their comparison of English, Dutch and Afrikaans. The scarcity in the use of *one* as an equivalent for man-impersonals may, as suggested by Malamud (2012), be attributed to the fact that *one* "reflects a more subtle relationship to the conversational participants" (2012:30) and, more generally, is losing ground as an impersonal mechanism. Malamud (2012) specifies that when speakers have to select an agent defocussing construction with the feature [human], [definite], [arbitrary] the favourite candidate in English would rather be the non-anaphoric third person plural than *one*. A similar phenomenon of disappearance is observed in Dutch, where *man* seems to make way for other impersonal mechanisms (Van Olmen & Breed 2018).

Even though man-impersonal constructions are available in a series of Germanic (and to a smaller extent non-Germanic) languages, they are not used with the same frequency (Van der Auwera 2012, Coussé & Van der Auwera 2012, Van Olmen & Breed 2018). This strengthens the belief that even though two or a small set of languages possess similar structures the use they make of it may greatly differs. Since French is considered as "the most Germanic of all Romance languages" (Van der Auwera 2012:29), the analysis of *on* in French – i.e., the Romance equivalent to 'man', will provide insightful information on whether it corresponds to similar equivalents in English as those pointed out for Germanic man-impersonal

As regards contrastive studies on other mechanisms than impersonal pronouns, Xiao et al. (2006)'s contribution on the canonical passive, a major work in the field, compares English and Chinese structures in written and spoken corpus-based data. Compared with English, Chinese possesses a wider range of possibilities to defocus the agent. Among these devices, Chinese

does have matching structures to express similar basic meaning as the English *be/get* + past participle. The authors highlight that, even though they can be considered as equivalent, these constructions differ to some aspects in the two languages (2006:24). They nevertheless note that the major differences concern their overall distribution of the structures rather than intrinsic distinctive properties. Similar findings were reported for the pair of language English and Swedish by Fredriksson (2016) since the author argues that these two Germanic languages function via similar patterns and their distinctive features are rather accounted in terms of frequency of these patterns.

In addition to differences in terms of distribution of the structures, genre is considered as another important criterion explaining differences in the use of passive structures (Xiao et al. 2006, Fredriksson 2016). Xiao et al. (2006) suggest the following example to illustrate this phenomenon. While the proportion of canonical passives is relatively high in formal genre in English (e.g., official documents or academic prose), these genres display the lowest proportion of passives in Chinese. This observation points to the crucial need to consider cultural practices when selecting data for contrastive analyses.

A last element that frequently emerged from contrastive studies regards the overt presence of the agent. Similarly to what is observed in English, all the languages studied in the body of contrastive literature on canonical passive, preponderantly resort to passives without explicit agent. This is observed whatever their degree of genetic similarity to English (Warburton 1975 on Greek, Xiao et al. 2006 on Chinese, Fredriksson 2016 on Swedish). This scarcity of passive clauses with overt agent is even more marked in some languages other than English.

### 3.2.2. Spanish and other languages

After a more Germanic-centred approach of agent defocussing mechanisms, I will now proceed to the review of contrastive studies involving Spanish data.

In a study comparing Spanish and English, Orts (2018) reports that, in legal opinion columns, which is a highly formatted genre, Spanish resorts far more frequently to impersonal structures than English. Impersonality in Spanish is mainly rendered through the use of passive constructions based on *se* and agentless passives. The author shows that English defocuses the agent mainly through two types of mechanisms: agentless passives and inanimate subjects in collocation with active verbs (e.g., *the law foresees that…*) (Orts 2018:809). These two strategies display a similar frequency of use. Based on these results, Orts (2018) suggests that the presence of the writer in legal columns is less marked in Spanish than in English, giving prominence to the world of ideas. As legal columns are classified among formal genres, a comparison of Orts' results and the results yielded by the analysis of our set of formal data will allow to capture a more fine-grained picture of the presence of the addresser in Spanish and English. Again, the study carried out by Orts (2018) reminds us, that, as already highlighted above, linguistic habits and norms vary according to the speaker's/writer's culture.

Though English and Spanish both display agentless passives among their preferred candidates to defocus the agent, Espinoza (1997) highlights that the uses of the passive voice in these two languages do not always match exactly. The main reason for this partial match is explained by the presence of the Spanish *se*-construction, an alternative used more frequently than the canonical passive in Spanish. The author suggests that aspect could be an important element impacting the choice of one or another structure. However, she does not provide corpus-based data on this issue. The present study tries to fill this lack of information and bring more insights into the role of aspect in the selection of impersonal constructions.

Since there is no English form-based equivalent to the *se*-construction, the preferred candidate to express the passive voice in Spanish, Durán (1993) tries to identify to which resource English relies on to convey similar content. He indicates that, in most cases, when the translator needs an option with an overt human subject or group of persons, English resorts to impersonal pronouns and to noun phrases with a generic meaning (1993: 110). The first-person plural will be chosen in cases the addresser is included in the proposition while the third person plural will be favoured in cases it is excluded. When the agent is totally omitted, English usually uses the periphrastic passive. The author adds that, in translation, the context plays a crucial role in guiding the selection towards one or another strategy (1993:110). *One* can also be used as to convey similar content as the impersonal *se*-construction. Nevertheless *one* is usually considered as a rather marked construction, more specific of formal written texts. When translators need an option to translate *se* with a reference to the speaker, the hearer and third persons, they would rather opt for the generic *you* rather than *one* (Durán 1993:113). The choice is thus rather guided by stylistic requirements than linguistic constraints.

In her study on newspaper discourse Martínez Caro (2002) notes that the impersonal use of a personal pronoun is a frequent strategy to defocus the agent, be it in English or in Spanish. English texts mainly resort to the pronoun *we* (64.34%), followed by *they* (23.08%) and to a smaller extent to other pronouns such as *you* or *one.* As Spanish is a pro-drop language, the omission of the pronoun, be it used in an impersonal way or not, is very common (Martínez Caro 2002: 75). The analysis of the grammatical phenomenon is thus based on the verbal inflection. The author finds similar trends regarding the English language, since the analysis of the verbal inflections shows that omitted pronouns are mainly first and third person plurals (64.58% and 25%, respectively). She highlights that the overt presence of personal pronoun used impersonally is relatively rare. *One* in English and its Spanish counterpart *uno*, instances of overt pronouns, usually perceived as formal strategies, do not reach a higher frequency than 3.5% (2002: 79). In addition to impersonal strategies based on personal pronouns, the author finds that indefinite pronouns such as *someone* in English and its Spanish counterpart *alguien* are a second important device to defocus the agent in English and Spanish. Similar frequencies of use were found in both languages. From this study it can be assumed that even though English and Spanish have different roots, they display a similar behaviour as regards their use of pronouns as impersonal strategies in newspaper discourse. From the contrastive literature on impersonal mechanisms in English and Spanish, it emerges that even though these languages pertain to a different language family, they display quite similar strategies when it comes to defocussing the agent. The main differences are rather subtle and differences for the canonical passive are mainly due to the fact that Spanish possesses two options (the construction with *se*

and the canonical passive) where English has only one. It can be added that differences between these two languages are also accounted for in terms of cultural habits (see for instance highly formatted genres) and not only in terms of linguistic devices.

Besides English, Spanish is usually compared with other Romance languages. I will first start with a review of studies which contrast Spanish with other languages than French and then close the section with the Spanish-French comparison.

Spanish is regularly compared to another closely related Romance language, that is Portuguese. Posio (2017) reports that Portuguese does not possess as many devices as Spanish to defocus agents. No equivalent mechanism based on the pronoun *uno* can be found in Portuguese and, compared to Spanish, Portuguese resorts far less frequently to the generic second person singular in spoken productions (Posio 2017:112). In addition to the difference in distribution, the use of the Portuguese second person is more restricted than the Spanish equivalent. To express similar content, even though the construction is available in Portuguese, speakers seem to prefer the noun phrase *a/uma pessoa* ('the/a person') (Posio 2017:12) or the first person plural. This means that even though two related languages possess a structurally similar construction, the actual use speakers make of it can vary. Again, cultural factors are suggested to explain the variation in behaviour. In her work on Spanish and Brazilian Portuguese de Aráujo (2015) suggests that passive and impersonal constructions with *se*, available in both languages, are often compared from a syntactic perspective, mainly based on the concordance. She argues that this approach is indeed effective when differences are obvious. When it comes to more subtle differences, for instance, in case of partial overlap between the structures, the syntactic approach alone cannot account for variations. The author suggests that factors such as the type of agent or the verbal aspect should receive greater attention in the contrastive process. Posio (2017) also highlights the importance of the tense when analysing impersonal mechanisms. These features will be considered in the present analysis, which provides solid corpus-based data that complement more grammar-based oriented work such as de Aráujo's study (2015).

Another combination frequently studied is Spanish and Italian (see e.g., Dobrovie-Sorin 1998, Mendikoetxea 2008, Giordano 2015). Both Spanish and Italian possess impersonal mechanisms based on personal pronouns to defocus the agent. Giordano (2015) reminds that, though there is a high degree of similarity between these two systems, the use one language makes of a pronoun does not always exactly match the use in the other language. This partial similarity may trigger problems for second language learners. An exact understanding of how impersonal mechanisms work in each language would allow to overcome this issue. Regarding passive constructions, Mendikoetxea (2008) suggests that the syntactic features of impersonal constructions based on *se* in Spanish and *si* in Italian can be explained with reference to agreement operations, to the nature of the argument in subject position - i.e., the presence of a null generic pronoun, and to verbal properties (e.g., tense) of the structures within which the clitic appears in (2008:290). Dobrovie-Sorin (1998) suggests that the use of clitics as impersonal strategies in Spanish, Italian and Portuguese differs from the one observed in French and Romanian. She puts forward that the ungrammaticality of French (87) and Romanian (86b)

impersonal structures are due to the fact that these languages lack a nominative *se*, the clitic only being grammatical in its accusative use.

(86) a. En esta escuela se castiga a los alumnos.
in this school SE punishes a the students
'In this school they punish the students.'

b. *In scoala asta se pedepse, te pe elevi. (same gloss as (a))

(87) *Il se procédera à une enquête.

(Examples and gloss Dobrovie-Sorin 1998: 405)

The author suggests that in French and English, languages lacking a nominative *se* or deprived from clitics, the content conveyed through impersonal *se* is best translated with other impersonal strategies such as *on* in French and *one* in English (1998:399). Since Dobrovie-Sorin's study (1998) is not corpus-based, the present investigation is expected to provide further elements as regards the properties and use of impersonal strategies in these languages.

The comparison of French and Spanish impersonal structures has aroused the interest of numerous researchers (see Lamiroy 1993, García Yebra 1997, Kluge 2012, Weber 2014 or Yao 2016). With a focus on pronoun-based strategies, Kluge notes that these two languages frequently resort to the generic second person in oral data. This strategy accounts for 20.9% in French and 24.7% in Spanish of the forms used to establish a generic reference that involves the interlocutor (2012: 305). The high frequency of this structure in oral data may be explained by the fact that it allows both addresser and the addressee to be situated and share the same mental space. Her fine-grained analysis of generic second person structures reveals that, even though largely used in both languages, they are much more frequent in Spanish than in French. Kluge (2012) attributes this difference of frequency to contextual and linguistic factors. She reports that the degree of familiarity and communicative intimacy between the participants largely impacts their language production (2012: 307). The address system and the range of impersonal devices a language possesses are impacting factors on the use of generic seconds. Adopting a translational approach, Kluge finds out that the presence of man-impersonal pronouns in the linguistic system affects the frequency of use of generic seconds. For instance, French and German, that do possess this alternative, would rely less on the use of generic second persons than English or Spanish, that do not have such construction (2012:308). Starting from the French man-impersonal *on*, Yao (2016) tries to identify which are the most adequate translations in Spanish. *On* is an indefinite pronoun that lacks a form-based equivalent in Spanish. To refer to indefinite content, the author explains that Spanish can rely on a series of devices, mainly the *se*-construction, the non-anaphoric third person plural and *uno.* The choice of the alternative largely depends on semantic and grammatical factors and the subsequent interpretation of *on* and (Yao 2016: 119). A main criterion in the selection process is degree of

focalization of the participant(s) and the inclusion/exclusion of the addresser (2016:128). A last important factor highlighted by Yao is the degree of formality.

In addition to pronoun-based strategies, Spanish and French both can resort to the periphrastic passive to express impersonal meaning. Though structurally similar, constructions based on the auxiliary verb *ser*/*être* + past participle differ in their use in Spanish and French, with a much more frequent use in French. Weber explains (2014: 307) that this variation can be attributed to the fact that the Spanish periphrastic passive is much more restricted both from a syntactic and semantic perspectives than its French counterpart. She reports that in Spanish these constructions mainly appear with perfective tense and verbs that express event and transitional processes (2014:307). The lower frequency of use can also be explained by the fierce competition between this construction and the *se*-construction, highly productive in Spanish. Spanish seems to only use a periphrastic construction for special needs, for instance rhythm or informative values (García Yebra 1997).

### 3.2.3. French and other languages

When revising the contrastive literature including French impersonal structures, one would be astonished by the scarcity of the studies conducted on the matter. Most of the time, these structures seem to either have been described with a monolingual perspective, that is contrasting two structures available in French (e.g., Lamiroy 1993, Gaatone 1998) or with a translation perspective. The revision of the existent contrastive literature on French impersonal constructions reveals that, in most cases, French language is investigated alongside Spanish, another closely related Romance language (see previous Section). Apart from the group of studies tackling the pair of languages Spanish-French, French is also compared to English. The comparison of French with English, pertaining to the Germanic group of languages, brings different insights into the understanding of agent defocussing mechanisms.

Ghenimi (2005) indicates that the frequency of use of the canonical passive form differs in French and in English. In French, the proportion of passives varies from 2.8% to 8.7% in written texts while this percentage falls to 1.6% in spoken language. In English passives are slightly more frequent as the author reports a percentage varying from 4% to 18% in written production, depending on the degree of formality of the texts. This should be borne in mind when the translator is engaged in translation issues. The author reminds that, even though languages can display similar grammatical schemes, the content expressed through them varies as each linguistic community has a different perception of the world, (2005:46). This may explain differences in variation in the use of an apparently similar grammatical phenomenon in two languages. Apart from differences in the distribution, Ghenimi (2005) also highlights the fact that English does not possess pronominal verbs. As illustrated in (88), *middle* passives in French would be translated with a periphrastic passive.

(88) *Ce livre se vend partout.*
'This book is sold everywhere'

(Example and translation from Ghenimi 2005:39)

Several authors point to the fact that some grammarians would even exclude the constructions built on a pronominal marker with passive or middle readings from the realm of passives (Ghenimi 2005, Raineri 2010). These constructions are however included in this study. Raineri (2010) compares the French construction *se faire* with the English *get*-passive, largely recognized as an alternative to periphrastic passives in some contexts. She argues that, though syntactically rather distant, these two forms display a series of similarities. The author mainly identifies the following features: a high semantic transitivity, a preference for an animate subject, a high degree of implication of the receiver in the action suffered, strongly associated to negative predicates, and a preference to be used in informal oral language (2010:388). These findings highlight the fact that structurally different constructions in two languages can share a relatively high degree of resemblance in their use, not self-evident at first sight.

Broadly speaking, the following four observations emerge from the review of the literature on contrastive works on impersonal structures: (i) structurally similar structures in two languages may display different frequencies of use and may be used in different contexts; (ii) conversely, distant structures can express similar content; (iii) verb-related features (such as aspect and tense) and the nature of participants appears as important distinctive features guiding the choice towards one or another structure; and (iv) external factors - i.e., the text type, the degree of formality and the cultural habits of each linguistic community are crucial factors in the choice of impersonal strategies. Our results will be discussed in light of these observations in order to examine the extent to which they hold true in the passive-like structures included in the present study.

# 4. Defining the theoretical framework

When working in corpus-based contrastive linguistics with data from various sources it is essential to clearly define the theoretical notions framing the investigation. One of the main assets (and challenges) of the present study is its ample variety regarding language production and parameters. This work includes data proceeding from both spoken and written modes (here used in the sense of *vehicle of the message*), with different degrees of formality and analyses nineteen parameters. Chapter 4 intends to delimit the notion of mode (4.1) and register (4.2) by summarizing the main issues usually debated in the literature in order to report their respective salient characteristics and to highlight their potential degree of overlap. Section 4.3 presents Hopper & Thompson's model of transitivity (1980) which paves the way for the annotation protocol.

## 4.1. Distance between speech and writing

The present section offers an overview of the spoken versus written nature of language production and develops to which extent speech and writing interact. It first intends to clarify the main terminology in 4.1.1. Then, section 4.1.2 presents the issue of primacy which, as will be commented, may provide some clues on why oral data have first been neglected in early contrastive studies (Chafe and Tannen 1987) and on the importance of including them in large-scale contrastive work such as the present one. The last two parts intend to situate speech and writing on a continuum, pinpointing their similarities and differences

### 4.1.1. Clarifying key terms

Language users have a wide range of devices at their disposal to express themselves and to adapt their linguistic realization according to the situational context and the requirements of their production. A first fundamental clarification must be made on the vehicle the interlocutor chooses to elaborate the message. Regarding this notion, Koch and Oesterreicher (2001) insist on the difference between what they name *code* and *medium* (also termed *mode* in the present investigation as it is the usual label in English). They understand the notion of *medium* as either the graphic or phonic realization of the linguistic production whereas the *code* refers to the aspect or 'allure linguistique' conferred to the production (Koch & Oesterreicher 2001: 585). Moreau (1997) also raises this issue of ambiguous terminology and specifies that the term *code* is sometimes used to refer to a language (in its trivial sense, e.g., English or Spanish) or to a system of transposition. Koch & Oesterreicher indicate that these two notions can be combined with each another and produce four options: "parlé phonique", "parlé graphique", "écrit phonique" and "écrit graphique" (2001:585). In this study, I will follow Koch and Oesterreicher's (2001) acceptation of the term *spoken* and *written* and use them to refer to their conceptual meaning.

### 4.1.2. Issue of primacy

Another frequent issue about language, understood here in its general acceptation, is the primacy of either spoken or written language. A clear understanding of the development of speech and writing and the relationship between them helps to understand the treatment these two dimensions receives in contrastive linguistics. Two trends seem to emerge in the literature. The first one supports the primacy of spoken language stemming from the idea that the origin of humanity is closely linked to spoken productions and oral traditions (Koch & Oesterreicher 2001). This first approach, called chronological primacy, is further supported through the following observations and reflexions (Stubbs (1980) cited in Alsaawi (2019)): (i) spoken comes before in the acquisitional development by children, (ii) spoken is innate and biological compared to written which requires explicit teaching processes to be acquired, (iii) spoken opposes conscious manipulation, (iv) literacy is considered as a recent event.

The chronological primacy of spoken over written tradition leads to consider written languages as a simple transcription of oral productions, which contributes to view written as a derived product of secondary importance (Biber 1988). This observation leads us to examine the issue of medium transferability (Lyons 1981, Moreau 1997, Koch & Oesterreicher 2001) which can be summarised as follows: "(…) it is possible to read aloud what is written and, conversely, to write down what is spoken (…)" (Lyons 1981:11). An illustration of this fundamental property of language could be reading juridical texts aloud or graphically transcribing a spontaneous conversation between friends (Koch & Oesterreicher 2001). A direct consequence to view written as a transcription of spoken is a loss of a certain amount of meaning and information (e.g., Halliday 1985). Even though the system used to transcribe spoken into written possesses established rules (for instance on the correspondence between phonemes and graphemes or how graphemes combine), information conveyed through paralinguistic communication, that is, intonation of voice, rhythm, pauses etc. which give clues on how the speaker says the message, may be lost in the transcription process. As a result, written transcriptions appear less elaborated than speech productions, which leads Moreau (1998: 210) to qualify written language as *parent pauvre* 'the poor relation' of language. This apparent disequilibrium may, however, to some extent, be counterbalanced. As mentioned by Rubin (1978:7), punctuation may be considered as a partial analogue which compensates for some prosodic parameters. The author adds that the different internal architecture of written productions may also intervene in compensating process. More recently, paralinguistic indications may also be rendered in written trough the additional use of signs or be annotated through specially dedicated software's. This way of proceeding may offer a useful alternative to compensate the loss of information resulting from transcription.

It is important to note that, even though the phenomenon of transcription of spoken production into written implies information loss, this transcription is of crucial help on two aspects. First, it facilitates and perpetuates the tradition of collective memory. As pointed out by Halliday (1985), humanity evolved and societies changed from a nomadic lifestyle, strongly associated with oral traditions to a sedentary life. This societal modification created new necessities which

can be correlated with the advent of written, a more static and permanent manner to pass down traditions. Second, as highlighted by Blanche Benveniste (1997: 24) oral realizations cannot be accurately analysed relying solely on the memory users have of them. Linguists, thus, need to resort to a graphic transcription of the spoken phenomena they are interested in.

A second position on the issue of primacy is the one considering written as of primary importance. Written realizations were the first to be studied, since, as explained by Chafe and Tannen (1987), oral productions were first neglected due to the lack of access and capability in terms of data collection. Some authors (see e.g., Akinnaso 1985 or Biber 1988) note that academic prose, regarded as the most literate written production in Western civilizations, became the norm, disregarding thus all the complexity of oral resources.

To close this section on language primacy it is worth mentioning that some societies are non-literate and, therefore, the spoken language performs all the functions usually endorsed by the written language in literate societies.

### 4.1.3. Spoken and written, divide or continuum?

From this idea of supremacy of written over spoken and vice versa originated the well-known, though controversial, dichotomy opposing the two codes. Akinnaso (1985:350) reports that previous research focusing on the comparison between spoken and written were prompted by the *idea* of differences rather than by the actual practice made of the two codes. This divide also owes its origin to researchers attempting to establish differences rather than pointing to potential similarities, which would have led to view the phenomenon as a continuum rather than a two-way distinction. As it is further developed hereunder, a mere dichotomy between spoken and written is far a reductive approach of the matter and precisely characterising the relationship between the two is an extremely complex challenge (Halliday 1985, Biber 1988, Blanche Benveniste 1997, Béguelin 1998).

Spoken and written productions can be placed on a continuum with various degrees of homogeneity governed by similar and dissimilar trends, to use a similar terminology as Béguelin (1998). She thus considers the phenomenon as a dynamic process. I will mainly focus on the general distinctions between spoken and written usually pointed out in the literature. It is important to note that since the present attempt to synthesize and offer suggestions is intended to fit the three languages at study in this investigation, unless otherwise stated, I will deliberately not provide detailed descriptions of specificities that would be characteristic of only one language. Such a synthesis is facilitated by the fact that the study involves two Romance languages and a Germanic one, the three pertaining to the Indo-European family. Spanish, French, and English share a certain socio-cultural connection with a similar surface structure from which some influences may, in some cases, be noted (for instance a common origin of words or borrowings). The three languages have a wide range of speakers around the world, display a high level of standardization and a long history of literacy. The elaboration of such a synthesis would have been less straightforward with much more distant languages (e.g., comparing tonal and atonal languages, Indo-European, and Austronesian languages).

### 4.1.4. Distinctions between spoken and written productions

Spoken and written can be distinguished on four major aspects. A first basic distinction concerns the vehicle used to convey the message. While spoken relies on phonemes, written resorts to graphemes. The physical representation and realization are thus different. However, on her study of French, Moreau (1997:2010) points out that phonemes and graphemes are defined in such a way that a parallel can be drawn between the two. Blanche Benveniste (1997) adds that the written mode, once in its final and permanent version develops along a linear axe, whereas the spoken mode offers the possibility of creating much more moves to, for instance, adapt, correct, or modify what has been said.

Besides this more practical difference, a second observation can be made about the distribution of the syntactico-semantic elements in the discourse. It is suggested that spoken and written do not, fundamentally, constitute two different codes but the difference would rather lie in the frequency of use of the grammatical elements available in the language (Béguelin 1998, Moreau 1997). Béguelin (1998) reports for instance more tensed verbs and a different use of deictic devices in oral data in French. Similarly, spoken English is said to be more verbally dense whereas written English would involve more nominal groups (Halliday 1985). In his wide-scope study on speech and writing in English, Biber (1988) investigates a comprehensive set of functional linguistic features and reports that no absolute difference could be identified between the two codes but rather identifies six dimensions alongside which speech and writing may differ in in the frequency of distribution of theses linguistic features:

a. Informational vs involved production
b. Narrative vs non-narrative concerns
c. Explicit vs situation-dependent reference
d. Overt expression of persuasion
e. Abstract vs non-abstract information
f. On-line informational elaboration

Regarding these dimensions, only three were identified as potentially displaying differences between spoken and written: the degree of involvement in the production (a), the explicitness of the reference (c) and the degree of abstraction of the message (e). He describes oral discourse as an involved production, situation-dependent reference, with non-abstract content while written and more literate discourse is characterized as a more informational production, explicit with elaborated references and abstract, technical information (Halliday 1985: 162). Regarding the other dimensions, no salient properties emerged and since, the characterizations overlap, the differences were rather attributed to the text type and other variables than to the nature of the code.

A third difference between the two codes is reflected on the way spoken and written productions build their complexity. Halliday (1985:76) reports an unfortunate tendency to consider spoken as *formless* and *featureless*, resulting from the practical difficulties in transposing spoken to written and the consequent loss of information. However, oral resources should not be considered as less complex or less structured than written productions (Halliday 1985). The

level of elaboration of both codes is rendered through a different architecture and structure. In an attempt to describe the complexity of spoken productions Halliday suggests that "they are intricate constructions of clauses, varying not only in the kind of interdependency (parataxis and hypotaxis) but also in the logical semantic relationships involved" (1985:86). This elaboration, in written productions, is rather achieved through higher levels of lexical density. The distinct architecture of the two codes is also a result of the difference in the acquisitional process, the methods of production, transmission and reception (Akinnaso 1982).

Fourth, considering the fact that there are no salient linguistic features characterizing spoken or written, it can safely be assumed that these two options are influenced, and sometimes constrained, by a series of distinct external variables. Indeed, spoken and written serves specific needs and are exploited for different communicative purposes (Halliday 1985, Biber 1988). Moreau (1997:235) notes that spoken is the preferred channel for familiar interactions on daily life topics while written is favoured to exchange on general interest topics between individuals sharing a low degree of familiarity. Though there is a relative tendency to resort to oral productions when communicating within the familiar sphere, orality and higher degrees of familiarity should not be automatically assimilated. Similarly, written discourses should be distinguished from literate discourses. Moreau (1997:235) concludes that even though there is an influence of the social norm, it does not entirely govern the code. Written productions with a high degree of informality are totally acceptable and vice versa, highly controlled, formal or academic spoken productions are possible. Biber notes that although spoken and written "(…) *can* be used for almost any communicative need, we do not in fact use the two forms interchangeably. Rather, depending on the situational demands of the communicative task, we readily choose one mode over the other (…)" (1988:8). In a similar vein, Halliday (1985) adds that some productions are usually either written or spoken as it would be pointless to use the other code. He illustrates this phenomenon by the fact that, generally, no one reads the regulations on the airline ticket aloud (1985:93).

These observations may bring us to consider written and spoken as two different possibilities, based on one and same semantic bases and lexico-syntactic system, to express a message. And both modes serve different purposes according to the requirement of the communicative situation. Akinnaso (1985) showed that when the (i) level of formality; (ii) topic and communicative task and; (iii) socio-cognitive functions are identical, both spoken and written share many of the linguistic patterns and features usually associated with one or the other code. When the case, the distance between spoken and written is highly reduced. Similarly, Halliday observed that "the distinction between speech and writing is becoming blurred as a consequence of modern technology" (1985:81), it is all the more relevant in the current context, when technology makes it possible, for example, to send vocal messages, implying a combination of *medium* and *code* which would have been inconceivable a few decades ago.

As was demonstrated, a mere binary distinction between speech and writing with primacy of one over the other would be misleading. Nor spoken or written should receive more importance. Both deserve careful attention and should be viewed as sharing a similar lexical, semantic and syntactic background system and being complementary since they fulfil different needs and

respond to different constraints. The potential differences are thus best accounted in terms of (a) practical realization of the message, (b) frequency of distribution and (c) functions rather than in linguistic features. When the term *dichotomy* is used it seems that it merely mirrors a difference in the medium and relies mostly on unfounded assumptions about the internal structure of the codes which fails to consider other key variables such socio-pragmatic functions, genres, registers or topic orientations. Apprehending the intrinsic relation of spoken and written as a continuum with various degrees of homogeneity is thus a more appropriate way to reflect its exact nature. Within this continuum it is suggested that different principles are at work, which "ensures that spoken and written language will never be totally alike (Halliday1985:44)" since duplicate of one and the same phenomenon would be useless. Among this principle is functional variation or register, which is described in the next section.

## 4.2. Register, a variational construct

### 4.2.1. Terminology

First coined by Reid (1956), definitions to characterize the notion of register and related terms abound. In some cases these definitions overlap, which regularly creates confusion. When it comes to defining the notion, Biber (1995), in line with Ferguson (1959) or Hymes (1964), Ure (1982), to only cite a few, suggests to view *register* as an umbrella term to cover "any variety associated with particular situational contexts or purposes" (Biber 1995: 1). Frequently related with *register* is the term *genre*. Sometimes considered as equivalents, sometimes viewed as two different notions, no general consensus has been reached to clearly distinguish the two terms (Biber 1995). According to Lee (2001) *genre* is situated on the level of text categorisation and the distinction between the different genres is solely based on external, that is, non-linguistic criteria (Biber 1995, Lee 2001). The identification of the exact difference between the two notions is, however, challenging since, as often pointed out, the two notions largely overlap. The main difference reported is a stronger association between *genre* and socio-cultural practices, ideology and power (Swales 1990, Bathia 1993, Lee 2001). Swales even views *genres* as being 'owned' by specific discourse communities (Swales 1990). Compared with *genre, register* is much more constrained by the situational context, with resulting specific linguistic features. It should, however, be noted that the distinctive factor between *genre* and *register* is rather their respective degree of association with cultural practice than the mere presence of an association, since some authors (Halliday 1978 or Agha 2004) consider that register variation is also culturally dependent and associated with social practices and the individuals potentially involved in them. Lee concludes that *register* and *genre* are, in essence, two different perspectives used to describe a same ground (2001:46). A last distinction that is worth mentioning, is the one between *register* and *domain*. *Domain* is a much wider notion since it refers to subject fields or, put in other words, to specific areas of activities in life such as, science, religion or art. To use Neumann (2013)'s term, the notion of *domain* is a category 'superimposed' on *register* since a *domain* contains different types of *registers*.

After this brief review of the fundamental terminology in register variation, I will turn now to the description of the typical features usually referred to in the description of *registers.*

### 4.2.2. Typical features of *register*

The notion of r*egister* is usually defined in non-linguistic terms since components such as the topic, the settings or the relation between the participants come into play. These factors, thus external to the language, constrain the format the message will take.

Within each specific register, Biber (1995), in a crosslinguistic perspective, identifies what he calls *register features* which are lexical and grammatical characteristics. While the main features used to define the notion of register are non-linguistic, register features are linguistic characteristics. Biber (1995: 94) gathers them into sixteen categories:

a. tense and aspect markers,
b. place and time adverbials
c. pronouns and pro-verbs
d. questions
e. nominal forms
f. passives
g. stative forms
h. subordination features
i. prepositional phrases, adjectives and adverbs
j. lexical specificity
k. lexical classes
l. modals
m. specialized verb classes
n. reduced forms and discontinuous structures
o. co-ordination
p. negation

The distribution of these register features serves as indicator to distinguish one register from another on the basis of linguistic characteristics. The presence of isolated register features is, however, not a decisive criterion to characterize a *register* but it is rather the recurrent co-occurrence of a series of features (Chafe & Danielewicz 1986, Biber 1995, Agha 2004) that would define a register. What characterize registers is thus the cumulative effects of several parameters. Biber further observes that some recurrent distribution of these combinations of features can be associated to the following dimensions (1995:104):

1. structural elaboration and complexity
2. lexical elaboration, complexity, and specificity
3. place and time reference
4. temporal organization of discourse (tense and aspect)

5. referential cohesion and explicitness
6. evidentiality, affect and stance
7. hedging and emphasis
8. interactivenesss and involvement

These dimensions may be considered, to some extent, as relatively universal since Biber (1995: 72) reports that, even though the languages included in his cross-linguistical investigations present structural differences, they serve similar purposes and display similar distributions across the registers. Xiao (2009) in an enhanced version of Biber's multidimensional model (1988, 1995) proofs the reliability and relevance of the approach in his analysis of registers and variation in world Englishes.
Besides these linguistic features pointed out by Biber (1995), several authors (see e.g., Gregory 1988, Halliday and Hassan 1989 or Neumann 2013) highlight the importance to consider the following three parameters when defining *register*: (i) the field of discourse (experiential domain and goal orientation), (ii) the tenor of discourse (agentive roles, social role relationship, social distance and appraisal), (ii) the mode of discourse (language role, channel and medium). The field of discourse covers the topic orientation of the message, the tenor specifies the relationship between the people engaged in the linguistic exchange and the mode of discourse brings information on the transmission methods (Neumann 20103: 15). While Biber's model was mainly based on salient pure linguistic features, these three types of parameters are considered as linked to metafunctions of the language (Halliday and Matthiessen 2004, Neumann 2013).

When it comes to classification, Biber (1995:7) apprehends the notion of register in terms of specialization. He categorizes the different types of registers alongside a continuum ranging from specific to general. For instance, methodology sections in psychology articles, a highly specific register, would be placed on one extreme of the continuum while fiction would be situated on the opposite extreme. A direct consequence of this continuous approach is the absence of discrete categories of registers. Biber (1995) highlights, nonetheless, that some registers are highly conventionalized. It is the case, for example, of legal documents which have well-defined norms. Within these constrained registers, there is little space for variation and, as a consequence, these registers form, to some extent, closer categories. In direct link with the dynamic approach of *register* is the notion of evolution. Registers are not fixed entities with permanent boundaries, they rather evolve over time. For instance, as pointed out by Agha (2004), "when prestige registers used by upper-class/caste speakers are imitated by other groups, the group whose speech is the sought-after variety often innovates in its own speech habits, seeking to renew or transform the emblem of distinction" (2004:25). This observation is particularly relevant when examining the phenomenon of formality.

Due to the high level of dynamicity discussed above and its cultural dependence, it can be safely assumed that establishing a comprehensive and universal list of the different types of registers is far too ambitious. As an illustration of some possible registers in English, Biber (1995) distinguishes the following registers for written and spoken modes: (i) written: press, religion, skills trades and hobbies, popular lore, biography, official documents, academic prose, fiction

and letters (each containing different sub-categories); (ii) spoken: face-to-face conversation, telephone conversation, public conversation, debates and interviews, broadcasts, spontaneous speeches, prepared speeches. Though these two lists are not exhaustive, they offer an overview of the main types of registers a language can possess. Another consequence of its dynamicity associated to social practices may result in difficulties to offer reliable cross-linguistical register comparisons. A simple and straightforward one-to-one correspondence cannot be established since each language is culturally dependent and may resort to different registers to express similar content.

Language production has the capacity to adapt to the contextual situation. It can be viewed as a communicational instrument containing a wide array of devices and strategies to convey a message under various circumstances and settings. Spoken and written, in the broad sense, were discussed as one possible type of variation of language productions. However, there is not a unique parameter at work when producing language. Another central principle governing the elaboration of a message is the situational variation. Broadly speaking, three main types of variations, occurring synchronically within one and the same language, can be identified: diatopic, diastratic and diaphasic variations (Koch & Oeisterrecher 2001).

Diatopic variation refers to geographical variations or *dialects* in Biber (1995)'s terms (e.g., Peninsular Spanish and Mexican Spanish, American English and British English). Diastratic variation, also called social variation by Halliday (1985) or *social dialect* (Biber 1995), corresponds to the language production of a social group. These first two types of variation are considered highly conventional (Biber 1995:1) compared to diaphasic variation, which forms the third main type of variation. Language productions also differ according to the function they serve. This phenomenon is termed diaphasic variation or *register*. Akinnaso highlights that, when outlining overall results based on spoken and written data, researchers often disregard three components: "(1) the mediating role of level of formality in the construction of discourse structures (see Irvine 1979); (2) the interplay between linguistic and rhetorical structures on the one hand and macro-organizational features, such as topic, setting, participants, and their roles/statuses, the nature of communicative task, etc. on the other; and (3) the socio-cognitive functions of discourse" (1985:330). These elements are key features in diaphasic variation which will be the focus of the present section. Since the terminology is far from being universal, I suggest addressing this issue first from the English linguistic research tradition and then further deepen the reflection with some comments on the Spanish tradition, whose terminological and practical approach to the matter seems to slightly differ.

### 4.2.3. Contrasting different terminological approaches

In Spanish the term *registro* ('register') seems to have undergone a specialization process and is commonly used to refer to formal and informal situational contexts and the general umbrella meaning, corresponding to the set of situational and culture-internal practices, observed in its English counterpart *register*, seems to have faded. To cover this notion, the term *registro* made way for *género* ('genre'). However, the overarching framework within which they are usually

used still seems similar in both traditions since, when studying register features, Albelda & Colomer (2006), in line with Briz (2010), observed two types of factors at play when characterizing registers: situational factors, considered as external to the language and primary factors, which define the variety in terms of communicative purpose. Within situational factors the authors include the following four features (2006). First, the degree of equality between the participants in terms of social and functional relationship. The most equal, the highest degree of colloquiality. Second, the common ground between the participants. The highest degree of shared knowledge and experiences, the most informal the communication. Albelda & Colomer (2006) explain that usually family, friends, persons from a same sports team or working colleagues share a high degree of knowledge within their specific social circle. Third, the settings within which the interaction takes place. The authors report that the more familiar the setting (such as a private house, a shop, a train station, a bar) the more informal the interaction. And fourth, the topic discussed. Daily-life subjects trigger a higher degree of colloquiality compared to more specialized topics.

As for primary factors, three features are at play:

1. the purpose of the interaction (transactional or interpersonal)
2. the degree of planification of the discourse
3. the formal/informal tone

These criteria partly overlap with those highlighted by Neumann (2013) since the topic discussed (situational factor) and the purpose of the interaction (primary factor) mirror the definition of the field of the discourse as, in both approaches, the features give indications on what the register is about. Similarly, the degree of equality and the shared common ground (situation factors) partially correspond to the tenor of the discourse since they all intend to capture the nature of the link between the participants. In Albelda & Colomer (2006) and Briz (2010), contrary to Neumann (2013) the mode of discourse does not seem to be a preponderant parameter in defining registers.

According to Albelda & Colomer (2006: 5) the central types of register variations are the formal and informal ones, which, seem to have become the primary meaning of the term *registro.*

Table (4), from Briz (2010:131), presents their main characteristics usually reported in the literature.

| **(+) Prototypical informal** | **(-) Peripherical informal** | **(-) Peripherical formal** | **(+) Prototypical formal** |
|---|---|---|---|
| *Informal inducing features* | | *Formal inducing features* | |
| (+) equality relation | | | (-) equality relation |
| (+) shared knowledge | | | (-) shared knowledge |
| (+) daily life settings | | | (-) daily life settings |
| (+) daily life topic | | | (-) daily life topic |
| *Typical features of informal register* | | *Typical features of formal register* | |
| (+) on line planification | | | (-) on line planification |
| (+) interpersonal purpose | | | (-) interpersonal purpose |
| (+) informal tone | | | (-) informal tone |

*Table 4. Features of formal and informal registers (Briz 2010)*

Informal register is thus characterised by an equal relationship between the participants and a high degree of shared knowledge. The interaction tends to take place in familiar settings and tackle daily life topics. The planification occurs on the run and the interaction is characterized by its interpersonal purpose and informal tone. Conversely, the prototypical formal interaction is characterized by the opposite features. This range of characteristics has to be considered as a dynamic continuum and not as a dual system. The absence of a feature does not entail the production to be considered as formal or informal, it only means it occupies a different place alongside the formality scale. Similarly, Briz (2002:19) adds that an isolated feature is not necessarily representative of one or the other register, but it is rather the combination of different features which confers the production its formal or informal character. The dynamic approach of the concept is also of preponderant importance since a conversation, usually considered as a prototype of informal registers, can resort to some features usually considered as formal. For instance, when an employee speaks with his/her boss within the framework of professional interactions.

Bennett (2009), in a tentative to summarise the key features of what is considered as formality in English, found a relative consensus among grammars. They usually agree on the following aspects: complex and longer sentences, which may be partly attributed to the planned character of the formal register, respect of the grammatical norms, avoidance of some interpersonal indicators.

In sum, the above observations make clear that the notion of register is far from being homogeneous and can be addressed from different perspectives. First, the terminology used to describe the issue within one and the same language may be ambiguous, and, various related terms are sometimes employed interchangeably to cover the same concept. It has also been observed that different languages do not always resort to similar terms. Whereas the English tradition adopts a rather ample approach of the phenomenon, the Spanish one seems to focus on the distinction between formal and informal features. Since, in the present investigation, one of the central features used to ensure the comparability of the data is the degree of formality, I

adopt the Spanish acceptation of the concept and use the term *register* to refer to language productions displaying different degrees of formality. The term *genre* is be used to refer to the text categorization. Second, the notion of register should be addressed within a multidimensional perspective since, as explained above, the concept can be described in terms of larger non-linguistic features such as mode, tenor and field which, in turn, consist in a series of different linguistic features.

To close this section I will call on Neumann's words which suggest that "people are never totally free in the way they use the language, but rather that recurring situations predetermine their use of language" (2013:2). Language users have a wide range of devices at their disposal to express their message but their freedom to select the medium, the code, the genre and the degree of formality is somehow constrained by situational communications and socio-cultural practices.

## 4.3. Transitivity, a multifactorial approach

This section is divided in two main parts. The first part presents Hopper & Thompson's transitivity hypothesis (4.3.1) while the second part suggests a series of comments on how it was received and presents more recent work on transitivity (4.3.2). These more recent contributions will be taken into account both in the annotation scheme and in the analytical part of the present study.

### 4.3.1. Exploring transitivity and Hopper & Thompson's approach

Transitivity is a grammatical phenomenon considered as involving two main notions: process, be it an action or an ascription, and participants, be it subject or complement (Halliday 1966:38). In transitive clauses, a transfer of process is observed from one participant to another (Hopper & Thompson 1980). From a syntactic point of view, a transitive verb receives an object complement and can be passivized. Basically, the transitive clause can have two core arguments: one argument with a referent accomplishing a particular activity and a second argument with the referent affected by this same activity. The intransitive clause would only consist of one argument.

Traditionally, the possibility of passivization is considered as the distinctive feature between transitive and intransitive clauses (Babcock 1970, Gross 1971). Nevertheless, numerous researchers, adopting different perspectives on the matter (cf. Hopper & Thompson 1980, Tsunoda 1985, Delancey 1987, Langacker 1987, Givón 1995, Campos 1999, Dixon & Aikhenvald 2000, Kemmer 2003, Malchukov 2006, Naess 2007), have demonstrated that the issue of transitivity is multifaceted and attempted to refine its definition and the different roles and applications of transitive systems. The framework used in this investigation is based on Hopper & Thompson's (1980) work as well as on Malchukov's (2006) contribution, which further suggests how Hopper & Thompson's parameters combine and offers a more refined scalar vision of Hopper & Thompson's transitivity proposal. Malchukov's (2006) refinements

will serve to structure the present analysis and to better understand the relationship between the different parameters of transitivity and the potential impact of this relationship on the distribution of our data. In their pioneering investigation on grammar and transitivity in world's languages, the authors propose a multifactorial approach to transitivity through the identification of the following ten grammatical parameters (1980:252). These parameters rank a construction as more or less transitive (see Figure 1):

| | | HIGH | LOW |
|---|---|---|---|
| A. | PARTICIPANTS | 2 or more participants, A and O | 1 participant |
| B. | KINESIS | action | non-action |
| C. | ASPECT | telic | atelic |
| D. | PUNCTUALITY | punctual | non-punctual |
| E. | VOLITIONALITY | volitional | non-volitional |
| F. | AFFIRMATION | affirmative | negative |
| G. | MODE | realis | irrealis |
| H. | AGENCY | A high in potency | A low in potency |
| I. | AFFECTEDNESS OF O | O totally affected | O not affected |
| J. | INDIVIDUATION OF O | O highly individuated | O non-individuated |

(Hopper & Thompson 1980: 252).

*Figure 1. Transitivity scale established by Hopper & Thompson (1980)*

- the action implies at least two participants (participant);
- the verb reflects an action (kinesis);
- "the action is viewed from its endpoint, i.e., a telic action (Hopper &Thompson 1980:252)" (aspect);
- the action is punctual and not inherently on-going (punctuality);
- the action is done on purpose (volitionality);
- the action is positive (affirmation);
- the action corresponds to a real event (mode);
- the action is performed by highly agentive participants (agency)[15] ;
- the object is totally affected by the action[16] (affectedness of the object);
- the agent is proper, human, concrete, singular, countable and referential/definite (individuation of the object).

Hopper & Thompson believe that this classification allows "to view transitivity as something more than just the presence vs. absence of a direct object but above all as a matter of degree

[15] Hopper & Thompson (1980) explain that the sentence *George startled me* is considered as highly agentive as George is a human that performed an action while in the sentence *The picture startled me* there is no perceptible action. Hopper & Thompson (1980) rather consider the situation as a "matter of an internal state (1980:252)." The second sentence has, therefore, a lower degree of agency.

[16] In *I drank up the milk* the object is totally affected while in *I drank some of the milk* the object is partially affected (Hopper & Thompson 1980:253)

(1980:5)". The degree of transitivity, calculation based on the number of parameters high in transitivity, increases according to how many of the following parameters the clause meets (Hopper &Thompson 1980:252). Tsunoda reminds that "these features can be manifested either morphosyntactically or semantically" (1985: 385). This classification allows to see whether the degree of transitivity or a certain category of parameters has an impact on the selection of impersonal structures.

It is hypothesized that transitivity is a factor impacting the choice of either one or another passive-like structure. Halliday suggests that the transitivity system construes the world of experience into a manageable set of process types (1994:106). This study attempts to determine to which extent transitivity constrains experiences where agents are defocused and to pinpoint potential differences between languages, modes and registers. Even though transitivity is a syntactic notion generally assumed to be universal, LaPolla et al. point out that actual intra- and inter-language differences can be found (2011:469). As pointed out by some researchers (see e.g., Martínez 2001), passive-like structures have not been much analysed in the light of the notion of transitivity and large contrastive corpus-based studies combining these elements are still scarce. Several works deserve however to be mentioned. Fernández's study (2007), a major work in that field, carried out a corpus-based analysis of the passive voice in Spanish with a particular emphasis on the periphrastic passive and the impersonal *se.* She demonstrates that these two constructions display noticeable differences in their relation to transitivity. Martínez (2001) provides us with a second important contribution to the study of transitivity and impersonal constructions. In her study on impersonality in the research article, she shows that the relation between transitivity and impersonality differs according to the sections of the research article. These differences are explained in terms of different need of abstraction, distance and negotiation between the participants. A third important work in the field is the one conducted by Clements (2006) on the relation between transitivity and non-anaphoric *se.* The author found a relative transitivity associated with that type of passive-like structure. Finally, Hurtado & Ortega-Santos (2019) make a significant contribution through their work on the relationship between transitivity and the use of *uno* in Colombian Spanish. They show that "a higher degree of transitivity disfavors the use of overt pronouns (…) for [the parameters] number of participants and kinesis, but not for mode, aspect, or individuation of the object (…)" (2019:36). They highlight that these findings largely corroborate Posio's proposal (2011, 2013) on the alternation of overt vs. null pronoun, where "low transitivity features favor the focusing of attention on the referent of the subject and thus increase the frequency of subject pronouns, while high transitivity features increase the likeliness of focusing attention either on the action expressed by the verb or on the referent of the object argument" (Posio 2011:25). The results of these studies will be discussed more in-depth in the light of those yielded by the present investigation.

### 4.3.2. Criticism of Hopper & Thomspon's work (1980) and other contributions

Although Hopper & Thompson's contribution in the domain of transitivity had a major impact in functional grammar (Fernández 2007), it has aroused various criticisms. This section displays the main remarks addressed to Hopper & Thompson (1980)'s approach and the position adopted in this investigation.

As pointed out, among others by Fernández (2007:168), in Hopper & Thompson's theory, a clause with a direct object can be considered as less transitive than a clause with only one participant. This may occur when a one-participant clause obtains a better grading in the transitivity scale established by the authors.

In his work *Remarks on transitivity* Tsunoda (1985) addresses four issues regarding Hopper & Thompson's (1980) framework in order to refine their notion of transitivity. More precisely, the author discusses the correlation and covariance of certain parameters, the relevance and rank of the ten parameters and the affectedness scale and its morphosyntactic correlates (1985:386). Tsunoda (1985) highlights that the strength of the correlation between parameters varies according to the type of feature. Parameters can indeed fall into different categories. For example, *volitionality* and *agency* are linked to the agent while *affectedness* and *individuation* concern the patient. Tsunoda adds that certain parameters or pairs of parameters can even "contradict each other and never co-vary" (1985:386). Therefore, he suggests that the transitivity hypothesis such as postulated by Hopper & Thompson is not so strong. Tsunoda also points out the need to refine certain parameters. For example, the parameter *affectedness* could be improved via the creation of a 'hierarchy of two-place predicates (1985:395)', which allows to underline the degree of affectedness. Hopper & Thompson did not give a different weight to their parameters and seemed to consider them as equally important. However, Tsunoda urges to establish a ranking according to "their relevance to morphosyntactic manifestations" (1985:393). He especially emphasizes the importance of the feature *affectedness* of the patient compared to *volitionality* or *agentivity*, which appear as irrelevant in his study. Unfortunately, the author does not offer any practical solution to this issue.

Malchukov (2006) attempts to respond to this need of recasting Hopper & Thompson's list of parameters. In line with Givón (1985), Malchukov suggests to organize them following their relation to the agent, the patient or the verb. He therefore establishes the following transitivity scale (2006:333) where A-features are related to the agent, V-features to the verb and O-features to the object or the patient (Figure 2).

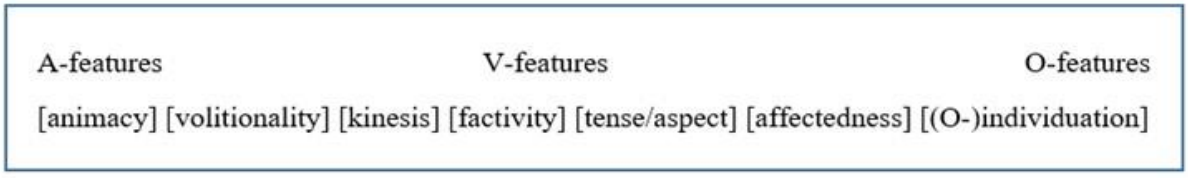



*Figure 2. Hopper & Thompson's transitivity scale modified by Malchukov (2006)*

He underlines that this scale would "represents selective semantic affinities between individual parameters through placement of the related parameters adjacently on the scale" (2006:333).

The author explains for example that *volitionality* is linked to *animacy* and that *affectedness* is closely related to *individuation.* Conversely, the "relation of affectedness to animacy is more subtle (…)" (2006:334). Features at the ends of the continuum are inherent to the semantics of the construction while features in the middle of the scale are related to the domain of the verb. Intermediate features serve as "'semantic bridges' linking different domains: kinesis is related both to A and V properties, while aspect is related both to properties of V and O (2006:334)".[17] In sum, Malchukov's work brings a refined classification of the transitivity parameters and highlights the importance to relate each parameter with its appropriate constituent. This classification will help to structure the present analysis and to shed light on the relation between Hopper & Thompson's parameters.

[17] For a detailed comment on Malchukov's transitivity scale see Malchukov, A. 2006. Transitivity parameters and transitivity alternations. *Case, valency and transitivity*, 329-357.

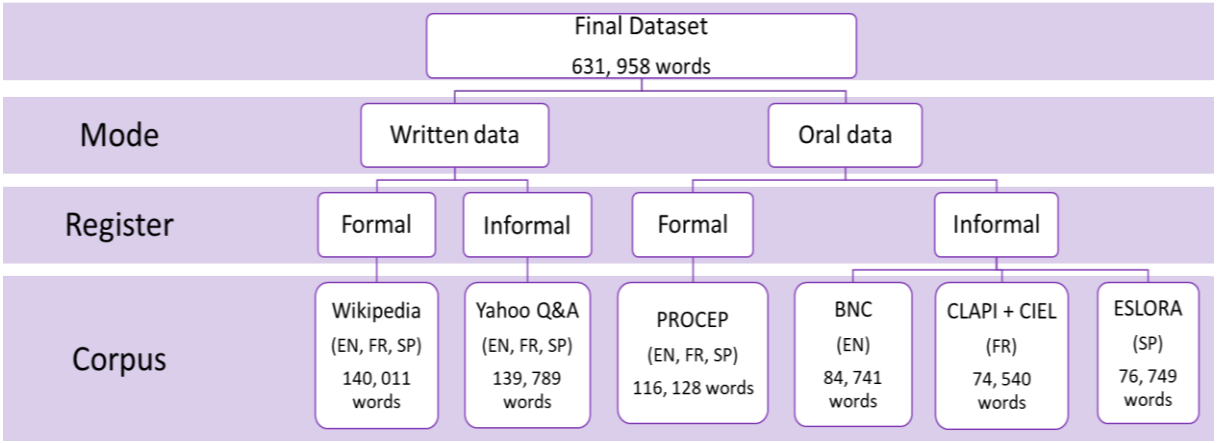
# 5. Data and methodology

To achieve the main objective pursued in this thesis, which is offering a contrastive overview of a selection of passive-like structures in three languages, a large and varied amount of corpus data is required. The present chapter aims to describe the corpora used to establish the panorama of the interaction between the selected passive-like structures

## 5.1. Data description

In order to offer a highly detailed overview of this issue, this investigation uses corpus data in Spanish, French and English, from oral and written modes. For both modes, formal and informal subcomponents were analysed. The final dataset is made of seven subcorpora suitable to establish comparisons:

- Yahoo-based Contrastive Corpus of Questions and Answers (henceforth Yahoo)
- Three dumps from the Wikipedia Database (henceforth Wikipedia)
- Proceedings of European parliamentary debates (henceforth PROCEP)
- British National Corpus (henceforth BNC)
- Corpus de la Langue Parlée en Interaction (henceforth CLAPI)
- Corpus International écologique de la langue française (henceforth CIEL)
- Español Lengua Oral (henceforth ESLORA)




*Graph 1. Overview of the final dataset composition*

Oral and written corpora display different characteristics and do not offer the same possibilities in terms of data collection and selection. This issue is discussed more extensively in 5.1.4. As explained in the next section, for a question of availability, structure and convenience, the written corpus Yahoo was used as the benchmark to create the other subcorpora required in the investigation and served as the basis of the pilot analysis. Each subcorpus will be described in more detail hereunder and a summary table will be presented at the end of the section (5.1.4). The first part of the section is devoted to the written corpora (5.1.1) and the second part sets out

the main characteristics of the oral corpora (5.1.2). Section (5.1.3) gives further information on the extraction of the data and the features of the final datasets.

### 5.1.1. Written data

To compile the written dataset used in this investigation, I made use of two different written sources: the digital forum Yahoo Questions and Answers and three dumps (one per language) from the Wikipedia Database.

#### Yahoo-based Contrastive Corpus of Questions and Answers

The *Yahoo-based Contrastive Corpus of Questions and Answers* (2009), compiled from the digital forum Yahoo (see De Smet 2009), served as the basis to determine the desired size of the subcorpora and the themes included in the present investigation. The availability of similar data in the three languages facilitated the compilation process to create comparable subcorpora. The Yahoo Q & A corpus includes information on a wide variety of themes. After verifying the availability of these themes in the three languages and ensuring that each of them displays a sufficient number of words to work with, four main subcomponents addressing the following issues have been designed: aircraft, dogs, botany and food habits. Each subcomponent for each theme accounts for around 10,000 words. This gives an approximate total of 50,000 words per language.

A main drawback of using such type of data is that, given the participative character of forum data, the geographic variation could not be controlled. The Spanish subcomponent dealing with food habits, contains, for example, texts on Mexican gastronomy. It can therefore be assumed that the Yahoo corpus does not only include Peninsular Spanish but also Latin-American varieties. Rasson (2017), on the indefinite Spanish pronoun *uno*, underlines that, while the regional variety has a direct impact on the frequency of use of *uno*, its influence on the different interpretational types the pronoun can receive is relatively limited. Following this line, it will be assumed that the variety has no significant impact on the type of data treated in this study. As the primary purpose is not to examine the frequency and the use of the passive-like structures with a focus on geographical variation, this characteristic will therefore not be considered.

#### *Wikipedia Database*

The formal written part of the data gathers files proceeding from dumps of the *Wikipedia* Database. Wikipedia provides access to specialized language for a wide range of topics in a multitude of languages. Similar data for the three languages could therefore easily be collected. The French part of the data comes from the Wikipedia-Corpus (2008) developed by the Université de Toulouse, the English part from the English Wikipedia corpus (Davis 2015) available via the corpus manager Sketch Engine and the Spanish portion from the Wikicorpus V.1.0 (Reese & al., 2010). A major drawback of this type of data is that there is no control on whether the text fragments are produced by a native-speaker of the target language or not.

However, this issue is partly overcome by including large amounts of data in the compilation process.

These corpora contain millions of words and, in some cases, their internal organization consisting of unclassified lists of hundreds of folders, did not allow to easily retrieve similar topics as those selected from the Yahoo Corpus. As manually examining all the files would have been extremely time-consuming, I used RStudio programming software to create a code that automatically indicates which are the files addressing the topics under study.

```
#------------------------------------------------------------------------
#Create a list with all text files names in a certain folder
f<-"wiki Spanish"
#Create list of text files with folder path
wikiFiles <- paste0(f,"/",list.files(f))
#Initiate template data frame to save results
out<-data.frame(file="a",vegetariano=0, vegetarianismo=0)
#Loop over every text file
for(i in 1:length(wikiFiles)){
tmp <- readLines(con <- file(wikiFiles[i])) #Load text file content
close(con)
# Number of article with at least 1
# o1 <- grep("vegetariano", tmp, ignore.case=T, value = F)
# o2 <- grep("vegetarianismo", tmp, ignore.case=T, value = F)
# Total occurrences in the file
o1 <- lengths(regmatches(tmp, gregexpr("vegetariano",tmp,ignore.case = T)))
#check every article
o2 <- lengths(regmatches(tmp, gregexpr("vegetarianismo",tmp,ignore.case = T)))
o1 <- sum(o1) #sum occurence per article
o2 <- sum(o2)
cat(paste0("\n\nIn file ",wikiFiles[i], "vegetariano appeared", o1))  #Print something on console
cat(paste0("\nIn file ",wikiFiles[i], "vegetarianismo appeared", o2))
out<-rbind(out, data.frame(file=wikiFiles[i],  vegetariano = o1,vegetarianismo = o2))
}
write.csv(out,paste("summary",f,".csv"),row.names = F)
#Set working directory
setwd("C:/Users/empierre/Documents/Doctorat/Corpora/Wikipedia/wikipedia Spanish")
#Have a look at what is in the working directory
dir()
#------------------------------------------------------------------------
```

*Figure 3. Example of a code allowing the identification of the text files containing the words vegetariano and vegetarianismo in the Spanish Wikipedia Corpus*

The code presented in Figure 3 allows to automatically generate a csv. file containing the exact number of occurrences a word related to the target topic appears and the name of the Wikipedia file in which it appears. Figure 4 illustrates a csv. file obtained while compiling the Spanish subcomponent dealing with food habits. One of the topics included in that section addresses the issue of vegetarianism. I inserted the key words *vegetariano* and *vegetarianismo* in the code

A1 | file

| | A | B | C | D |
|---|---|---|---|---|
| 1 | file | vegetariano | vegetarianismo | |
| 2 | wiki Spanish/raw.es.tar.063 | 44 | 22 | |
| 3 | wiki Spanish/raw.es.tar.197 | 8 | 1 | |
| 4 | wiki Spanish/raw.es.tar.660 | 6 | 1 | |
| 5 | wiki Spanish/raw.es.tar.153 | 6 | 0 | |
| 6 | wiki Spanish/raw.es.tar.081 | 5 | 0 | |
| 7 | wiki Spanish/raw.es.tar.640 | 4 | 4 | |
| 8 | wiki Spanish/raw.es.tar.217 | 3 | 8 | |
| 9 | wiki Spanish/raw.es.tar.249 | 3 | 3 | |
| 10 | wiki Spanish/raw.es.tar.610 | 3 | 2 | |
| 11 | wiki Spanish/raw.es.tar.464 | 3 | 1 | |
| 12 | wiki Spanish/raw.es.tar.105 | 3 | 0 | |

*Figure 4. Example of an csv. file allowing the identification of food habits in the Spanish Wikipedia Corpus*

and the program automatically detected that these words occurred respectively 44 and 22 times in the file 063.

Once all the names of the files with the relevant material were obtained, I manually opened the files containing the most words I am interested in and selected the texts dealing with the target topics. The collected texts were gathered to form each subcomponent of the written formal dataset.

As is the case for the Yahoo corpus, the part extracted from the Wikipedia Database accounts for (46,326 words in English, 46,162 in French and 47,523 in Spanish. As explained in the description of the Yahoo Q&A corpus, the geographical variation could not be controlled in the Wikipedia corpus. There is no guarantee either that the text fragments are produced by a native speaker of the target language. However, insofar as possible, I intended to include data produced in contexts allowing comparability, except for the informal oral corpus as will be explained in the next section.

### 5.1.2. Oral data

As written and oral modes display different features, the final dataset was supplemented with oral data. This section details the sources used for the Spanish, English, and French subcorpora.

#### Corpus Español Lengua Oral

The Spanish oral informal part relies on data from the corpus Español Lengua Oral (ESLORA 2) compiled by the Grupo de Gramática del Español from the University of Santiago de Compostela[18] (see Barcala, M. et al. 2018 or Vázquez Rozas et al. 2020 for a detailed description of the corpus). The selected data consist in 16 spontaneous conversations (44 participants). These conversations were recorded between 2007 and 2015. Given the linguistic situation of the area were the recordings took place, that is Galicia, a specific attention was paid in the selection of the conversations. The metadata available with the corpus allowed to select only the participants whose main language is Castilian Spanish and to ensure that they use it on a daily basis. Participants whose native and main language were Galician or another official language of Spain were discarded. As far as possible and because the material allows it, I tried to include data from various sociolects, ages and genders. Conversations mainly address themes of the daily life and totalize 76, 749 words.

#### British National Corpus

The data used for the English oral informal part were extracted from the British National Corpus (Mc Enery et al. 2014). The BNC, originally compiled by the Oxford University press, is a large online data base containing more than 100 million words of texts from different genres. The part selected for the present study consists in 15 spontaneous conversations with a total of 38

[18] I am extremely grateful to the University of Santiago de Compostela for granting me the access to these data during my research stay in 2018.

participants. An exact balance between male and female participants could be obtained. Due to the internal organization of the data in the BNC interface, the selection of the themes could however not be controlled. Conversations tackles topics such as daily life, family, travel sickness, sawing woods, politicians or sports. The selected recordings totalize 84,741 words.

### Corpus de la LAngue Parlée en Interaction and Corpus International écologique de la langue française

Collecting French informal oral data was the most arduous task. As Crible points out "spoken French is less well documented, with smaller or specific corpora that do no meet the requirements of size and representativeness of a reference corpus (…) (2017:85)." The data used for the French informal spoken part come from two different existing corpora: the Corpus de la LAngue Parlée en Interaction (CLAPI) and the Belgian part of the data available in the Corpus International Ecologique de la Langue Française (CIEL-F). The CLAPI part contains 9 private informal interactions recorded in real situations between 2006 and 2009. It includes for example conversations during pre-dinner drinks between friends or during a family lunch. A total of 31 participants took part in these conversations. Due to the scarcity of data available in French, topics could not be controlled. Unfortunately, the CLAPI corpus does not contain a sufficient quantity of data which meet the requirements of this investigation. The remaining French informal oral data comes therefore from the Belgian part of the CIEL-F corpus (see Gadet et al. 2012 for a more detailed description of the corpus). One conversation was selected to complement the data already selected from the CLAPI. The interaction takes place during a family meal and involves five persons. The conversation was recorded in 2011.

The final set of French oral informal data used in this thesis accounts for 74,540 words.

### *Proceedings of European parliamentary debates*

The data used for the formal oral part consist in extracts from a wide range of European parliamentary debates. The main strength of this type of data lies in their availability in the three target languages. This allows to increase the comparability of the material and reinforce the validity of the study. Parliamentary debates are considered as a *hybrid genre* (Cabasino 2010) or a *written-to-be-spoken* (Laanemets 2013) genre since they both include spontaneous oral features and features considered as more frequent of the written mode (Manole 2014). Nevertheless, as Neumann & Hansen-Schirra (2012) point out, speeches in parliamentary debates, when prepared, are thought and written with the intention to be pronounced. We expect thus that it would still come closer to oral features than written ones. Laanemets (2013) highlights that the communicative purpose and the situational context of this oral genre differ in many ways from day-to-day conversations. The purpose of parliamentary debates is to argue and convince through the use of a more formal and standardised language. The primary function does not serve to reaffirm social bonds or exchange information as it is the case in informal conversations (2013:73). The present study allows to explore to which extent the use of passive-like structures in parliamentary debates comes closer to either oral or written prototypical genres. Even though the EUROPARL corpus, an existing corpus gathering data from the

European parliament, offers a strong basis to work with for researchers who wish to analyse European parliamentary debates, I decided to select my own data from the material available on the European Parliament website. Three reasons justify this methodological choice: (i) obtain more up-to-date data; (ii) ensure that all the discourses were pronounced by a native speaker of the target language; (iii) make sure that, via the viewing of the videos available alongside the transcription, each contribution was produced spontaneously that is, not read. Unfortunately, the material available does not allow to rely only on entirely spontaneous contributions. To overcome this issue, I proceeded as follows. I collected all the spontaneous speeches fulfilling the desired criteria and then supplemented them with non-spontaneous speeches where needed. In order to include as comparable data as possible as to those from spontaneous contributions, the non-spontaneous data set consists in contributions during which the Member of Parliament only glances at the documents he/she prepared to support his/her intervention. Table 5 presents the composition of the different subcomponents.

| **Language** | **Theme** | **Spontaneous** (no. of words) | **Non-spontaneous** (no. of words) | **Total No. of words** |
|---|---|---|---|---|
| **Spanish** | Animal | 2462 | 4545 | 7007 |
| | Environment | 1939 | 7224 | 9163 |
| | Food | 10042 | 2695 | 12737 |
| | Transport | 5122 | 3913 | 9035 |
| | | | | 37942 |
| **French** | Animal | 2506 | 4592 | 7098 |
| | Environment | 4302 | 3499 | 7801 |
| | Food | 13916 | _ | 13916 |
| | Transport | 5522 | 3547 | 9069 |
| | | | | 37884 |
| **English** | Animal | 1542 | 7043 | 8585 |
| | Environment | 3261 | 4106 | 7367 |
| | Food | 13881 | _ | 13881 |
| | Transport | 7209 | 3260 | 10469 |
| | | | | 40302 |

*Table 5. Overview of the different subcomponents in the PROCEP corpus*

Given all the parameters and the constraints at play in this study, a perfect balance between male and female Members of Parliament was not reached. As can be seen from Table 5, the subcorpus for each language contains approximately 40,000 words. All the data used were extracted from debates conducted between 2011 and 2019.

In the compiling process, I try as far as possible to control the topics I selected to obtain similar data to those included in the written corpora. The issues addressed during the parliamentary debates allowed to tackle four main topics: animals (such as animal well-being or veterinary

drugs), environment (including for example issues about woods, storms and hurricane, climate change), food (agriculture, fishing quotas) and transport (road, sea or air transport).

### 5.1.3. Features of the final datasets and extraction method

The original corpora are of varied sizes and some are extremely large and, therefore, impossible to make use of their entirety for the purpose of a study like the present one. Hence the need to sample them and created comparable subcomponents based on the following selection of criteria:

- Registers
- Size
- Time of creation
- Themes

To obtain the final size of the material used, every corpus was subject to a process of cleaning to eliminate irrelevant words or clauses. Irrelevant material includes for example comments from the transcribers indicating that a person laughs or indicating that part of the recording is inaudible. First, I determined which words had to be discarded. This was done using the transcription guidelines, most of the time available together with the corpora. Example (89) and (90) illustrate a series of words and clauses that were weeded out in the French part of the Wikipedia Corpus and in the BNC.

(89) *Cet article est une ébauche concernant. Vous pouvez partager vos connaissances en l'améliorant. (Comment ?).* ' (WIKI FR)
'This section is a draft concerning. You can share your knowledge by improving it. (How?).

(90) *<vocal desc="yawn" />* (BNC)

In (89) it is an automatic message generated by the website and, in (90), a comment of the transcriber indicating that a person is yawning during the interaction. Once the list of irrelevant material was designed, through a combination of automatic searches, I was able to determine how many occurrences of each word or sequences of words had to be subtracted from the total amount of words. This allows to have an overview of the exact number of words produced during the interactions.

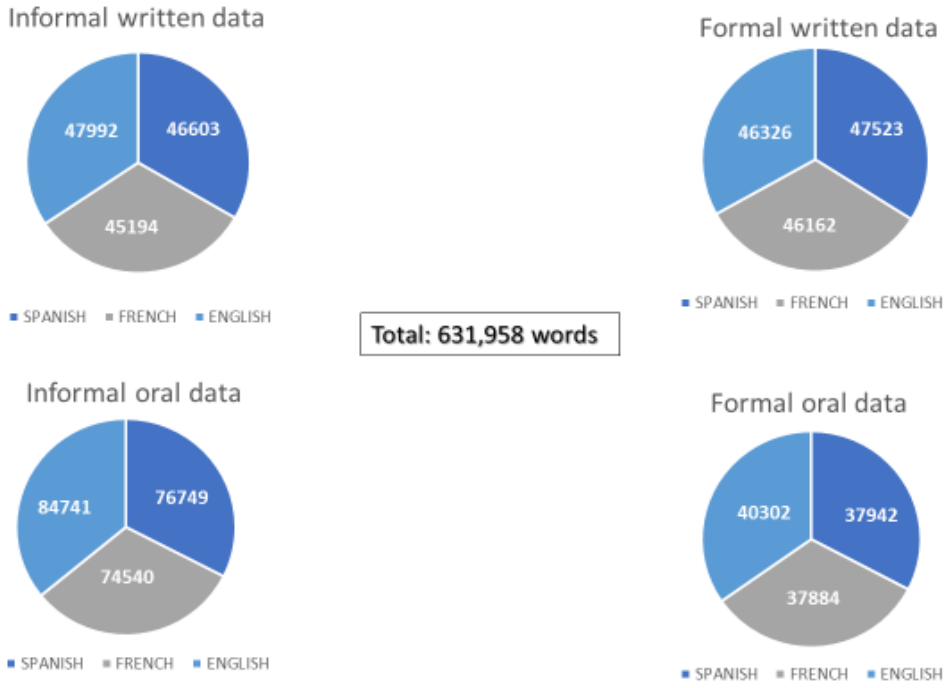



*Figure 5. Number of words per language through the different registers and modes*

Figure 5 shows the total number of words per language, after the cleaning process, through the different registers and modes. It can be observed that, within each register and mode, a perfect balance in terms of size was reached between the three languages. The exact same size between the different subcorpora could not be maintained as the pilot analysis revealed that passive-like structures are less frequent in the informal oral data. As a consequence, the size of informal oral subcorpora had to be increased in order to obtain a sufficient number of occurrences of less frequent structures. The results of the present investigation are therefore be presented through the use of the relative frequency per ten thousand words.

### 5.1.4. Overview of the final datasets

While the same source could be used to collect the data in the three languages for both written formal and formal corpora and the oral formal one, the compilation of the informal oral part relies on several sources.

| Corpus | Language | No. of words | Date | No. of participants | Themes | Control on the language |
|---|---|---|---|---|---|---|
| **Yahoo Corpus-based Question and Answers** | English<br>French<br>Spanish | 139,789 | 2009 | _ | + | _ |
| **Dumps from Wikipedia** | English<br>French<br>Spanish | 140,011 | 2008, 2010, 2014 | _ | + | _ |
| **PROCEP** | English<br>French<br>Spanish | 116,128 | 2011 - 2019 | 47<br>36<br>32 | + | + |
| **British National Corpus** | English | 236,03 | 2014 | 38 | _ | + |
| **Clapi** | French | | 2006 - 2009 | 31 | _ | + |
| **Ciel-B** | | | 2011 | 5 | _ | + |
| **ESLORA** | Spanish | | 2007 - 2015 | 44 | _ | + |

*Table 6. Overview of the features of the corpora*

As appears from Table 6, the final sets of data are nearly balanced with respect to the total size, except for the oral informal part as previously explained. All the texts or recordings can be considered as rather up-to-date productions as they were produced and compiled between 2006 and 2019. Regarding the oral part, the recordings were manually selected, the number of participants could be controlled. A minimum of 32 participants per language was attained in each register. Using such a range of participants helps to avoid idiosyncratic linguistic features that would potentially bias the results. As far as the themes are concerned, through automatic and manual manipulations, it was possible to select substantially similar themes. As indicated in Table 6, they could not be controlled in informal oral data. Several reasons explain this limitation: (i) the scarcity of the data available, especially for the French component; (ii) the features of the mode, as controlling the topics addressed in spontaneous conversations goes in contradiction to the exact nature of this kind of interaction; (iii) the metadata and the internal structure of the original corpora. The selected themes aim at ensuring, when possible, maximum comparability through the subcorpora. Nevertheless, they were selected on the available material and, as the study only includes four themes, they do not pretend to offer a complete representation of the whole language. A last information provided by Table 1 is the potential control on the language of production. A manual revision and selection guarantee that the texts and interactions are produced by native speakers of the target languages for all the oral subcorpora. Unfortunately, the written part of the data does not offer the possibility of verifying this criterion. It is important to consider this inherent limitation when interpreting the results, specifically for the English language as it is widely recognized as a lingua franca and used by non-native speakers.

### 5.1.5. Extraction process

All the corpora created were converted into txt. files. Occurrences of each construction were then extracted from these files by a combination of automatized searches. This process was done via UNITEX software. This program allows to create graphs that correspond to the linguistic phenomenon under study. Thus, I created twelve graphs corresponding to the passive-likes I am interested in. Due to their syntactic characteristics and specificities each language requires its own graphs. Figure 6 illustrates the graph created in UNITEX to extract the periphrastic passive in English.

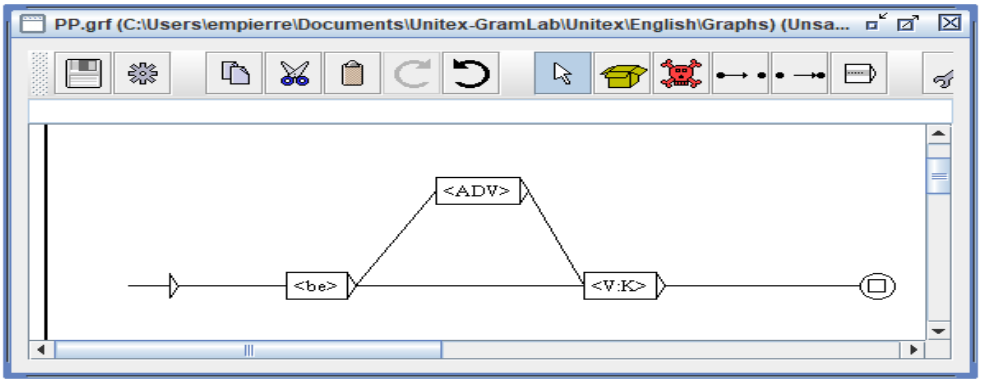



*Figure 6. Example of graph used to extract the periphrastic passive in English*

This graph allows to identify all the forms of the verb *be* (<be>) followed by a past participle (<V:K>) with (91) or without (92) an adverb (<ADV>) inserted between the auxiliary and the past participle. All the graphs created for the other passive-like structures are available in Appendix 1.

(91) *Aircraft **are typically designed** to withstand greater G-forces than the humans flying them.*

(92) *Where are the flying cars? (…) <answer> They already exist. They **are called** airplanes.*

To ensure the software retrieved all the relevant structures, the recall rate was calculated during the pilot analysis stage. The recall rate allows to measure "the proportion of relevant materials retrieved" (Salton 1989: 284). The automated retrieval method proved to be effective as a rate of approximately 95% was obtained.

In a second step a manual revision and selection was operated. Hand-editing enables to disambiguate certain forms and to weed out irrelevant data. The disambiguation process was necessary for all the passive-like constructions. The software does not enable to distinguish, for example, the non-anaphoric third person plural forms from the anaphoric ones, the passive construction based on the reflexive pronoun *se* from the purely reflexive uses of the pronoun, the numeral-based indefinite pronouns *uno* in Spanish and *one* in English from their exclusively numeral uses. Cases of personal *on* (according to Fløttum et al. (2007)'s taxonomy presented in 2.3.2) where discarded. When occurrences were considered as ambiguous, it was decided to include them in the set of data.

Similarly, a manual review was carried out for the periphrastic passive to ensure all the occurrences were purely passive constructions and not adjectival constructions such as (93) where *ouverte* ('open') describes a state and not an action performed by an agent.

(93) *La porte est ouverte (et non pas fermée).*
'The door is open (and not closed).'

(Example Desclés & Guentchéva 1993 :79)

Thoroughly analysing a linguistic phenomenon like passive-like structures implies the need to access the whole context within which these structures appear. This ample need of context is a further reason which justifies the manual selection of the data. The automatic selection only allows a restricted number of words occurring after and before the structure and does not consider full sentences.

## 5.2. Annotation protocol

In order to ease the annotation process, the creation of comparative crosstables and the execution of the statistical analyses, each language, mode, register and passive-like structure was attributed a specific code. These codes can be found in Appendix 2. Similarly, a coding system has been established to annotate the nineteen parameters under study in this investigation. The present section proposes a careful examination of these parameters and presents the coding scheme established for each of these features. The section is divided into three main parts. The first two parts are based on Hopper & Thompson's model and describe the parameters according to their relation to either the verb (5.2.1) or the participants (5.2.2). This structure is largely based on Malchukov's proposal. It has to be noted that section (5.2.1) has been complemented with three parameters considered as bringing relevant information as to the description of passive-like structure: *tense, verb mode* and *type of process. Tense* and *verb mode* are usually not considered as key factors in studies on transitivity. However, Laanemets (2013) indicates in her study on Danish, Norwegian, and Swedish passive voices that the parameter *tense* is central in the selection of the type of passive as some tense forms are agrammatical with certain passive constructions. Considering this finding and the fact that *tense* and *verb mode* are somehow connected to the parameter *aspect*, we assume that their inclusion yields further insights into the understanding of the aspectual-temporal dimension of passive-like structures. The analysis of the *type of verb process* adds a semantic layer to our study, a dimension which is absent from Hopper & Thompson's (1980) approach. Finally, section (5.2.3) describes the parameters related to the linguistic context.

The parameters are either illustrated with examples from the literature or, when more clarification is needed as to how the parameters are understood in the present study, examples from our data are included. All the examples extracted from our data appear in their original form and, subsequently, can contain grammatical or spelling errors.

### 5.2.1. Verb-related parameters

The present section offers an in-depth description of Hopper & Thompson's parameters related to the verb and illustrates how they are understood and classified in this investigation. The section also includes the description of the three additional parameters included in this study.

#### *5.2.1.1. Kinesis*

Among the different properties affecting the degree of transitivity, Hopper & Thompson (1985) highlight the important role played by two closely related parameters *kinesis* and *punctuality*. However, the authors do not provide a precise and detailed enough definition of these two notions. In order to be as precise as possible in the annotation process, a more elaborate account of the distinction between the dynamic/stative and punctual/durative contrasts are developed hereunder.

Hopper & Thompson differentiate dynamic from stative verbs on the basis of the transferability of the action. When the verb is dynamic, the action "can be transferred from one participant to another" (1980:252) whereas when the verb is stative no transfer occurs. Various attempts to develop syntactic tests to distinguish stative from dynamic verbs on more precise criteria have already been made (for example see Dowty 1979). However, these tests do not always prove to be efficient as, in some situations, a verb traditionally classified as stative may be used with a dynamic sense. Compare (94) and (95) where the verb *have* has a stative reading in (94), with *have* understood as *to possess* whereas in (95) *have* dinner is understood as *to eat*, acquiring then a dynamic sense. To overcome this issue, Quirk et al. (1985) rather use the terminology *dynamic/stative situation types*, where the categorization between dynamic and stative considers the whole situation and not merely the verb, instead of *dynamic/stative verbs*.

(94) The chair *has* beautiful carved legs.
(95) We *have* dinner at Maxim's quite frequently.

(Examples from Quirk et al. 1985: 178)

Vendler's seminal model (1967) of lexical aspect proposes a distinction of four verb types: states, activities, accomplishments and achievements. In this model, the first type would be considered as stative and the other three as dynamic situations. This traditional taxonomy of lexical verb types has been challenged by several scholars (for example Van Valin 2006 on the semantic properties of achievements and accomplishments or Fábregas & Marín 2017 on non-dynamic eventive situations). In their influential work on the description of the English language Quirk et al. (1985) suggest a fine-grained division of the different verb types. Their work will serve as the basis for the present section.

The main characteristic to distinguish the two situation types is the following one. A verb with a dynamic meaning is associated with event decomposition (e.g., Quirk et al. 1985 and Gisborne 2001). The verb can be broken down into a sequence of separate events such as in example (96). The verb *have driven* refers to several separate actions of driving, occurring at different

moments. A stative meaning would rather refer to an uninterrupted state such as in (97). In this example, the decomposition is not possible. The verb *known* refers to an uninterrupted state.

(96) *I **have driven** sports cars for years.*
(97) *I **have known** the Penfolds all my life.*

(Examples Quirk et al. 1985: 178)

Within the label *stative situation type* Quirk et al. (1985: 200) includes *qualities* as in (98) where the Canadian nationality is a feature of the person, *states* as (99) where *tired* describes the actual state of Mary, and *stances* where stances refer to situation considered as less permanent (e.g., live, stand or lie).

(98) *Mary is Canadian.*
(99) *Mary is tired.*

(Examples Quirk et al. 1985: 200-201)

Quirk et al. (1985: 203) specify that the category labeled *states* also includes *private states.* They distinguish 4 types of private states: intellectual states (e.g., know, believe, think, suppose, wonder, imagine, realize, understand), states of emotion or attitude (e.g., intend, wish, want, like, disagree), states of perception (e.g., see, hear, feel, smell, taste, appear) and states of bodily sensation (e.g., hurt, ache, tickle, itch, feel cold).

(100) *I heard the bell ring.*
(101) *I could hear the bells ringing.*

(Examples from Quirk et al. 1985: 200-201)

They add that, for perception verb types, when the perceiver is in subject position, "we use the modal can/could followed by the verb of perception. The simple present or past would express dynamic meaning, so that the act of perception would be seen as an event with a defined beginning and endpoint (1985:203)". In example (100), the emphasis lies on the event while in example (101) the modal auxiliary 'could' co-occurs with the perception verb, conferring a stative sense to this latter. The process of hearing is viewed as continuing over a period of time.

Quirk et al. (1985) highlight two major characteristics particular to stative verb types. First, when used with such meaning, these verbs are considered as incompatible with progressive tenses. They may appear, however, with the progressive in specific situations. This use implies a change of interpretation (e.g., Leech 2004 on meaning and verbs). Second, stative verb types adopt a sense of temporariness when used in the simple present. In (102) the verb refers to a clearly limited period while in (103) there is no such time limitation. Example (103) can take on a habitual meaning. Stative situations are by nature considered as durative.

(102) *The manager is away this afternoon.*
(103) *The manager works this afternoon.*

A last feature commonly used to characterize a dynamic verb is its ability to confer a habitual or frequentative interpretation to the situation when used in the present (Vázquez Rozas 2006:6).

(104) *A Ana le **gusta** Juan (*todas las semanas/*cada tarde/ *con frecuencia).*
'Ana **likes** Juan (*every weekend/ *every evening / *frequently).'
(105) *Ana **visita** a su abuela (todas las semanas /cada tarde/con frecuencia)*
'Ana **visits** her grandmother (every weekend/ every evening / frequently).'

In example (104) the habitual interpretation is blocked while in sentence (105) the action can occur frequently.

In order to determine whether the type of situation, namely dynamic versus stative, within which the passive-like occurs has a key role in the distribution of these structures, dynamic verbs receive the code 1 and stative verbs the code 0 in the annotation process.

#### *5.2.1.2. Punctuality*

When defining a dynamic situation, Quirk et al. (1985) distinguish between punctual and durative verb types. A durative dynamic situation includes: (i) activities (106-107), (ii) goings-on defined as "activities carried out by inanimate forces" (1985:207) (108-110), (iii) processes, that is, situations which imply some kind of change (111) and (iv) accomplishments where the action has a specific goal (112).

(106) *Jill **was writing/working/singing** etc.*
(107) *The children **are playing** chess.*
(108) *It **is raining.***
(109) ***Is** your watch **working**?*
(110) *The weather **is getting** warmer.*
(111) *Our economic prospects **are** now **improving.***
(112) *Jill **is knitting** herself a sweater*

(Examples Quirk et al. 1985: 207-208)

A punctual dynamic situation can be of two categories: momentary and temporary situations. Both situations include verbs of events and actions. The difference between momentary events (113) and acts (114) is that the former is not agentive. Quirk et al. (1985) distinguish these categories of situations from transitional events and acts. Again, the difference between transitional events (115) and acts is (116) the non agentivity of events.

(113) *The tops of the trees were waving in the wind, and the branches were shaking and **knocking** against the side of the house.*
(114) *John **was nodding** his head.*
(115) *The train **is arriving** at platform 4.*
(116) *I **am stopping** the car at this garage.* (Examples Quirk et al. 1985: 208-209)

Punctual verbs receive the code 1 and durative ones the code 0.

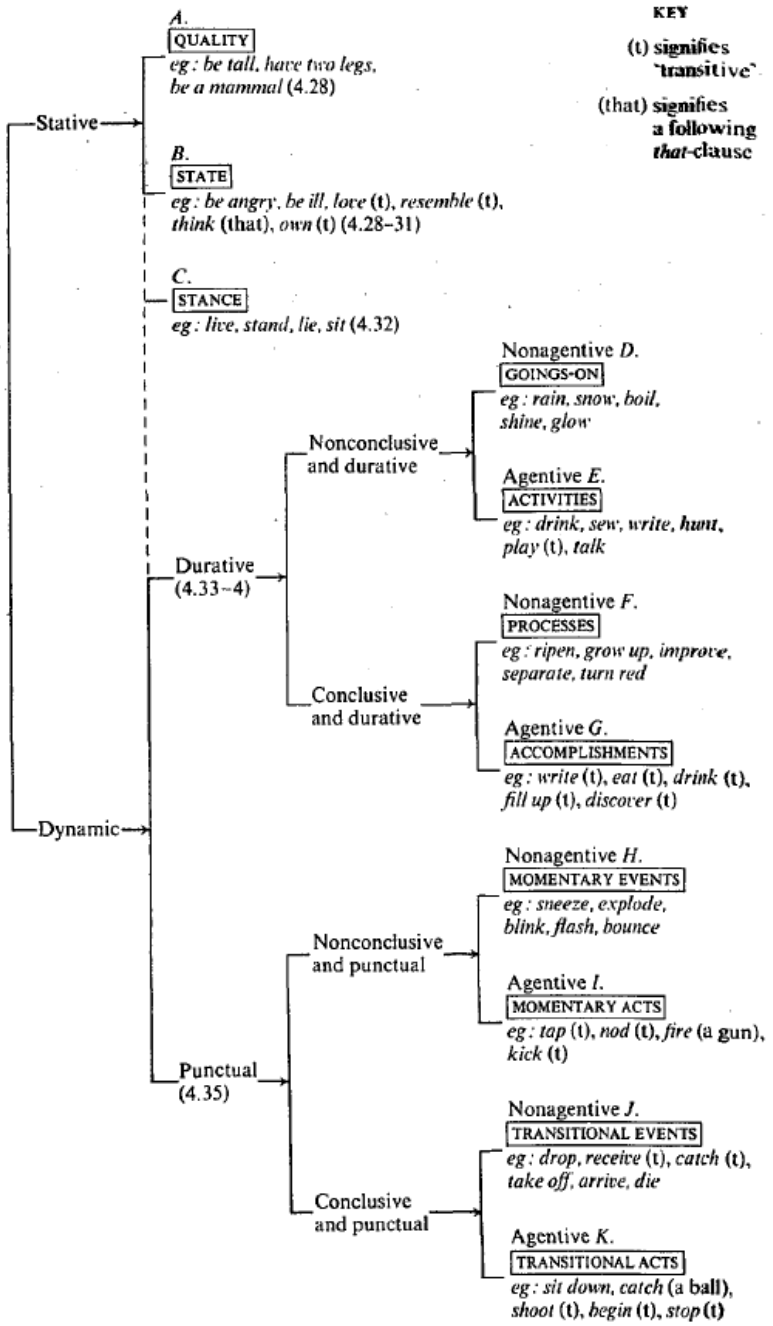



*Figure 7. Classification dynamic vs stative verbs (Quirk et al. 1985)*

Figure 7 from Quirk et al. summarises the different categories of dynamic and stative verbs as well as their relation to punctuality.

#### *5.2.1.3. Aspect*

The relationship between tense and aspect has considerably evolved through time and literature abounds of attempts to describe the notion of aspect (Rojo 1990). Suggesting a detailed

definition of the notion of aspect seems to remain a challenge for grammarians as this concept may differ not only according to which linguistic tradition the researcher is part of but also from one language to another. According to Porzing (1927) in Bache, a key concept in defining aspect is the 'speaker/writer's view of the action or situation described' (1981:5). Most researchers consider that, in English, aspect can be either perfective or imperfective. Comrie (1976:4) differentiates the two in the following terms: "(...) the perfective looks at the situation from outside, without necessarily distinguishing any of the internal structure of the situation, whereas the imperfective looks at the situation from inside, and as such is crucially concerned with the internal structure of the situation (...)". Other grammarians, among which Hopper & Thompson (1980), rather understand the notions of perfective and imperfective in terms of completion: perfective denoting completion whereas imperfective denoting incompletion. Put differently, the action can either be viewed from its endpoint (telic action) or still in process (atelic action) (Hopper &Thompson 1980:252). Within this approach, aspect is conceived in semantic terms rather than in morphological ones (Hopper & Thompson 2006:35) Romance languages resort to similar systems to render the notion of aspect. The most frequent terms used to refer to this notion in French are 'perfectif' and 'imperfectif' or 'accompli' and 'inaccompli', whose exact meaning and acceptation vary from one author to another. In French, according to Riegel et al. (1994), the terms 'perfectif' and 'imperfectif' refer to the semantic aspect (considered as Aktionsart[19] in the German linguistic tradition (Wilmet 1991)) of the verb rather than to the grammatical aspect. To illustrate his claim Riegel et al. (1994) use the verb *sortir* ('to go out') which can only be conceived as a completed process when the action of *sortir* is totally realized, namely when the person crossed the doorstep and is out (1994:293). The perfective or imperfective nature of the verb, in this approach, is thus intrinsically linked to the lexical meaning of the verb. In the grammatical approach, the opposition *accompli/inaccompli* is rather manifested through the use of different tenses. Wilmet (1991) adopts a similar view on *aspect* with a semantic approach, which involves the nature of the verb, and a formal one, rather based on the formal characteristics of the verbal argument. In Spanish, the notion of aspect is usually referred to with the terms 'perfectivo' and 'imperfectivo'. Rojo (1990:33) explains that basically, these terms refer to the opposition between an action seen as finished and an action still ongoing, which matches Comrie's view (1976) of aspect.

When referring to *aspect*, this study adopts the completion approach of the concept. The following two examples, extracted from our data, illustrate a perfective form (117) and an imperfective one (118) as they are understood and classified in the present investigation.

perfective: 1

(117) ***Se perdieron*** *viviendas y negocios* (PROCEP SP)
'Homes and businesses **were lost**.'

imperfective: 0

(118) *Tampoco* ***se puede tolerar*** *que pretendan impedir que los trabajadores del transporte reciban al menos el salario mínimo del país donde operan (...).* (PROCEP SP)

[19] Some authors dissociate the notion of *aspect* from *Aktionsart* (see. e.g Bach 1982). For detailed information on these two notions and an ample discussion on this often debated issue see among others Goedsche 1940, Bach 1982, Sasse 1991 or Ramsay 1996.

'Nor can **it be tolerated** that they seek to prevent transport workers from receiving at least the minimum wage of the country where they operate.'

#### *5.2.1.4. Volitionality*

Hopper & Thompson (1980) assume that a clause will receive a higher degree of transitivity when the action is carried out purposefully compared to clauses expressing accidental actions. DeLancey already mentioned a relationship between agentivity and volitionality (1990). Examples (119) and (120) illustrate the contrast between volitional and non-volitional verbs. While in (119) the decision to place the plant in full sunlight is deliberate, the confusion in (120) is unintended.

Volitional verb : code 1

(119) *le yucca d intérieur ne **se met** pas en plein soleil (1mn de soleil est suffisant) par contre il doit rester à la clarté* (YAHOO SP)
'the indoor yucca cannot **be placed** in full sunlight (1mn of sun is enough) however it must remain in the light (YAHOO).'

Non-volitional verb: code 0

(120) *El daño por los agentes físicos **se puede confundir** por ataque biótico, pero la carencia de muestras visibles de los hongos, insectos, o perforadores marinos (...).* (WIKIPEDIA SP)
'The damage by physical agents **can be confused** by biotic attack, but the lack of visible samples from fungi, insects, or marine drillers (...).'

It is assumed thus that volitionality impacts the mechanism selected to defocus the agent. Existing studies already show a connection between volitionality and impersonal structures (see e.g., Arnett 1997, Primus 2010 for German, Beliën 2016 for Dutch or Fernández 2007 for Spanish).

#### *5.2.1.5. Affirmation*

According to Hopper & Thompson's approach (1980), the polarity of a clause contributes in determining its degree of transitivity. Occurrences received the code (1) when affirmative and the code 0 when negative. All types of negation are taken into account. For instance, complex negations such as *ne… plus* in French or negative adverbs such as *never* in English are also included and codified with the code 0.

Affirmative sentence: 1

(121) *Madam President, I would like to thank the Members of the House for their support and for highlighting the **issues that have been raised**.* (PROCEP)

Negative sentence: 0

(122) *Toutes ces méthodes prennent du temps, ont leurs limites et il arrive qu'un aéroport doive fermer pour de longues périodes car* ***on ne peut plus dégager*** *les pistes sans y passer un temps fou.* (YAHOO)
'All of these methods take time, have their limits, and sometimes an airport has to close for long periods of time because the runways **can no longer be cleared** without spending a lot of time on them (YAHOO).'

(123) *Many other rules of religious orders contained similar restrictions of diet, some of which even included fowl, but fish* ***was never prohibited****, as Christ himself had eaten fish (Luke 24:42-43).*

#### *5.2.1.6. Dichotomy realis/irrealis*

When clauses encode real events, Hopper & Thompson (1980) consider that they receive a higher degree of transitivity. The authors distinguish between the 'realis' and 'irrealis' ways of encoding an event. The irrealis mode corresponds to a situation where the action or the event did not occur or "is presented as occurring in a non-real (contingent) world" (1980:252). I therefore understand the concept of 'irrealis' in a broad sense of the term and not as merely counterfactual. For the purpose of this study the addition of a negative particle, the use of future and conditional tenses and the use of the subjunctive and modal indicators such as exemplified in the following sentences are defining criteria to distinguish realis clauses from irrealis clauses.

Realis: code 1

(124) } *Aujourd'hui* ***on recherche*** *davantage le transport de masse au moindre coup, avec des paris différents selon Boeing (787 pour lignes "longues et fines : long courriers avec relativement peu de passagers) et Airbus (A380). (Yahoo)*
'Nowadays **we are looking** more for mass transport at a low cost, with different bets according to Boeing (787 for "long and thin" routes: long haul with relatively few passengers) and Airbus (A380).'

Irrealis: code 0

(125) *Plants from the ground that are not washed properly may contain B12 from bacteria in the soil, often from faeces ; drinking water* ***may also be contaminated*** *with B12-producing bacteria (…). (WIKIPEDIA)*

(126) *your training route will reflect your ability - only the cream become Fighter pilots{S} C you* ***will be offered*** *another non-flying trade, but you will be free to resign your commission should you wish to do so. (YAHOO)*

While example (124) illustrates a passive-like structure classified as realis, thus actualized, anchored in the real world, examples (125) and (126) encode irreality since (125) contains the modal indicator *may* and (126) resorts to a future tense.

Note that Hopper & Thompson's use of the term *mode* does not exactly match the term *modality* usually found in the literature to refer to the fact "that a situation is represented as actualizing in a nonfactual world" (Declerck 2011: 22). To avoid any confusion the present investigation uses the label *dichotomy realis/irrealis* to refer to the abovementioned parameter. The notion of *verb mode* will further be described in Section 5.2.1.8.

All the parameters commented above are part of Hopper & Thompson's theory (1980). The last part of the section details three parameters that complement those suggested by Hopper & Thompson (1980). The first two parameters – i.e., tense (5.2.1.7) and verb mode (5.2.1.8) – are two inter-related syntactic parameters linked to the verb phrase. The third parameter, namely the type of verb process, which allows to add an additional semantic layer to the analysis is described in (5.2.1.9).

#### *5.2.1.7. Tense*

Grammarians usually consider that there are two tenses in English, namely the present and the past (Dowty 1979, Quirk et al. 1985, Huddleston and Pullum 2002, Kaufmann 2005). Yet, according to Lyons (1977), 'the past' and 'the non-past'. This division reflects the grammatical features of the verb rather than actual temporal nuances encoded in the real world (De Capua 2008:167). A second approach considers the future tense in addition to the present and the past, thus distinguishing three tenses (Comrie 1982, Thomson & Martinet 1986, Gennari 2003, Jespersen 2003). This tripartite division of tense is adopted to better understand the temporal dimension within which the passive-like structures. Distinguishing present from future actions and states may be relevant when studying agent-defocussing mechanisms and may bring further information as their respective use. No further distinction between the tenses will be provided, as this investigation does not aim at providing an exhaustive description of how the conjugation system works in the three languages, nor at offering exact translational equivalents of tenses in the three languages. A number of tenses, despite being comparable from a formal point of view, differ in their usage. As pointed out, for example, by Vinay & Darbelnet for French and English, "(…) the list of tenses of the two languages is about equal; however, the distribution of the tasks they have to perform is not at all comparable. Moreover, the fact that English can express every tense in the progressive form, gives its conjugation system a different orientation" (1995:130).

Given this complexity, the tripartite coding scheme used here serves as a basis to situate verbs in the three main dimensions of the time scale and identify whether there is a correlation between the type of passive-like structure selected and its realization either in the present (code 1), the past (code 2) or the future (code 3). Two specifications should, however, be laid down. The first one concerns the status of the conditional which, depending on how it is used, can either be considered as a tense, especially when the idea of condition is absent and the conditional is used in a subordinate clause such as in (127), or as a mode (exemplified in (128)) (see Grevisse 2008). This phenomenon is particularly marked in French.

(127) *Il m'a dit qu'il reviendrait ce soir.*
'He told me he would come back tonight.'

(128) *Elles [= des maisons] nous DIRAIENT des choses à pleurer et à rire, si les pierres parlaient.*
'They [= the houses] would told us things to cry and to laugh at, if the stones talked.'

(Examples Grevisse 2008: 889)

Although grammar books tend to present the conditional as a mode (Merle 2001: 10), the issue seems to be almost settled since the main view tends to abandon the trend to consider the conditional as a mode and rather treats it as a tense from the indicative (Riegel et al. 1994, Abouda 1997, Merle 2001, Grevisse 2008). This approach is the one adopted in the present study. Although debatable, the choice was thus made to include conditional occurrences under code 1 or 2 according to the interpretation of the situational context. Since the pilot analysis revealed that passive-like structures occurring with conditionals are relatively infrequent in our data set, no further distinction will be made.

Similarly, and as highlighted by Vinay & Darbelnet (1995: 133), the imperfect (*imparfait* in French and *imperfecto* in Spanish) is, sometimes considered as an aspect. Nevertheless, as the present classification system does not aim to establish a precise comparative synthesis of how tense, aspect and modality are connected in the three languages, occurrences of such type are included in the past category (code 2).

A second clarification has to be made regarding the classification of future tenses. The expression of future values can be realized through the use of a periphrasis including a verb form in present tenses as exemplified in (129) and (130).

(129) ***Se van** a celebrar pronto las elecciones.* (Bosque & Demonte 1999: 3366)
**'**Elections will be held soon'

(130) ***On va** partir en Espagne le 27 juillet 1994.* (Helland 1995:19)
'We will go to Spain on the 27th of July 1994.'

Several researchers (Werner 1980 or Vet 1993 among others) specify that, in the past, this construction was not considered as a future tense but rather classified as a present tense. To facilitate the annotation process, this work considers the syntactic form of the periphrasis rather than its semantico-pragmatic intention and therefore classifies it as an instance of a present tense.

#### *5.2.1.8. Verb mode*

Establishing a consistent taxonomy to classify the different devices available to express modality in the three languages presents a considerable challenge.

The starting point of this classification is grounded in the work of Riegel et al. (1994), which distinguishes 5 modes in French, i.e., the indicative, subjunctive, imperative, infinitive and participle. In their classification, the authors associate the gerund with the participle (1994: 287). This classification will be slightly modified and adapted to the crosslinguistic needs of this study. It is indeed difficult to match the three languages since they rely on different classification systems. The gap can be particularly wide between Romance languages, which present fairly similar systems, and the English language. Although we acknowledge that classifying modality is a much more complex issue, we will restrict ourselves to the categories presented in Table 7.

| Verb Mode | Code |
|---|---|
| Indicative, assertive | 1 |
| Subjunctive | 2 |
| Modal auxiliaries | 3 |
| Others | 4 |

*Table 7. Classification of the parameter verb mode*

Table 7 summarises the system used to codify modality. All the tenses pertaining to the indicative or assertive mode (past, present or future) are marked under code (1). The subjective is assigned code (2). It seems important to codify the subjunctive mode separately, as it is largely used in Spanish and, to a lesser extent, in French. The use of the subjunctive is less widespread in English and similar nuances tend to be rendered through modal auxiliaries, which represent the third category (code 3). This class gathers the auxiliaries *can*, *may*, *might*, *should*, *could* and *would* in English and the three modal verbs in French *pouvoir*, *devoir* and *falloir* and their counterparts *poder* and *tener que* in Spanish. It is hypothesized that other modes like the gerund, the infinitive or the imperative will rather be infrequent in our data set and are, therefore, grouped under the category 'other' (4).

#### *5.2.1.9. Type of verb process*

The last parameter added to complement those suggested by Hopper & Thompson's (1980) is the type of verb process. The analysis of the type of verb process allows to add a lexical dimension to the present study, which is absent in Hopper & Thompson's work. The analysis of the different types of verbs relies on a classification available in the ADESSE database (Alternancias de Diátesis y Esquemas Sintáctico-Semánticos del Español)[20] designed by the University of Vigo (Spain). The database recognizes six main categories of verbs, i.e., mental, relational, material, verbal, existential and modulation. Tables 8 and 9 present and illustrate the

[20] For more detailed information on the classification, the taxonomy is available online: http://adesse.uvigo.es/data/clases.php. (last accessed 16/05/21)

sub-classification of these six central categories. The code they receive in this study is specified in parentheses.

| **Mental 1** | **2 Relational** | **3 Material** |
|---|---|---|
| Sensation (11)<br>*to love, to desire* | Attribution (21)<br>*to be, to cost* | Space (31)<br>*to go, to sit* |
| Perception (12)<br>*to see, to hear* | Possession (22)<br>*to have, to pay* | Change (32)<br>*to break, to create* |
| Cognition (13)<br>*to believe, to know* | | Other facts (33)<br>*to rain, to touch* |
| Election (14)<br>*to chose, to decide* | | Behaviour (34)<br>*to cry, to eat* |

*Table 8. Types of processes (1)*

| **4 Verbal** | **5 Existential** | **6 Modulation** |
|---|---|---|
| Communication (41)<br>*to tell, to speak* | Existence (51)<br>*to occur, to exist* | Causation (60)<br>*to help, to allow* |
| Assessment (42)<br>*to criticize, to congratulate* | Phase-Time (52)<br>*to begin, to last* | Disposition (61)<br>*to dare, to try* |
| Sound emission (43)<br>*to sing, to howl* | Life (53)<br>*to kill, to live* | Acceptance (62)<br>*to accept, to reject* |
| | | Verbs of support (64)[21] |

*Table 9. Types of processes (2)*

The parameters described so far are related, though in a different manner, to the verb. The next part of this chapter is devoted to the description of the participant-related parameters.

### 5.2.2. Participant-related parameters

This section details the four aspects related to the participants in a clause. First it defines the notion of participant (5.2.2.1). Second, it discusses the notion of animacy of these participants both from the perspective of the agent and the patient. Third, it details the notion of affectedness of the patients and, lastly, the degree of individuation of the patients.

#### *5.2.2.1. Participant*

It is assumed that the number of participants implied in a clause may be a factor in the selection of the passive-like construction. Quirk et al. define the notion of participant in the following manner: "Every clause describes a situation in which a number of participants are involved. By PARTICIPANTS, we understand entities realized by noun phrases, whether such entities are

[21] Verbs of support are verbs that combine with nominal clauses to form semantically complex phrases (ADESSE) which confer them another meaning than the one conveyed by the base verb.

concrete or abstract" (1985:740). As pointed out by Vázquez Rozas (2006), Hopper & Thompson (1980) do not offer an explicit definition of the notion of participant as used in their work. Following the position adopted by Vázquez Rozas (2006) in her study of Spanish *gustar*-type verbs[22], in addition to the agent and the direct object, we will also consider indirect objects as participants. The *agent* is here understood as the actor, instigator or initiator (Siewierska 2008:7) and the *patient* as "an O[bject] which is in fact the 'receiver' of the action in a cardinal transitive relationship" (Hopper & Thompson (1980:252). In the present study, *they* and *on* have then been classified as agent. We acknowledge that these R-impersonal can in some cases endorse a semantic role other than the one of agent in a strict sense (some uses tending towards an experiencer role were observed). Note, however, that, while these R-impersonals can fulfil a patient role, this was not observed in our datasets. The choice to classify these forms as agents (rather than adopting a more fine-grained perspective) makes it possible to compare the passive-like structures on a common ground regarding agents and is mainly made in view of the larger comparative purposes of this study. Though a more detailed distinction of their semantic role goes beyond the scope of the present study, information on their status can also be obtained through the analysis of the type of verb process, presented in the previous section.

The participant may be overtly realized in the clause or retrievable from the verbal flexion in case of pro-drop languages (131).

The action involves one participant: code 0

(131) *No creo que el hombre halla pisado nunca la Luna, en 1969* ***llegaron*** *sin problemas (según la version oficial) y hoy en dia todos los coetes que se mandan tienen fallos al despegar, cómo se explica esto? (YAHOO)*
'I don't think that man has ever stepped on the moon, in 1969 **they arrived** without problems (according to the official version) and today all the rockets that are sent have failures when they take off, how do you explain this?'

The action involves 2 or more participants: code 1

(132) *{S}Monsieur le Président, Madame la Commissaire, les choses sont simples.{S}* *Elles* ***ont été exposées*** *par mes collègues.* (PROCEP)
'Mr President, Commissioner, things are simple. They **have been outlined** by my colleagues (PROCEP).'

Note that, following Hopper & Thompson coding system (2006:31), only noun phrases will be included under the label *participants* in this investigation. Complement clauses such as in (133 and 134) introduced by *why* or the conjunction *que* ('that') are not counted as participants even though the complement clause functions as direct object of the verb. Since these clauses do not

[22] *Gustar* in Spanish differs from the English verb *to like* as "whereas *like* codes as subject the entity that experiences a certain feeling, and as object the stimulus responsible for that feeling, *gustar* expresses the experiencer though an indirect object (or dative) and the stimulus through the subject" (Vázquez Rozas 2006: 80).

consist in a sole noun phrase, the analysis of the animacy of the patient and the parameters *individuation* and *affectedness of the patient* would not be applicable.

(133) *I was wondering why I hadn't heard from him*

(Example from Hopper & Thompson 2006:31)

(134) *H1 y notaba que se acercaba / y claro era ya de noche / y por los focos Pausa H1* ***se veía*** *que no / que no era Pausa*
*H1 por la forma Pausa H1 y resulta que era uno PausaH1 que Pausa (ESLORA SP)*
'H1 and note that he was approaching / and of course it was already night / and by the lights Pause
H1 **we could** see that no / no Pause
H1 by the form Pause H1 and it turns out it was one Pause'

#### *5.2.2.2. Animacy*

Hopper & Thompson assume that "participants high in agency can affect a transfer of an action in a way that those low in Agency cannot" (1980:252). Nevertheless, they do not provide any delimiting criteria to define what it is understood by high and low in Agency. In order to establish clear bases for the annotation process of this parameter I call on Fauconnier's (2012) approach of animacy.

Agents are typically animate but Fauconnier points out that the notion of animacy "(...) is not just a simple matter of living versus non-living" (2012: 37). She creates several categories to classify entities considered as being inanimate since a series of elements functioning as agent can be of different nature than either human or animal: organizations, countries, machines, supernatural or mythical beings, abstract concepts, forces of nature, and physical objects. As a highly detailed classification of the type of agent is not the main focus of this investigation and is not considered as a key element in the literature on passive-like constructions, I will reduce the classification to two general categories: animate (further subdivided into human and animal)

and inanimate (subdivided into three: organization, country or city and others). This taxonomy (see Figure 8) is also applied to patients and receives the same codes.

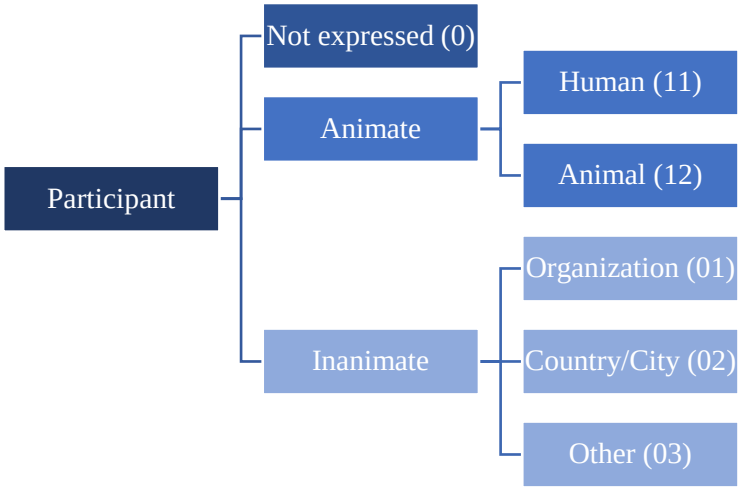



*Figure 8. Classification of participants with their corresponding code in parentheses*

For the periphrastic passive, only prototypical agent complements introduced by the preposition *by* in English, *por* in Spanish and *par* in French are taken into account. As pointed out by Delbecque (2014) agent complements can be introduced by different types of prepositions. This idea is further developed in Section 5.2.4.6 and receives a specific coding system.

Some agent complements, called *janus agents* (Svartvik 1966:105), allow two different readings of the passive construction such as exemplified in (135).

(135) *Oil will **be replaced** by coal.*

(Example from Svartvik 1966: 105)

Either the sentence can be understood as *coal will replace oil*, where *coal is considered as the doer*, or as *[they] will replace oil by coal*, where a third entity, totally defocused, is considered as performing the action (Svartvik 1966:105). Even though in one of the potential interpretations the adjunct does not function as the doer of the action, janus agents are included in the present study as they fit the definition of instigator suggested by Siewierska (2008).

#### *5.2.2.3. Affectedness of the patient*

According to Hopper & Thompson (1980) when the patient is totally affected by the action, the degree of transitivity will be higher. In the clause *I drank up the milk* the patient is totally affected while in *I drank some of the milk* the patient is partially affected (Hopper & Thompson 1980:253).

Totally affected: code 1

(136) *Le rhizome **peut être fixé** sur un élément du décor comme une pierre ou une racine à l'aide d'un fil en nylon. (WIKIPEDIA)*

'The rhizome **can be attached** to a decorative element such as a stone or a root with the help of a nylon thread.'

Partly affected: code 0

(137) *Aux États-Unis, plus de la moitié de l'eau consommée **est utilisée** pour la production de bétail. (WIKIPEDIA)*
'In the United States, more than half of the water consumed **is used** for livestock production.'

This parameter is not applicable to clauses lacking an object, in which case, the code 99 is used.

#### *5.2.2.4. Individuation of the patient*

When Hopper & Thompson (1980) designed their framework to calculate the degree of transitivity, they believed that a patient that is individuated receives a higher degree of transitivity. They listed six features that form a highly individuated object[23]. Such patient should be proper, human or animate, concrete, singular, countable and referential or definite. As the authors did not suggest a scale to measure the degree of individuation nor a threshold, in our data, the occurrences that meet four of the above-mentioned criteria were coded as highly individuated. As the authors did not explicitly define the notion of object, it has been decided that in case of the presence of both a patient in direct and an indirect object positions, the priority is given to the one in direct object position. When there is only an indirect object and no direct object, the patient in indirect object position is taken into account.

Individuated : 1

(138) *Te h**an mentido** vilmente, eso es falso.* (YAHOO SP)
'**They have lied** to you vilely, that is false.'

High transitivity features: human, concrete, singular, referential

Non-individuated: 0

(139) *Mr President, there is quite adequate legislation on animal transport: the problem is that it **isn't being enforced**, and here we find a cultural difference.* (PROCEP EN)

High transitivity features: singular, countable

This parameter is not applicable to clauses lacking a patient such as in (140). If so, the code 99 is used.

(140) ***Il est recommandé de** la rempoter dans un mélange bien drainé, composé de sable et de terreau de rempotage.* (YAHOO FR)

[23] For the purpose of this analysis, the notion of object is understood as the entity that is affected by the process encoded in the verb. In the sentence *Drinking water may also **be contaminated** with B12-producing bacteria (...)* the nominal clause *drinking water* will be considered as the object.

‘**It is recommended** to repot it in a well-drained mixture of sand and potting soil.’

### 5.2.3. Context-related parameters

When dealing with passive-like structures many researchers highlight the importance of taking into account contextual factors (e.g., Monge 1954, Fløttum et al. 2007, Gjesdal 2008, Landragin & Tanguy 2014, Gast 2015, Yao 2016, De Cock 2020).

What is considered here as the linguistic context varies according to the register and the modality. For the Yahoo Question and Answers corpus, when the passive-like structure is used in an answer I consider both the answer and the question (as illustrated in 141), as some clues guiding the analysis of the passive-like structure included in the answer may be influenced by elements already introduced in the question.

(141) *<question>¿Podemos afirmar que la gastronomía mexicana no está dentro de las más saludables?Me refiero a gastronomía, no a dieta, lo cual es diferente. Esto lo pregunto porque es evidente que en la comida mexicana tradicional no se hace mucho uso de vegetales o se combinan mal; no hay muchas infusiones, como tés, los cuales son antioxidantes, lo mismo que las verduras. Asimismo, hay platillos mexicanos que hacen uso de manteca de cerdo para la cocción, la cual es grasa saturada. Espero su opinión. <question/>*
*<answer> No se si eres mexicano, pero evidentemente tienes muy pocos conocimientos de la comida mexicana, no* ***se cocina*** *con manteca, eso solo lo llevan los tamales por ejemplo (...) <answer/>* (Yahoo SP)
‘<question> Can we say that Mexican cuisine is not among the healthiest? I am referring to gastronomy, not diet, which is different. I ask this because it is obvious that traditional Mexican food does not make much use of vegetables or they are poorly combined; there are not many infusions, such as teas, which are antioxidants, as well as vegetables. Also, there are Mexican dishes that use lard for cooking, which is saturated fat. I look forward to hearing your opinion. <question/>
<answer> I don't know if you are Mexican, but you obviously have very little knowledge of Mexican food, it **is not cooked** with lard, that's only in tamales for example (...) <answer/> ‘

For the Wikipedia Corpus and the PROCEP corpus I look at approximately four to five sentences before and after the structure. In some cases, it is not possible to have such extension as the impersonal structure is used at the beginning or the end of the intervention.

This section examines different factors of the linguistic context that might help identify the potential hidden agent and influence its interpretation. According to the literature, six factors have been listed and, according to their presence/absence, are considered as playing a potentially decisive role in the identification process. These factors are the following: the presence of a person reference (5.2.3.1), the use of an adverb or adverbial complement (5.2.3.2), the presence of a temporal/spatial restriction (5.2.3.3), the presence of another passive-like

structure (5.2.3.4), the type of context in which the passive-like structure is used (5.2.3.5), and the presence of a prepositional phrase (5.2.3.6). These contextual factors are described and exemplified hereunder.

#### *5.2.3.1. Person reference in the context*

The co-occurrence of the passive-like structure with person references can provide valuable indications on which participants (the addresser, the addressee or a third entity) are potentially involved in (or discarded from) the impersonal process. Several authors have already demonstrated the impact of person deixis on the interpretation process of impersonal constructions (Hugo Rojas 2011: 164, De Cock 2014, Lavandera 1984, Company and Pozas Loyo 2009: 1206), either leading to downgrade the scope of the reference or, on the contrary, to provide a more generalizing scope.

Person deixis is commonly expressed through pronouns, possessive affixes, vocatives and agreeing predicates (Levinson 1983: 70). In this dissertation, the expression 'person reference' also includes non-deictic forms, mainly nouns, which, compared to deictic ones, "maintain their referentiality also when isolated from the discourse production" (De Cock 2014: 23). When person references to several participants are present in the context, the one assumed to influence the identification process of the agent the most is considered.

| **Person reference** | **1st person singular** | **2nd person singular** | **3nd person singular** | **1st person plural** | **2nd person plural** | **3nd person plural** | **Absence of reference** |
|---|---|---|---|---|---|---|---|
| **Code** | 1 | 2 | 3 | 4 | 5 | 6 | 0 |

*Table 10. Coding system for the parameter 'person reference'*

Table 10 displays the codes that have been assigned to the different forms of person references. The absence of reference is also considered and receives the code 0. Examples 142 and 143 illustrate two cases where the passive-like structure co-occurs with a first person singular reference.

(142) *hola que tal mi pregnta es como es la vida de un piloto, si es divertida, presionada, yo se que si **uno** hace lo q le gusta no sera aburrido pero me gustaria saber de alguien que ya lo aya experimentado y vivido. (Yahoo SP)*
'Hello, my question is how is the life of a pilot, if it is fun, stressful, I know that if **you** do what you like it will not be boring but I would like to know from someone who has already experienced and lived it.'

(143) So I cannot endorse the report **as it was drafted**, but nor will I vote against it. {S} (PROCEP EN)

Note that the presence of a person reference in the context facilitates the interpretation but does not necessarily imply that that participant in particular is involved in the passive-like structure.

#### *5.2.3.2. Adverbial complement*

Several authors have already highlighted the importance of adverbial constructions to analyse the use of impersonal constructions (Legendre 1990 in French; Hernanz 1990, Bidot Martínez 2008 in Spanish). Adverbs are usually defined as an heterogeneous class of invariable words (Hernández Alonso 1974, Barrenechea 1978) which can take on different forms and can express a wide range of meanings. Usually, this grammatical category functions as a circumstantial complement of the verb, but, less frequently, may also co-occur with an adjective or another adverb (Alarcos Llorach 2000). The adverb either consists in one word or two or more items. Both types of adverbial formations are included in the present study. Two options are generally considered to classify adverbials. They are either based on their meaning or their function (e.g., Alarcos Llorach 2000, Gómez Torrego 2015). The classification adopted in this study is grounded in a semantic framework and considers the following central categories as impacting the interpretation and selection of a passive-like structure, each time illustrated in the three languages: *time, place, mode* or *others* (such as quantity, affirmation, doubts etc.).

The adverb can be placed in-between the two parts of the past participle, such as *requete* in (144).

(144) *Segun la información oficial es que la Luna ya fue explorada al 100 porciento, o sea no hay nada nuevo que ver, o buscar. Por ello el desarrollo aeroespacial se ha destinado a marte y otros planetas. Ya el satelite natural* ***se ha requete estudiado.*** (YAHOO SP)
'According to official information, the Moon has already been explored 100 percent, so there is nothing new to see or look for. That's why aerospace development has been directed at Mars and other planets. The natural satellite has already been studied a lot.''

In case of co-occurrence of two adverbs such as in (145) the item closest to the verb was considered. Though debatable, this choice enabled a constant analysis of the phenomenon.

(145) *Algo que otros países, también en esas regiones, no tienen, y que aquí jamás se denuncia.* (YAHOO SP)
'Something that other countries, also in those regions, do not have, and which is never denounced here.'

Given the ample scope of this study, only the presence of any type of adverbial construction related to the passive-like construction is indicated by means of the following coding system: presence of an adverb: 1 and absence of an adverb: 0.

#### *5.2.3.3. Spatio-temporal restriction*

When analysing impersonal structures several authors (e.g., Hugo Rojas 2011, Kluge 2012 or De Cock 2014) have pointed out the key role of space-builders. De Cock (2014: 113) reports that these items can take on three different forms: (i) physical places such as *en Moscú* in (146),

(ii) conceptual places such as *debate complete* in (147) or (iii) temporal space *en los minutos finales* in (148).

(146) *Es la situación en Moscú.*
Is the situation in Moscow.
'It's the situation in Moscow.'

(147) *Estamos en un debate completo.*
Are-1pl in a debate complete.
'We are in a complete debate.'

(148) *Estamos ya en los minutos finales y son las puntualizaciones (…).*
Are-1pl already in the minutes final and are-3pl the making-points (…).
'We are already in the final minutes and it's time to make the final points (…).'
(Examples from De Cock 2014: 114)

The use of such indicators can provide relevant information on the participants that are potentially included in the interaction. It is assumed that the presence of an adverbial construction or of a space-builder may impact the selection of the impersonal structure and facilitate its interpretation, either restricting its range or functioning as a generalizing operator. Similarly to the methodological choice adopted for the annotation of adverbial complements, since the focus of the present work is much larger than the study of space-builders, only the presence (code 1) versus absence (code 0) of the three forms of space-builders reported by De Cock (2014) will be coded.

#### *5.2.3.4. Presence of another passive-like structure*

The presence of another passive-like structure (be it the same as the one analysed or another type of passive-like structure) in the linguistic context is assumed to influence the process of interpretation (see for instance Fernández Soriano & Táboas Baylín 1999 or Bidot Martínez 2008). The impersonal strategies taken into account are those studied in the present investigation as well as the generic use of the second person singular.

Presence of another passive-like structure: 1

(149) *Pomelo es lo q **se conoce** como toronja. En algunos paises hsipanoparlante les llaman pomelos en otros toronjas.(YAHOO)*
'Grapefruit ['pomelo'] is what **is known** as grapefruit ['toronja']. In some Spanish-speaking countries they called them grapefruits ['pomelo'] in other countries grapefruits ['toronjas']. (YAHOO)

Absence of other passive-like structures: 0

The codification process allows to identify whether the passive-like structure co-occurs (code 1) or not (code 0) with another impersonal strategy. A further coding layer distinguishing the exact type of impersonal strategy was, unfortunately, beyond the scope of this study.

#### *5.2.3.5. Type of context*

The type of context in which the passive-like structure appears and the content of the discourse (production) can also influence the choice of the structure and the interpretation of the latter. The analysis of the discourse content can provide clues on how to interpret the structure as, as pointed out by several researchers in European languages, lexical devices can guide towards different types of interpretations (e.g., Bermúdez 2002 on evidentiality in Spanish, Rasson 2018 on *uno* or Fernández 2007 on impersonal constructions).

In a similar approach as the one adopted in Rasson's (2018) work, three types of contexts can be distinguished, i.e., generic/indefinite (Examples 150 and 151), restrictive, or neutral (absence of any indicator). Various grammatical elements may play a role in determining the type of context, such as adverbs or adverbial phrases, nouns, or nominal phrases favoring a generic interpretation, as illustrated in examples (150) and (151).

(150) *[e- que siempre que hay] / siempre que hay una crisis / **se resuelve** con una guerra (ESLORA)*
'[e- that whenever there is] / whenever there is a crisis / it is resolved with a war'

(151) *Por una parte, este tipo de vegetarianos consideran que, al ser los recursos naturales limitados, **se ahorra** si el ser humano se alimenta directamente de los vegetales, en lugar de utilizarlos como comida para los animales. (WIKIPEDIA)*
'On the one hand, this type of vegetarians considers that, since natural resources are limited, it can be saved on if the human being directly feeds him/herself with vegetables, instead of using them as food for animals.'

A restrictive context is typically favored by the presence of personal experience, advice, request, criticism (Rasson 2018: 90), that is, situations that highly involve either the speaker or the hearer's face. In (152) the speaker congratulates the hearer which thus involves a subjective judgement from the speaker. In the same example, the speaker uses the imperative *colócalo* ('put it') to gives advice to the interlocutor. In (153) and (154), the passive-like constructions appear in a context where the speakers explain personal experiences they have had with dogs.

(152) Glaviole: *te felicito por la desición que has tomado, aunque no se si el pequeño árbol podrá recuperarse, no **se pierde** nada con probar. Después que lo laves bajo el chorro de agua, colócalo en una maceta que no sea muy grande y ve cambiándolo cada tanto a una más grande, para que no sufra un cambio muy brusco y se pierda tu pino. Suerte y saludos (YAHOO)*

‘Glaviole: I congratulate you for the decision you have taken, although I do not know if the small tree will be able to recover, nothing is lost by trying. After you wash it under the stream of water, put it in a pot that is not very big and go changing it every now and then to a bigger one, so that it does not suffer a very sudden change and your pine is lost. Luck and greetings’

(153) *Bueno yo a mis perros los coonsidero parte de mi familia, aclaro que no ocupan en mi corazon el lugar que ocupa mi hijo, pero son parte importante en mi vida, bueno no contemplemplo mi vida sin tener a mi lado un peludo amigo. Eso me* ***enseñaron*** *en casa, una casa sin perrito no es un hogar completo* (WIKIPEDIA)
‘Well, I consider my dogs part of my family, I clarify that they do not occupy in my heart the place that my son occupies, but they are an important part in my life, well I do not contemplate my life without having a furry friend by my side. That's what I was taught at home, a house without a dog is not a complete home (WIKIPEDIA)’

(154) *No, do not kick your dog!{S} That is abuse and it wont solve anything.{S} One of my dogs used to do the same thing, but he would go lay back down in the living room after* ***being told*** *to leave the second or third time.{S} Shortly after he had gotten obedience training and I hadn't had that problem since* (WIKIPEDIA)

The absence of indicator is coded as (0), generic/indefinite contexts as (1), and restrictive contexts as (2). This coding process allows to verify whether certain passive-like structures will display a preference to occur in a generic/indefinite, neutral or respective contexts.

#### *5.2.3.6. Presence of a prepositional phrase*

It is assumed that prepositional phrases provide valuable clues to facilitate the interpretation process and identify the possible agentive force involved in the action. Prepositional phrases are codified according to Delbecque’s work on the passive voice in Spanish (2014). The author underlines the important semantic role played by prepositional phrases which function as oblique complements in passive constructions. Delbecque (2014: 219) considers prepositions as a device for agenthood management and suggests that other prepositions than the prototypical preposition *por* (*by* in English and *par* in French) can be used to express agentivity (Example 155). This alternation of prototypical and non-prototypical prepositions is also observed in French (Gross 1975, Leeman 2008, Le Querler 2006).

(155) *Il* ***était accompagné*** *par des Indiens, n’est-ce pas ? (Les aventures de Tintin, Le temple du soleil, p. 17)*
‘He was accompanied by Indians, wasn't he? (The Adventures of Tintin, The Temple of the Sun, p. 17)’

(156) *Il* ***était accompagné*** *d’Indiens, n’est-ce pas* ?

‘He **was accompanied** by Indians, wasn’t he?’

(Le Querler 2006: 130)

It is nevertheless acknowledged that these prepositions do not endorse identical semantic roles like the prototypical proposition *por*. Following Delbecque’s line of research and, in addition to the traditional agentive preposition (157), I codify instrumental, source and locative prepositions (illustrated in (158), (159) and (160) respectively).

(157) *El libro de Vargas Llosa **fue presentado** ayer por el escritor cubano Cabrera Infante.*
‘The book of Vargas Llosa **was presented** yesterday by the Cuban writer Cabrera Infante’.

(158) *La paz no **se conseguirá** con las armas.*
‘Peace will not **be obtained** with weapons.’

(159) *Debe de **estar rodeado** de amigos.*
‘He must **be surrounded** by friends.’

(160) *La decisión **fue tomada** en Bruselas.*
‘The decisión **was taken** in Brussels.’

Examples and translations from Delbecque (2014: 217-8)

Since the taxonomy developed hereunder is grounded in Delbecque’s work (2014), the four Spanish types of prepositions were taken as a starting point to decide which prepositional complements are included in the French and English parts of the analysis. It should be noted that the instrumental category represented by the preposition *con* ‘with’ in Delbecque’s classification was supplemented by the preposition *en* as illustrated in example (161) and by the preposition *de,* and by extend, *a partir de.* These adjustments were made after the pilot analysis revealed examples such as (161) and (162).

(161) *Se explica en ingredientes frescos (no se utilizan*) (YAHOO)
‘This **is explained** by fresh ingredients (they are not used)’

(162) L*os que **se obtienen** a partir de los animales. Estiércol de caballo y vaca guano de las aves, palomina de las palomas,etc. (YAHOO)*
‘Those **which are obtained** from animals. Horse and cow manure, poultry guano, pigeon droppings, etc.’

(163) *Creo que todo esto son simplemente conceptos humanos que se han hecho para distinguir la gran cantidad de productos naturales de nuestra alimentacion; entendiendo esto desde que se cria la planta hasta que se cocina o **se aliña** para consumirla. (YAHOO)*
‘I believe that all these are simply human concepts that have been made to distinguish the great amount of natural products in our food; understanding this from the time the plant is raised until it **is** cooked or **seasoned** to be consumed.’

Example (161) illustrates a case where *en* functions as a prototypical preposition to express the source. A fifth category has been added to classify other types of prepositions, regardless of their nature. Example (163) accompanied by the preposition *para,* which plays the semantic

role of goal). Code 0 has been attributed to cases where the passive-like structure is used without any prepositional clause.

| | Spanish | French | English | Code |
|---|---|---|---|---|
| **Agentive** | Por | Par | By, through | **1** |
| **Locative** | En | Dans, en, à, sur | In, on, at | **2** |
| **Source (provenance, possession)** | De, a partir de | De, à | Of, from | **3** |
| **Instrumental** | Con, En | Avec, à, en, de | With | **4** |
| **Others** | | | | **5** |
| **Absence of prepositional phrase** | | | | **0** |

*Table 11. Coding system in the three languages for prepositional phrases*

When developing the present taxonomy, the aim was not to provide exact semantic equivalents for the Spanish propositional phrases studied by Delbecque (2014), but rather to include the main prepositions fulfilling similar semantic functions that is, agentive, locative source and instrumental. The prepositions included in Table 11 can thus not be considered as fixed translational equivalents. The aim was not to provide an exhaustive list of the possible options available to express these functions, but only to include the main ones. This taxonomy was tested during the pilot analysis phase and proved to be effective, as it appeared to be encompassing all the possible options encountered.

It can be observed that some prepositions, for example the preposition *de* in French, are found in several categories and do not immediately evoke a specific semantic notion. When this is the case, these prepositions are termed *empty* (Riegel et al. 1994, Gaatone 2001, Leeman 2008). They are viewed as linking two items and, due to their usually high frequency of use, their semantic role varies according to the context within which they appear, since they do not convey a fixed common basic meaning. Since these *empty* prepositions do not convey a semantic meaning related to the agenthood management, they are classified as *others.*

To be selected, the preposition and the clause it introduces has to be situated directly after the verbal clause, or could be separated by a nominal clause such as in example (164).

(164) ***Ont été propagées*** *le long des grandes routes européennes par le ex 4560* (WIKI FR)
'**Were propagated** along the major European routes by the ex 4560'

(165) *The term ' ' vegan ' '* ***was coined*** *in 1944 by Donald Watson when he co-founded the Vegan Society in England* (WIKI EN)

When two prepositions co-occurred, I selected the one assumed to be most prototypical to express the agentive force, as illustrated in (165). Even though the preposition *in* is situated directly after the verb, the preposition *by* was favoured for the analysis since it is the typical agentive preposition.

Though this system allows to obtain fine-grained information on the type of preposition employed in passive-like structures and their possible contribution to agenthood management, it presents a major limitation. Cases of locative preposition such as *en Bruselas* (160), repeated here as (166) are annotated both as the presence of a locative preposition and as a use of a spatio-temporal restriction, the parameter presented in 5.2.3.3.

(166) *La decisión **fue tomada** en Bruselas.*
'The decisión **was taken** in Brussels.'

A footnote is included in Chapters 7-9 to specify the number of overlapping cases.

#### *5.2.3.7. Summary*

To close this Chapter on the description of the factors assumed to impact the selection of passive-like structures, Figure 9 provides a summary of the nineteen parameters (described above and) used in the present analysis.

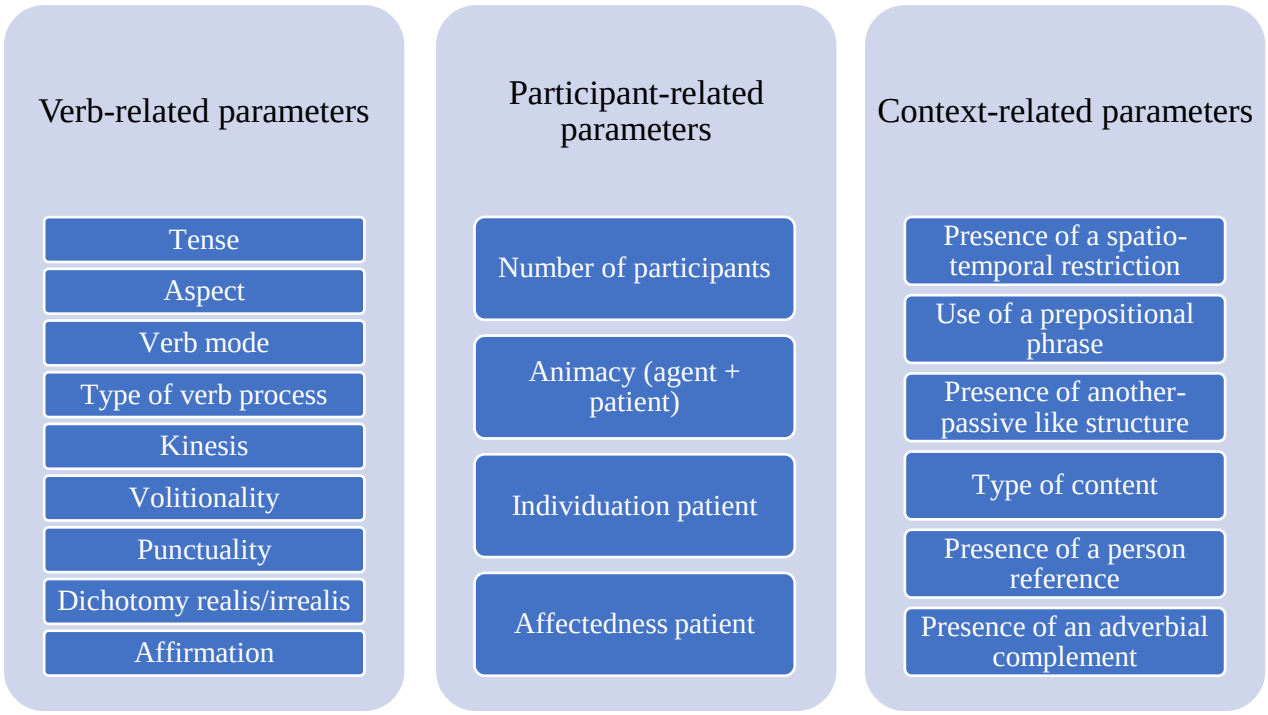



*Figure 9. Classification scheme of the nineteen parameters under study*

This taxonomy is based on Malchukov's proposal (2006) to classify the parameters according to their relation to either the verb, the subject or the object. Some adaptations have been implemented to better fit the aims of the study. Malchukov's *subject* and *object* categories have been grouped under the label *participant* and, since this work adopts a semantic approach to passive-like structures, the phenomena are studied from the role of *agent* and *patient.* An

additional category has been included to group contextual-related parameters, a specificity of this study.

## 5.3. Interrater agreement

The coding scheme presented in the previous section (5.2) was applied manually to all the data. Objectivity in codifying procedures is of crucial importance to ensure the validity of the study. As highlighted by Neumann (2013:45) "the more a given study relies on human interpretation, the more important this concept becomes." To ensure the reproducibility of the analysis and verify the viability of the annotation scheme an expert coder annotated approximately 10% of the data (mainly written data). Note that when the passive-like structure was rather infrequent and fewer occurrences were found in the corpora this percentage was higher or all data where annotated in order to obtain a sufficient amount of data annotated by the second coder. Once the interrater annotation process was carried out, a kappa score (Cohen 1960) was calculated through the statistical software R. This statistical test allows to calculate the degree of agreement between two coders and thus the subsequent reliability of the coding scheme. The calculation was automatically run via a R code, adapted for each parameter[24].

| **Parameters** | **1st phase** | | **2nd phase** | |
|---|---|---|---|---|
| | Kappa score | % agreement | Kappa score | % agreement |
| **Participant** | 0.368 | 76.8 | 0.81 | 95.8 |
| **Kinesis** | 0.562 | 80 | 0.717 | 88.4 |
| **Punctuality** | 0.052 | 72.6 | 0.947 | 97.9 |
| **Affirmation** | 0.883 | 98.9 | _ | _ |
| **Type process** | 0.819 | 84.2 | _ | _ |
| **Mode** | 0.642 | 94.7 | _ | _ |

*Table 12. Examples of kappa scores*

Table 12 contains an example of some kappa scores obtained for the construction based on the reflexive pronoun *se*[25]. It emerges from the interrater analysis that some kappa scores were considered as very strong (e.g. $\kappa$=0.88 for the parameter *affirmation*) while others appeared to be rather weak (e.g. $\kappa$=0.052 for the parameter *punctuality*). It is assumed that two possibilities may explain weak kappa scores. A first reason can be due to a possible lack of clarity in the definition of the parameter. When both the kappa score and the percentage of agreement resulted to be relatively low, I redefined and better delimitated the parameter and the other coder performed a second phase of annotation. This method was carried out for the first three

[24] An example of R code and the steps I followed to obtain the kappa scores can be found in Appendix 3. I am grateful to Thomas François for his precious help in making the R code and his suggestions to solve the problems that emerged.

[25] The complete tables containing all the kappa scores and the percentages of agreement for all the structures are available in Appendix 4.

parameters of Table 12 (*participant*, *kinesis* and *punctuality*) and, as it can be observed, proved to offer excellent results.

A second reason explaining low kappa scores could be the fact that some parameters exhibit little variation. For example, in *se*-constructions, the parameter *mode* (the last parameter in the Table) obtains a relatively low score ($\kappa$= 0.64). This parameter displays little variation as it was mainly codified as realis. Due to the little amount of variation, the probability to obtain the correct answer by chance is much higher, which implies that, in case of disagreement between the two annotators, the penalty is more severe. To counter this phenomenon, an option in R software allows to automatically calculate the percentage of agreement between the two annotators. The percentage of agreement was calculated to obtain more information in case of weak kappa scores. This method showed that even though the parameter *mode* obtained a relatively low score, the percentage of agreement between the coders was substantial (94.7%). This means that the delimitation of the parameter is appropriate and does not need to be reconsidered but rather than the kappa calculation method is very sensitive in some particular cases. The computation of kappa scores combined with percentages of agreement was thus the best way to ensure the reliability of the taxonomy.

## 5.4. Statistical tests

In order to support our findings, the present work relies on two statistical methods: Multiple Correspondence Analyses (5.4.1) and chi-square tests (5.4.2).

### 5.4.1. Multiple Correspondence Analyses

The quantitative part of the present work is based on the Multiple Correspondence Analysis (MCA) method. MCA is a descriptive method that allows to analyse the patterns of relationships among categorical variables (Abdi & Valentin 2007, Le Roux & Rouanet 2010). The aim of such graphical technique is to "display geometrically the rows and columns of the data table – where rows represent individuals and columns the categories of the variables – in low-dimensional space, so that proximity in the space indicates similarity of categories and of individuals" (Le Roux & Rouanet 2010:vii). This method particularly suits binary variables (e.g., perfective vs imperfective) but can also accommodate quantitative variables as long as they are recoded as nominal variables with different levels (Abdi & Valentin 2007). Individuals and variables are projected on a two-axis graph.

In the present work, an individual refers to one occurrence of an impersonal structure, i.e., one line in the data table, and a variable refers to one of the nineteen parameters under study. The interpretation of these axes corresponds to encountering the similarity between the elements on the right part of the graph and those on the left, highlighting the opposition and contrast between the two extremes (Benzécri 1992:405). Gatti & Maroni (2004:4) highlight the following two main advantages of MCA: (i) synthetic and rational representations of data since MCA provide

an immediate and visual picture of the pattern of relationships among the variables and (ii) this technique also allows a critical evaluation of data and helps to get rid of irrelevant information. Though particularly useful when examining complex data sets, MCA also suffer some limitations. Gatti & Maroni explain (2004) that the results are strongly affected by the samples analysed, which means that, for instance, anomalous variables or variables with a totally different behaviour considerably impact the geometrical representation. It happens to be the case with our set of data, which is affected by the features of the French man-impersonal *on*. This issue will be discussed in more detail in Chapter 6.

The final material used to compute MCA had to suffer a recoding process since some of the cells in the data table, which correspond to the categories within the parameters (e.g., the different types of verb processes), have observed values less than five. These categories have thus been grouped to encounter this issue. Hereafter are listed the modifications that have been undergone in order to carry out the MCA. These modifications, however, are specific to the statistical tests and do not affect the annotation scheme and the analytical description of the results.

- For the general factor map presenting the variables for the three languages, the constructions *get, uno* and *one* are excluded as these three constructions totalize a very restricted number of occurrences. In the context of the present work, there is no sufficient justification for grouping them together or with another impersonal structure.
- The four subcategories of the parameter *mode* have been grouped in two : indicative mode (category 1) and other (category 2)
- The parameter *type of process*, originally consisting in twenty categories, was recoded into six major categories, according to the nature of the type of process: *mental*, *relational*, *material*, *verbal*, *existential* and *modulation*.
- The parameters *agent* and *patient* have been divided into three categories: animate (1), inanimate (2) and absence (0) instead of the initial six presented in Section 5.2.
- Prepositional phrases have been grouped in three categories instead of the six presented in the methodology: absence (category 0), prepositional phrases with a possible agentive force (category 2) and other prepositional phrases (category 5).
- The use of a person reference in the context has been divided in two categories: absence (category 0) and presence (category 1) instead of the initial six listed in the Section 5.2.
- The parameter *type of content* has been removed as it displays extremely little variation.
- Occurrences which were impossible to codify due to for instance unfinished sentences in oral interactions or background noise have been removed.

### 5.4.2. Chi-square tests

In addition to Multiple Correspondence Analyses, this investigation also relies on chi-square tests, a commonly used statistical test in corpus linguistics (McEnery et al. 2006) to support the

interpretation of the findings and to strengthen their validity[26]. The use of a chi-square test is among one of the most popular statistical methods "for answering questions about the association or difference between categorical variables (Franke et al. 2012)." More concretely, in our data we are interested to know whether the frequencies obtained for two parameters are significantly different or follow a uniform distribution. Prior to the calculation, a null hypothesis has to be formulated. An example of a null hypothesis tested in this study would be the absence of difference between the variables under consideration, - i.e., whether the variables are equally frequent and the observed differences are only attributed to random variations. The computation of the chi-square allows to test the null hypothesis. Obtaining a significant chi-square, that is, above the recommended probability threshold of 0.5, leads to reject the null hypothesis. In other words, the variables have a different distribution. Tallarida & Murray point out that this method "is applicable to many situations in which experimental frequencies are compared to theoretical frequencies based on a hypothesis" (1987:140).

[26] For an ample discussion on the use of chi-square tests in corpus linguistics see for instance McEnery et al. 2006, Bestgen 2017, Litosseliti 2018.

# 6. General overview of the data

This chapter presents the overall frequencies of use of the different passive-like structures. It aims to offer some preliminary answers to research questions 1 and 2 that is, how languages and registers and modes impact the selection of impersonal constructions. These answers will further be completed through Chapters 7 to 9 which analyse in much more detail how the structures function. It is important to bear in mind that, as explained in the methodological part, this study includes five impersonal mechanisms and do not pretend to gather all the available devices to defocus the agent in the studied languages. A second important point to remember when interpreting the results is that not all the selected structures are available in the Spanish, French and English. Comparative results are given, and hold true, only for the studied phenomenon in the present investigation.

The chapter is divided in three main parts. It first reports on the overall raw and relative frequencies of the structures (section 6.1). This section includes two points: the general distribution of the structures by language and by register and mode. In the second part of the chapter (section 6.2), I enrich the discussion by examining and comparing the type/token ratio of each structure and establish a top three of the most frequent verbs. The last part explores the overall results of the general Multiple Correspondence Analyses (6.3).

## 6.1. Overall distribution

Before delving into the examination of the overall distribution of the structures through the data, it is relevant to examine the existing literature on the matter. As already explained in the previous chapter, there is a profusion of studies discussing the topic of impersonality. Even though these studies adopt different approaches (see for instance the various typological works led by Siewierska) and may not always be corpus-based, they may provide valuable insights into the frequency of use of passive-like structures and into what is usually stated in the body of literature.

Gaatone reports (1998:34) that, in any language production, canonical passives are generally by far less frequent than active sentences. English canonical passive structures account for 4% (informal written data) to 18% (formal written data) of the total number of sentences produced (Givón 1990). In French, this proportion sharply drops as Hupet & Costermans (1976) report an interval of 3.8% to 8.7% of canonical passives for the written mode and Greidanus (1990) only finds 1.6% of these structures in oral data. The variation in the frequency of use of canonical passives also indicates that modes and registers impact their productivity. These studies point to the fairly low frequency of the passive phenomenon, especially in French, in comparison to active alternatives. Several studies on Spanish confirm this finding and indicate that most Spanish grammars and grammarians highlight the fact that the passive voice is a rather infrequent phenomenon (e.g., Fernández 2007 or Laslop & Díaz 2010). Comparing cross-linguistically the frequency of use of periphrastic passives, García Yerba (1990) specifies that English is the language relying most on that structure, followed by French and then Spanish.

As already mentioned, the passive is not the only grammatical device to defocus the agent. Quesada (1997:41) notes that the use of the Spanish canonical passive is substantially decreasing over time, especially in spoken language, in favour of other impersonal mechanisms. The author indicates the *se*-construction as the most common alternative strategy. It seems largely agreed among researchers that the Spanish *se*-construction has become the favourite candidate, over the periphrastic passive, to defocus the agent and to express passive meanings (RAE 1973, Calvo Montoro 1983, Lamiroy 1993, Maldonado 1999, Mendikoetxea 1999, Dieck 2016, Cabañas Maya 2006, Tremblay 2006, Laslop & Díaz 2010 to only cite a few). Gili Gaya (1995) even uses the terms 'general repugnance' to describe the infrequent use of the canonical passive in Spanish.

> "Además de esta repugnancia general al uso de la pasiva, se producen numerosas interferencias expresivas con el significado del verbo copulativo y con las oraciones llamadas de pasiva refleja e impersonales, las cuales hacen retroceder día a día el empleo de la pasiva con ser" (Gili Gaya 1995:111).
>
> "In addition to this general repugnance to the use of the passive, there are numerous expressive interferences with the meaning of the copulative verb and with the so-called reflexive and impersonal passive constructions, which make the use of the be-passive decline day by day" (Gili Gaya 1995:111).

García Yerba (1997: 348) adds that Spanish would only rely on the canonical passive when an active option could not fulfil some specific grammatical requirements such as the sake of clarity or rhythm. This fierce competition with the *se*-construction could be a factor explaining the low frequency of use of canonical passive structures in Spanish, compared to French or English (Weber 2014). Dieck (2016) indicates that the use of the Spanish canonical passive is more common in written data while the *se*-construction seems to be largely employed in both modes. The author, however, points to the lack of solid comparative studies to support that claim (2016:153).

Some strategies are less well-documented than the periphrastic passive and the *se*-construction. The two impersonal strategies relying on a numeral, *one* and *uno* have drawn less attention than those discussed so far. On Spanish, Rasson (2017) finds that *uno* is a structure with a relatively low frequency of use (1,98 occurrences per ten thousand words). Her findings also indicate an impact of the mode and the register on the frequency of the structure, *uno* being more used in oral informal interactions than in digital forums or academic writing.
On the French man-impersonal *on*, Yao (2016) indicates that though this device is largely used in informal and familiar contexts, it is not totally absent from the formal register. No other relevant information on the frequency of that impersonal structure has been found.
Even though many studies on the possible mechanisms available to defocus the agent have already been conducted, very few provide quantitative information on the frequency of use of these different devices. This absence of information may be due to a lack of large-scale contrastive studies on passive-like structures. The following section intents to fill this gap and provide solid corpus-based contrastive data.

I will first describe the distribution of the structures according to the language and then break the results down according to mode and register.

| | | English | | Spanish | | French | |
|---|---|---|---|---|---|---|---|
| | | raw | pttw | raw | pttw | raw | pttw |
| **Structure** | 3PP | 186 | 8.48 | 498 | 23.85 | 109 | 5.34 |
| | GET | 49 | 2.2 | _ | _ | _ | _ |
| | ON | _ | _ | _ | _ | 995 | 48.83 |
| | ONE | 9 | 0.41 | _ | _ | _ | _ |
| | PP | 1415 | 64.5 | 325 | 15.56 | 857 | 42.05 |
| | SE | _ | _ | 968 | 46.36 | 73 | 3.58 |
| | uno | _ | _ | 5 | 0.24 | _ | _ |
| **Total** | | 1659 | **75.63** | 1796 | **86.01** | 2034 | **99.81** |

*Table 13. General distribution of the data by structure and language*

Table 13 indicates that the English *get*-passive is a relatively infrequent phenomenon. The low frequency of *uno* and *one*, already highlighted by Martínez Caro (2002), is confirmed in the present study. In addition, it is interesting to note that *uno* and *one* obtain a fairly similar frequency of use (0.24) and (0.41). A larger set of data would be valuable to confirm a comparable behaviour in terms of frequency. As already stated, a series of impersonal mechanisms, apparently similar in their forms, are available in English as well as in French and Spanish. This is the case of the periphrastic passive and the non-anaphoric third person plural. One can wonder whether these structures display a similar frequency of use across the three languages. Table 2 shows that this assumption does not hold true. For the periphrastic passive, a sharp gap between the languages can be observed. The English periphrastic passive obtains the highest relative frequency (64.5) and the Spanish one the smallest with a score of 15.56 occurrences pttw. The French construction scores in-between as it shows 42.05 occurrences pttw. The high frequency of the periphrastic passive in English may partly be explained by the dynamicity of the language. Compared to French, English is usually considered more dynamic, reflecting the action in its development (Van Hoof 1989:62). As will be discussed in Chapters 7-9, periphrastic passives are preferred candidates to express dynamic actions while R-impersonals, dominant in French, include far more stative verbs. Van Hoof (1989) also indicates that, whereas French tends to imply the agent in the description of the events, English prefers to report on events and situations without making any reference to the agent. This implication of agents in French may be realized by R-impersonals which, though they reduced the referentiality of the agents, still maintain them in topic positions. Even though a similar frequency of use of periphrastic passive could have been expected in French and Spanish since they are both Romance languages, results point to a different behaviour. They indicate that French periphrastic passives are used nearly three times more frequently than their Spanish counterparts. Regarding the non-anaphoric third person plural, the construction is much more widespread in Spanish (23.85) than in French and English, which display a similar frequency (5.34 and 8.48, respectively).

A look at the *se*-construction, employed in Spanish as well as in French, also provides insights into the frequency of use of passive-like structures. Even though both constructions have an apparently similar form, the use the two languages make of it greatly differs, with much more constraints applying to the French structure. These restrictions seem to largely affect the productivity of the structure since the proportion found for the Spanish construction is by far higher than the one obtained for the French equivalent. While the Spanish *se*-construction totalizes 46.36 occurrences pttw, the French one only resorts to 3.58 of *se*-constructions pttw. This result largely supports the highest productivity of the Spanish structure, usually reported in the literature. It is particularly striking to observe the frequency of use of the Spanish *se*-construction, that is, 46.36 can be compared to the scores found for the French periphrastic passive (42.05) and the man-impersonal (48.83).

In sum, the analysis has revealed that each language studied in the present investigation has its own preferred impersonal device(s), namely the man-impersonal and the periphrastic passive for French, the *se*-construction for Spanish (and, in a smaller proportion, the non-anaphoric third person plural) and the periphrastic passive for English. The relative frequencies reveal that even though an impersonal mechanism is available in different languages, these languages do not resort to it at the same frequency. Different factors such as the presence of other productive impersonal mechanisms or major syntactic constraints may account for the heterogeneity in the frequency of use. The latter point will further be developed and examined in Chapters 7 to 9.

In search for a comprehensive view on the frequency of use of passive-like structures, the present section examines the variation according to registers and modes. The analysis goes back and forth between language variation and register/mode variation since language proves to be a key factor in the distribution. This way of proceeding allows to obtain the most fine-grained material to establish the panorama of the passive-like structures understudy.

First, even though this mechanism is not the most frequent in English, I will briefly mention the case of the get-passive. In the set of data used for this study, English mainly resorts to the get-passive in informal contexts and more precisely in informal oral productions, which totalize 32 out of the 49 occurrences found. The structure is nearly absent from the formal register. These results should nevertheless be confirmed with a wider set of data.

The *se*-construction was proven to be the favourite impersonal device in Spanish. The literature usually reports a preference for the Spanish *se*-construction to appear in formal and distancing contexts, although, as highlighted by some researchers (see e.g., Quesada 1999 or De Cock 2020), this mechanism is also present in spoken and more informal situations. Nevertheless, Dieck (2016) points to the lack of large corpus-based studies to precisely document the preferred modes and registers of *se.* The present work brings detailed information as for the written versus spoken nature and the degree of formality within which *se*-constructions occur, which we believe contributes to filling the gap identified by Dieck (2016).

| | Se-constructions | | | |
|---|---|---|---|---|
| | Spanish | | French | |
| | raw | pttw | raw | pttw |
| **Yahoo** | 279 | 59.87 | 19 | 4.2 |
| **Wikipedia** | 337 | 70.91 | 46 | 9.96 |
| **Ciel** | _ | _ | 4 | 0.54 |
| **ESLORA** | 122 | 15.89 | _ | _ |
| **PROCEP** | 230 | 60.61 | 4 | 1.05 |

*Table 14. Overall frequencies for se-*constructions *in French and Spanish*

Table 14 shows that the structure is used both in written and oral data and in formal and informal registers. It can, however, be noted that it is less productive in informal corpora. This tendency is especially marked in oral data. As for the French *se*-construction, results indicate that the construction is mainly specific of written productions, especially formal written ones. This might be the main similarity between Spanish and French *se*-constructions as for their distribution

| | Man-impersonal *on* | |
|---|---|---|
| | raw | ptw |
| **Yahoo** | 213 | 47.13 |
| **Wikipedia** | 98 | 21.23 |
| **Ciel** | 618 | 82.9 |
| **PROCEP** | 66 | 17.42 |

*Table 15. Overall frequencies for the man-impersonal construction*

The rather productive French man-impersonal *on* is used in both registers and modes, with a marked preference to be employed in informal oral data (82.9 occurrences pttw) and, in a smaller proportion informal written data (47.13 occurrences pttw) (see Table 15). In formal contexts, both written and oral modes obtain a similar score of respectively 17.42 and 21.23, which means that, even though the structure is less frequent than in informal data, it is not totally absent from the formal register. The presence of *on* in parliamentary debates may partly be explained by the fact that this construction can be used "to decrease intrusiveness as when it is used to tone down a request" (Viberg 2010:142). A comparison with the data obtained for the *se*-construction reveals that, registers and modes do not impact these two structures in the same way since the *se*-construction is more common in formal data while *on* is more largely used in informal contexts. And it is undeniable that *on* and *uno* and *one* have a different behaviour in terms of frequency and thus cannot be used as translational equivalents.

| | Periphrastic passive | | | | | |
|---|---|---|---|---|---|---|
| | **Spanish** | | **French** | | **English** | |
| | raw | pttw | raw | pttw | raw | pttw |
| **Yahoo** | 53 | 11.37 | 171 | 37.84 | 285 | 59.38 |
| **Wikipedia** | 213 | 44.82 | 416 | 90.12 | 596 | 128.65 |
| **BNC** | _ | _ | _ | _ | 160 | 18.88 |
| **Ciel** | _ | _ | 55 | 7.38 | _ | _ |
| **ESLORA** | 6 | 0.78 | _ | _ | _ | _ |
| **PROCEP** | 53 | 13.97 | 215 | 56.75 | 374 | 92.79 |

*Table 16. Overall frequencies for the periphrastic passive in Spanish, English and French*

The main aspect to highlight regarding the use of the periphrastic passive is that this structure seems to be specific of the written mode (see Table 16). This assumption is valid for the three languages. It can also be noted that the periphrastic passive is more frequent in the formal register than in the informal one, be it in written or oral corpora. While the tendency to use the structure in formal oral interaction is particularly marked in French (56.75 occurrences pttw) and English (92.79 occurrences pttw), Spanish obtains a smaller result with 13.97 occurrences pttw. This finding corroborates Laanemets' study on Scandinavian languages (2013:78), which indicates that the passive voice is more frequent in debates than in conversations.

| | Non-anaphoric third person plural | | | | | |
|---|---|---|---|---|---|---|
| | **Spanish** | | **French** | | **English** | |
| | raw | pttw | raw | pttw | raw | pttw |
| **Yahoo** | 134 | 28.75 | 32 | 7.08 | 65 | 13.54 |
| **Wikipedia** | 4 | 0.84 | 3 | 0.65 | 8 | 1.73 |
| **BNC** | _ | _ | _ | _ | 113 | 13.33 |
| **Ciel** | _ | _ | 74 | 9.92 | _ | _ |
| **ESLORA** | 358 | 46.64 | _ | _ | _ | _ |
| **PROCEP** | 2 | 0.53 | 0 | 0 | 0 | 0 |

*Table 17. Overall frequencies for the non-anaphoric third person plural in Spanish, French and English*

As for the non-anaphoric third person plural, in the three languages, the structure is more widely used in informal data with a tendency to appear more frequently in the written mode (see Table 17). It is, however, interesting to note that this mechanism is relatively frequent in Spanish informal oral data. Such phenomenon is not observed in the other two languages.

Coming back to what is usually said in the literature, the following observations can be made. First, as pointed out by Laanemets (2013) and confirmed in our data, the productivity of the structures seems to be linked to the degree of formality. By way of illustration, the periphrastic passive in French is much more present in formal registers, whereas active options, that is, the two French R-impersonals, are dominant in informal situations. Second, it is frequently reported that English makes ample use of periphrastic passives. This claim is totally verified in the present data since, compared to French and Spanish, English is the language that makes most use of the periphrastic passive. Moreover, it seems to be the preferred device to defocus the

agent in English. Third, researchers widely acknowledge the preponderant use of the *se*-construction in Spanish, which supplants the periphrastic passive when it comes to expressing a passive meaning. Again, the present results confirm this claim since the *se*-construction is twice as frequent as the periphrastic passive. Dieck (2016) observes that the Spanish *se*-construction is frequent in both modes and registers but is, nevertheless, calling for more precise information. The present study indicates that the construction is indeed commonly used in both modes and registers. It further shows that informal oral data are those that resort the least to this type of device since they contain the construction approximately four times less than the other modes and registers. A slight tendency to use a higher number of *se*-constructions in written data was also observed. Fourth, as regards the man-impersonal *on*, few studies have provided information on its frequency of use. Yao (2016) indicates that it is more frequent of informal contexts. This study yields more detailed results on its behaviour. They confirm that *on* is more typical of the informal register, especially the oral informal one. It is, however, important to note that *on* is also a common possible impersonal construction in more formal registers. It can be concluded that the present findings largely corroborate what has been found in the literature and allow to provide more detailed information.

Summarizing the main observations of this section, it can be said that even though several structures have an equivalent in the three languages their distribution is, most of the time, substantially different. It is clear that both the language and the register and the mode are factors affecting this distribution. A further analysis will be carried out to determine whether the structures understudy are prone to grammatical and/or pragmatic restrictions, which could possibly explain these differences in the frequency of use. Nevertheless, before turning to the core of the analysis of grammatical features, the next section examines the type token ratio of each impersonal mechanism.

## 6.2. Type/token ratio

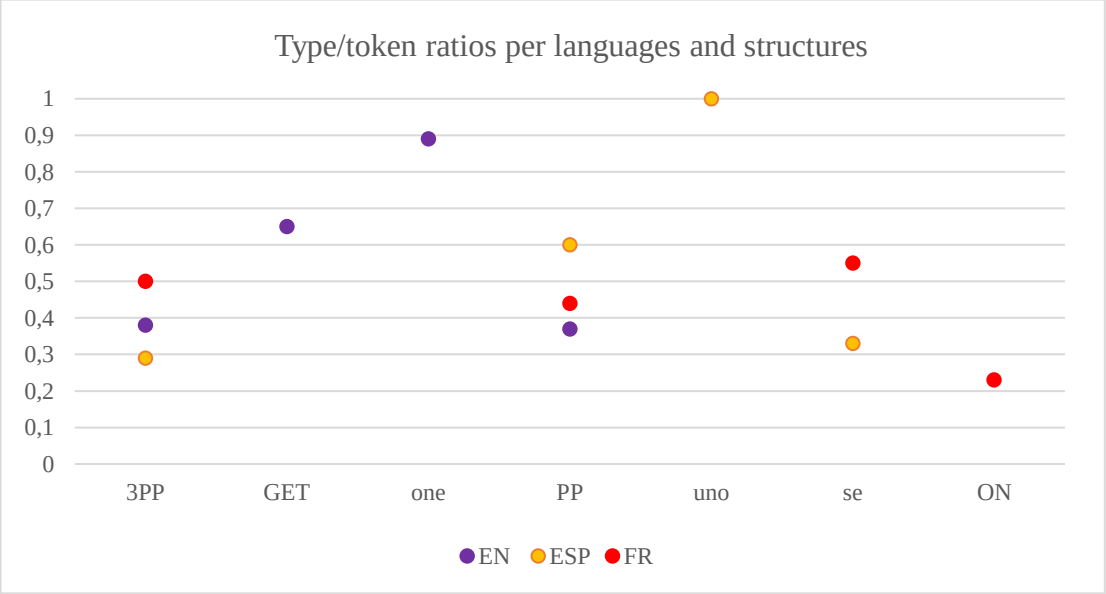
When examining a grammatical phenomenon such as impersonal constructions, computing the type token ratio of the structures can add valuable information about their respective behaviour. The type token ratio, which is the number of different words divided by the total number of words contained in a production (see Biber 1988) allows to measure the lexical density. The closer the ratio is to 1, the higher degree of lexical richness the production contains. More precisely, in this case, it will allow to calculate the lexical richness of the verbs used in the six types of passive-like constructions studied. Such a measure will provide complementary information both on the general behaviour of the structures and on their behaviours within the three languages and the two registers and modes.

It seems more relevant to report the results by language since some verbs, similar in form in all the three languages, may convey a different meaning in each language (cf. false friends) and one language may resort to different verbs to express a meaning rendered by only one verb in another language. Finding equivalence between the verbs in the three language would result in a colossal task and is beyond the scope of this study.

| | Total no. of structures (tokens) | No. of different verbs (Types) | T/t ratio |
|---|---|---|---|
| **English** | 1659 | 561 | 0.34 |
| **Spanish** | 1796 | 496 | 0.28 |
| **French** | 2034 | 546 | 0.27 |

*Table 18. Type token ratios per language*

Table 18 presents the ratio of the number of different verbs according to the languages. These figures do not consider the different types of structures. It can be observed that French and Spanish, the two Romance languages, obtain a type/token (T/t) ratio remarkably similar (0.28 and 0.27, respectively). As for English, it obtains a somewhat higher score with a ratio of 0.34. This means that when it comes to impersonal structures, English displays a slightly higher verbal density compared to the other two languages. These findings are further detailed in Graph 2, which includes the T/t ratios for the different structures in each of the three languages. The figures for *uno/one* and the *get*-passive are only included for informative purpose. The low level of productivity of these constructions seems to considerably impact the T/t ratio and may explain their extremely high scores compared to the other passive-like structures.

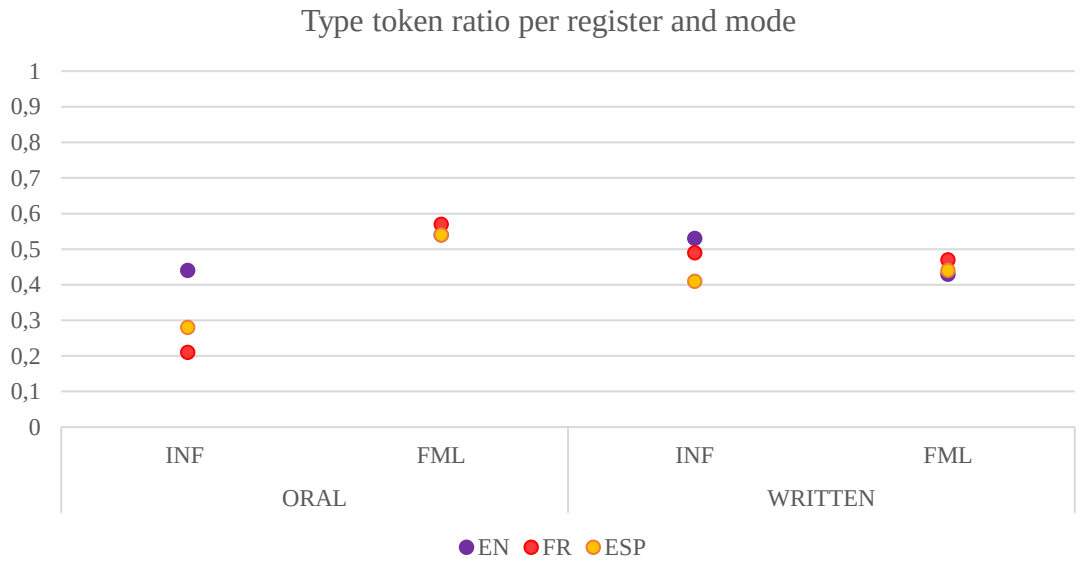



*Graph 2. Type/token ratio per language and structure*

As can be observed from Graph 2, the type of structure affects the lexical density. First, a look at the periphrastic passive reveals this construction as having the highest lexical richness since its ratio varies within an interval ranging from 0.37 in English to 0.6 in Spanish. The non-anaphoric third person plural obtains an interval consisting of smaller ratios since it ranges from 0.29 in Spanish to 0.5 in French. In other words, this means that the use of the non-anaphoric third person plural relies on a smaller variety of verbs than the periphrastic passive.

Language can be regarded as a second key factor affecting the lexical density. The language displaying the highest lexical density is French, except for the periphrastic passive. It is

interesting to note that *on*, the only man-impersonal construction analysed in this study, scores the lowest in terms of lexical density. Leaving aside the *get*-passive and *one* since these structures are fairly infrequent in our data, it can be observed that English never displays the highest lexical density level. It can be assumed that this phenomenon may be due to the high frequency of occurrences of the structure. However, the close examination of the relative frequencies shows that French periphrastic and the man-impersonal *on* display similar figures and the periphrastic passive obtains a higher T/t ratio. It is rather suggested that mode and register are responsible for this low T/t ratio. As seen below, *on* is particularly used in oral and informal registers. Since oral productions are said to have a lower lexical density compared to written productions (Halliday 1985), the mode may explain this result. This leads us to consider the impact of mode and register on passive-like constructions.




*Graph 3. Type token ratio per register and mode*

This graph presents the T/t ratio of the number of different verbs in each mode and register. It does not consider the variable 'structure'. For the formal register, both in written and oral modes, figures suggest fairly similar ratios in the three languages, that even overlap in oral formal data. A slight tendency to resort to a wider lexical variety can, however, be noted in French. The similarity between formal oral and written data may be partly explained by the genre productions included in formal oral corpora, i.e., proceedings of European parliamentary debates. Even though, inasmuch as possible, the corpora consist of spontaneous productions, the interventions were probably planed in advance. Though they are prepared to be spoken, these talks might contain features usually attributed to written production, such as the degree of variation in the choice of the verb. This phenomenon of similar ratios between written and oral registers is, however not observed in informal data. A wider diversity between the languages can be noted, English being the language to possess the highest lexical density in informal written and oral data. A possible explanation of this tendency is the fact that informal

situations are usually considered as less contrained by universal prescriptive rules than formal settings.

On the whole, written productions contain a greater variety of verbs (interval ranging from 0.41 to 0.53) than oral ones (interval ranging from 0.21 to 0.57). Results also point to more homogeneity in formal and informal written data (smaller interval) than in oral data (wider difference between the lowest and the highest ratio). In other words, oral data resort more frequently to the same verbs when it comes to agent defocussing.

The computing of T/t ratio also allowed to identify which are the most frequent verbs used in each structure in each language. Table 19 offers a top three of these verbs for the most frequent structures, that is, the periphrastic passive, the non-anaphoric third person plural, the man-impersonal *on* and the *se*-construction[27].

| | **English** | **French** | **Spanish** |
|---|---|---|---|
| **3PP** | to have (20)<br>to do (17)<br>to be (9) | avoir (20)<br>faire (10)<br>mettre (8) | decir (54)<br>tener (52)<br>dar (31) |
| **PP** | to use (78)<br>to call (42)<br>to make (33) | utiliser (40)<br>prendre (16)<br>dire (15) | utilizar (10)<br>fundar (8)<br>conocer (8) |
| **SE** | x | * | hacer (52)<br>llamar (45)<br>encontrar (25) |
| **ON** | x | faire (98)<br>avoir (79)<br>être (70) | x |

*Table 19. Top three most frequent verbs for the non-anaphoric third person plural, the periphrastic passive, the se-construction and the man-impersonal on*

First, results show that, for the periphrastic passive, the verb *to use* is the most frequent in the three languages. For the non-anaphoric third person plural, a similarity between the languages is also be noted with the verbs 'to have', *avoir* in French and *tener* in Spanish and 'to do' and *faire* in French, which may have semantically similar uses in the three languages. Spanish, however, displays a slightly different behaviour since the top three verbs includes *decir* ('to say') and *dar* ('to give'), which do not appear in the top three of English and French. The difference may be due to the fact that this passive-like structure has a higher frequency of use in Spanish as indicated by the calculation of the relative frequencies.

Table 19 indicates that most often the top three contains auxiliary verbs and/or what can be called basic verbs, that is, verbs that are not domain-specific and may be considered as having a polysemous nature. Aarts & Meyer (1995) established a corpus-based list of the most frequent words occurring in English speech and writing. They found the following 15 top verbs: *come,*

[27] Note that for the *se*-construction, information has only been provided for the Spanish data since the French structure is far less productive and hardly no verbs were found to occur more than three times.

*do, find, get, give, go, have, know, look, make, say, see, take, think* and *use*. Though this list was created for English, it can be noted that it also holds true for French and Spanish. Blanche Benveniste (2002:18) stresses that due to their polysemy these verbs might endorse multiple grammatical uses (the auxiliaries *avoir* 'to have' and *être* 'to be' being the prototypical example) and can thus be employed in a wide range of syntactic constructions. The author suggests the following verbs as the most frequent verbs in French: *avoir* 'to have' and *être* 'to be', *faire* 'to do', *dire* 'to say' and *devoir* 'to have to'. This list concord with Aarts & Meyer (1995)'s findings. Blanche Benveniste (2002), however, specifies that these results are based only on oral data.

It can be concluded that, regarding the lexical diversity of the top three verbs, passive-like constructions show a rather homogeneous behaviour since no major language specificities could be identified at this stage. These high frequency verbs, as well as all the other verbs found in passive-like constructions, will be closely examined in the language-specific analytical chapters (Chapters 7, 8 and 9) with a fine-grained analysis of the types of verb processes.

## 6.3. Exploring statistical results

This section presents the overall graphs obtained from the Multiple Correspondence Analysis tests. Three types of graphs are examined : a biplot with the variables, a factor map with all the passive-like structures and a factor map presenting the distribution of the occurrences of passive-like structures according to mode and register.

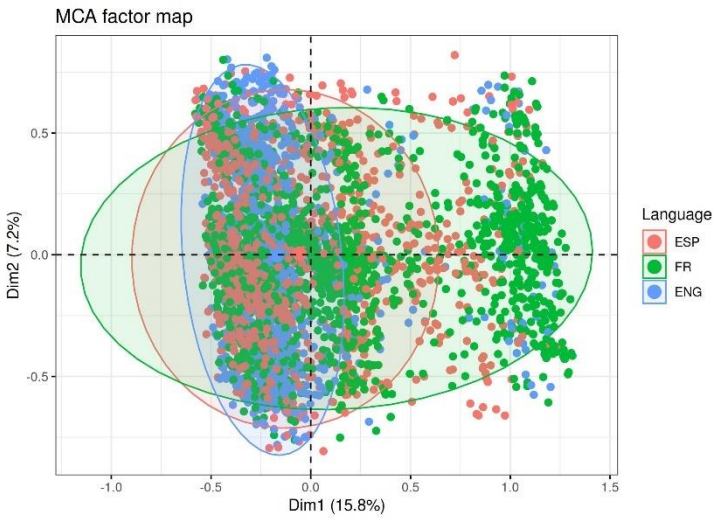



*Figure 11. Distribution of languages, including the man-impersonal ON*

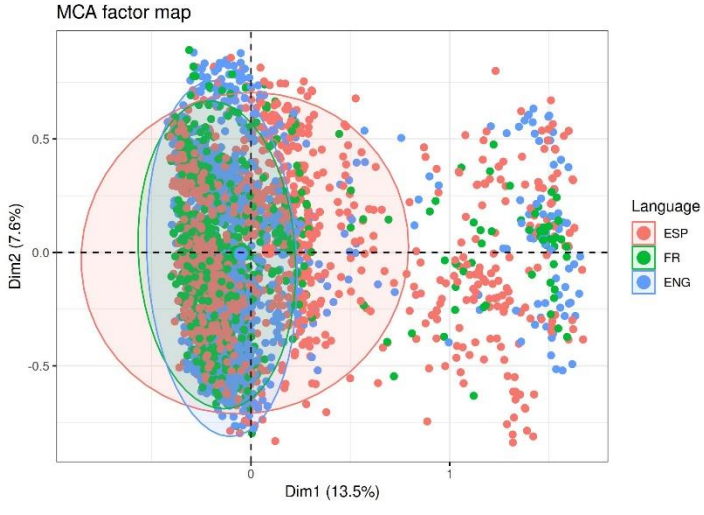



*Figure 10. Distribution of languages, excluding the man-impersonal ON*

Figures 10 and 11 present the distribution of the data with a special focus on the languages, where Spanish occurrences of passive-like structures are represented in red, French ones in green and English ones in blue. Two versions of the factor maps, with the man-impersonal *on* included in the computation of the data (Figure 11) and without (Figure 10), are presented above. As can be seen, this passive-like structure has an important impact on the distribution of the data. The inclusion of *on* tends to distort the confidence ellipsis of the French language,

making it developing much more horizontally than the ellipsis representing Spanish and English data. In this visualisation, French appears as the most distinct language as for passive-like constructions. A look at Figure 10, which visualises the data excluding *on*, suggests a completely different picture. The confidence ellipses of French and English largely overlap while Spanish data appear to be the most different ones. The comparison of Figure 10 and 11 shows that the French man-impersonal *on* has a specific behaviour compared to the other passive-like structures included in this study. This divergence can be partly due to the fact that English and Spanish do not possess a similar feature-based defocussing mechanism.

As the purpose of this study is to obtain as a complete picture of the different systems as possible, it has been decided to work with the version of the graphs including *on*. This will allow to obtain a comparable basis for the contrastive analysis of this work. The graphs excluding the man-impersonal *on* are available in Appendix 5 for informative purposes.

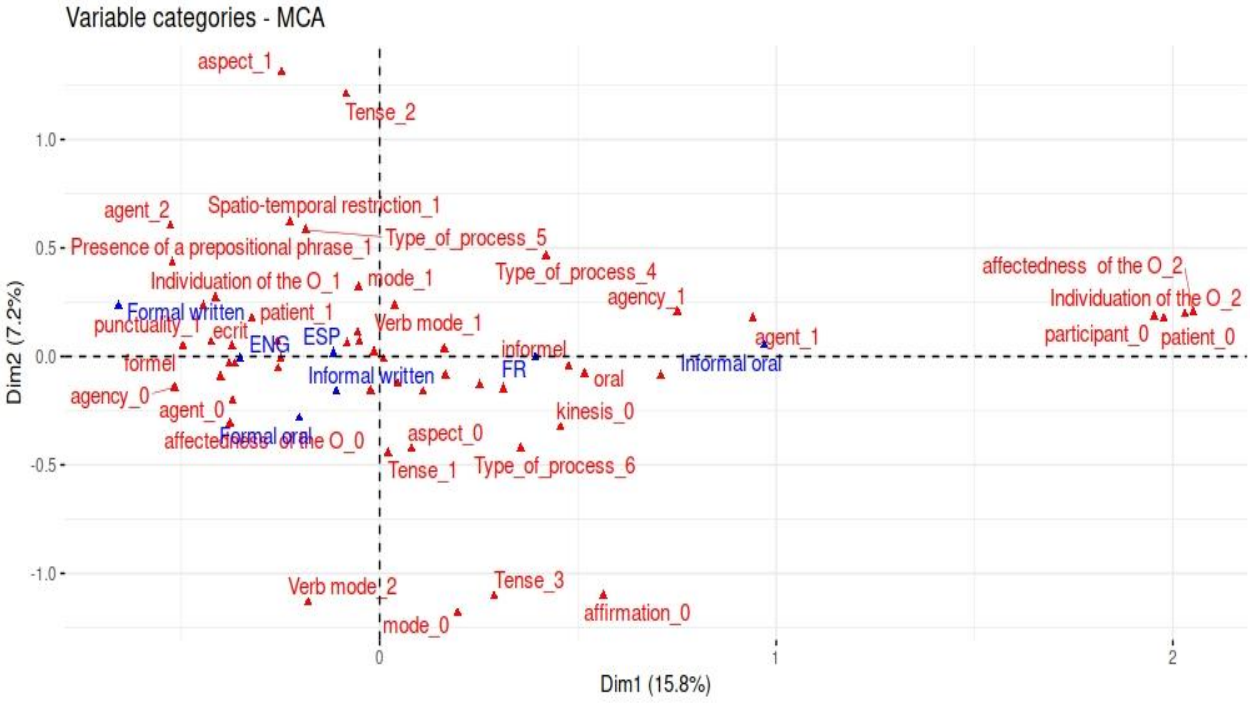



*Figure 12. Distribution of the nineteen parameters*

Figure 12 allows to visualise how the nineteen parameters (also called *variable* in the figure), represented in red, are distributed in the three languages and in the different modes and registers (represented in blue). Their location on the map represents their degree of affinity, that is, parameters located next to each other tend to have a similar behaviour while those situated far from each other differ more considerably as their distribution. Figure 1 thus also makes it possible to determine which parameters are considered as particularly contributing in explaining the variation between the passive-like construction included in this study. To be considered as salient features, variables have to be located between 0.5 and the extreme of the axis. Variables occupying a position on the extremes are likely to contribute the most in explaining the variation between the passive-like structures. As is clear in the figure, informal written and informal oral data develop alongside the horizontal axis. Parameters related to the number of participants, patients (category 0; absence) and their degree of individuation and

affectedness (category 2; totally affected) as well as agent (category1; animate) are expected to be distinctive features as they are located at the extreme right of the axis. Greater attention will be devoted to their behaviour in informal oral data since they are placed relatively close to each other on the map. We can also expect that French passive-like structures will display a particular association with this set of features as it is the only language situated on the right-hand side of the figure. However, its position, fairly close to the centre of the graph, leads us to assume that the strength of the link will be more limited than the one expected between this set of features and informal oral data. Kinesis, situated on the 0.5 limit will also be included in the analysis. Since English and Spanish are situated on the other side of the X-axis, other types of associations can be expected for these two languages. On the left-hand side of the horizontal axis, two variables are placed on 0.5: agent (category 2; inanimate) and the presence of spatio-temporal restrictions (category 1; presence). As is indicated by Figure 12, these parameters are expected to be key elements in formal written data, and more especially in English and Spanish. Finally, no extreme values are found on the left-hand side of the horizontal axis, as opposed to the right-hand side of the axis.

Let us now interpret the results suggested for the Y-axis. First, it can be observed that, while informal data tend to develop alongside the horizontal axis, formal data are rather located alongside the vertical one. At the bottom of the Y-axis, four parameters can be found: Mode, with a capital letter, which refers to the verb mode (category 2; *others*), tense (category 3; future), affirmation (category 0; negative clause) and mode, with a lower case, which stands for the dichotomy realis/irrealis (category 0; *irrealis*). As these parameters are situated between 1 and 2, they are expected to play a preponderant role in the analysis. Aspect (category1; perfective) and tense (category 2; past), located at the extreme top of the Y-axis, are the other two parameters considered as particularly salient. The presence of a prepositional phrase and the type of process (category 4; *verbal* and 5; *existential*) will also be examined as they are situated on the 0.5 threshold.

Out of the nineteen variables, the following ones will then serve as common basis for this work and receive particular attention: (i) verb-related parameters: aspect, tense, verb mode, type of process, kinesis, dichotomy realis/irrealis and affirmation; (ii) participant-related parameters: number of participants, agent, patient, individuation and affectedness of the patient and (iii) context-related parameters: presence of a spatio-temporal restriction and use of a prepositional phrase. The other parameters, considered as less salient by the Multiple Correspondences Analyses, will be briefly discussed in a separate paragraph at the end of each section. Since this general graph combines the three languages, the next three chapters, which present the analysis of the results for French, Spanish and English data is introduced by language-specific MCA graphs which make it possible to obtain a more precise picture of how data vary within that language and, if needed, to add parameters that contribute to explaining the variation within one language.

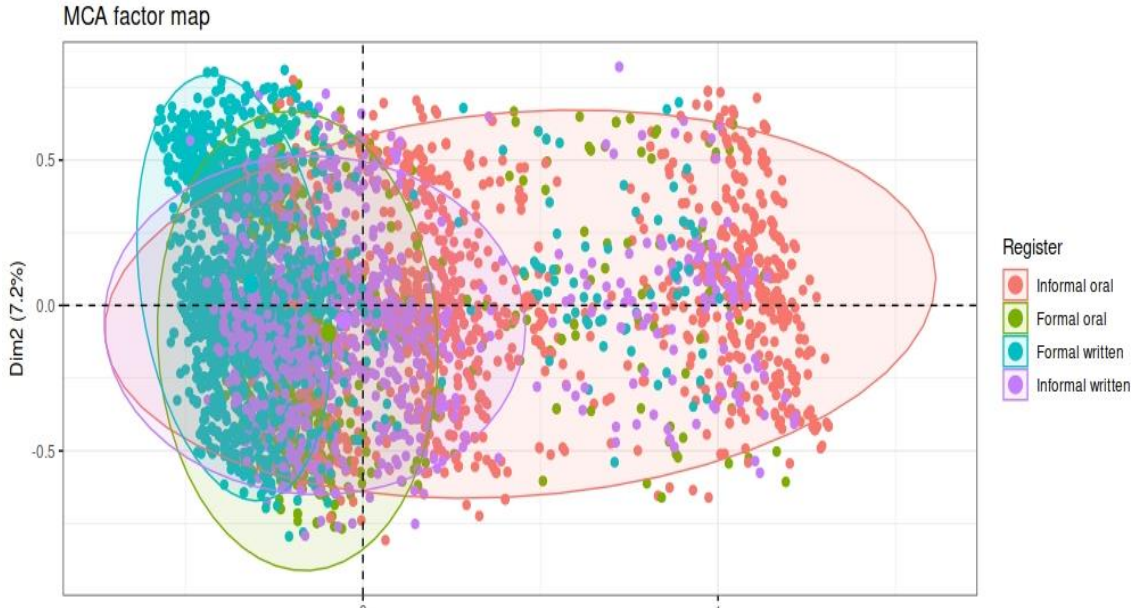



*Figure 13. Factor map for mode and register*

Figure 13 provides information on how the register and the mode combine in our data. It shows that formal written data (blue dots on the graph) are concentrated on the left-hand side of the graph. Its confidence ellipsis largely overlaps the one of formal oral data (represented in green). Besides, a similar form of ellipsis can be observed for informal and formal oral data since they both develop alongside the Y-axis. A look back at Figure 12 shows that the vertical axis corresponds to types of processes (category 5; existential), agent (category 2; inanimate), spatio-temporal restriction (category 1; presence), presence of a prepositional phrase (category 1; presence). Informal written data (represented in purple) rather develop slightly more alongside the X-axis than what has been observed for formal data. This means that informal written data share characteristics both with formal written data and with informal oral data. As can be seen in Figure 13 informal oral data, represented in red, are those with the most heterogeneous profile. Unlike the other three data sets, informal oral data largely develop alongside the right-hand side of the X-axis. This area of the graph shows close relation with variables linked to the number of participants and to the nature of the patients.

# 7. Passive-like structures in French

This first analytical chapter aims to describe the system of passive-like structures in French. It contains five distinct sections. The first section (7.1) reports on the results of the multiple correspondence analysis (MCA) computed with R. The next three sections discuss the results obtained for verb-related, participant-related and context-related parameters. An exhaustive description of the parameters related to the verb is presented in Section 7.2, the parameters related to participants are developed in Section 7.3 and the features related to the context are examined in Section 7.4. Within each section, the contributing parameters (selected after the general MCA including the three languages, as explained in Chapter 6) are first presented in a general graph that includes the four passive-like structures. This enables a profile of each structure to be drawn and compared. The chapter ends with a portrait of the system of passive-like structures in French (7.5).

This first approach does not consider register or mode variation. This information is incorporated later through a detailed analysis of the periphrastic passive and the man-impersonal *on*.

| **Structure** | **Raw frequency** | **Relative frequency (pttw)** |
|---|---|---|
| **Periphrastic passive** | 857 | 42.05 |
| ***Se*-construction** | 73 | 3.58 |
| **Non-anaphoric third person** | 109 | 5.34 |
| **Man-impersonal *on*** | 995 | 48.83 |

*Table 20. Raw and relative frequencies per ten thousand words (pttw) in French data*

As can be seen in Table 20, the total number of occurrences of *se*-constructions (73) and non-anaphoric third person plurals (109) is relatively low. Breaking down the results into mode and register variations would not, therefore, make it possible to obtain reliable information regarding the distribution of the data in these two passive-like structures. The general frequencies for the *se*-construction and the non-anaphoric third person plural are, nevertheless, provided for each register and mode at the end of each section, but for informative purposes only. Quantitative analysis of the influence of mode and register on passive-like structures, therefore, focuses on the periphrastic passive and the man-impersonal *on,* which, once the results have been broken down, still contain a relatively high number of occurrences, enabling us to examine the impact of registers and modes.

## 7.1. Panoramic overview

Multiple correspondence analysis (MCA) graphs make it possible to filter the findings obtained and keep the most contributing parameters, or set differently the parameters that most contribute to explaining the differences and the variation within the data set. Three types of representations are commented: (i) a biplot, which shows the distribution of the variables (here, the parameters studied), (ii) factor maps with point clouds, which pattern the structures, modes and registers, and (iii) a plot, which draws the potential correlation(s) between the variables.

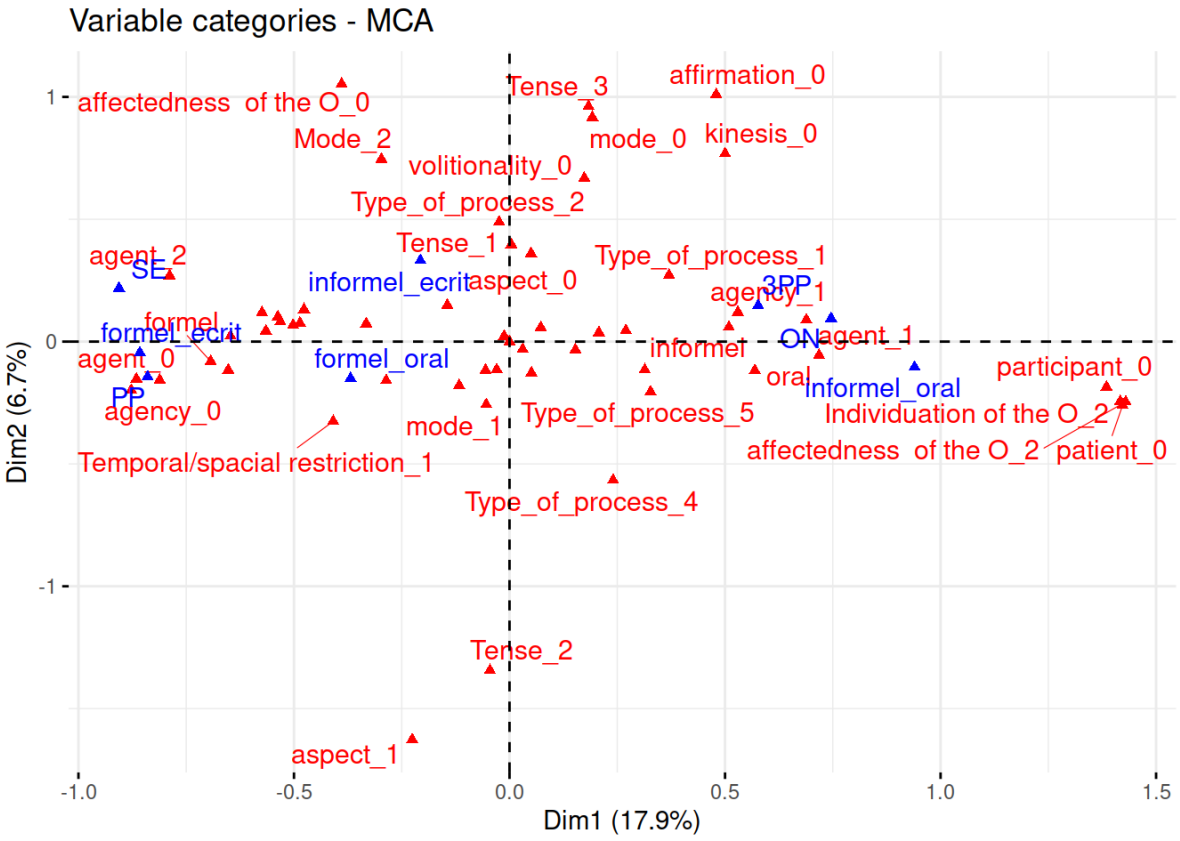



*Figure 14. Distribution of parameters in French passive-like structures*

The visualisation above (Figure 14) includes the structures, the registers and modes (in blue) and the different parameters annotated (in red). The parameters positioned on the extremes of the axes are suspected to contribute most to variation. The distance between two parameters represents their degree of association. Parameters with similar behaviour will thus be grouped together. *Participants* (category 0; one-participant clause), *patients* (category 0; absence of a patient) and *affectedness* and *individuation* of the patient (category 2; individuated), placed on the extreme right of the plot, appear as associated and contributing variables, especially in informal oral data. It can also be observed that this group of variables will be essential in the variation of the man-impersonal *on* and non-anaphoric third person plurals. In MCA, negatively correlated variables are placed at the exact opposite of the plot. A look at the graph reveals that the corresponding variables, negatively correlated with parameters associated with participants

and patients, are those related to *agency* (extreme left). *Agency* will be particularly important when explaining the behaviour of the periphrastic passive and the *se*-construction. This group of parameters will also be closely associated with formal written productions and, to a lesser extent, formal oral and informal written situations.

Let us now look at the vertical axis. It can be observed that *aspect* (category 1; perfective) and *tense* (category 2; past) are situated at the bottom of the axis, while, negatively correlated with these two variables, we find *tense* (category 3; future) and *mode* (category 0; irrealis). At the top right-hand corner of the biplot are positioned the parameters *affirmation* (category 0; negative clause) and *kinesis* (category 0; stative). At the top left-hand corner, *affectedness of the patient* and *mode* are expected to explain more variation in informal written data. Variables positioned close to the origin, i.e., the centre of the plot, such as *type of process*, will contribute less to explaining the variation within the data.

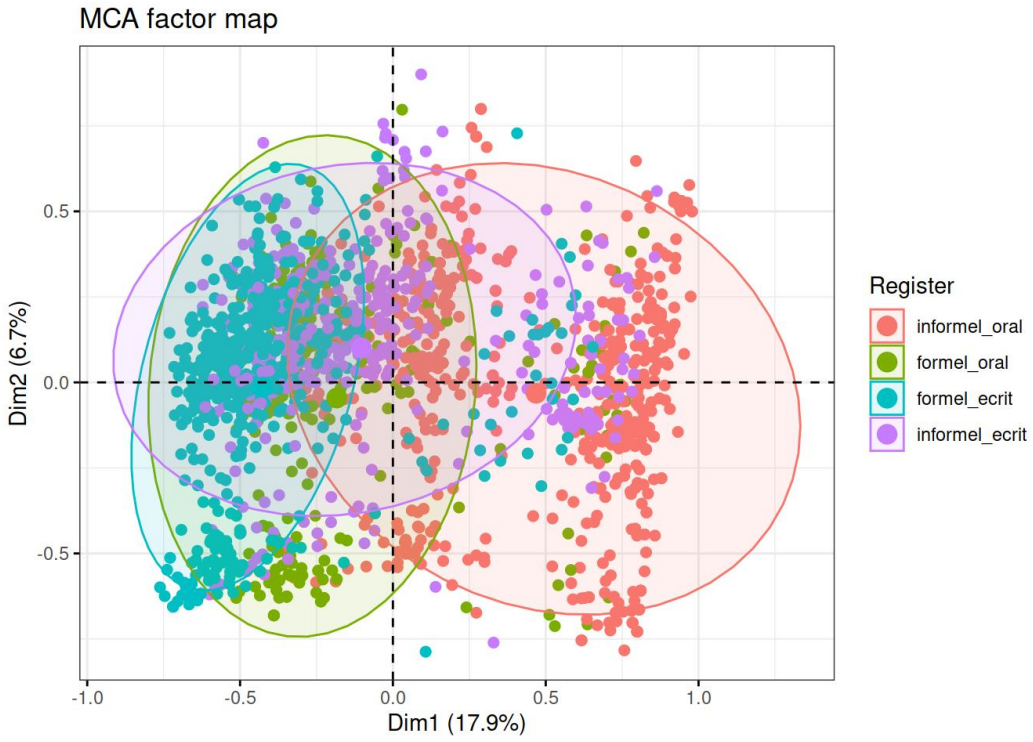



*Figure 15. Factor map for mode and register in French passive-like structures*

The factor map in Figure 15 makes it possible to gather and visualise all the individuals, i.e., all the occurrences of passive-like structures in the two registers and modes. Concentration ellipses are added around each group to better visualise the different patterns. The graph highlights the fact that formal written data, situated on the left-hand side on the plot, are the most homogeneous type of data. Formal oral data appear as slightly less homogenous. However, they display a similar pattern and largely overlap formal written data. Compared with formal written data, formal oral data share more characteristics with informal written and oral productions. The characteristics of passive-like structures used in informal oral data are the least homogenous. These individuals are situated on the right-hand side on the plot, while individuals in formal oral and written and informal written tend to be situated on the left-hand

side. Rather distinct behaviour will thus be expected for constructions used in informal oral data.

There are already some indications about the behaviour of registers and modes in the previous figure (information reported in blue in Figure 15). A more detailed picture can be obtained by projecting Figure 1, which visualises the parameters, onto the point cloud above. Written data, both formal and informal, will be better accounted for in terms of *agency*. In addition to *agency*, informal written data will also be influenced by *aspect* (category 0), *tense* (category 1) and *type of process* (category 2), as shown by the points gathered along the top part of the vertical axis. Formal oral data (represented by the green dots) cover a large part of the centre of the graph, which means they share a series of characteristics with the other three groups. Nevertheless, they also extend into the bottom left of the map. This small area corresponds to *temporal/spatial restriction* (category 1). Special behaviour regarding the use of temporal/spatial restrictions can thus be expected for formal oral data. Informal oral data, as already explained above, will be impacted by parameters related to patients (right-hand side of Figure 15).These parameters will receive further attention in the following sections.

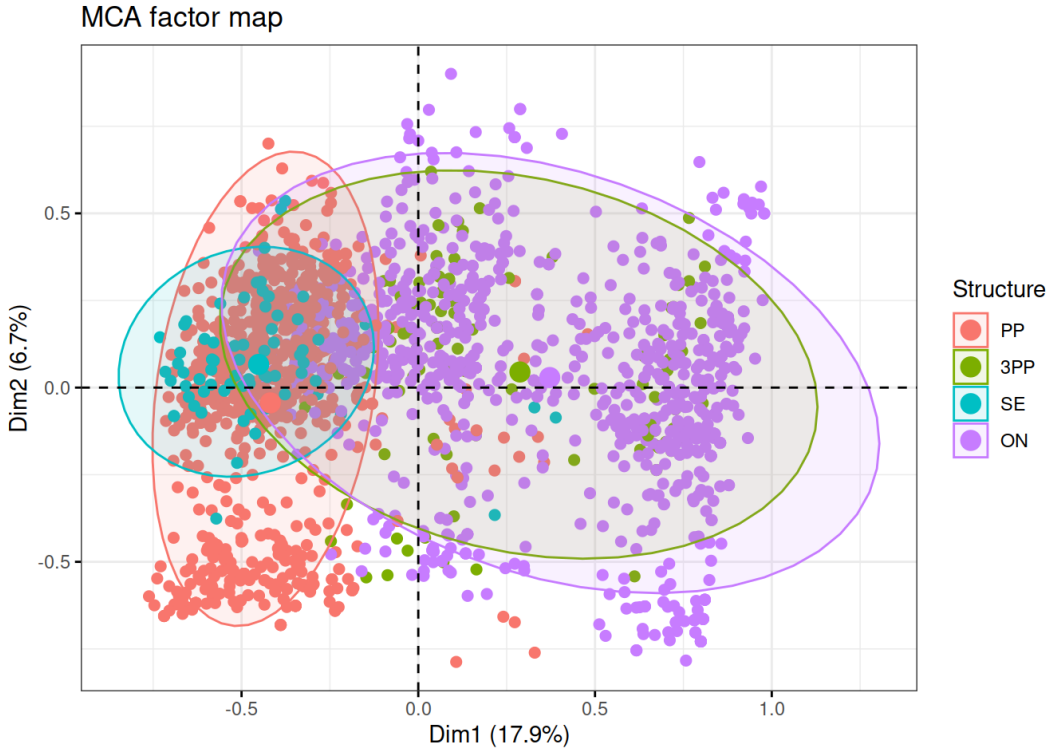



*Figure 16. Distribution of the four passive-like structures in French*

This factor map helps to get a general overview of the distribution of the different structures. Overall, the ellipses indicate two main pairs: the man-impersonal *on* and the non-anaphoric third person plural, with a greater presence on the right of the map, and the periphrastic passive and the *se*-construction occupying the left-hand side of the map. It is interesting to note that the four constructions overlap, meaning that they share a certain number of properties. This is not always the case, as will be seen for Spanish (see Chapter 8). The periphrastic passive almost entirely overlaps the *se*-construction. It can be assumed that these two constructions will be similar and that the *se*-construction is more restricted in its use and may be considered a

specialisation of the periphrastic passive. Detailed analysis of these constructions will help in creating a clear picture of this issue. As these constructions are situated on the left-hand side of the map, a stronger association with the written mode is expected. Indeed, a look back at Figure 14 showing the variables (see information in blue in Figure 14) indicates that written productions are situated on the left-hand side of the biplot, i.e., in the same position as the periphrastic passive and the *se*-construction in Figure 16. The ellipsis of the man-impersonal *on* encompasses and goes beyond the area covered by the non-anaphoric third person plural. This means that in numerous situations both constructions will share parameters but some will be much more specific to the man-impersonal *on*.

Figure 17 shows the potential correlation(s) between the parameters in the French data set. To be considered correlated, two (or more) parameters have to be situated close to each other on

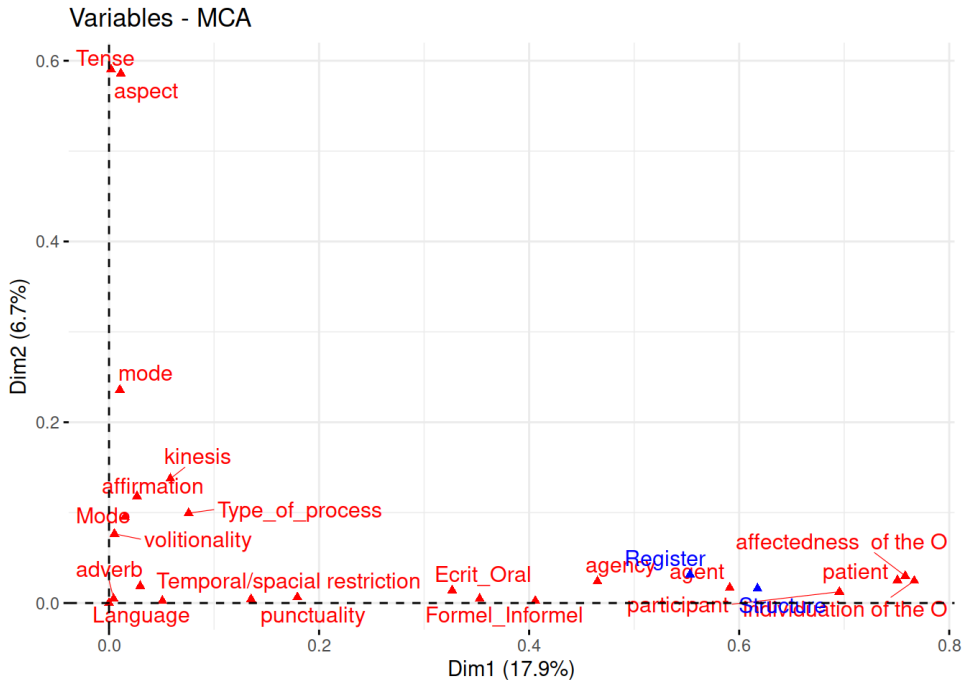



*Figure 17. Correlation(s) in French data*

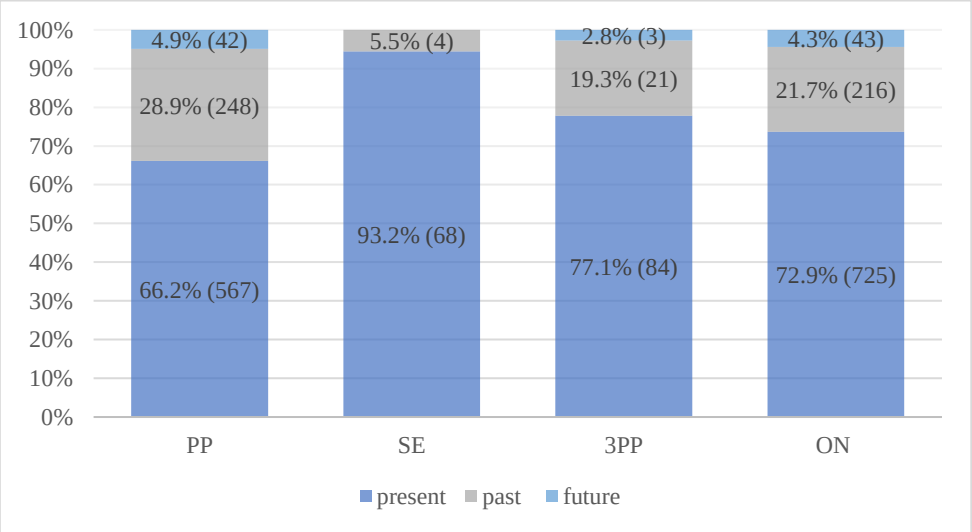
the graph and between the range 0.5 and 1 on the graph. For the French impersonal system, three groups of parameters can be observed. On the vertical axis, the variables *tense* and *aspect*, situated at 0.6, form a first group of parameters that can be considered correlated. A second pair of correlated parameters can be observed on the horizontal axis: participant and agent, both situated between 0.5 and 0.6. A last group of variables, situated between 0.7 and 0.8, are those related to the use of patients: *patient*, *affectedness* of the patient and *individuation* of the patient. Particular attention will be devoted to this phenomenon of correlation when discussing the results.

## 7.2. Verb-related parameters

This section describes the main contributing features related to the verb, as revealed in the MCA: *tense* (7.2.1), *aspect* (7.2.2), *mode* (7.2.3), type of verb process (7.2.4), *kinesis* (7.2.5), *realis/irrealis* dichotomy (7.2.6) and *affirmation* (7.2.7).

### 7.2.1. Tense

A key feature that helps explain the variation between the four passive-like structures is the time period within which the verbal process takes place: present, past or future. Graph 4 offers an overview of the distribution of this tri-partite division of temporal reference across the four passive-like structures in French.




*Graph 4. Distribution of tenses in French passive-like structures*

Present tenses, as illustrated in (167) with the verb *mangent* ('eat') in the non-anaphoric third person plural, predominate in the four passive-like structures but the proportion varies according to the structure.

(167) *<answer> ben non ce n'est pas de la tarte puisque c'est du jell-o !..... dans la cuisine chinoise il y a aussi un désert qui ressemble à du jell-o ( aussi mauvais !... )* ***ils le mangent*** *avec une cuillère en porcelaine* (YAHOO FR)
'well no, it's not pie since it's jell-o ! ..... in Chinese cuisine there is also a dessert that looks like jell-o (just as bad!... ) **they eat** it with a porcelain spoon'

Interestingly, the two defocussing mechanisms involving a personal pronoun, i.e., the non-anaphoric third person plural and the man-impersonal *on*, have a remarkably similar profile: between 73% and 77% of present tenses, approximately 20% of past tenses and a small percentage of future tenses. Tense is therefore not the parameter that would differentiate these two passive-like constructions. However, results reveal significant differences between the distribution of tenses within these two structures, on the one hand, and the periphrastic passive

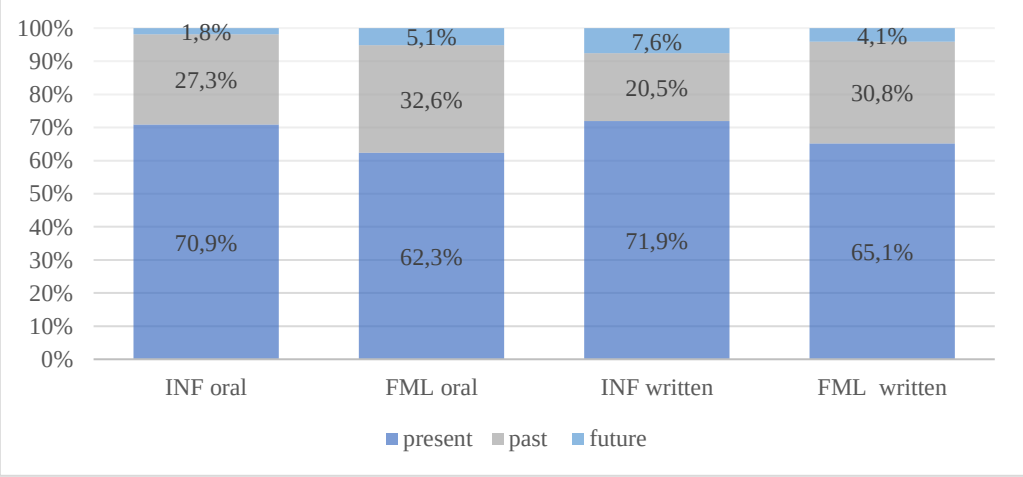
and the *se*-construction[28], on the other. Even though the MCA revealed that the *se*-construction largely overlaps the area covered by the periphrastic passive, Graph 5 indicates that these two structures diverge in their use of tenses. It can be observed that while the *se*-construction almost exclusively relies on present tenses, this proportion dramatically falls, though remains above the 50% threshold, for the periphrastic passive (66.2%). A chi-square test reveals that this difference in the proportion of present tenses is significant ($\chi^2 = 24.5625$, $p < .05$). In his cross-linguistic approach of Swedish, Danish, and Norwegian impersonal constructions, Viberg (2010) also notes that one type of impersonal construction, the *s*-passive available in Danish and Norwegian, is much more productive in present tenses. The author puts forwards that the usage pattern of the construction, that is, to primarily express a general norm or a generic meaning contributes to explaining the high use of present forms. A similar explanation based on the meaning and the interpretation of the passive-like construction could explain the higher presence of present tenses in French *se*-construction. In (168), it is a general norm that parsnips are sown from February to June.

(168) *Le panais **se sème** de février à juin et se récolte 4 mois après le semis*. (WIKI FR)
'Parsnips **are sown** from February to June and harvested 4 months after sowing.'

This observation remains, however, tentative as the number of se-constructions in this study is relatively restricted.

Our results further corroborate Viberg's findings (2010) that periphrastic passives tend to be used to reflect specific situations or events, especially with past tenses. As shown in Graph X, the French periphrastic passive is the passive-like construction that contains the highest number of past tenses, preferred candidates to express specific events. This result is further supported by the high presence of perfective forms in the periphrastic passive, as will be detailed in the next section.

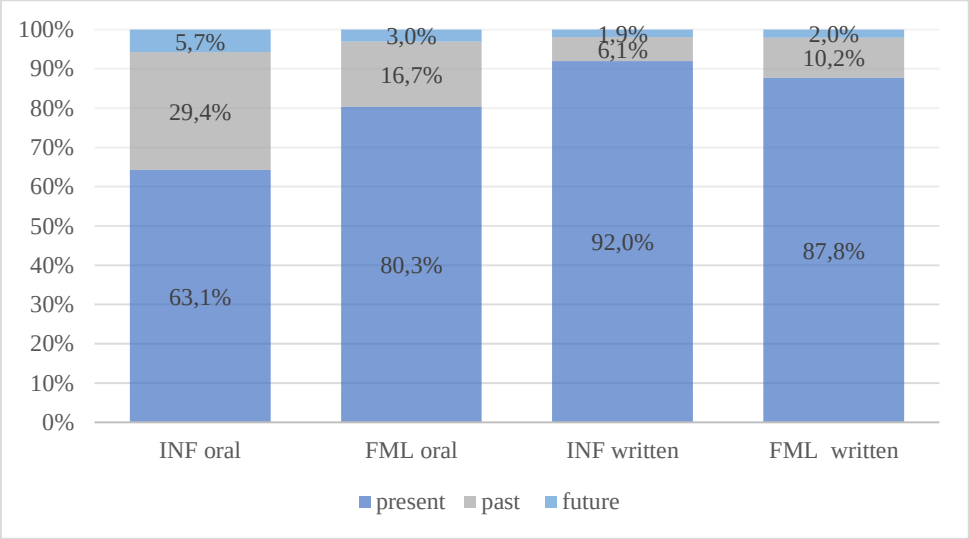
Let us now consider the influence of register and mode on the parameter *tense* by examining periphrastic passives and man-impersonals *on* more closely.




*Graph 5. Distribution of present, past and future tenses in French periphrastic passives*

[28] The difference in the proportion of present tenses is significant for PP vs ON ($\chi^2 = 12.3703$, $p < .05$), PP vs 3PP ($\chi^2 = 5.8964$, $p < .05$), SE vs 3PP ($\chi^2 = 9.1353$, $p < .05$) and SE vs ON ($\chi^2 = 15.4686$, $p < .05$).

Graph 5 breaks down the results for the periphrastic passive and gives a more precise view of the distribution according to mode and register. In the two modes and registers, present tenses account for approximately two-thirds of the data. A closer look at Graph 5 points towards the influence of the degree of formality, since the frequency of present tenses seems to be higher in informal data than in formal data. Chi-square tests reveal, however, that the difference between the two registers is not significant, in either mode. Past tenses are used slightly more in formal data (30.8% in written data and 32.6% in oral data), but there is no significant difference compared with the percentages obtained for informal data. It can thus be assumed that tenses in periphrastic passives are little impacted by modes and registers and have fairly stable behaviour.



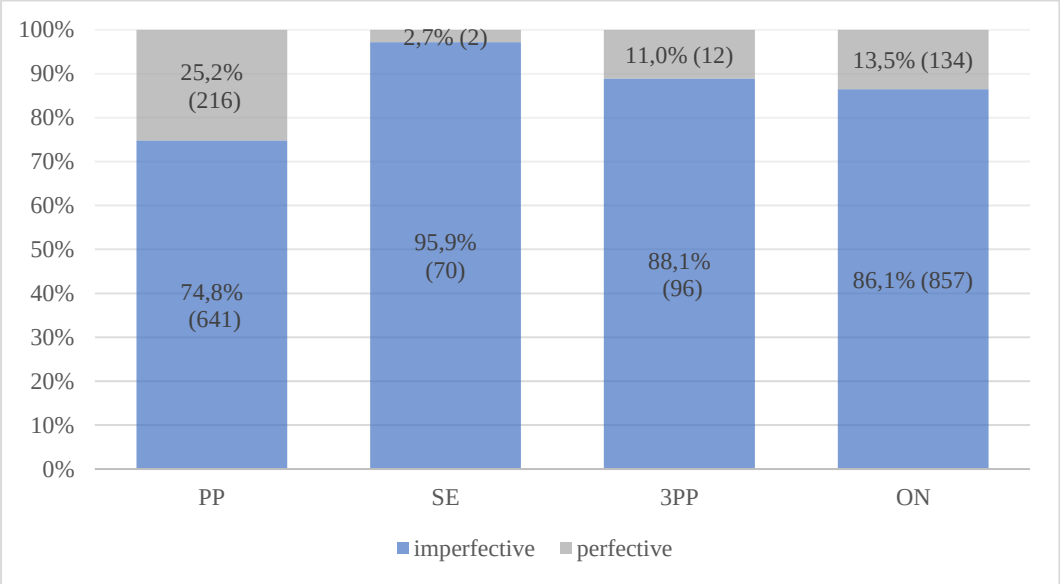

*Graph 6. Distribution of tenses in the French man-impersonal on*

The man-impersonal *on* occurs mainly with verbs in present tenses (72.86%; 725 occurrences), followed by past tenses (21.71%; 216 occurrences) and, to a lesser extent, future tenses (4.32%; 43 occurrences)[29]. Compared with periphrastic passives, it can be noted that present tenses are slightly more frequent with *on*, at the expense of past tenses. Regarding mode variation, the results reveal that past tenses are more frequent in oral data, especially in informal oral, than in written data ($\chi^2 = 60.6195$, $p < .05$) (see Graph 6). Oral contexts thus seem more appropriate to relate past experiences, while written data appear to be more anchored in the present situation. It is premature to affirm that the tenses used in passive-like structures are impacted by register. Even though there are significantly more past tenses in informal oral data compared with formal oral data ($\chi^2 = 5.8291$, $p < .05$), no such difference between formal and informal written data was found. While present and past tenses seem to be better candidates to vary across mode and register, the proportion of future tenses does not vary substantially.

[29] The final percentage may not total 100%, as some occurrences could not be annotated. This may occur, for instance, in oral data when the speaker does not finish a sentence or pronounces the construction in an incomprehensible or inaudible manner.

### 7.2.2. Aspect

Graph 7 shows the distribution of perfective and imperfective aspects in the four different structures of the French impersonal system.



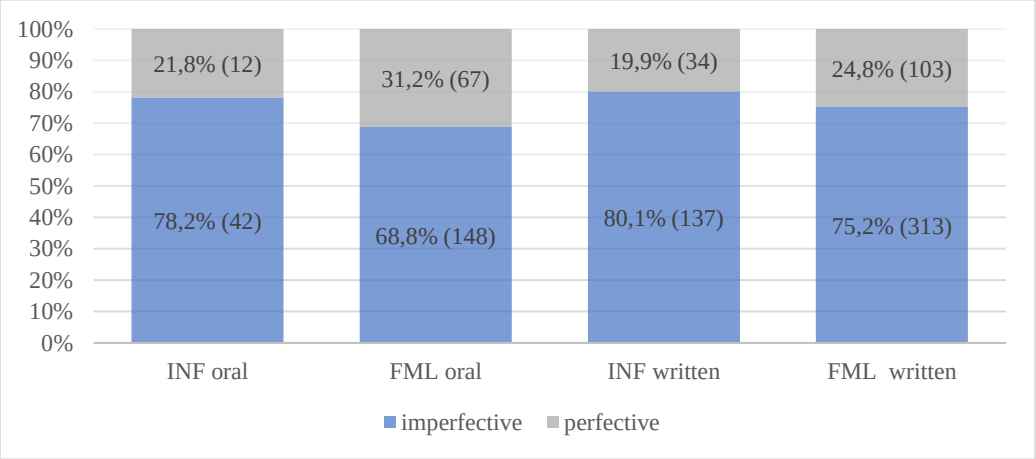

*Graph 7. Distribution of perfective and imperfective forms in French passive-like structures (parameter aspect)*

It can be observed that, for the four structures, the imperfective aspect largely predominates (see Graph 7). In examples (169) and (170), the verbs *se plaît* ('like') and *est présenté* ('is presented') both refer to a reality that is still ongoing at the time of speaking.

(169) *Monsieur le Président, Monsieur le Commissaire, qu'il est bon de se taper sur l'épaule en se félicitant du travail bien fait.{S} De COP en COP,* ***on se plaît*** *à dépeindre l'Union européenne comme étant à la pointe du combat climatique.* (PROCEP FR)
'Mr President, Commissioner, it is good to pat oneself on the back to congratulate oneself on a job well done. {S} From COP to COP**, people like** to portray the European Union as being at the forefront of the fight against climate change.'

(170) *Tel que le texte* ***est présenté*** *et amendé par la commission de l'environnement, de la santé publique et de la sécurité alimentaire et par la commission de l'agriculture et du développement rural, on ne pourra pas s'assurer que des produits issus d'animaux clonés ou de leurs descendants n'arrivent pas jusque dans nos assiettes.* (PROCEP FR)
'As the text **is presented** and amended by the Committee on the Environment, Public Health and Food Safety and the Committee on Agriculture and Rural Development, it will not be possible to ensure that products from cloned animals or their descendants do not reach our plates.'

As is the case for the parameter *tense*, the man-impersonal *on* and the non-anaphoric third person plural behave similarly to *aspect*, with 86.1% and 88.1%, respectively, of the imperfective aspect. This result is, nevertheless, expected since analysis revealed that these two parameters are highly correlated. The imperfective aspect is more common in structures with a higher proportion of present tenses, such as the *se*-construction, while the perfective aspect predominates in structures resorting more to past tenses, such as the periphrastic passive. This correlation was expected since a series of past tenses focuses on the termination of an event in a time period occurring before the moment of speaking. These tenses, such as the *passé simple* in French, correspond to the perfective aspect. This ability to focus on termination is not available with present tenses.

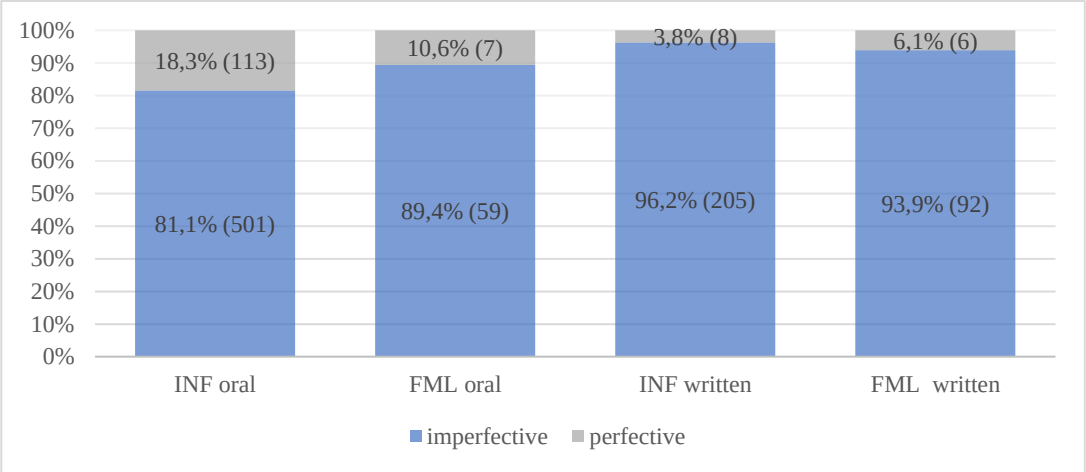



*Graph 8. Distribution of imperfective and perfective forms in French periphrastic passives*

Breaking down the results into formal/informal and written/oral data reveals that, for the periphrastic passive, the proportion of verbs in the perfective aspect varies according to the register and the mode, where the perfective is used in 20% to 31% of cases. Interestingly, as can be observed in Graph 8, the perfective aspect seems to be more frequent in formal data than in informal data. Example (171) illustrates a case of the perfective aspect with a periphrastic passive. The action of *ont été plantées* (‘has been planted’) is seen as totally completed.

(171) *Par ailleurs, les consoudes **ont été plantées** le long des chemins de pèlerinage pour qu'elles soient disponibles partout pour un usage médicinal.* (WIKI FR)
‘In addition, comfrey **has been planted** along the pilgrimage routes so that it is available everywhere for medicinal use.’

However, chi-square tests reveal that the difference is not significant, in either of the two modes. The proportion of clauses in the imperfective aspect is also similar in the two registers. The behaviour of the parameter *aspect* can thus be considered stable, with no influence of the mode or the register. While aspect across the periphrastic passive does not vary greatly, the picture is different for the construction with *on.*



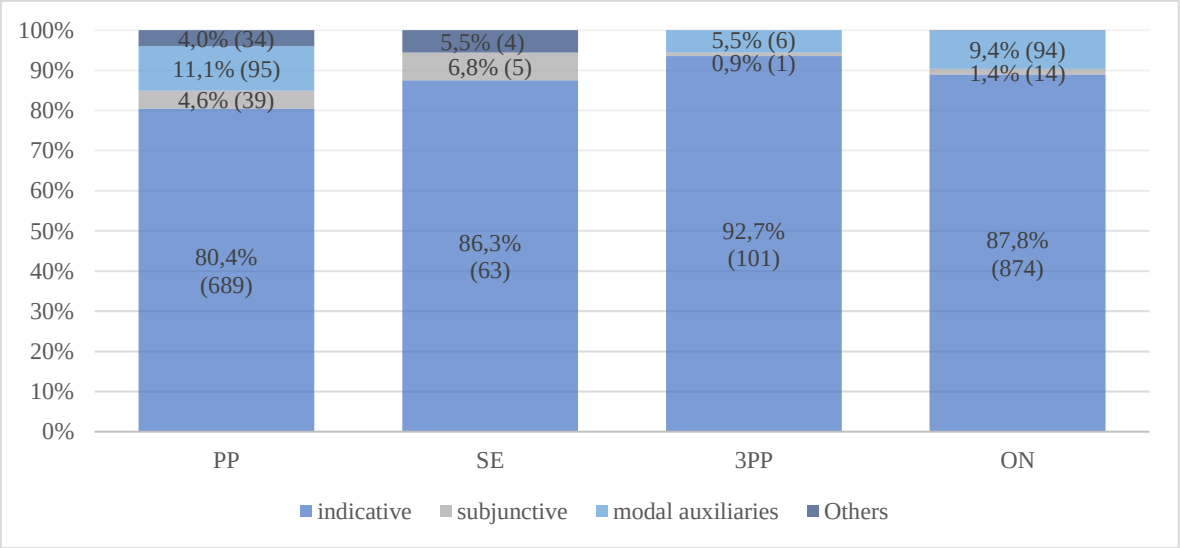

*Graph 9. Distribution of imperfective and perfective forms in French man-impersonals* on

As can be observed in Graph 9, oral data appear to be better candidates for the perfective aspect. Results indicate a significantly higher use of perfective aspects in oral informal data than in written informal data ($\chi^2 = 27.1658$ $p < .05$). Even though there is a similar tendency for formal data, the difference is not significant in our set of data. This absence of a significant difference regarding *aspect* between formal oral and formal written data may be attributed to the nature of the communicative interaction, which is more normalised in formal contexts. Moreover, the formal oral corpus comprises proceedings of the European Parliament, productions that, as discussed in the methodology section, may have been prepared beforehand. This prior preparation of the oral discourse may be a factor leading to the inclusion of features usually associated with written productions. Such a resemblance was not observed for the feature *tense.* No differences between registers could be found for the man-impersonal *on.*

One final relevant piece of information emerges from a comparison of the frequency of the perfective aspect in periphrastic passives and the man-impersonal *on,* i.e., the two passive-like structures most used in our set of data. Formal data resort far more to perfective aspects with periphrastic passives than with man-impersonal *on.* This tendency is also observed for informal written data, but not for informal oral data, where both structures have a similar distribution.

### 7.2.3. Verb mode

The parameter *verb mode* has been divided into four categories: indicative, subjunctive, modal auxiliaries and others.




*Graph 10. Distribution of verb modes in French passive-like structures*

Graph 10 shows how verb modes are distributed across the four structures studied in French passive-like structures. It can be observed that in all constructions, the indicative mode is the preferred type of modality to express the speaker's attitude towards the world (see example (172)).

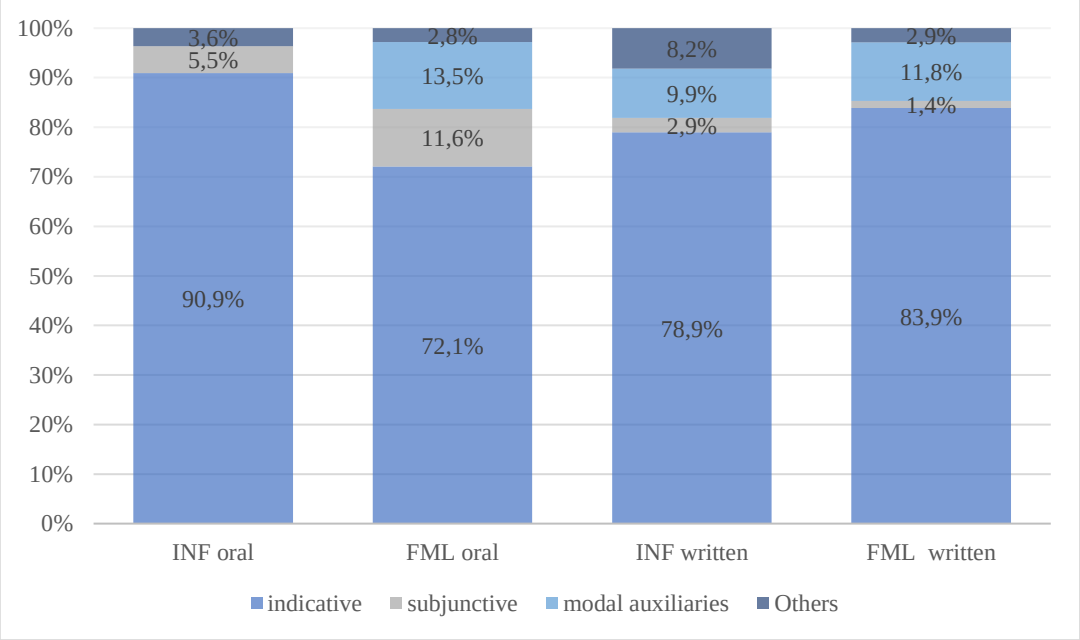
(172) *Je pense – et c'est le sens des amendements que j'ai déposés, et qui **n'ont pas été adoptés** par la commission INTA, qui seront mis aux voix demain – qu'il y a un problème concernant le déclenchement des clauses de sauvegarde.* (PROCEP FR)
'I think - and this is the sense of the amendments I tabled, which **were not adopted** by the INTA Committee, and which will be put to the vote tomorrow - that there is a problem regarding the triggering of safeguard clauses.'

It can, however, be noted that the distribution and the frequency of use of other types of modality vary according to the structure, the periphrastic passive being the most prone to variation. As can be seen, periphrastic passives are those with the lowest proportion of indicatives. It is interesting to note that the two defocussing mechanisms based on an impersonal subject pronoun are those with the highest proportion of indicatives. The difference is significant for the pairs involving the periphrastic passive, i.e., periphrastic passive vs non-anaphoric third person plural ($\chi^2 = 11.1238$, $p < .05$) and periphrastic passive vs man-impersonal *on* ($\chi^2 = 25.9572$, $p < .05$). Viberg (2010) indicates that the Swedish man-impersonal frequently occurs with a modal or in a conditional or temporal clause. In French, the man-impersonal *on* obtains a percentage of 9.4% of modal auxiliaries, a value very similar to the one obtained by the periphrastic passive. However, Graph 10 shows a slight tendency for modal auxiliaries to

be more used with a man impersonal than with a non-anaphoric third person plural. This result would corroborate Viberg's observation on the use of modals in man-impersonal constructions. As a similar trend can be observed in the periphrastic passive, it can be assumed that, when used as a defocussing mechanism, *on* comes closer to the prototypical agent-defocussing mechanism.

The non-significance of the results for pairs involving *se*-constructions may be partly attributed to the restricted number of occurrences obtained for that construction.

Having given a panoramic overview of verb mode, let us refine the analysis and break the results down by registers and modes. For the *se*-construction, this reveals that more than half of the occurrences (46 out of 73) are used in formal written data. Due to the scarcity of occurrences, only this combination of register and mode will be briefly commented. Results reveal the following distribution of modality: indicative 82.6% (38 occurrences), modal auxiliaries 10.9% (5 occurrences) and others 6.5% (3 occurrences). Similarly, the breakdown for the non-anaphoric third person plural reveals that 67.9% of the occurrences (74 out of 109) are used in informal oral data and that the non-anaphoric third person plural mainly resorts to the indicative mode (91.9%). The analysis of modality across register and mode variation will thus be centred on the periphrastic passive and the man-impersonal *on*, which have a much higher number of occurrences.




*Graph 11. Distribution of verb modes in French periphrastic passives*

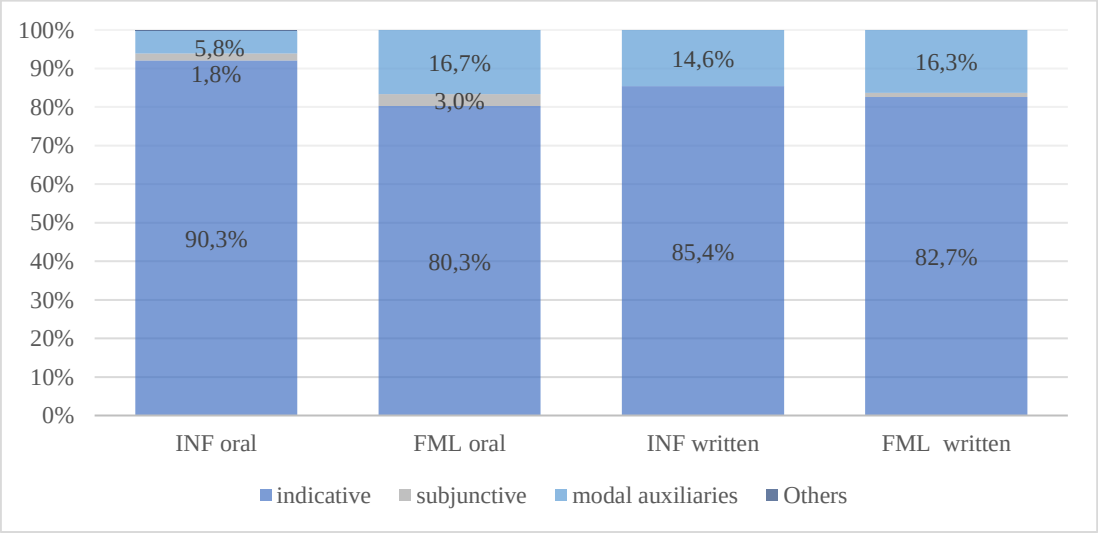
As can be seen for periphrastic passives, no clear influence of the mode and register can be noted as far as the use of the indicative is concerned (Graph 11). The main interesting result that emerges is the higher proportion of indicatives in informal oral data, while formal oral data resemble written productions much more. Formal oral data have a higher level of subjunctives (see example (173)).

(173) *Le Parlement a déjà voté la nécessité ici de faire en sorte que la plus-value* ***soit répartie*** *entre les producteurs et que les agriculteurs ne soient pas victimes des entreprises et de la grande distribution.* (PROCEP FR)
'Parliament has already voted on the need to ensure that the profits **are distributed** among producers and that farmers do not fall victim to companies and supermarkets.'

This could be explained again by the nature of the data (proceedings of the European Parliament), where the interventions may require much more caution and be conceived in terms of possibility and probability, whereas informal oral data, i.e., conversations between family members and/or friends, are much more anchored in the real world rather than that of possibility and potentiality, thus relying much more on indicatives.

(174) *{S} Si ce programme* ***doit être financé*** *– et il devra sans doute l'être, au titre de l'aide sociale, du Fonds social européen – il faudra alors compenser.* (PROCEP FR)
'If this programme is to be financed - and it will probably **have to be financed**, under social assistance, from the European Social Fund - then we will have to compensate.'

Similarly, informal oral is the only category that does not rely on modal auxiliaries. This can be explained once again by the fact that informal oral interactions are much more related to real experiences, while formal oral interactions and written productions seem to be better suited to contain periphrastic passives with modal auxiliaries, such as in (174).



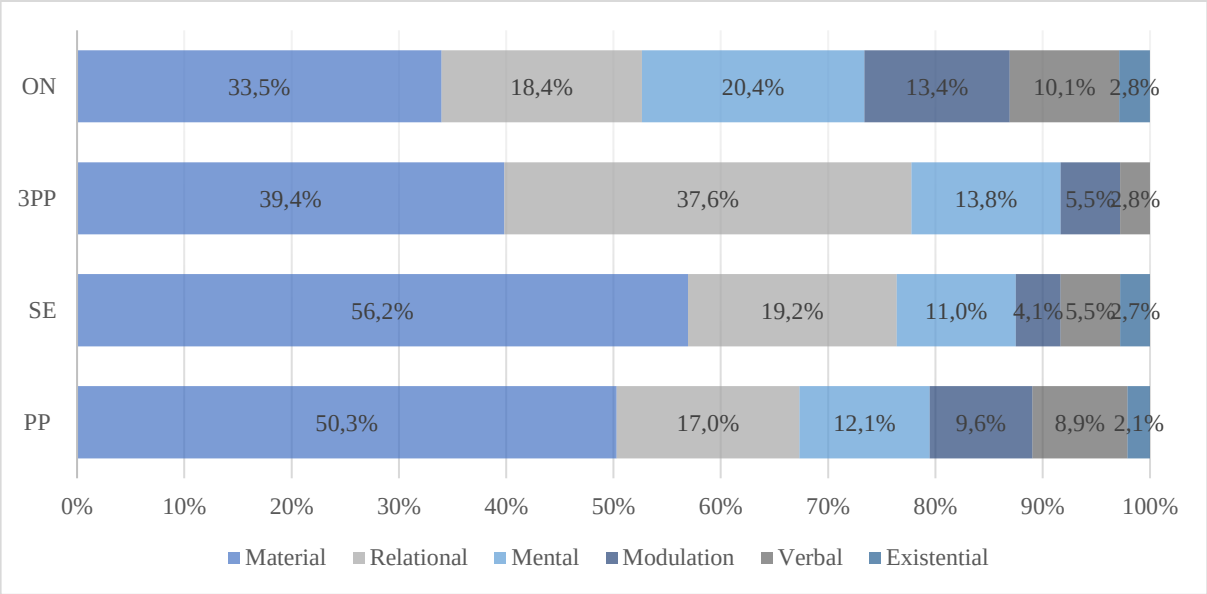

*Graph 12. Distribution of verb modes in French man-impersonals* on

As far as constructions with the man-impersonal *on* are concerned, comparison of the two registers reveals that there are slightly more indicatives in informal data than in formal data. The difference is, however, only significant in oral data ($\chi^2 = 10.399$ $p < .05$). Conversely, mode does impact the distribution of the data. Interestingly, as was the case for the periphrastic passive, formal oral resembles the profile of written data, be it formal or informal. Statistical tests did not reveal any significant difference between the use of indicatives in formal oral,

formal written and informal written. This similar behaviour may be linked to the nature of the productions (proceedings of the European Parliament), as was the case for the parameter *aspect.* As parliamentary debates are planned speech and sometimes considered as a written-to-be-spoken genre, the overlap with some characteristics of the written mode is to be expected.

### 7.2.4. Type of verb process

The nature of the type of verb process can be regarded as a fundamental element to consider in syntactic analysis. To our knowledge, the type of verb process included in defocussing mechanisms has not been studied. We assume that the type of verb used in a passive-like structure determines the choice of the passive-like structure. Six main categories are usually reported to classify verbal arguments: *material processes*, *relational processes*, *mental processes*, *modulation processes*, *verbal processes* and *existential processes*. Graph 13 shows how the distribution of these types is organised within each passive-like structure. It also enables the structures to be compared and highlights their specific features. We will see that, although the four structures share common ground, such as wide use of material processes, each structure has some distinctive characteristics.




*Graph 13. Type of verb process in French passive-like structures*

It can be seen from Graph 13 that two categories largely predominate: *material* and *relational* processes (illustrated with the periphrastic passive in (175) and (176), respectively).

(175) *Les pâtisseries, les crêpes et d'autres préparations culinaires traditionnelles peuvent* ***être confectionnées*** *sans œufs ni produits laitiers*. (WIKI FR)
'Pastries, pancakes and other traditional culinary preparations **can be made** without eggs and dairy products.'

(176) *Cela est dû au fait que, pour produire de la viande, de l'eau doit être utilisée dans la production de nourriture pour animaux, qui doit* ***être donnée*** *aux animaux pendant leur vie entière.* (WIKI FR)

‘This is due to the fact that, in order to produce meat, water has to be used in the production of animal feed, which **has to be given** to the animals throughout their lives.’

Material and relational processes, nevertheless, occur in relatively different proportions according to the passive-like structure with which they are used. Let us first consider material processes, which refer to physical actions and can be divided into four subcategories: *space*, *change*, *behaviour* and *others*. In the periphrastic passive and the *se*-construction, material processes account for more than half of the verb processes, 50.3% for the periphrastic passive (401 occurrences) and 56.2% for the *se*-construction (41 occurrences). In periphrastic passives, the type of material process categorised as *changes*, such as in (177), predominates (187 occurrences), followed by *others* (116 occurrences) and *space* (98 occurrences). In (177), the verb *sont détruits* (‘are destroyed’) implies a change of state, from normal to no longer existing or existing in a severely damaged state.

(177) *Quatre de ces vaisseaux **sont détruits** par des accidents, la plupart pendant leur transfert au hangar*. (WIKI FR)
‘Four of these vessels **are destroyed** by accidents, most of them during their transfer to the hangar.’

Example (178) illustrates a verb classified as *others*. This rather heterogeneous category includes verbs referring to any physical type of action that does not meet the criteria to be related to *changes*, *space* or *behaviour*.

(178) *Un zeppelin désigne un aérostat de type dirigeable rigide, de fabrication allemande, mais **est souvent utilisé** dans la langue populaire pour désigner n'importe quel ballon dirigeable*. (WIKI FR)
‘A zeppelin refers to a rigid airship, of German manufacture, but **is often used** in popular language to refer to any airship.’

In (178), the verb *utilisé* (‘used’) denotes a physical manipulation but the patient, here the noun zeppelin, does not suffer any modification.

(179) *Les présidents des groupes, notamment de la droite et des socialistes, se sont opposés à cette commission d’enquête et ont bloqué la possibilité que nous avions de vérifier les conditions réelles dans lesquelles les animaux **sont transportés***. (PROCEP FR)
‘The chairmen of the groups, notably the right and the socialists, opposed this committee of enquiry and blocked the possibility for us to check the actual conditions under which animals **are transported**.’

The verb *sont transportés* (‘are transported’) in (179) illustrates the category *space*, where the animals are moved from one place to another.

Similarly to what was found for the periphrastic passive, the most common subcategory of material process in *se*-constructions is *changes* (26 occurrences).

(180) Le panais **se sème** de février à juin et se récolte 4 mois après le semis. (WIKI FR)
'Parsnips **are sown** from February to June and harvested 4 months after sowing.'

The process of sowing parsnips in (180) implies a change of state from a seed that becomes an element in the ground with the potential to develop into a plant. The other subcategories of material processes in *se*-constructions have only a very limited number of occurrences.

Compared with the periphrastic passive and the *se*-construction, the man-impersonal *on* (33.5%) and the non-anaphoric third person plural (39.4%; 43 occurrences) use fewer material processes. As regards the man-impersonal *on*, verbs belonging to the sub-category *others,* which include, for instance, processes related to contact, control, use or activity, are the most represented within this category. Example (181) illustrates a material verb classified as *others*, the verb *utiliser* 'to use'.

(181) *Les clarifiants et stabilisants du vin, www.vinexpert.com. Consulté le 03 juin 2008. « Or on sait que la bentonite aide à éliminer cette matière colorante lorsque* ***on l'utilise*** *pour traiter le vin avant un collage à la gélatine. » (…)* (WIKI EN)
'Wine clarifiers and stabilisers, www.vinexpert.com. Accessed on 3 June 2008. "Bentonite is known to help remove this colourant when **used** to treat wine before fining with gelatin."'

The most frequent sub-category of material verb process for the non-anaphoric third personal plural is *spaces,* e.g., in (182). In this example, the woman is hired by the company where she did her trial. As the employers decided to keep her in their company, the person will thus maintain the same position and occupy the same spot in the organisation. This notion of keeping the person in the same location leads us to classify example (182) in the category *space.*

(182) (*0.9) mais bon/ .h ils la gardent (.) elle l'a su elle a eu son mois d'essai* ***ils la gardent*** *(0.8) en plus/ (0.6) cette semaine elle nous a appelés/ (0.5) pour nous dire t` sais qu` normalement quand tu fais une formation comme ça/ tu es prise automatiquement tu es prise dans l'entreprise/* (CIEL FR)
'(0.9) well/ .h they're keeping her (.) she knew she had her month's trial **they're keeping** her (0.8) in addition/ (0.6) this week she called us/ (0.5) to tell us you know that normally when you do a training course like this/ you're automatically taken on in the company/'

So far we have seen that material processes are the most frequent type of verb process in the four structures, albeit in different proportions and with different types of subcategories predominating. Let us now turn to relational processes, which appear to be the second most frequent type of verb process, except for the man-impersonal *on.* Relational verbs can be classified into two subcategories: *possession* and *attribution*.

In the non-anaphoric third person plural, relational verb processes are nearly as frequent as material processes since they account for 37.6% of the cases (41 occurrences). Verbs of *possession,* such as in (183), are the most frequent (26 occurrences).

(183) *pour le visa tourisme australien il faut juste s'enregistrer sur internet aupres du site des visas et payer 20$ avec CB de frais d'enregistrement,* ***ils te donnent*** *un numero que tu donnes quand tu arrives en australie c'est tout. (YAHOO FR)*
'for the Australian tourist visa, you just have to register on the internet on the visa website and pay $20 registration fee by debit card, **they give** you a number that you give when you arrive in Australia that's all.'

In (183), someone receives an entity, in this case a number, and becomes the owner of that entity. In the three other passive-like structures, the proportion of relational verbs is much lower, as they range between 17% and 19.2%. It is thus assumed that a high use of relational verbs is characteristic of the non-anaphoric third person plural and differentiates it from the other three structures, especially from the other pronoun-based passive-like structure with which it usually shares a wide range of features.

Mental verbs occupy the third most important position in all passive-like structures except for the man-impersonal *on*, the only structure in which these verbs are the second most frequent category (20.4%; 203 occurrences). Unlike the other three structures, the man-impersonal *on* has more occurrences of mental verbs than relational verbs. Mental verbs governed by man-impersonals are either verbs of *perception*, *cognition* or *sensation*. This result concurs with Fløttum et al. (2004:103)'s claim about the preferential use of *on* with cognitive and perceptive verbs in scientific research articles. Our results make it possible to generalise Fløttum et al. (2004)'s finding and suggest that a high frequency of mental verbs is a defining characteristic of the French-man-impersonal.

While verbs of *perception* (184) (92 occurrences) and *cognition* (185) (77 occurrences) dominate in the mental category, the use of verbs of *sensation* (34 occurrences) remains rather limited compared with the other two categories.

(184) *{S}Aujourd'hui* ***on peut voir*** *un de ces appareils au National Air and Space Museum situé à Washington D.C. (États-Unis) (...)* (WIKI EN).
'Nowadays, **we can see** one of these aircraft at the National Air and Space Museum in Washington DC (United States) (…).'

(185) *{S}* ***On connaît*** *la formule de Talleyrand: «Quand je me regarde, je me désole, quand je me compare, je me console», mais méfions-nous, car si nous sommes tous désolants, ce sont nos enfants et nos petits-enfants qui seront désolés.* (PROCEP EN)
'**We know** Talleyrand's formula: "When I look at myself, I feel sorry, but when I compare myself with others, I am reassured", but beware, because if we are all sorry, it is our children and grandchildren who will be sorry.'

As regards the other three passive-like constructions, a different subcategory of mental verbs predominates. For the periphrastic passive construction, the analysis reveals that verbs of *cognition* are the most common (44 occurrences), while verbs of *sensation* (8 occurrences) predominate in the non-anaphoric third person plural. No clear pattern emerged for the *se*-construction, as very few occurrences of mental verbs were obtained. These results show that each passive-like structure displays a preference, on the one hand, regarding the proportion of

mental verbs, with the man-impersonal *on* resorting to them far more, and, on the other, regarding the sub-categories of mental verbs used.

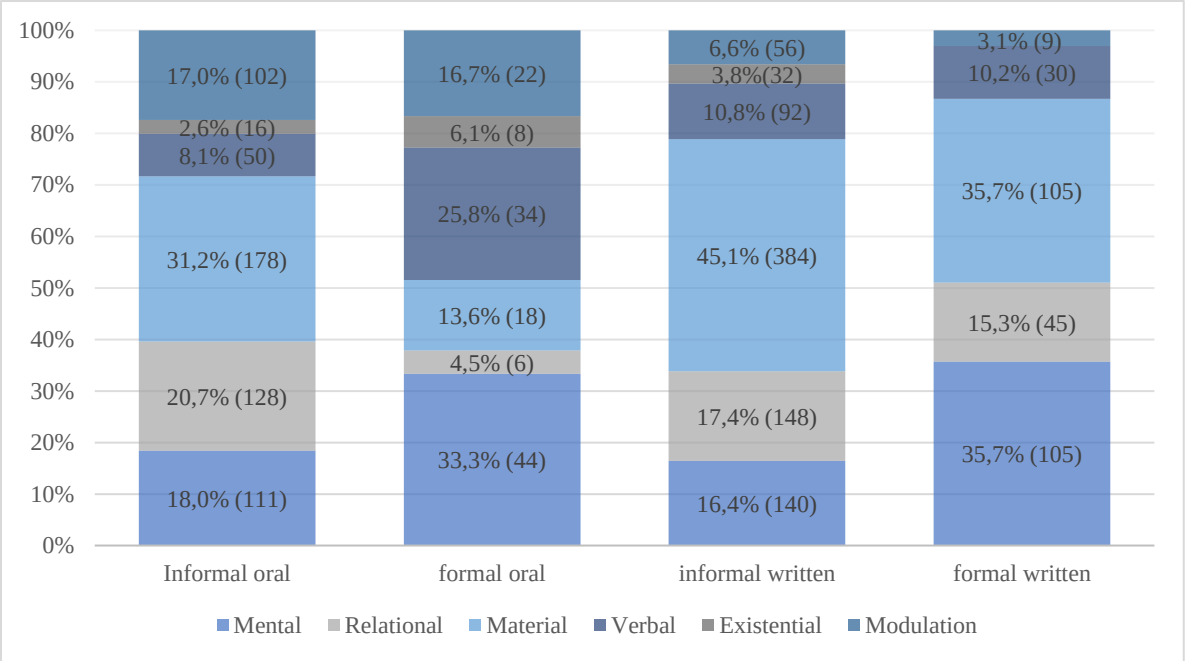
The presence of the other three types of verb processes, *modulation*, *verbal* and *existential*, is relatively restricted in the four structures. Existential verbs are even absent from non-anaphoric third person plurals. The only relevant result regarding these categories is the presence of modulation verbs in man-passive-like constructions, which score fairly high (13.4%; 133 occurrences) compared with the other passive-like structures. Example (186) illustrates a case of modulation verbs from the sub-categories *acceptation.* In (186), Members of Parliament cannot agree with the fact that people speculate on food issues.

(186) *{S}* ***On ne peut pas accepter****, sur la question alimentaire, que des spéculateurs viennent spéculer sans tenir compte des conséquences que cela peut avoir puisque il s'agit de la nourriture du monde et de l'humanité*. (PROCEP)
‘{S} **We cannot accept**, on the food issue, that speculators come and speculate without taking into account the consequences that this may have, since we are talking about feeding the world and humanity’.

In short, we have seen that material processes are the most frequent type of verb, followed by relational verbs, except for the man-impersonal *on.* The periphrastic passive and the *se*-construction have a fairly similar profile, while the non-anaphoric third person plural and the man-impersonal *on* have quite a specific profile. The non-anaphoric third person plural shows a balance between material and relational verbs, while the man-impersonal *on* is the only structure to have such a high proportion of mental and modulation verbs. The frequent use of mental verbs in the man-impersonal may partly be explained by the nature of the structure, involving an overt human agent in subject position. Mental verbs, which describe processes of sensing such as thinking, imagining, liking, etc., frequently imply a human-like participant that feels, thinks, or perceives (Zhang 2017:67). A man-impersonal consisting of a human agent can thus be considered as a better candidate for mental verbs than a periphrastic passive, which, as will be detailed in 7.3, is more prone to include inanimate entities as agents.

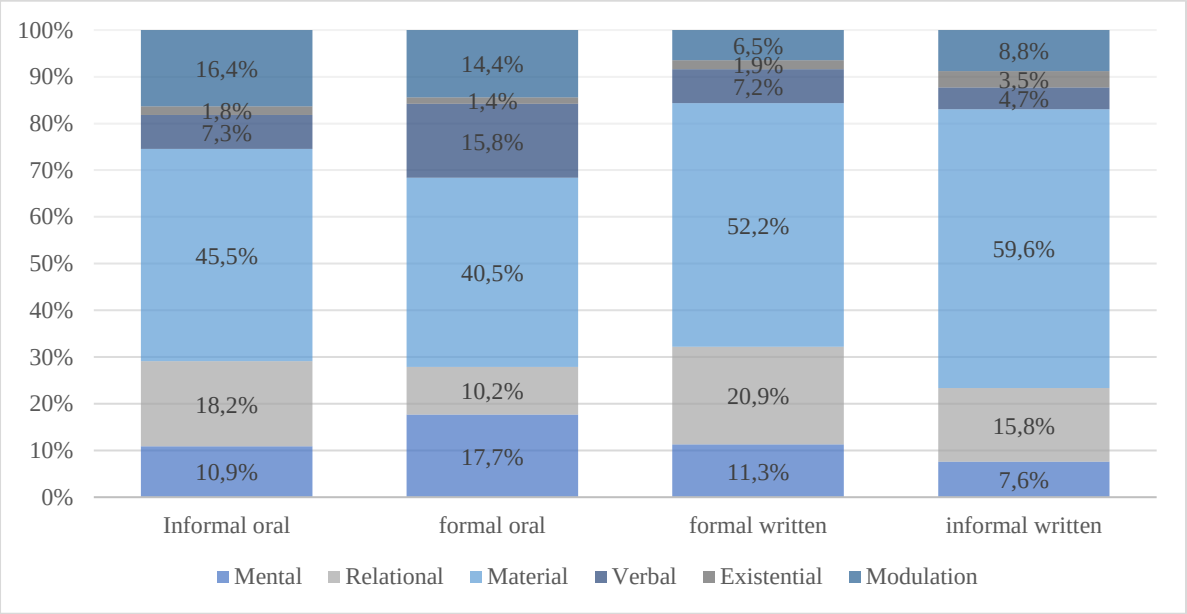
Now that we have given a panoramic overview of the distribution of type of verb process, the second part of the analysis will focus on the influence of mode and register variation. As the overall number of occurrences obtained for the non-anaphoric third person plural and the *se*-construction is rather low, making it difficult to obtain a sufficient number of occurrences to analyse the influence of mode and register once the results have been broken down, the next part centres solely on man-impersonals and periphrastic passives.

The following results emerged for the man-impersonal *on.* Graph 14 shows that, within each register and mode, the six types of processes are distributed differently.




*Graph 14. Distribution of verb processes in French man-impersonals*

It is interesting to note that the man-impersonal *on*, a passive-like construction usually associated with informal oral contexts, has quite a different profile when used in formal oral situations. While material verbs predominate in the other combinations of mode/register, mental (33.3%; 22 occurrences) and verbal (25.8%; 17 occurrences) verbs account for nearly 60% of the occurrences in formal oral data. Mental and verbal types of processes engage the speaker on the cognitive and communication level, while material processes are much more related to physical and concrete situations. This reflects a different use of the man impersonal constructions when the speaker shapes language production in formal oral situations. In political discourse, as pointed out by Zhang (2017: 65), "speakers have to measure their words with special caution to interact with people, to express their attitudes and judgments and to influence the viewpoints and behaviour of the audience." We hypothesize that a possible way to achieve this is the use of mental verbs, that allow the expression of situations and events from the senser's perspective rather than placing the focus on the hearer and possibly threatening his/her face. Formal written data share this high proportion of mental verbs (35.7%; 35 occurrences) with formal oral data but also rely heavily on material verbs (35.7%; 35 occurrences), which is not the case for formal oral data. Such observations lead us to assume that the degree of formality greatly impacts the type of verb process in the impersonal *on.* The rather specific profile of formal oral data that could be observed for the man-impersonal *on* is, nevertheless, not found for the periphrastic passive.




*Graph 15. Distribution of verb processes in French periphrastic passives*

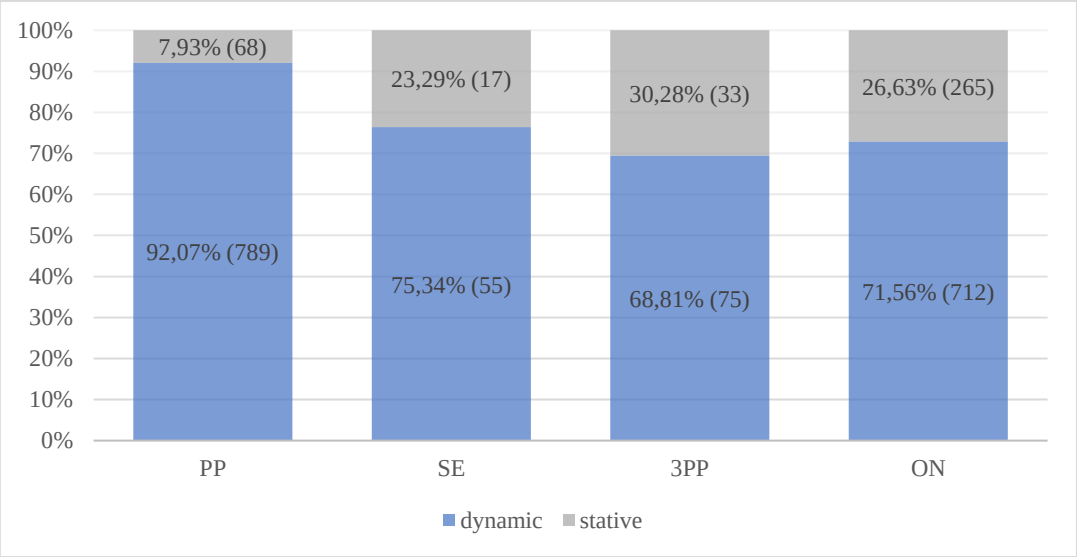
Graph 15 on the periphrastic passive shows that formal oral data use significantly more mental and verbal types of processes. The difference is, nevertheless, slightly less marked than what was observed in the data for the man-impersonal *on*. This could be explained by the fact that periphrastic passive constructions are typical of situations considered to be more formal. As a consequence, the periphrastic passive does not need to adapt its characteristics and the type of verb process when used in formal contexts, as we observed for the man-impersonal *on*. Contrary to what is observed for *on*, variation in types of processes within periphrastic passives comes from the mode rather than the degree of formality (see, for instance, the significant difference between the use of mental ($\chi^2 = 4.9442$, $p < .05$) and verbal processes ($\chi^2 = 11.509$, $p < .05$) in formal oral and written data). In Graph 15, oral data show a higher proportion of modulation verbs and a smaller proportion of material verbs compared with written data. Interestingly, both in the periphrastic passive and *on*, verbal types of processes are more frequent in formal oral data ( parliamentary debates) than in the other three combinations. The figure even amounts to 25% of the total of verbs in periphrastic passives. This type of process which tends to engage more the speaker in the discursive process and helps to reflect on the content can be considered as characteristic of passive-like constructions used in French formal oral situations.

A brief qualitative exploration of the non-anaphoric third person plural and the *se*-construction reveals the following results. In our data set, a greater diversity in the types of processes is found in formal written data (11 subcategories) compared with informal written data (5 subcategories) for the *se*-constructions. Oral productions hardly resort to *se*-constructions in French (8 occurrences), so the impact of the mode and register on the type of verb process cannot be measured in our data sets. As for the non-anaphoric third person plural, an influence of the mode is suspected but a larger data set would be needed to verify these assumptions. As we obtained only a very limited number of occurrences of the non-anaphoric third person plural for formal data (3 occurrences), no further observations can be made regarding the influence of mode and register.

Summarising our observations on the *type of verb process* parameter, we can conclude that this is a key feature in passive-like structures. The four structures have either specific affinities with certain types of verb processes or particular behaviour linked to mode and register influence. While the periphrastic passive and the *se*-construction have quite a similar profile, largely relying on material processes, two types of verb processes prevail in the non-anaphoric third person plural (*relational* and *material*) and the distribution of verb processes is much more balanced in the man-impersonal *on.*

### 7.2.5. Kinesis

Graph 16 illustrates the feature *kinesis*, which compares dynamic and stative verbs.




*Graph 16. Distribution of dynamic and stative verbs in French passive-like structures (parameter kinesis)*

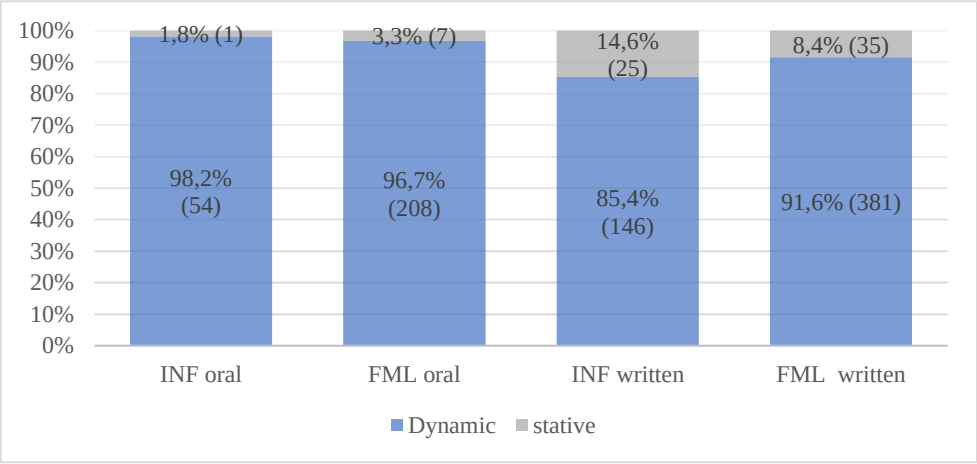
It can be seen that, even though dynamic verbs are the most frequent in the four passive-like structures, their proportion differs according to the type of structure. While the non-anaphoric third persona plural and the man-impersonal *on* once again show a similar proportion of dynamic verbs (68.8% and 71.6%, respectively), results for the periphrastic passive suggest a considerably different picture since this structure mainly resorts to dynamic verbs (92.07%), such as in (187), where the action of *préparés* ('prepared') requires a dynamic process.

(187) *Les suppléments de B12 tels que les pilules de vitamines* ***sont souvent préparés*** *à partir de déchets d'abattoir et ne conviennent donc pas aux végétaliens, même si le nombre de marques ne contenant pas de produit animal augmente.* (WIKI FR)
'B12 supplements such as vitamin pills **are often prepared** from slaughterhouse waste and are therefore not suitable for vegans, although the number of brands that do not contain animal products is increasing.'

This high use of dynamic verbs corroborates with Collins's (2012:295) work which suggests that the passive typically has dynamic meaning.

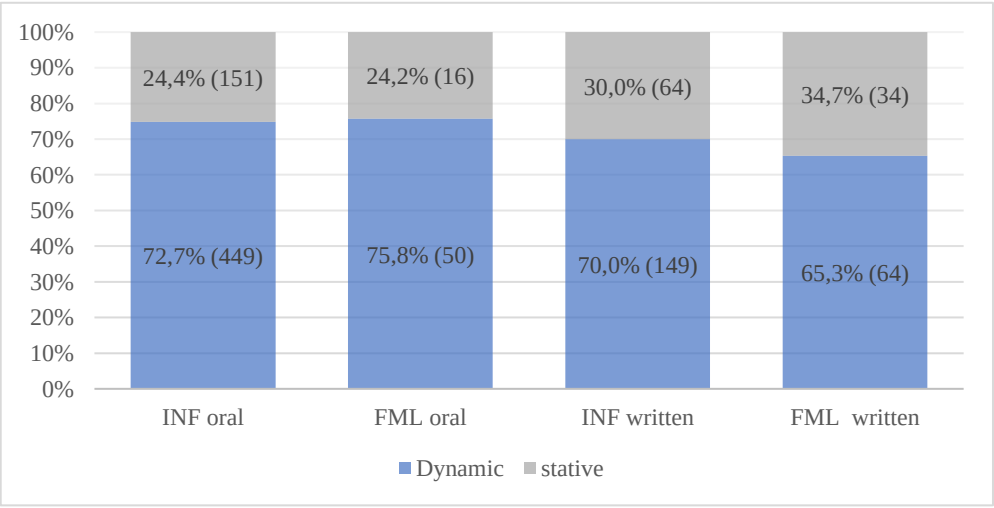
Even though the *se*-construction has a higher number of dynamic verbs than the non-anaphoric third person plural and the man-impersonal *on*, there is no significant difference. It can therefore be suggested that the periphrastic passive rather stands alone as regards its *kinesis* behaviour. When French speakers need to describe a state rather than an action, they tend to rely on the *se*-construction, the non-anaphoric third person plural or the man-impersonal, whereas the results reveal that speakers would almost exclusively express dynamic events in periphrastic passive.

Let us now focus on the periphrastic passive and the man-impersonal *on* and examine the distribution of dynamic and stative verbs in light of register and mode variation.




*Graph 17. Distribution of dynamic and stative verbs in French periphrastic passives*

Graph 17 shows the distribution of results for the periphrastic passive, indicating that stative verbs are more frequent in written data than in oral data. Chi-square tests reveal that the difference is significant between informal written and oral data ($\chi^2 = 6.6988$, $p < .05$) and between formal written and oral data ($\chi^2 = 6.0688$, $p < .05$). Register can also be considered a factor explaining the variation in the proportion of dynamic verbs. Results suggest that informal interactions contain significantly more stative verbs in written data than in formal productions ($\chi^2 = 5.0868$, $p < .05$). The influence of register is, nevertheless, much more limited in oral data since results reveal that there is no significant difference between formal and informal data.




*Graph 18. Distribution of dynamic and stative verbs in French man-impersonals* on

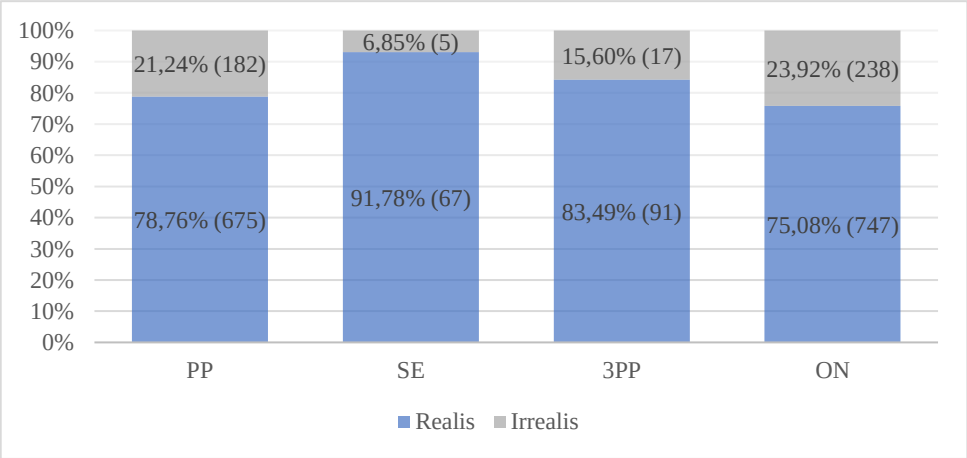
The analysis reveals that for the man-impersonal *on*, dynamic verbs (188) are far more frequent than stative ones (Graph 18). Percentages of dynamic verbs are situated in an interval ranging from 65.3% to 75.7%.

(188) *Les jeunes pousses de panais sauvage peuvent agrémenter beaucoup de recettes (...). **On peut** les **utiliser** dans tous les plats de légumes, les sauces (sauce tomates aux panais, pour les pâtes par exemple), les omelettes, gratins, etc.* (WIKI EN)
'Young wild parsnip shoots can be used in many recipes (...). **They can be used** in all vegetable dishes, sauces (tomato sauce with parsnips, for pasta for example), omelettes, gratins, etc.'

Chi-square tests reveal that for *on*, even though dynamic verbs are used slightly more in oral mode, there is no significant difference, depending on the mode or the register.

### 7.2.6. Realis/irrealis dichotomy

The analysis of the *realis/irrealis* dichotomy enables us to understand how the situation expressed through the passive-like construction is anchored in the real world.




*Graph 19. Distribution of realis and irrealis clauses in French passive-like structures*

The percentages in Graph 19 reveal that the four structures vary in terms of their anchorage in reality. The construction with *se* almost exclusively encodes real events and situations, such as in (189). The verb *se dit* ('is said') is used in the indicative present, in the affirmative and is not accompanied by any modal indicators.

(189) *Quel adjectif qualifie la taille d'un avion?**se dit** de la taille d'un avion (Yahoo FR)*
'What adjective describes the size of an aircraft? **is said** about the size of an aircraft'

The periphrastic passive and the man-impersonal *on* are the two structures that have the highest proportion of irrealis clauses - i.e., events and situations in a non-real world. Examples (190)

and (191) illustrate irrealis clauses with these two defocussing mechanisms. In (190), irreality is expressed through a negative particle, while in (191), irreality is encoded through the use of a conditional tense.

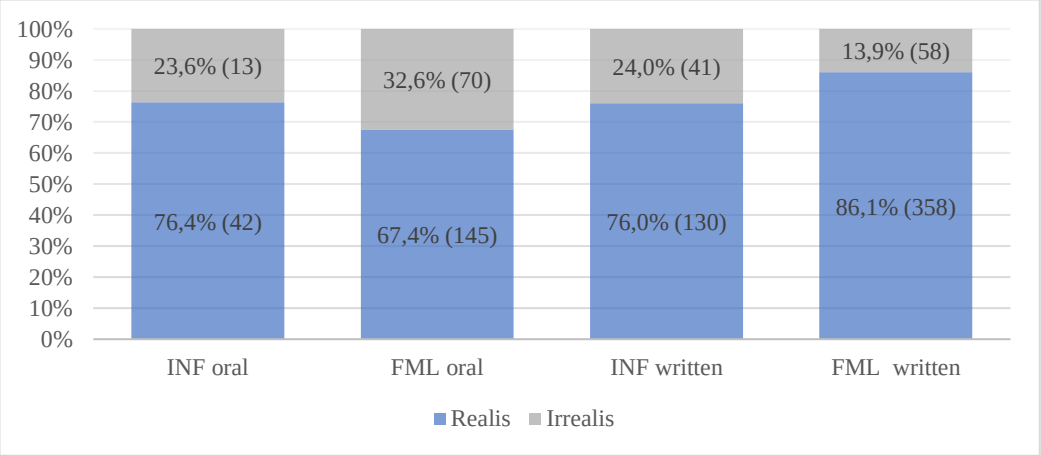
(190) *Les hampes défleuries certainement, pour éviter que la plante utilise toutes ses capacités à faire une graine dont* ***on n'a pas*** *besoin.* (Yahoo FR)
'Stems with no flowers certainly, to avoid the plant using all its ability to make a seed that **is not needed**.'

(191) *Il était initialement prévu que les milliers des PW-5* ***seraient commandés*** *et que différents producteurs pourraient en acquérir la licence, réduisant encore les coûts par les économies d'échelle jusqu'à environ 20 000 dollars US.* (WIKI FR)
'It was originally planned that thousands of PW-5s **would be ordered** and licensed to different producers, further reducing costs through economies of scale to around US$20,000.'

Surprisingly, and contrary to what has been observed in our data so far, the non-anaphoric third person plural and the man-impersonal *on* differ in how they encode events. The man-impersonal *on* is used significantly more in irreal situations (23.9%), such as in (192), than the non-anaphoric third person plural (15.6%) ($\chi^2 = 3.8592$, $p < .05$). In (192), the man-impersonal *on* is used in the negative, thus expressing a situation that cannot happen.

(192) *{S}En effet,* ***on ne peut pas,*** *d'un côté,* ***oeuvrer*** *à l'amélioration des conditions de vie des citoyens et, de l'autre, faire une exception au transport routier en renforçant la flexibilité des temps de repos.* (PROCEP FR)
'Indeed, **one cannot**, on the one hand, **work** to improve the living conditions of citizens and, on the other, make an exception for road transport by increasing the flexibility of rest periods.'

This difference in the behaviour of the two R-impersonals confirms Viberg (2010)'s suggestion on Swedish where man-impersonals tend to appear more frequently in contexts referring to irrealis situations, involving modals or conditional clauses.




*Graph 20. Distribution of realis and irrealis clauses in French periphrastic passives*

Graph 20 provides details of how periphrastic passives encode reality through different registers and modes. It can be seen that formal oral data contain the highest proportion (32.6%) of irrealis

clauses. This finding can be explained by the nature of the data included in this sub-corpus. In parliamentary debates, a more cautious style, including modal auxiliaries or conditional tenses, is to be expected.

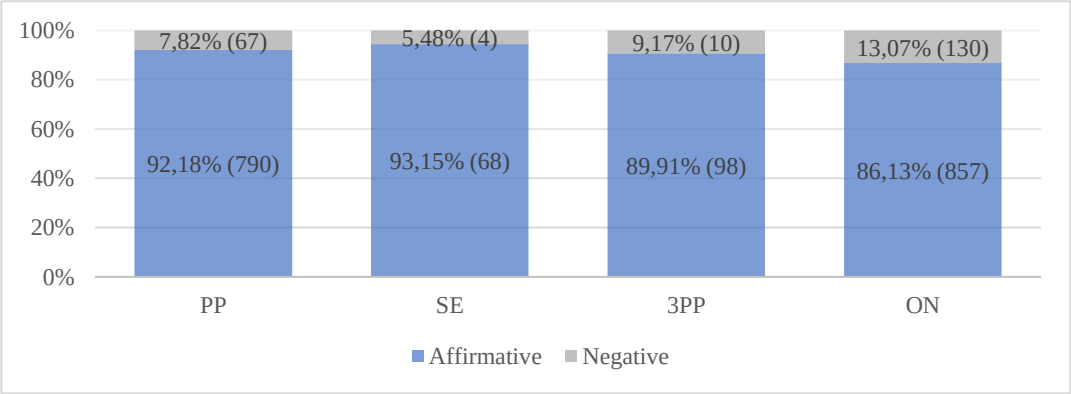
In French, the analysis of man-passive-like constructions reveals that neither register nor mode impacts the nature of the situation. *On* encodes real events in approximately 75% of cases, as illustrated in (193), for example. *On* is used with a verb in the present tense and in the affirmative, which means the situation is real and not hypothetical.

(193) *Certains crudivores peuvent être végétariens ou végétaliens* ***: on parle*** *à ce moment de crudi-végétariens ou crudi-végétaliens.* (WIKI FR)
'Some raw food may be vegetarian or vegan: **we** then **talk** about raw vegetarian or raw vegan.'

The absence of a higher proportion of irrealis clauses in formal oral data for *on*, as was observed for the periphrastic passive, can be explained by the fact that *on* is an impersonal mechanism that is relatively infrequent in French formal oral productions.

### 7.2.7. Affirmation

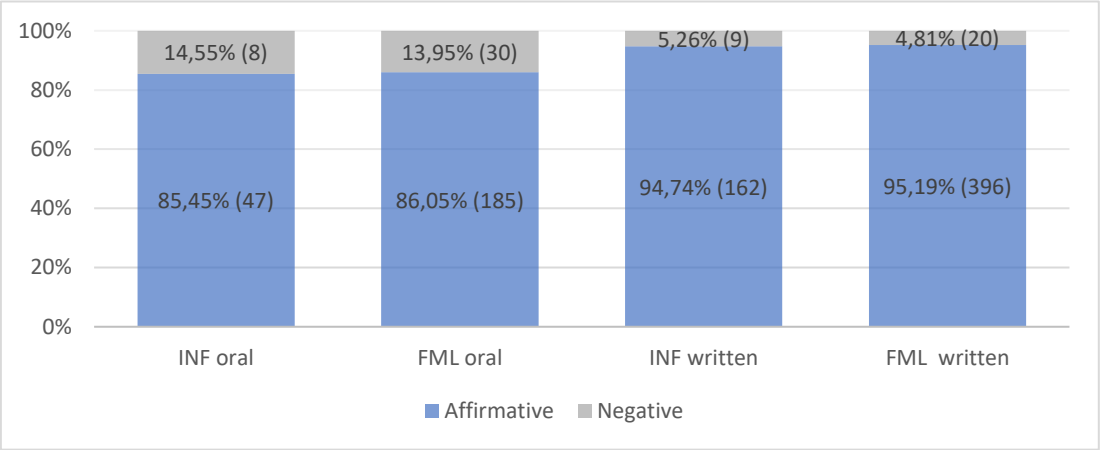
Closely linked to the *realis/irrealis* dichotomy is the parameter *affirmation*.




*Graph 21. Distribution of affirmative and negative clauses in French passive-like structures*

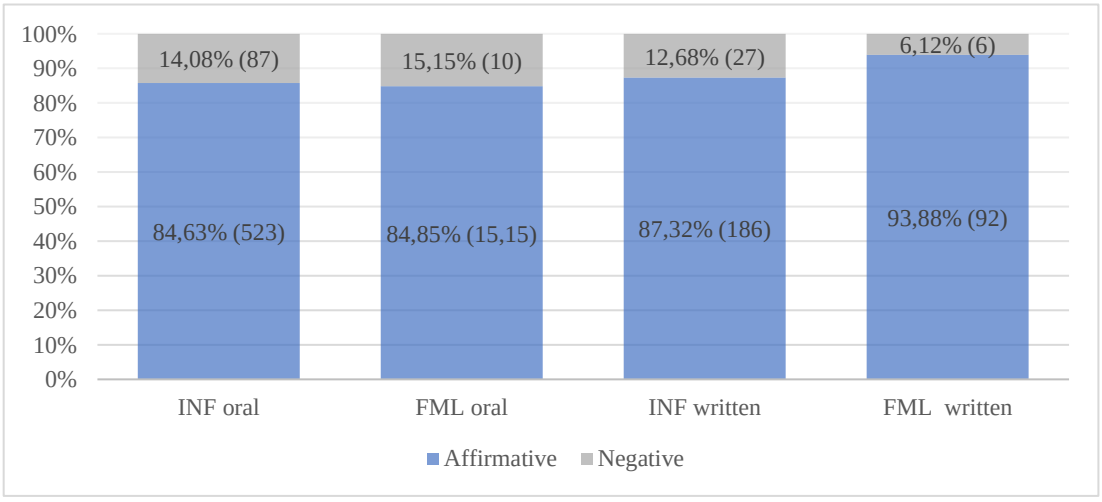
Unsurprisingly, affirmative clauses largely predominate in the data (see Graph 21). The proportion is slightly lower in passive-like constructions based on a pronoun, especially for the man-impersonal *on*, illustrated in (194) though the difference between the two constructions is not significant.

(194) *Même le président de Goldman Sachs, qu'****on ne savait pas*** *si soucieux de la planète, y est allé de sa petite condamnation.* (PROCEP)
'Even the CEO of Goldman Sachs, whom **we did not know** was so concerned about the planet, expressed his condemnation.'




*Graph 22. Affirmation in French periphrastic passives*

The analysis shows that for the periphrastic passive, negative clauses are mainly used in oral data (Graph 22). This suggests that speakers will resort to other strategies, such as a different grammatical architecture, to refer to negative situations or events when expressing impersonal value in written productions. As for register variation, no influence could be observed for the parameter *affirmation* in periphrastic passives.




*Graph 23. Affirmation in French man-impersonals* on

Compared with the periphrastic passive, the results show a slightly different picture as regards the influence of mode and register on the man-impersonal *on* (see Graph 23). *On* is mainly used in affirmative clauses, with quite a similar proportion of clauses in oral and written modes and in informal and formal registers, except for formal written data. While affirmative clauses account for 84.6% to 87.3% in the other three sub-categories, formal written data contain the highest number of affirmative clauses (93.9%). This means that when the man-impersonal *on*, a structure that tends to be associated with informal oral contexts, is used in formal written productions, its behaviour diverges from the other modes and registers. The higher frequency of affirmative clauses in formal written data was also found for the periphrastic passive. It can be suggested that, when speakers have to express negative content in formal written situations, they tend to adapt their discourse and opt for a positive clause.

### 7.2.8. Other parameters

This section briefly summarises the results obtained for the other verb parameters, i.e., *volitionality* and *punctuality*, which the MCA found to be less salient than the parameters discussed so far. The analysis revealed that volitional verbs predominate in our data and account for approximately 82% to 88% of the occurrences, except for the construction with *se*. This construction tends to include fewer volitional verbs, as the proportion falls to 72.6%. As for the parameter *punctuality*¸ the non-anaphoric third personal plural and the man-impersonal *on* both use non-punctual verbs in around 80% of cases. The picture is quite different for the periphrastic passive and the *se*-construction, since the proportion of non-punctual verbs falls to 65.6% and 42.5%, respectively.

### 7.2.9. Summary

To sum up so far, the analysis has yielded the following results concerning the key parameters related to the verb in French passive-like structures. First, it can be said that the periphrastic passive is, generally, the most distinct profile compared with the other three structures. This is the case, for instance, with the features *tense*, *modality* and the *realis/irrealis* dichotomy. It was also observed that the non-anaphoric third personal plural and the man-impersonal *on* usually have a substantially similar profile. The only verb-related parameter where these two structures greatly differ is the parameter *realis/irrealis*.

Second, a correlation between several parameters could be observed. This is the case for *tense* and *aspect*, where present tenses and imperfective aspects predominate and have similar proportions. A link between the parameters *realis/irrealis* and *affirmation* was also found. This finding was to be expected since a negative clause is a possible device for expressing irrealis content.

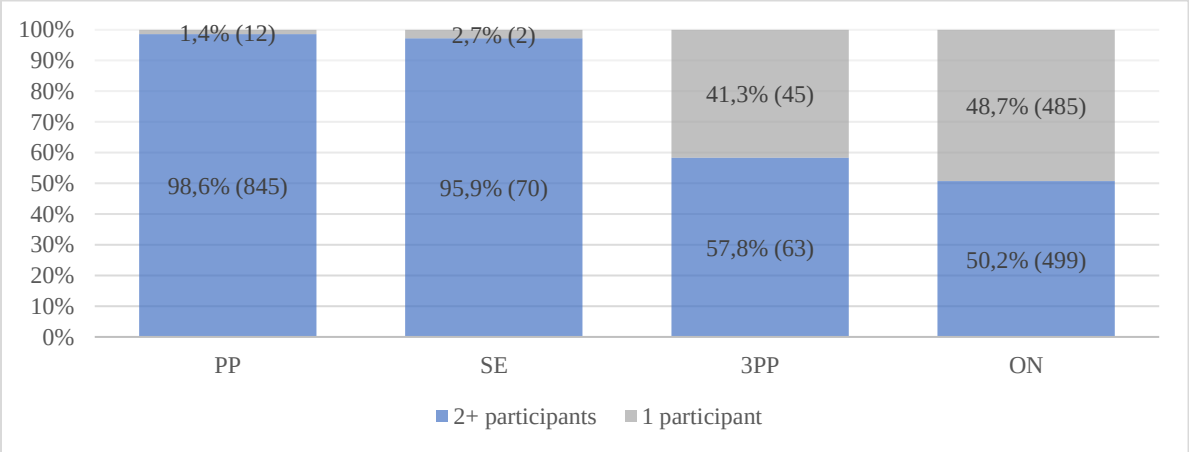
Third, the analysis has uncovered interesting findings as regards the influence of mode and register. As far as verb-related parameters are concerned, it was noted that mode is the major influence on the distribution of the parameters. Whereas the impact of the mode could be seen in nearly all the parameters studied, the influence of the degree of formality is much more limited and punctual. Results have shown that with respect to register, formal oral data vary the most.

## 7.3. Participant-related parameters

In the following, we analyse the parameters related to participants. As explained in Chapter 5, under the label *participant* we include agents and patients. This section first opens on a discussion on the number of participants in the passive-like structure (7.3.1), then focuses on the use of an explicit agent (7.3.2), the presence of a patient (7.3.3) and the individuation and affectedness of the patient (7.3.4).

### 7.3.1. Number of participants

The number of participants included in the passive-like construction may be a factor influencing the variation between the four passive-like structures. Graph 24 gives an overview of the distribution of one participant-clause versus two or more participant-clauses. The participants may be overt or defocused as will be further illustrated.




*Graph 24. Distribution of one- and two-participant clauses in French passive-like structures*

As can be observed, the distribution of participants greatly varies according to the structure. While the periphrastic passive and the *se*-construction almost exclusively resort to two-participant clauses, the proportion of one- and two-participant clauses is much more balanced in passive-like constructions based on a pronoun. The exclusive use of two-participant clauses with the periphrastic passive and the *se*-construction is explained by the syntactic construction of these two structures. As the patient is promoted in the argument slot usually occupied by the agent (namely the subject slot), in the majority of the cases, there is at least a patient and an agent (defocused (195) or not (196)) in periphrastic passives.

(195) *J'adore le mont d'or, et si vous avez l'occasion de venir dans le Haut Doub (là ou **il est fabriqué**), vous pourrez en acheté de qualité exeptionnel (...). (YAHOO FR)*
'I love the Mont d'Or, and if you have the opportunity to come to Haut Doub (where **it is made**), you will be able to buy exceptional quality.'

(196) *Tu dis que c'est une chienne très gâtée , elle doit **être surprotégée** par sa maîtresse ou son maître et se retrouvant sans elle ou lui dans un lieu inconnu avec une inconnue ça la met dans un état d'angoisse pas possible* . (YAHOO FR)
'You say that she is a very spoiled dog, she must be overprotected by her mistress or her master and being without her or him in an unknown place with a stranger it puts her in a state of anxiety not possible.'

In (195) one participant is explicitly stated, namely the third person singular pronoun *il*, standing for the Mont d'Or cheese, while the other participant, the entity supposedly implied in

cheese making, is left unspecified. This example contains thus one overt and one defocused participant. Conversely, example (196) illustrates a clause with two explicit participants: the pronoun *elle* (‘she’) and the nominal phrase *sa maîtresse ou son maître* (‘her mistress or her master’).

One-participant clauses seem to occur exclusively when the periphrastic passive is constructed with an expletive *il* (‘it’), illustrated in (197), where the only participant in the clause is the defocused agent performing the action, in this case the verb *noté* (‘noted’).

(197) *Il doit cependant **être noté** qu'une étude japonnaise a montré l'existence d'une corrélation entre la consommation accrue de certains produits animaux et la baisse de certaines formes de maladies cérébrovasculaires (…)* (WIKI FR)
It should **be noted**, however, that a Japanese study has shown a correlation between increased consumption of certain animal products and a decrease in certain forms of cerebrovascular disease

In a similar vein, the specific architecture of the *se*-construction, with the presence of an object in subject position, is a factor which can contribute to explaining the predominance of two-clause.

As far as the non-anaphoric third person plural and the man-impersonal *on* are concerned, results show that they nearly equally resort to one-and two-participant clauses. Though the use of a one participant-clause appears as slightly more marked for *on*, the difference between the two passive-like structures is not significant ($\chi^2 = 2.2634$, $p < .05$). Compare examples (198) and (199) where, in addition to the impersonal subject *on* and *ils*, the clauses are constructed with an object *le (*‘him’) and *le risque* (‘the risk’) to examples (200) and (201), where there is no object, the subject being the only participant in these two occurrences. It is thus assumed that the number of participants the speaker wants to mention and integrate in the clause constrains the choice of passive-like constructions.

(198) *{S}JUL il il est pas payé à l'heure hein (.)il est payé en fait*
*{S}JUL [`fin ]*
*{S}ROM [soumis] au cadre les:: forfaits (cadre)*
*{S}JUL ouais*
*{S}JUL [forfait cadre ouais]*
*{S}ANN [ouais mais **ils le ] payaient** quand même le dimanche même si i`s vont travailler* (CIEL FR)
‘{S}JUL he he is not paid by the hour eh (.)he is paid in fact
{S}JUL [`end ]
{S}ROM [submitted] to the framework the:: fixed rates (framework)
{S}JUL yeah
{S}JUL [executive package yeah]
{S}ANN [yeah but **they ] still pay** him on Sunday even if he goes to work’

(199) ***On gère** le risque que l'on a le plus de "chances" de voir se produire.{S} Ainsi, la neige est présente chaque hiver en abondance dans les pays nordiques, ils*

*sont donc équipés pour : métareil, système de dégivrage, habitude des personnels. (YAHOO FR)*

'**We manage** the risk that is most "likely" to occur.{S} So, snow is present every winter in abundance in the Nordic countries, so they are equipped with: equipment, de-icing system, personnel habits.'

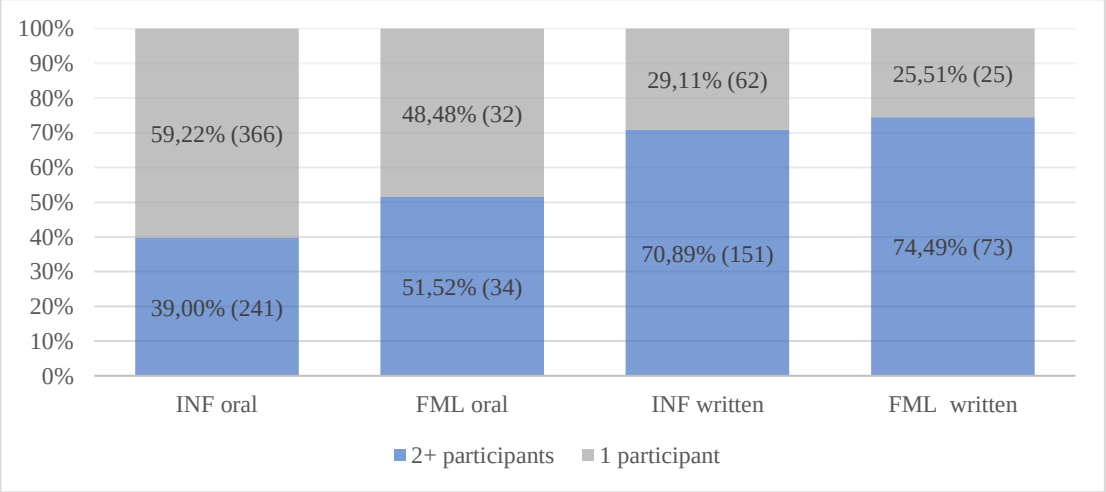
(200) *Ce sont les journalistes qui doivent être contents: encore une nouvelle à la con à raconter pendant un quart d'heure...même la guerre en Israel est désiroire poue eux, comparé à la neige en france.{S} Comment **font ils** au Groenland?* (YAHOO FR)

'It's the journalists who must be happy: again a stupid news to tell during a quarter of an hour... even the war in Israel is derisory for them, compared to the snow in France.{S} How do **they do** in Greenland?'

(201) *Pour ce qui est des humains...je pense que c'est la même chose **on dit** que quand on aime quelqu'un on doit accepter de le laisser partir...c'est un peu le problème de l'euthanasie (YAHOO FR)*

'As far as humans are concerned...I think it's the same thing **we say** that when we love someone we must accept to let them go...it's a bit the problem of euthanasia'

Turning to register and mode variation, Graph 25 presents the distribution of one and two-clause participants for the man-impersonal *on*.




*Graph 25. Distribution one- and two-participant clauses in French man-impersonals*

The number of participants occurring with a man-impersonal *on* varies considerably, especially depending on the mode. Graph 25 indicates that in written data, the structure is accompanied by at least two participants in 70.9% to 74.5% of the cases. Example (202) illustrates such construction with two participants, where the two participants are the subject pronoun *on* and the object *different species*.

(202) {S} *Plusieurs des espèces sont appelées « pinceau indien », traduction littérale de l'anglais « Indian paintbrush » {S}* ***On a dénombré*** *près de 200 espèces différentes*. (WIKI FR)
'Several of the species are called "Indian paintbrush" {S} Nearly 200 different species **have been counted**.'

This higher number of participants in written data may be linked to a higher level of grammatical elaboration. Written and oral data rely on different types of architecture to express discourse complexity. While written data build the complexity through more complex structures, oral data rely on the interdependency of the clauses and their logical semantic dependence and relationship (Haliday 1985:86). In written data, register does not impact the number of participants as the proportions are similar both in formal and informal productions. Compared to written data, oral data resort far less to two- or more participant clauses. As can be seen in Graph 25 informal oral productions prefer one-participant clauses as in (203) since they amount for 59.2% of the data set. Whereas one-participant clauses predominate in informal oral data, the distribution of the two types of clauses is much more balanced in formal oral data, occupying an intermediate position between written data and informal oral ones. An influence of the register can thus be observed for oral data, which was not the case in written data.

(203) *Ce manque de volonté politique est renforcé par le confort de se dire qu'****on fait*** *mieux que les autres.* (PROCEP FR).
'This lack of political will is reinforced by the comfort of telling ourselves that **we are doing** better than others.'

Due to an almost exclusive use of two-participant clauses with periphrastic passives (98.6%), no relevant analysis of the possible impact of mode and register on the distribution of one- and two-participant clauses could be performed for this structure.

As participants can either be in subject – with agent role- or in object – with a patient role-positions, I will further analyse the agent (7.3.2) and, in a second step, the patient (7.3.3) with a focus on the notion of *animacy*.

### 7.3.2. Agent

Three out of the four passive-like structures included in this analysis do not allow much variation regarding the agent. The non-anaphoric third person construction and the man-impersonal *on* always imply a human animate agent in subject position in our data. There is thus no possible variation regarding the type of agent for these two structures. Similarly, as usually reported in the literature and confirmed by our data, the *se*-construction hardly contains an overt agent. This section focuses thus on the periphrastic passive, which allows much more variation as to the presence of the agent

In line with numerous previous research on the topic (Fernandez, 2006; Shibatani, 1985; Siewierska, 1984), the present set of data shows that agentless periphrastic passives, that is constructions without any explicit agent introduced via the typical agentive oblique

complement *par* ('by'), constitute the predominant type of periphrastic construction in French (85.5%). This type of construction is exemplified in (204).

(204) *Les jacinthes font partie des fleurs comestibles et **sont utilisées** en cuisine et en confiserie.* (WIKI FR)
'Hyacinths are among the edible flowers and **are used** in cooking and confectionery.'

The focus will now be placed on overt agents in periphrastic passives. I will first offer the main results, regardless of register and mode variations and, in a second step, precise the picture by commenting on their distribution according to these two types of variations.

A first distinction can be drawn between animate, illustrated in (205) by the proper name Georges Ohsawa, and inanimate agents, exemplified in (206) with the nominal clause *les débris* ('debris').

(205) *L'approche macrobiotique **fut définie** par Georges Ohsawa.* (WIKI FR)
'The macrobiotic approach **was defined** by Georges Ohsawa.'

(206) *{S}Les grilles n'apporteraient rien parce qu'elle seraient encore moins solides que les aubes de réacteur et **serait obstruées par les débris**, sans parler que les morceaux de grilles métalliques projetés dans les moteurs.*. (YAHOO FR)
'The grids would not bring anything because they would be even less solid than the jet engine blades and **would be clogged** with debris, not to mention the pieces of metal grids thrown into the engines.'

Table 21 gives an overview of the different types of animate and inanimate overt agents.

| | **Agent** | **Occurrences** | **%** |
|---|---|---|---|
| **Animate** | Human | 32 | 25.80% |
| | Non human | 4 | 3.2% |
| **Inanimate** | Organization | 20 | 16.1% |
| | Country/City | 7 | 5.6% |
| | Inanimate Other | 61 | 49.2% |
| | **Total** | 124 | 100% |

*Table 21. Classification of overt agents in French periphrastic passives*

Results reveal that the use of inanimate agents predominates as they account for 70.9% of the data (88 occurrences). Within inanimate agents, the category *others* is the most frequent one (61 occurrences out of 88). This category is rather heterogeneous as it contains all types of inanimate agents that do not fall in the categories *organization* or *country/city*, which, as will be explained below can be, to some extent, related to collectives of people. The category *others* includes concrete inert elements such as the engine in example (207) or more abstract concepts as price volatility in example (208).

(207) *En combinant des pièces déjà existantes _ **l'engin est propulsé** par un moteur ordinaire (YAHOO FR).*
'Combining existing parts _ the machine **is powered** by an ordinary engine.'

(208) *Elle **est aggravée** par la volatilité des prix et la volatilité des prix est une conséquence de la financiarisation de l'économie mondiale. (PROCEP FR)*
'It **is aggravated** by price volatility and price volatility is a consequence of the financialization of the world economy.'

The second most frequent category as for inanimate agents includes organizations (20 occurrences) as in (209). The agent *la Commission* ('the Commission') is one of the main institutions of the European Union. Countries and cities, illustrated in (210), only have an agentive role in 5.6% of the cases. In the literature these two types of agents are classified as inanimate (see Fauconnier 2012). They are, nevertheless, connected to human agents to the extent that organizations and countries/cities are entities which stand for groups of people. _

(209) *Je déplore que l'initiative citoyenne sur la vivisection **n'ait pas été suivie** par la Commission.* (PROCEP FR)
'I regret that the citizens' initiative on vivisection **was not followed** by the Commission.'

(210) *bourg en bresse rayonne plus sur l'ouest du département parce que là c'est polarisé par bellegarde et la suisse enfin bellegarde notamment **est polarisée par genève** quoi j'ai dit (…)* (CIEL FR).
'bourg en bresse radiates more on the west of the department because there it is polarized by bellegarde and switzerland, well bellegarde in particular **is polari**zed by geneva what I said (…).'

A look at animate agents reveals that the category animate human is the most frequent one (25.8%; 32 occurrences). And, overall, it is the second most frequent category of agents, after the category inanimate *others*.

(211) *Monsieur le Président, Madame la Commissaire, les choses sont simples.{S} Elles **ont été exposés** par mes collègues*. (PROCEP FR)
'Mr President, Commissioner, things are simple {S} They **have been set out** by my colleagues.'

(212) *Spike le Bulldog est un personnage des dessins animés Warner Bros Looney Tunes et des Merrie Melodies.{S} Il **a été créé** par Friz Freleng et sa première apparition date de 1952 dans le dessin animé Tree for Two*. (WIKI FR)
'Spike the Bulldog is a character from the Warner Bros Looney Tunes cartoons and Merrie Melodies.{S} He **was created** by Friz Freleng and first appeared in the 1952 cartoon Tree for Two.'

Human agents can be expressed through the use of common nouns such as *mes collègues* ('my colleagues') in (211) or through proper names as *Friz Freleng* in (212). Animate non humans,

animals in our data, is the least frequent type of agent in periphrastic passives (see example 213).

(213) *Le pistage **est aussi pratiqué** par les chiens de chasse*. (WIKI FR)
'Tracking **is also practiced** by hunting dogs.'

According to Shibatani (1985), when the agent is expressed in the periphrastic passive, the occurrence moves away from the prototypical instance of the passive. Periphrastic passives with overt agents are sometimes called *non-basic* (Keenan & Dryer 1981) in opposition to agentless passives considered as the prototype or the basic periphrastic passive. Though the patient is promoted in the topic position and receives more attention than in an active clause, the presence of an overt agent impedes the defocusing process to be considered complete. The salience of the patient is therefore diminished compared to a prototypical agentless periphrastic passive. When the agent is expressed, several researchers indicate that it refers to a well-known figure or a person occupying a specific position. It is also reported that the agent tends to be expressed in situations presented as negative (Davison 1980, Fernández 2007). A brief qualitative overview of our French data confirms that a vast majority of overt agents appears in one of these two situations. The situation can either be considered adversative due to the semantics of the verb included in the periphrastic passive or due to the general linguistic context within which the passive-like construction appears. In (214) the use of the agent is prompted by the semantic of the verb *ravag*é ('devastate') which presents the hurricane as the agent responsible for the destruction of the archipelago.

(214) *{S}Monsieur le Président, Monsieur le Commissaire, la semaine dernière les Antilles ont été ravagées par l'ouragan Irma.* (PROCEP FR)
'{S}Mr President, Commissioner, last week the Antilles were devastated by hurricane Irma.'

Within this category, our data include verbs such as *menacer* (to threaten), *aggraver* (to worsen), *endommager* (to damage) or *infecter* (to infect).

(215) *En 1963, à Washington, il a dit: «Nos problèmes **ont été créés** par l'homme et nous pouvons donc les résoudre*. (PROCEP FR)
'In 1963, in Washington, he said: "Our problems **were created** by man and we can therefore solve them.'

The agent may also appear in negative situations including a verb that does not convey a negative value as exemplified in (215). Deprived of any context, the verb *créer* (to create) does not trigger any negative effect. However, used in the specific context of (215) with the presence of the nominal phrase *les problèmes* (the problems), the periphrastic passive and its agent *l'homme* ('man') turn out to appear in an adversative environment.

Our data concur with the literature on the fact that, when the agent is expressed and the situation is not regarded as problematic, the agent is either a famous figure, someone occupying a specific position in the society or an organization, institution or politic entity. Example (216) illustrates such situation with *the Kaiser* as agent.

(216) *Les premiers raids **sont approuvés** par le Kaiser en janvier 1915*. (WIKI FR)
'The first raids **were approved** by the Kaiser in January 1915.

The following noun phrases are other illustrations of these categories: the *NASA*, *le consulat de France* (‘the French consulate’), *les Etats Membres* (‘Member States), *le ministre* (‘minister’) ou encore *l’homme d’affaire* (‘the business man’).

The qualitative revision of the French data reveals a limited number of periphrastic passives with an overt agent which is not a well-known person nor an institution or organization and when the passive-like construction does not appear in an adversative situation. A closer examination of these occurrences (3 out of 69 overt agents) suggests textual cohesion as the main factor explaining the use of a periphrastic passive. The following excerpt (217) illustrates such functional domain.

(217) *Bien qu'elles puissent être compensées par la consommation de compléments alimentaires, l'équilibre alimentaire est plus difficile à atteindre et les questions éthiques liées à la supplémentation, notamment aux méthodes de synthétisation, font débat chez les végétaliens […]. {S}La vitamine B12, ou cobalamine,* ***est synthétisée p****ar des bactéries.{S} Hormis la consoude, une plante toxique pour le foie[23], qui peut extraire cette vitamine du sol[24], les végétaux n'en contiennent pas[25][26] sauf en cas de contamination*. (WIKI FR)

‘Although they can be compensated for by supplementation, dietary balance is more difficult to achieve and the ethical issues associated with supplementation, including methods of synthesis, are debated among vegans[….]. {S}Vitamin B12, or cobalamin, **is synthesised by** bacteria.{S} Apart from comfrey, a plant that is toxic to the liver,[23] which can extract this vitamin from the soil,[24] plants do not contain it[25][26] except in cases of contamination.’

The use of the periphrastic passive enables the noun phrase *La vitamine B12*, the central theme of the paragraph, to be placed in topic position. The presence of the agent allows to provide details on the type of instigator synthesizing the vitamin while presenting the information as secondary. The use of a passive option instead of an active one enables to maintain the focus on the main theme of the discussion (*Vitamin B12*) by promoting it to topic position, even though, as patient, it suffers the situation. This scenario, however, remains rather infrequent in our French data. As will be detailed in Chapter 8, the functional domains of Spanish overt agents seem to slightly differ from what has been observed for French.

Let us now consider the distribution of overt agents through modes and registers. Table 22 allows to see in which mode and register overt agents are used.

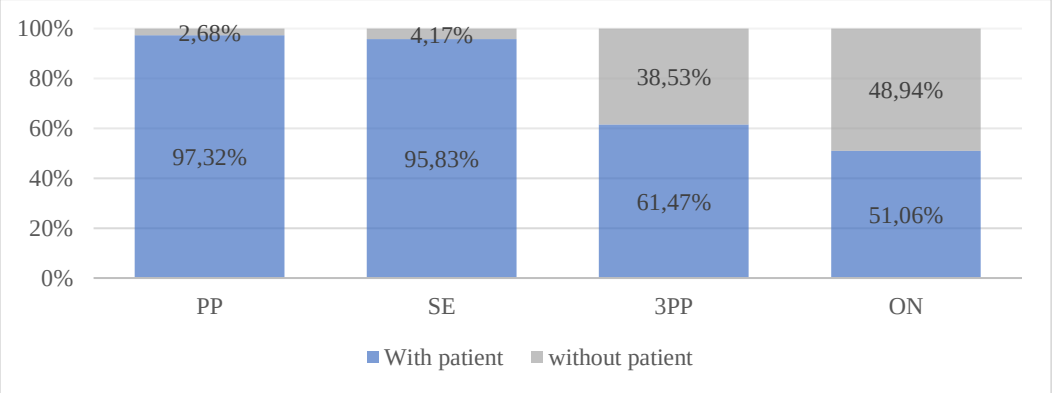
| Register | Occurrences | % |
|---|---|---|
| **Informal Oral** | 5 | 4.8% |
| **Formal Oral** | 30 | 24% |
| **Informal Written** | 24 | 19.2% |
| **Formal Written** | 65 | 52% |
| Total | 124 | 100% |

*Table 22. Distribution of overt agentsin periphrastic passives by mode and register*

It can be observed that 52% of the expressed agents are used in formal written data. It can be assumed that this high frequency of overt agents in formal written data is linked to the higher level of grammatical complexity generally attributed to formal and written contexts. The periphrastic passive, which usually occurs without overt agents, appears as grammatically more complex in formal data since the second participant of the two-participant clause is expressed. A similar explanation, calling on the lower degree of syntactic complexity, can be put forward to explain the low number of overt agents in informal oral data, which are those to least resort to explicit agents (4.8%). Informal written data and formal oral data both contain 24% of the overt agents. They are both situated in an interim position between informal oral data with a lower level of grammatical elaboration, thus a smaller proportion of periphrastic passives with overt agents and formal written productions, with a higher complexity and, as a consequence, a higher use of explicit agents. Now  a clearer picture of the types of agents have been establish, let us examine patients.

### 7.3.3. Patient

Patients can occupy a subject position as in the periphrastic passive and the *se*-construction or an object position as it is the case for the man-impersonal *on* and the non-anaphoric third person plural. The presence of a patient is, however, not a requirement for these two constructions.




*Graph 26. Presence of a patient in French passive-like structures*

Graph 26 shows that the periphrastic passive and the *se*-construction are almost exclusively constructed with a patient. This was expected due to the intrinsic grammatical properties of these two constructions e.g., – constructions where the object is promoted in subject position and suffers the action such as illustrated in (219). The cases where there is no patient are those with expletive pronouns, already developed with example (197) (repeated here as 218) in section 7.3.1 for the parameter *participant* .

(218) *Il doit cependant* ***être noté*** *qu'une étude japonnaise a montré l'existence d'une corrélation entre la consommation accrue de certains produits animaux et la baisse de certaines formes de maladies cérébrovasculaires (...)* (WIKI FR)

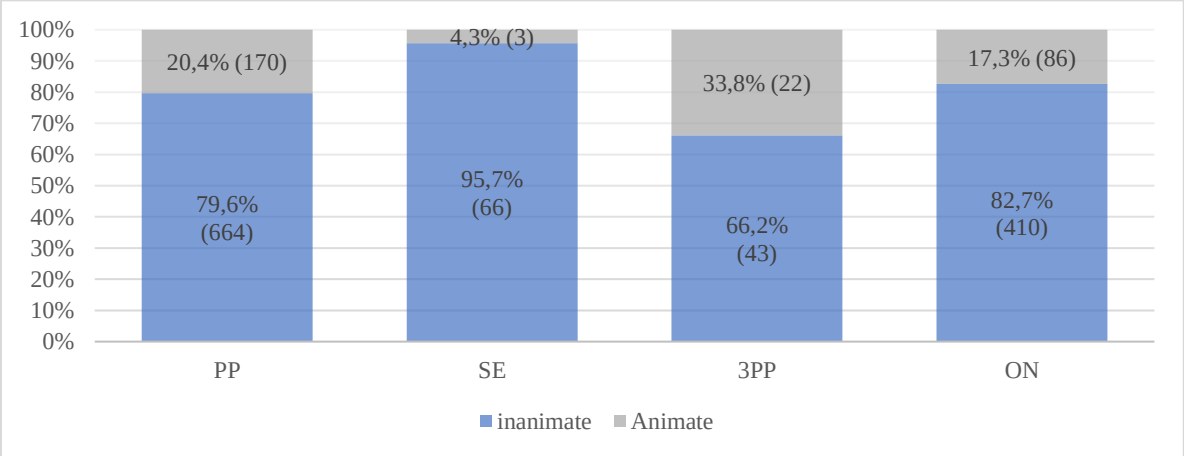
‘It should **be noted**, however, that a Japanese study has shown a correlation between increased consumption of certain animal products and a decrease in certain forms of cerebrovascular disease (…)’

(219) *Il faut que les territoires – que ce soit l'Espagne, la Grèce ou la France et les îles comme la Corse, la Sicile et la Sardaigne –* ***soient protégés*** (PROCEP FR).
‘Territories - be it Spain, Greece or France and islands such as Corsica, Sicily and Sardinia - **must be protected**.’

Compare (219), where Spain, Greece and France and Corsica, Sicily and Sardinia are patients and undergo the action of being protected, with (218), where the pronoun *il* (‘it’) does not substitute any concrete entity and does not suffer any action. Such cases of periphrastic passives with no patient are nevertheless rare in French (2.68%; 23 occurrences).

The man-impersonal *on* and the non-anaphoric third person plural are less frequently constructed with a patient since results indicate the absence of a patient in respectively 51.1% and 61.5% of the cases. This result is a direct consequence of the fact that *on* and non-anaphoric third person plurals are frequently used in one-participant clauses, thus leaving empty the argument slot of the patient.

Let us now focus on occurrences of the four passive-like constructions when constructed with a patient. The distribution animate versus inanimate patient will be first discussed and, in a second step, these two categories will be further detailed.




*Graph 27. Distribution of animate and inanimate patients in French passive-like structures*

The figures in Graph 27 indicate that inanimate patients are the most frequent in the four passive-like constructions (see Examples 220 to 223).

(220) *Les suppléments de B12 tels que les pilules de vitamines* ***sont souvent préparés*** *à partir de déchets d'abattoir et ne conviennent donc pas aux végétaliens* (…). (WIKI FR)
‘B12 supplements such as vitamin pills **are often prepared** from slaughterhouse waste and therefore not suitable for vegans (…).’

(221) *<question> {S}Combien de temps à l'avance faut-il se présenter à l'aéroport pour un vol international? (...) <answer> {S}Si le vol est vers les États-Unis ou à partir des États-Unis, prévoir 3 heures: on ne sait jamais quel nouveau règlement* ***ils vont inventer*** *ni comment ils vont interpréter les règlements actuels.* (YAHOO FR)
'<question> {S}How far in advance do you have to be at the airport for an international flight? (...) <answer> {S}If the flight is to or from the US, plan 3 hours: you never know what new regulations **they will invent** or how they will interpret the current ones.'

(222) *S}En plus il y a la rosette, un truc qui* ***se consomme*** *en buvant du Beaujolais mais j'aime pas trop ça.* (YAHOO FR)
'In addition, there is the rosette, something that **can be eaten** while drinking Beaujolais, but I don't really like it.'

(223) *Zinc : Dans le monde végétal, le zinc se trouve principalement dans les noix et amandes.{S}* ***On en trouve*** *également dans les produits laitiers. (*WIKI FR)
*'*Zinc: In the plant world, zinc is mainly found in nuts and almonds {S} **It is also found** in dairy products.'

The proportion, nevertheless, considerably varies according to the construction in which patients are used. Whereas the *se*-construction is almost exclusively built with inanimate patients (95.7%), the other three structures appear as much more flexible regarding the nature of the patients. Though the analysis has revealed so far that the man-impersonal *on* and the non-anaphoric third person plural share a large number of characteristics, they seem to diverge as to their use of patients ($\chi^2 = 10.0739$, $p < .05$). The higher use of animate patients in non-anaphoric third person plural structures should, however, be considered with caution as figures for this structure remain low. With our data sets, it can be suggested that the (in)animate nature of the patient would be a distinctive feature in French passive-like structures.

Table 23 offers further details regarding the categorization and the distribution of patients.

| Structure | Animate human | Animate non human | Organization | Country/City | Others |
|---|---|---|---|---|---|
| **Periphrastic passive** | 9.5% (79) | 10.9% (91) | 1.4% (12) | 1.1% (9) | 77.1% (643) |
| ***Se*-construction** | - | 4.3% (3) | - | - | 95.6% (66) |
| **Non-anaphoric third person** | 29.2% (19) | 4.6% (3) | - | - | 66.1% (43) |
| **Man-impersonal *on*** | 9.9% (49) | 7.5% (37) | 0.4% (2) | 0.2% (1) | 82.6% (407) |

*Table 23. Classification of patients in French passive-like structures*

As for animate patients, the periphrastic passive resorts to human and non-human patients in comparable proportions (9.5% and 10.9%, respectively). A relatively similar result is observed for the man-impersonal *on*, though a slightly smaller proportion of animate non human patients is noted (9.9% and 7.5%). A further interesting result is the relatively high proportion of animate humans in the non-anaphoric third person plural, a characteristic that none of the other passive-like structures display. In (224) the non-anaphoric third person *ils accueillent* ('they welcome') is accompanied by the personal pronoun *t'* ('you', which is classified as a human animate patient.

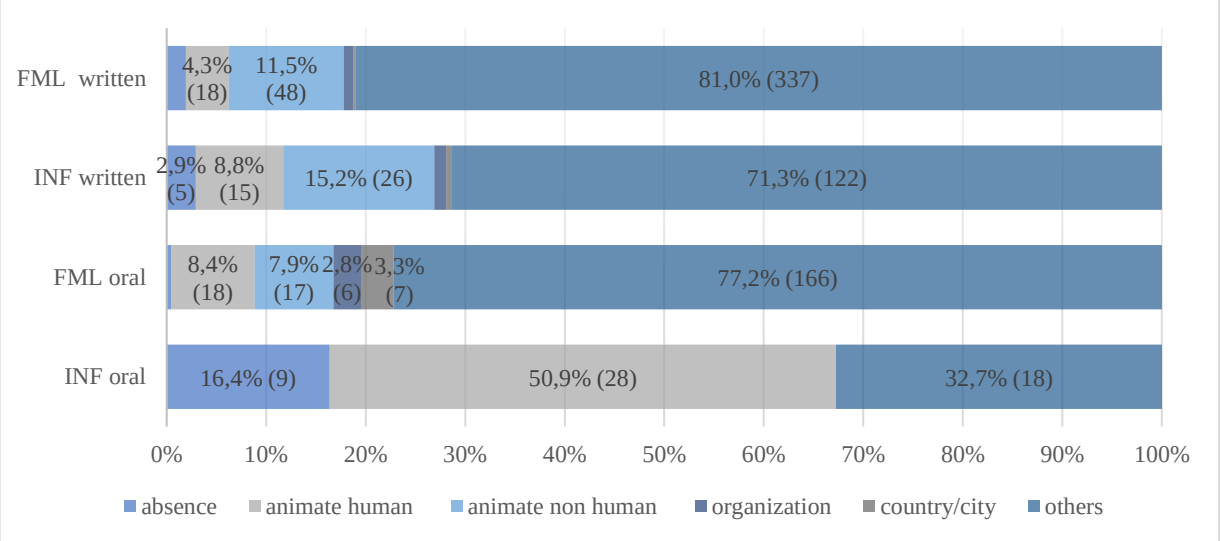
(224) *{S}J (rire) ben oui c'est pas juste pour le plaisir de pas assister à la cérémonie en même temps ce genre de cérémonie enfin j'ai tellement enfin j'espère j` sais pas au japon ça doit être pareil mais en prépa et tout les pauvres cérémonies d'accueil et tout en général*
*S}M ah je pense que si **ils t'accueillent** vraiment comme il faut au japon*
*{S}J oui ça doit être pas pareil (...)* (CIEL FR)
'{S}J (laughs) well yes it's not just for the pleasure of not attending the ceremony at the same time this kind of ceremony well I have so much well I hope I don't know in Japan it must be the same but in preparatory and all the shabby welcoming ceremonies and everything in general
{S}M ah I think **they really welcome** you properly in Japan
{S}J yes it must be different (...)'

A last information provided by Table 23 is the extremely high use of inanimate patients classified as *others*. This is particularly the case for the *se*-construction, where they account for 95.6% (66 occurrences). This phenomenon may be explained by a higher number of syntactic restrictions suffered by this structure as explained in Chapter 2 (section 2.2).

(225) *Mais, franchement, si les discours sur le climat et sur l'environnement avaient dû sauver la planète, ça **se saurait**.* (CIEL FR)
'But, frankly, if speeches about climate and environment were to save the planet, it **would be known**.'

In (225) the demonstrative pronoun *ça* ('it'), which refers to the fact that speeches do not save the planet, is classified as an inanimate patient in the category *others*.

Turning to the influence of register and mode on the use of patients in the four passive-like structures, results suggest what follows. As was the case for verb-related parameters in the previous section, the focus will be placed on the periphrastic passive and the man-impersonal *on* as they totalize a much larger number of occurrences. For the non-anaphoric third person plural and the *se*-construction, which display too small raw frequencies to allow a relevant examination of register and mode variation, the overall tendencies are the following ones. Results show that, as far as the *se*-construction is concerned, which only contains two categories of patients as can be seen from Table 23, all animate non human patients (3 occurrences) occur in written formal data. As for the non anaphoric third person plural, results indicate that animate patients are rather specific of informal oral data (20 occurrences).




*Graph 28. Distribution of patients in the French periphrastic passive*

The distribution of patients in periphrastic passives is quite different according to the mode and register. As can be seen from Graph 28, informal oral data are the ones that most resort to periphrastic passives without patients, that is, as explained below, with the expletive *il* ('it'). This data set is also the one containing the largest proportion of human beings as patient, as in (226) with the personal pronoun *elle* ('she') which stands for the sister of one of the speakers.

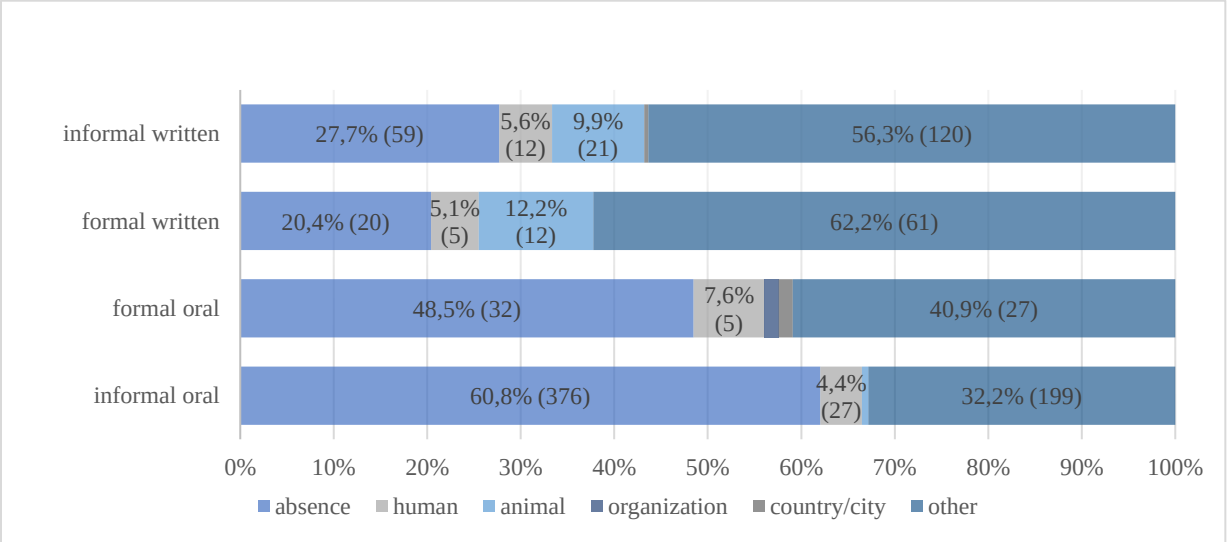
(226) *{S}BEA ouais/*
*{S}ELI j` t'ai dit* ***elle a été prise*** *aux impôts (elle a quand même) une chance elle euh j` te dis {S}BEA hm hm {S}* (CIEL FR).
'*{S}*BEA yeah/
{S}ELI I told you **she was taken** to the tax office (she still has) a chance she uh I tell you
{S}BEA hm hm {S}'

This tendency to resort to a high number of human patients and to relatively few inanimate *others* is not found in the other three combinations of mode and register sets of data. Vázquez Rozas puts forward that the anthropocentric and subjective character of conversational language contributes in explaining the high presence of human referents in that type of discourse (2006:411).

Graph 28 also brings to light that the use of animate non human patients (see example 227 with *bétail* 'livestock' as patient) appears to be rather specific to the written mode, informal (15.2%) and formal (11.5%). It can be assumed that, in spoken situations, when animate non human entities suffer any type of process, French speakers resort to another construction than the periphrastic passive.

(227) {S} Cela a provoqué la création de fermes géantes, dans lesquelles le bétail **est confiné**. (PROCEP FR)
'This has led to the creation of giant farms, in which livestock **are confined**.'

It can be observed that the profile of formal data as for the distribution of the different types of patients resembles what was found for written production. As will be seen, the picture is totally different for the man-impersonal *on*.




*Graph 29. Distribution of patients in French man-impersonals on*

Graph 29 presents[30] the distribution of the different types of patients within each mode and register. Overall, it can be observed that a patient is much more used in written data than in oral data. The absence of a patient ranges from 20.4% to 27.7% in written data whereas it accounts for 48.5% to 60.8% in oral data It can also be noted that this absence of patient seems to be a phenomenon more specific to informal data, where other strategies to include patients are suspected. These two observations can be linked to a general higher level of syntactic precision and complexity in written and formal productions.

When there is a patient, the following findings can be observed. Human beings with the semantic role of patient are represented in similar proportions (between 4.4% and 7.6%) in both modes and registers. Patients pertaining to the category *animal* are only present in written data. The absence of animals as patients in formal oral data was unexpected as a subcorpus dealing with animal welfare was included in the formal oral data set. A possible explanation for this result is the low frequency of use of the construction *on* in parliamentary proceedings (formal

[30] For a question of legibility, the percentages and raw frequencies for categories obtaining less than 2% are not tagged on the graph. It is the case for the categories *organization* and *country/city* in formal oral data, for *country/city* in informal written data and for *animal* in informal oral data.

oral data). However, the frequency of use of this passive-like construction in formal written data is equally low and animals account for 12,2% (12 occurrences) of the total number of patients. It can be assumed that when speakers in formal contexts have to orally express themselves on matters affecting animals they will resort to different mechanisms, for instance noun complements as *commerce d'animaux* in (228).

(228) ***On peut se réjouir*** *qu'il dénonce dans le même temps la corruption et le rôle majeur du commerce d'animaux en ligne*. (PROCEP FR)
'**We can be pleased** that at the same time it denounces corruption and the major role of the online animal trade.'

As for the distribution of patients included in the category *others*, Graph 29 reveals that they received a double influence, from the mode and from the register in which they appear. First, this type of patient is more used in written data than in oral ones. The mode seems thus to be an important element to consider. In addition to mode, register seems also to affect the use of this category of patient. It can be observed that the proportion of patients is higher in formal data that in informal data.

The analysis of patients with a special attention to mode and register revealed that these two components impact not only the absence/presence of a patient but also its category. The influence of mode and register, though under different forms, was observed both for the periphrastic passive and the man-impersonal *on*. The analysis of patients will be further deepened with a discussion on the degree of individuation and affectedness of the patients.

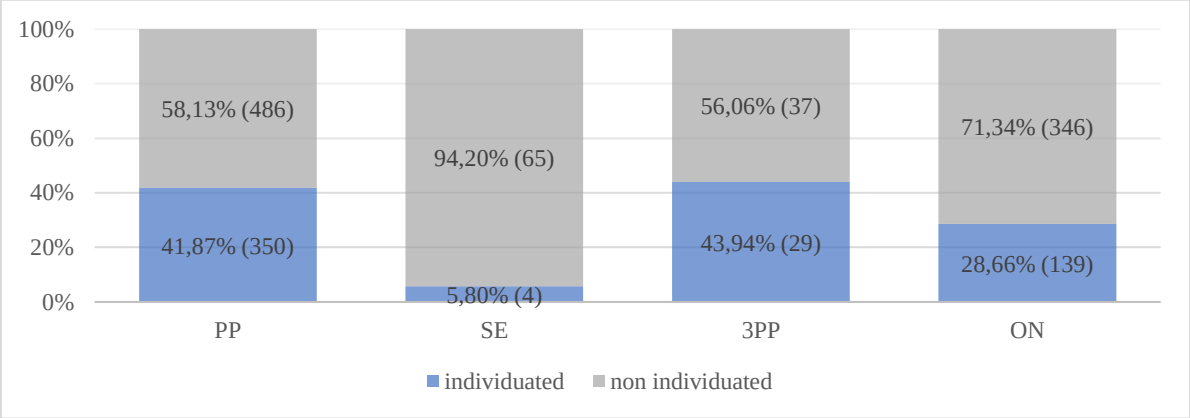
### 7.3.4. Individuation and affectedness of the patient

Patients can either be individuated, that is experiencing a sense of individuality, separated from other entities, or non-individuated. Hopper & Thompson (1980) list six features of an individuated patient: proper, human or animate, concrete, singular, countable and referential or definite. Patients meeting these criteria, or part of them, are better candidates to high transitivity clauses. As explained in the methodological section (Chapter 5), since Hopper & Thompson did not specify the number of parameters a patient should meet in order to be considered as individuated, we set the minimum to four out of the six features.

Results show that non individuated patients predominate in the four passive-like constructions, though in different proportions.

(229) {*S*} *Il faut que cela baisse, et vite, si l'****on veut respecter*** *les engagements pris dans le cadre de la COP21 pour sauver la planète, le climat et l'humanité.* (PROCEP EN)
'This has to decrease, and quickly, if **we want to respect** the commitments made at COP21 to save the planet, the climate and humanity.'

In (229) the patient *les engagements* ('the commitments') is classified as non-individuated since its characteristics are the following ones: common noun, inanimate entity, plural, abstract, countable and definite. Only the last two criteria out of the six pertain to the features considered as highly individuated by Hopper & Thompson, which leads us to classified *les engagements* as non-individuated.




*Graph 30. Distribution of individuated and non-individuated patients in French passive-like structures*

While the *se*-construction resorts almost exclusively to non individuated patients (94.2%), the man-impersonal *on* uses them in approximately two third of the cases (71.3%). Although the periphrastic passive and the non-anaphoric third person plural are those that contain the smallest percentages of non-individuated patients, the figures remain above the 50% threshold (58.1% and 56%, respectively). In (230) the patient *1 100 chauffeurs des pays de l'Est*, used with the periphrastic passive, is considered as individuated since it complies with the criteria *human*, *concrete*, *countable* and *definite*. Although it does not meet the same criteria as we saw in (229), Example (231) can also be considered as individuated. The patient *le* ('it'), which stands for the Chinese dessert, has the following individuated features: *concrete*, *countable*, *single* and *definite*.

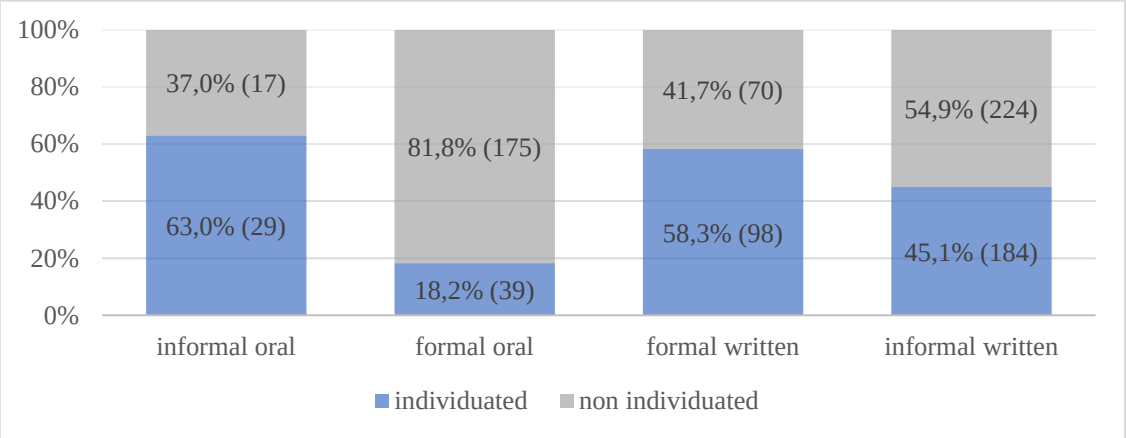
(230) *Ce sont 1 100 chauffeurs des pays de l'Est qui* ***auraient été exploités*** *dans des conditions de travail indignes et pour un salaire de misère*. (PROCEP FR)
'They are 1,100 drivers from Eastern countries that **were exploited** in undignified working conditions and for a poverty wage'.

(231) *<answer> ben non ce n'est pas de la tarte puisque c'est du jell-o !..... dans la cuisine chinoise il y a aussi un désert qui ressemble à du jell-o ( aussi mauvais !... )* ***ils le mangent*** *avec une cuillère en porcelaine* (YAHOO FR)
'well no it's not pie since it's jell-o ! ..... in Chinese cuisine there is also a desert that looks like jell-o (as bad !... ) **they eat** it with a porcelain spoon'

While the difference between the periphrastic passive and the non-anaphoric third person plural is not significant. A high proportion of non-individuated patients seems specific to the *se*-construction and the man-impersonal *on (*especially marked for the former). This result could be link to the fact that these two constructions contain a high proportion of inanimate patients, better candidates to be classified as non-individuated as they lack the feature 'animate', one of the six properties listed by Hopper & Thompson (1980) to be considered as individuated.

Compared to the *se*-construction and *on*, the periphrastic passive and the non-anaphoric third person plural display a much more balanced distribution of individuated and non-individuated patients. Individuation can thus be considered as a distinctive feature between the passive-like structures.

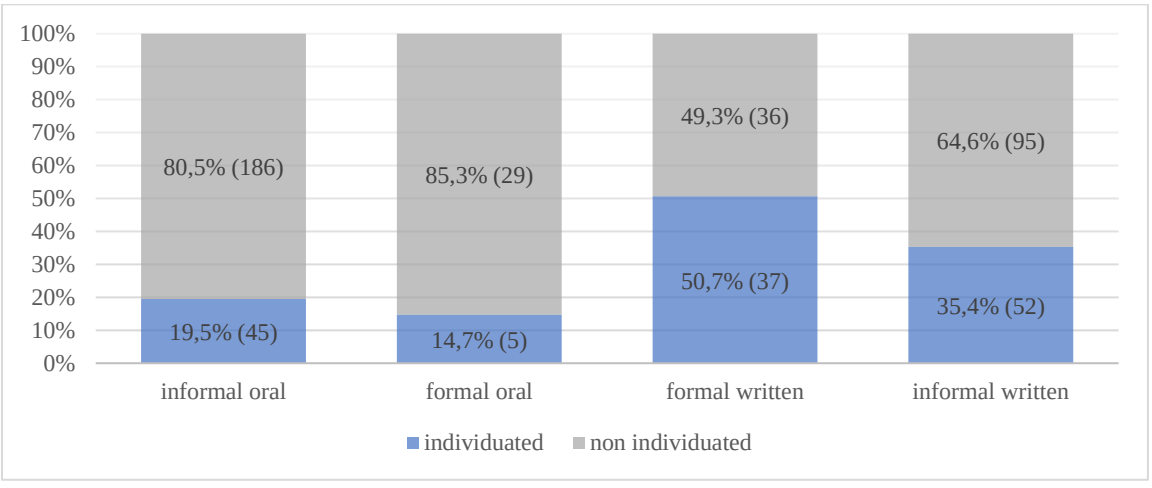
Zooming in on the influence of register and mode on the use of individuated patients, it appears from Graph 31 that the periphrastic passive seems to show great variation in oral data.




*Graph 31. Distribution of individuated and non-individuated patients in French periphrastic passives*

While non-individuated patients largely predominate in formal oral data, they only account for 37% in informal oral data. Individuated patients imply most of the time human patients. The low proportion of individuated patients in formal data means that , when human patients suffer any type of event or situation in formal contexts, the periphrastic passive is not the favoured construction. A significantly higher level of non-individuated patients is also found in informal written data compared to formal written data ($\chi^2 = 8.3419$, $p < .05$). It can be thus concluded that the degree of formality of the situation largely impacts the use of individuated patients. Nevertheless, since the number of periphrastic passives in informal oral data is still fairly low, only tentative conclusions can be formulated as for the oral mode.

Graph 32 shows a totally different picture as for the influence of register and mode on individuated patients in French man-impersonals.




*Graph 32. Distribution of individuated and non-individuated patients in French man-impersonals on*

We have just seen that for the periphrastic passive, the degree of formality accounts for a large part of the variation. For the man-impersonal *on*, the key element that influences the distribution of individuated patients appears to be the mode. Graph 32 indicates that non individuated patients largely predominate in oral data with no significant difference between formal and informal registers data ($\chi^2 = 0.4414$, $p > .05$). To a smaller extent, we can also observe an influence of the register, but only in written data since informal written data use significantly more non individuated patients than formal written data ($\chi^2 = 4.7467$, $p < .05$).

In addition to the degree of individuation patients can be also be characterized by the degree of affectedness. A patient can either be totally affected or partly affected. Compare (232) and (233). In (232) only half of the water is used for livestock production, not the entire amount of water available. Conversely, in (233), the whole Belgium is affected by the case. Literature usually reports that patients in passive voice are highly affected (see Myhill 1992, López & Falk 1999 or Dam 2001). Myhill (1992) puts forward that the degree of affectedness of the patient is one of the features that would allow to distinguish passive from active clauses. Results confirm that, in French, nearly all occurrences of passive-like structures are totally affected (between 97.2% and 100%).

(232) *Aux États-Unis, plus de la moitié de l'eau consommée* ***est utilisée*** *pour la production de bétail.* (WIKI FR)
'In the United States, more than half of the water consumed **is used** for livestock production.'

(233) *Monsieur le Président, Monsieur le Commissaire, comme vous le savez, la Belgique* ***est secouée*** *par un dossier judiciaire qui vise un important transporteur routier.* (PROCEP FR)
'Mr President, Commissioner, as you know, Belgium **is shaken** by a legal case which targets a major road haulier.'

This exclusive use of affected patients was expected since a typical characteristic of impersonal mechanisms is the focus on the patients, conferring them greater importance. What was less expected is the total affectedness of patients in man-impersonal and non-anaphoric third person constructions, which are active constructions. From our data, it is suggested that when active constructions are used impersonally, they behave like typical passive voice constructions with totally affected patients.

Since total *affectedness* of the patient appears as the norm in the passive-like structures studied in this investigation, no influence of the mode or the register could be observed.

### 7.3.5. Summary

Summarizing the behaviour of participant-related parameters, the following results can be put forward. The number of participants included in the passive-like construction varies according to the structure. Two groups can be distinguished: the periphrastic passive and the *se*-construction on the one hand, which are usually accompanied by two or more participants, and

the man-impersonal *on* and the non anaphoric third person on the other, within which the number of two- and one-participant clause are much more balanced. The division of the four structures into two groups can be explained by the specific syntactic architecture of the periphrastic passive and the *se*-construction, placing a patient in focus position. This different frequency of use of one- and two-participant clauses is directly linked with the presence of a patient. Patients are far less used in non-anaphoric third person plurals and constructions with *on* since these two structures are less syntactically restricted as for the use and the position of the patient. As for the nature of patients they are mainly inanimate with a particularly high proportion of patients belonging to the category inanimate *others.* A strikingly high proportion of animate patients was, however, found for the non-anaphoric third person plural. The analysis also reveals that participant-related parameters are considerably impacted by the mode and the register.

## 7.4. Context-related parameters

This section develops two parameters related to the context within which the passive-like structure appears. The presence of a spatio-temporal restriction is analysed in section 7.4.1 and the use of a prepositional phrase is examined in 7.4.2. The other parameters related to the co-text but considered by the MCA as less fundamental in explaining the variation between the structures will be briefly presented in 7.4.3.

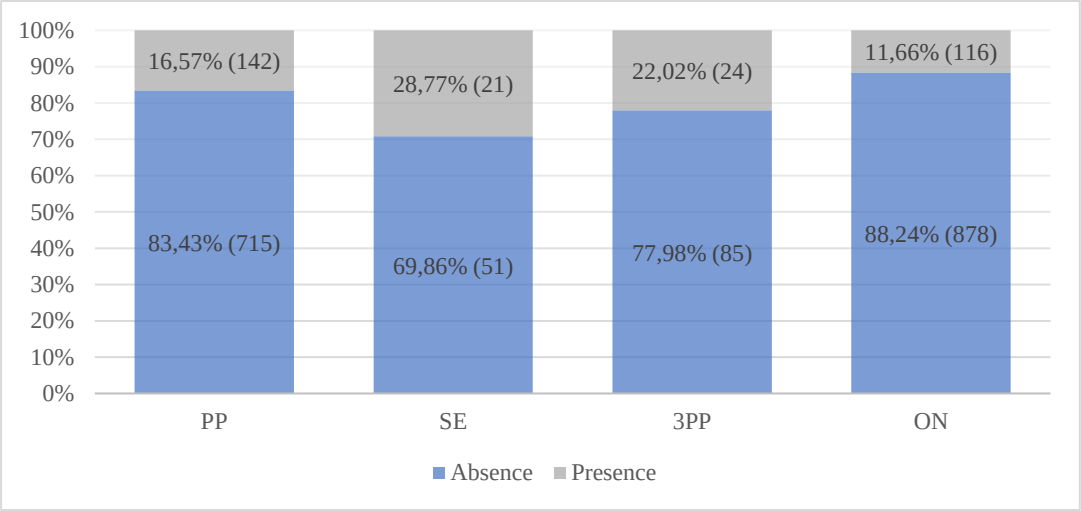
### 7.4.1. Presence of a spatiotemporal restriction

The presence of spatio-temporal restrictions can help in identifying the agentive force in passive-like structures as they restrict either the geographical space (234) or the time period (235) within which the passive-like structure occurs.

(234) *Ce légume est particulièrement commun en Chine, où **il est connu** sous le nom de gaosun ou jiaobai*.(PROCEP FR)
'This vegetable is particularly common in China, where **it is known** as gaosun or jiaobai.'

(235) *Aujourd'hui, les élevages **sont souvent dominés** par une seule variété: la poule Leghorn est présente dans tous les pays du monde (…).* (PROCEP FR)
'Today, farms **are often dominated** by a single variety: the Leghorn hen is present in every country in the world.'

Example (234) illustrates a use of a spacebuilder with the periphrastic passive, where the name of the vegetable *gaosun* or *jiaobai* is especially used in China. By using this structure, speakers construe the agents as any person living in China rather than as specific persons. In (235), the passive-like structure co-occurs with the adverb *Aujourd'hui* ('Today'), which allows to situate the action in time: the Leghorn dominates farms nowadays and not in the past.



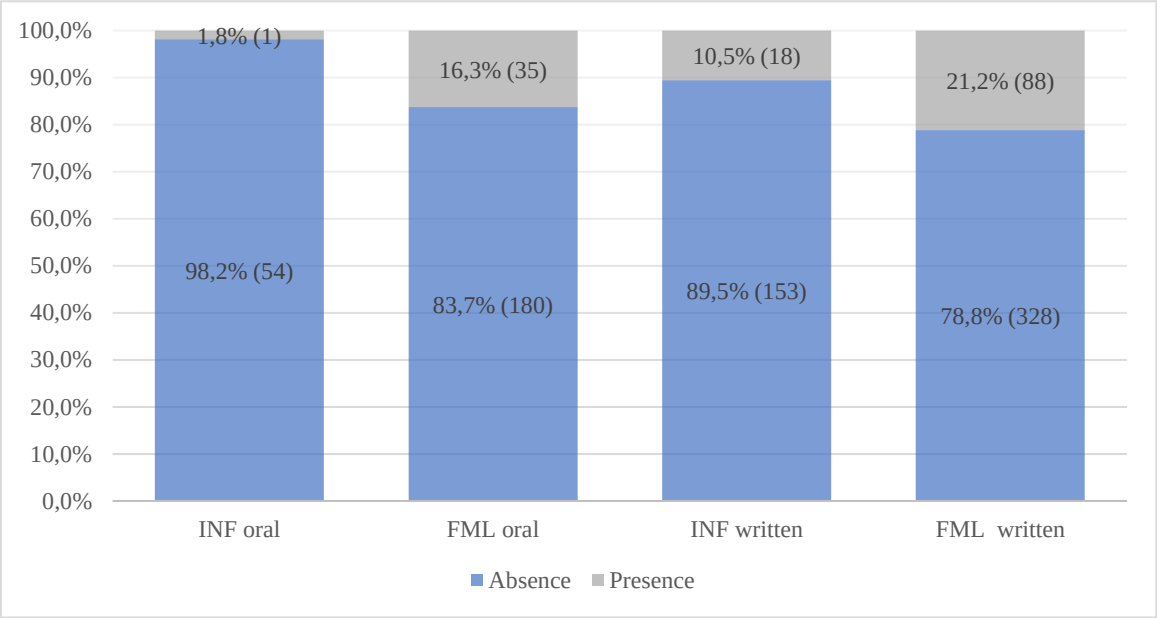

*Graph 33. Presence of a spatiotemporal restriction in French passive-like structures*

The *se*-construction is the passive-like structure that most uses spatio-temporal restrictions such as in (236). The action of *se dit* ('is said') is restricted to the media and is not something that everyone reports. The high presence of these elements could be explained by the fact that the *se*-construction hardly uses overt agents compared to the other French passive-like structures. The use of spatio-temporal restrictions can be a strategy to provide additional information on the individuals potentially included in the situation or the event.

(236) *Loin d'être l'apanage d'une horde de bobos vegans confite dans son confort matériel, comme il* ***se dit*** *dans les médias, la question du bien-être animal recèle en elle le devenir d'une population mondiale en explosion.* (PROCEP FR)
'Far from being the prerogative of a horde of vegan bobos confined to their material comforts, **as they say** in the media, the question of animal welfare conceals within it the future of an exploding world population'.

So far we have seen that for several parameters the periphrastic passive and the *se*-construction display a similar behaviour. This is not the case regarding the use of spatio-temporal restrictions since the *se*-construction uses these items significantly more compared to the periphrastic passive ($\chi^2 = 7.2856$, $p < .05$). A similar phenomenon can be observed for the pair involving the two constructions based on a pronoun since the non-anaphoric third person plural tends to use significantly more spatio-temporal restrictions than the man-impersonal *on* ($\chi^2 = 9.4924$, $p < .05$). It can be thus be assumed that the use of spatio-temporal items is a distinctive parameter since each passive-like structure has its own specific preference.

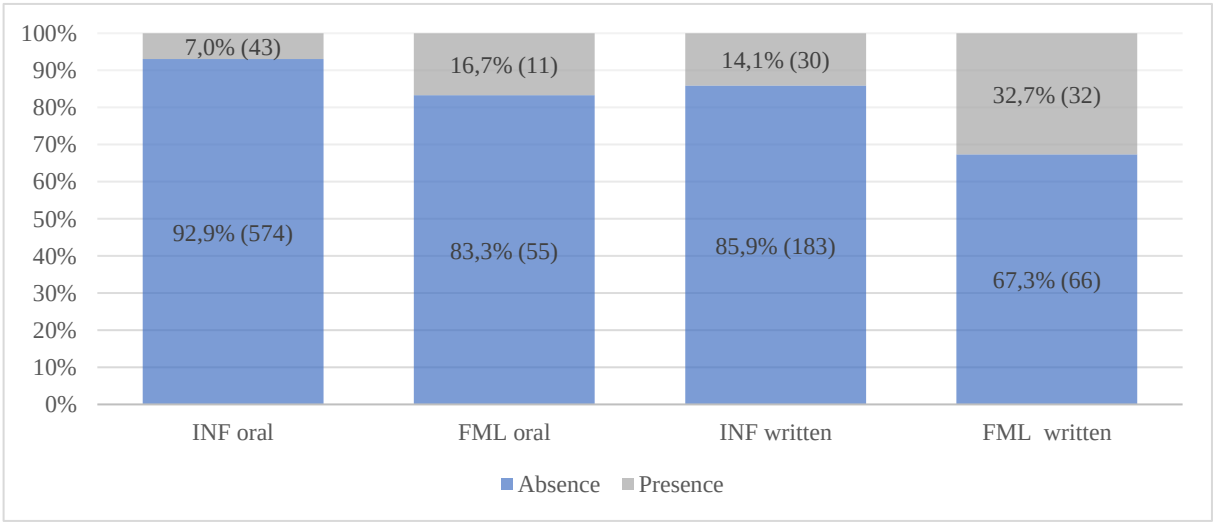
As shown in Graph 34, the use of spatio-temporal restrictions in periphrastic passives differs according to the mode and the register.




*Graph 34. Distribution of spatiotemporal restrictions in French periphrastic passives*

Both in oral data and in written data, the use of spatio-temporal restrictions is more specific of formal situations. This difference is significant for oral data ($\chi^2 = 7.9257$, $p < .05$) and for written data ($\chi^2 = 9.25$, $p < .05$). From Graph 36 it can be also seen that written data tend to use substantially more spatio-temporal restrictions than oral data. The difference is, however only significant in informal data ($\chi^2 = 4.0982$, $p < .05$). It can be concluded that the use of spatio-temporal restrictions in periphrastic passives suffers a double influence, from the register and, to a smaller extent, from the mode. This finding supports Steiner's claim (2017) that location and time builders are characteristic of written productions (2017:33). The high proportion of spaciotemporal devices in formal oral data can be explained by the fact that parliamentary debates may contain features usually associated to written modes.

Let us now examine whether a similar impact is observed for the man-impersonal *on*.



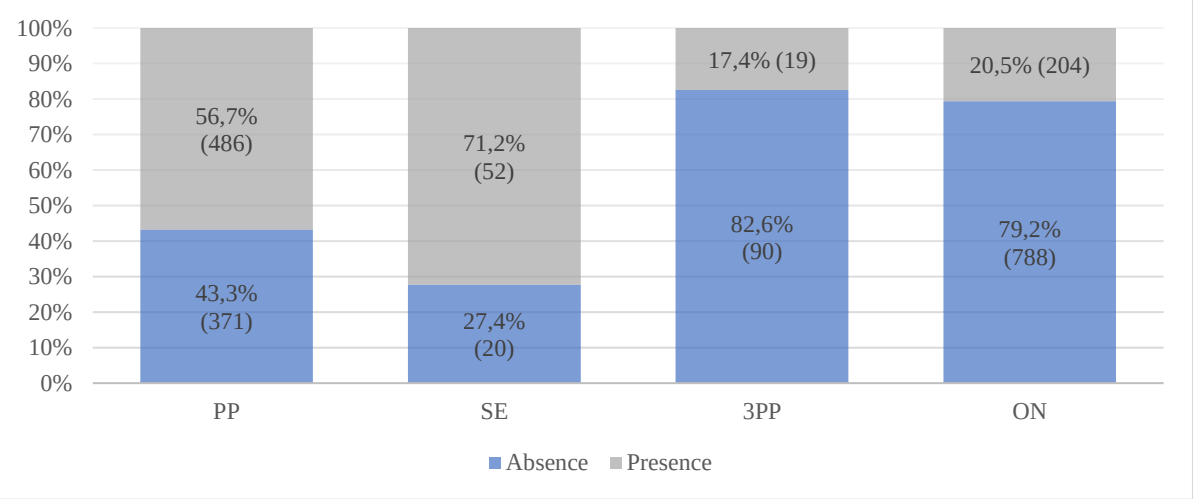

*Graph 35. Distribution of spaciotemporal restrictions in French man-impersonals*

Graph 35 provides information about the distribution of spatio-temporal restrictions with the man-impersonal *on*. As was the case for the periphrastic passive, the use of spatio-temporal restrictions is associated with formal situation. This holds true for oral data ($\chi^2 = 7.7005$, $p < .05$) and for written data ($\chi^2 = 14.4988$, $p < .05$). Written contexts are also prone to contain more restrictions since data show that man-impersonals use significantly more spatio-temporal items in formal ($\chi^2 = 20.1667$, $p < .05$) and informal ($\chi^2 = 9.9933$, $p < .05$) written data. Similarly to what was observed for the periphrastic passive, the elements in the co-text of the man-impersonal *on* are impacted by the degree of formality and the mode.

It can thus be said that formal and written contexts are more prone to give supplementary information as far as the temporal and geographical delimitation are concerned. These results are in line with Steiner's (2017) observations and confirm Mair's (2006) claim that "the fact that writing usually invloves communication across space and time will make it necessary to spell out references to the context of situation which may be left implicit in speech." Spatiotemporal elements help to define the scope of the passive-like structures in French. It can be hypothesized that in more informal contexts this information is either provided differently or omitted since literature has shown that interactions in informal contexts usually involve a higher degree of shared common ground between the participants. In this case, the necessity to specify this category of information may appear as superfluous or irrelevant.

### 7.4.2. Presence of a prepositional phrase

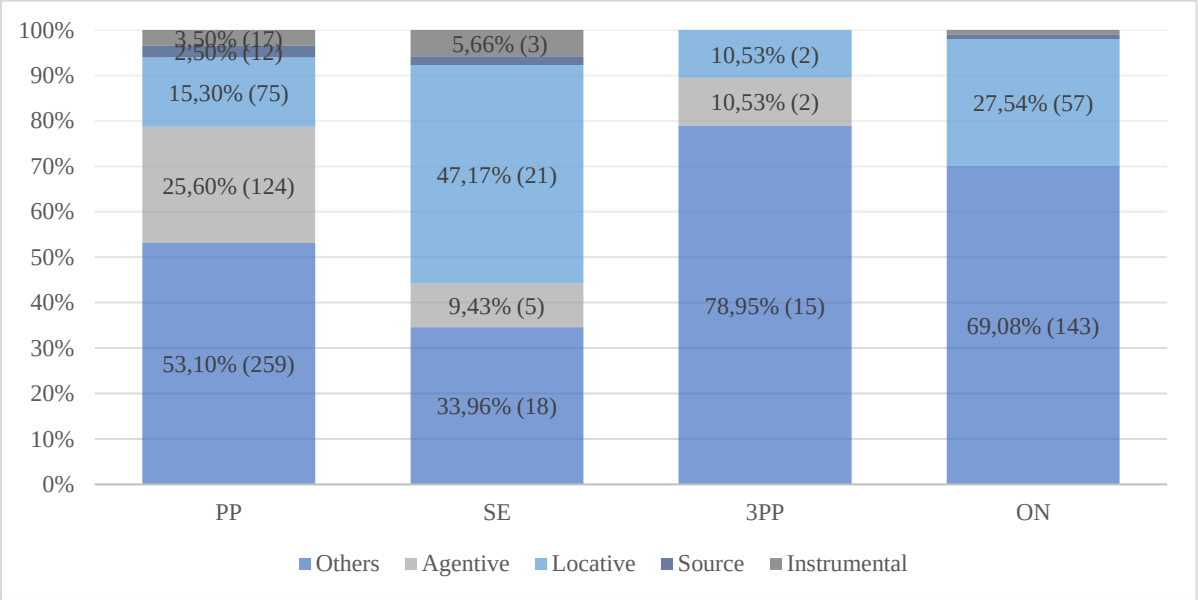
Agenthood management may be expressed by other prepositions than the typical agentive preposition *par/por/by* as shown by Delbecque (2014) for Spanish. Locative, source and instrumental prepositional phrases can also, to a certain extent, express nuances of agentivity. In this analysis, Delbecque's proposal is also extended to French and English.




*Graph 36. Presence of prepositional phrases in French passive-like structures*

Graph 36 shows that passive-like structures do not use a prepositional phrase in the same proportion. Prepositional phrases are relatively infrequent in the non-anaphoric third person plural and the man-impersonal *on* (17.4% and 20.5%, respectively whereas they dominate in

the periphrastic passive (56.7%) and the *se*-construction (71.2%). This tendency is especially marked for the *se*-construction. When they resort to a prepositional phrase, the four structures display quite a different profile. Graph 37 provides the distribution of these prepositional categories[31] across the four passive-like structures.




*Graph 37. Distribution of prepositional phrases in French passive-like structures*

Again, the four constructions can be analysed in two groups, the periphrastic passive and the se-construction on the one hand and the non-anaphoric third person plural and the man-impersonal on the other. Whereas the category *others* largely predominates in R-impersonals, the distribution of the prepositional phrases is much more contrasted in the periphrastic passive and the *se*-construction. Let us first examine the pair formed by the periphrastic passive and the se-construction. For the periphrastic passive results reveal that the two main categories of prepositional phrases are the category *others* (53.1%) and the typical agentive one as illustrated in (237). Though the category *others* is dominant, the periphrastic passive employs that type of preposition far less than R-impersonals. Regarding the profile of the *se*-construction, Graph 37 indicates that locative prepositions as illustrated in (238) with the preposition *dans* ('in'), predominate.

(237) *Cette règle **fut relâchée** par le concile du Vatican de 1965, mais la plupart des trappistes adhèrent encore à l'enseignement originel sur le végétarisme. (WIKI FR)*
'This rule **was loosened b**y the Vatican Council of 1965, but most Trappists still adhere to the original teaching on vegetarianism.'

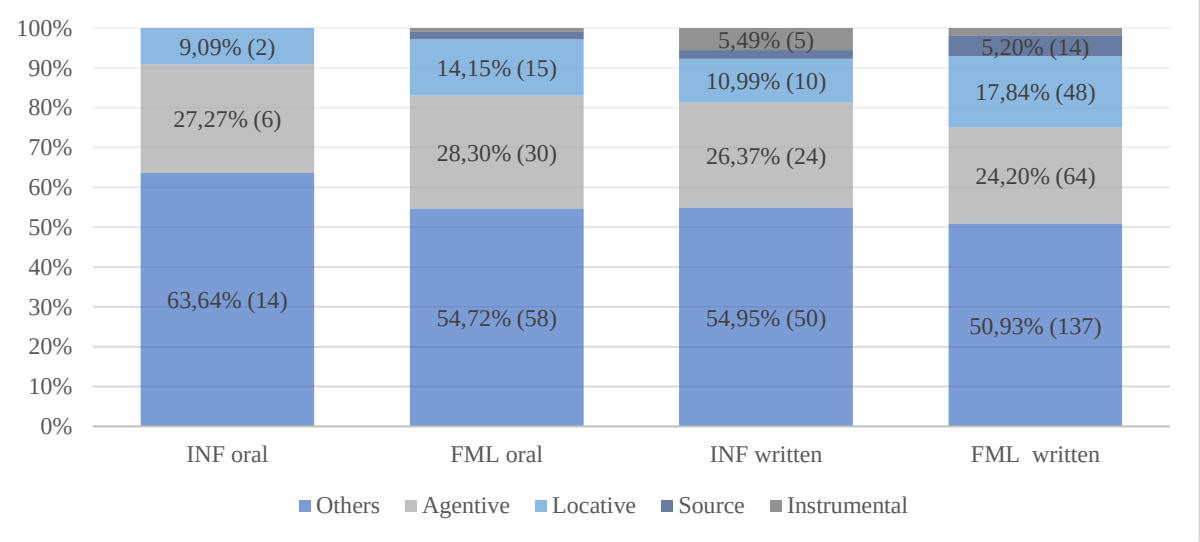
(238) *En France, la plante se trouve facilement dans les grandes jardineries, les magasins d'ameublement ou les supermarchés. (YAHOO FR)*

[31] Locative prepositional phrases and spatio-temporal restrictions overlap in 38 (out of 75) cases for the periphrastic passives, 34 (out of 57) cases for the man-impersonal and 17 (out of 21) cases for the *se*-constructions. No cases of overlap were found for the non-anaphoric third person plural.

‘In France the plant can easily be found in large garden centres, furniture shops or supermarkets.’

As agentive prepositions are relatively rare in *se*-constructions, it can be assumed that locative prepositions endorse part of their role. This difference in the use of prepositional phrases to express agentive nuances is thus a distinctive feature of the periphrastic passive and the *se*-construction. Let us now consider the two constructions based on a pronoun. Interestingly, Graph 37 shows that typical agentive prepositions are absent from constructions with *on* and hardly present in non-anaphoric third person plural (10.5%; 2 occurrences). In man-impersonal *on*, precisions about potential agents are rather brought through the use of locative prepositions (27.5%; 57occurrences). This trend does not seem to apply for the non-anaphoric third person plural since locative prepositions only account for 10.5% (2 occurrences). Since the category *others*, which does not express any agentive value at all, predominates in non-anaphoric third person plurals, it can be concluded that this type of passive-like construction does not resort to prepositional phrases to bring information on the defocused agent. Finally, source and instrumental prepositional phrases are hardly used with French passive-like structures.

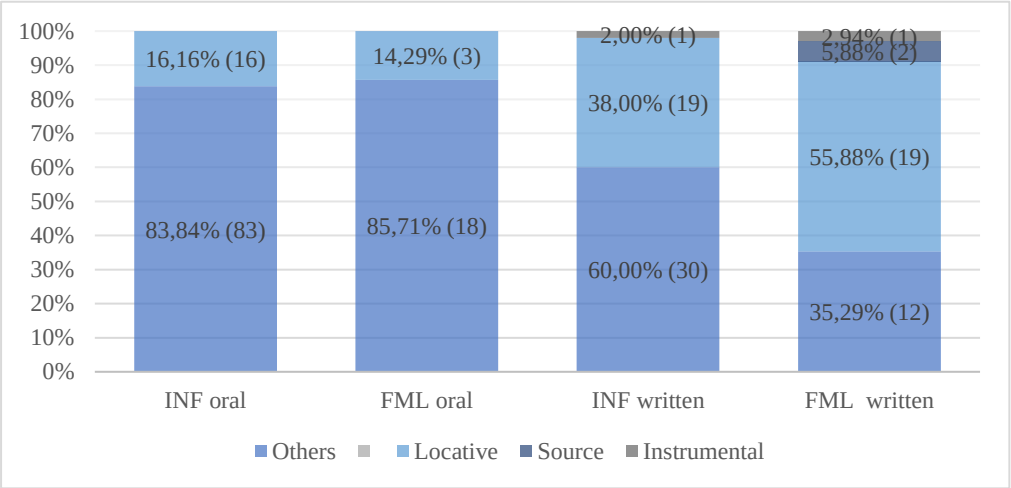
Regarding the impact of register and mode on the distribution of prepositional phrases, Graph 38 shows that no significant difference can be observed.




*Graph 38. Distribution of prepositional phrases in French periphrastic passives*

The main propositional complement is the category *other*, thus not conveying any agentive force in the four mode/register combinations. It can however be noted that locative prepositions seem to be more frequent in formal data and especially in formal written data (17.8%). Since the figures remain low, these are only tentative conclusions.

A totally different picture emerges for the man-impersonal, partly linked to the fact that *on* is often the agent. As we have seen above, *on* does not resort to the typical agentive preposition *by* but rather relies on locative prepositions.




*Graph 39.Distribution of prepositional phrases in French man-impersonals*

Graph 39 indicates that these locative prepositions are characteristic of written data. Particularly formal written data seem to be the data set that resorts most to these prepositions. Nevertheless, the difference is not significant ($\chi^2 = 2.6124$, $p < .05$), which may be attributed to the restricted number of occurrences obtained for these categories. Although in very small proportions, formal written data is also the only data set to resort to source and instrumental prepositions. This suggests that written data and, more precisely, formal written data rely more frequently on prepositions serving agenthood management than oral data.

### 7.4.3. Other parameters

In addition to the use of spatio-temporal restrictions and prepositional phrases, other parameters may also contribute, though to a smaller extent, in the identification process of the agent. This section briefly summarises the results obtained for the following parameters: the use of another passive-like construction, the presence of a person reference in the context, the use of an adverbial complement and the type of content. Overall results suggest that the four passive-like structures tend to co-occur with another impersonal mechanism, be it the same as the one analysed or another one. In (239) the periphrastic passive *est cultivé* ('is grown') occurs in the linguistic context of the *se*-construction *se trouve facilement* ('is easily found').

(239) *Cette plante **est cultivée** en tant que plante d'ornement pour ses feuilles belles et brillantes.{S} En France, la plante **se trouve** facilement dans les grandes jardineries, les magasins d'ameublement ou les supermarchés.* (YAHOO FR)
'This plant **is grown** as an ornamental plant for its beautiful, shiny leaves.{S} In France, the plant **is** easily **found** in large garden centres, furniture shops or supermarkets.'

The only structure within which this tendency is less marked is the *se*-construction since it only appears together with another impersonal mechanism in 45.2% of the cases. As for the type of

content, the four passive-like structures almost exclusively (between 92.6% and 99.1%) appear in neutral discourse fragments, that is with no clues guiding towards either a generic or a restrictive interpretation. Moreover, the French passive-like structures are rarely surrounded by an adverb or an adverbial phrase. The only structure that appears slightly more frequently with such devices is the *se*-construction (35.6%; 26 occurrences) such as in Example (239) above, which appears with the adverb *facilement* ('easily'). Finally, results show that the presence of a person reference in the context, as illustrated in (240), varies according to the structure. In (240) the presence of a reference to the *selectionneurs* 'breeders' helps to understand who makes efforts to recreate the breed.

(240) *S} Après quelques siècles, des efforts* ***se multiplient*** *pour recréer la race hovawart, notamment par des selectionneurs comme Bertram König, son fils Kurt F.{S}König, Alwin Busch, J.A.Becker et Theo Gräb (WIKI FR)*
'{S} After a few centuries, **efforts were increased** to recreate the Hovawart breed, notably by breeders such as Bertram König, his son Kurt F.{S}König, Alwin Busch, J.A.Becker and Theo Gräb.'

Each passive-like structure is accompanied by a person reference in different proportions, the group formed by the periphrastic passive 13.3% and the *se*-construction 28.8% scoring lower than the non-anaphoric third person plural (49.5%) and the man-impersonal *on* (72.2%).

### 7.4.4. Summary

The analysis of the two main contributing context-related parameters revealed what follows. Spatio-temporal restrictions, which help to define the space and time within which the passive-like structure is anchored, display different frequencies in the four passive-like structures. The division of the four passive-like structures into two groups, the periphrastic passive and the *se*-constructions versus the non-anaphoric third person plural and the man-impersonal *on*, does not seem to apply for this parameter. As for the use of prepositional phrases, the pairs match since far more prepositional phrases can be found with the periphrastic passive and the *se*-construction. The most frequent preposition contributing to agenthood management in these two structures is a prototypical agentive *by* in periphrastic passives and locative oblique complement in *se*-constructions. Conversely, for the man-impersonal *on* and the non-anaphoric third person plural, prepositional phrases belonging to the category *others*, which lacks agentive force, largely predominate. This lower use of agenthood management complements may be attributed to the fact that the man-impersonal *on* and the non-anaphoric third person plural involve an overt human agent in subject position. The need for additional information regarding the possible agent, provided by an oblique complement, may thus be less essential than in the periphrastic passive and *se*-constructions. It also has to be noted that the presence of a *by*-phrase in periphrastic passive is one of the intrinsic and defining syntactic characteristics of the structure. Gast (2015) reports that in Germanic man-impersonals are sensitive to the context, especially when interpreting the impersonal strategy or finding translational equivalents. Our study revealed that the man-impersonal *on* is not the structure to resort to most to contextual elements. Two suggestions can be offered to explain this finding: a different behaviour of French man-impersonals, rather productive in that language, compared to Germanic man-

impersonals or a different limitation of the notion of context. Finally, the mode and the register are again preponderant factors in the variation of context-related parameters.

## 7.5. Portraying the French system

The detailed description of the verb-related, participant-related and context-related parameters supplemented by an analysis of the influence of mode and register allows to compare and draw a complete portrait of the four French passive-like structures studied in this work. Table 24 summarises all the syntactico-pragmatic features and highlights in bold the specificities of each French passive-like structure, based on chi-square calculations.

| | | Periphrastic passive | *Se*-construction | Non-anaphoric third person plural | Man-impersonal *on* |
|---|---|---|---|---|---|
| Verb-related parameters | Tense | **present (66.2%)** | **present (93.2%)** | present (77.1%) | present (72.9%) |
| | Aspect | **imperfective (74.8%)** | **imperfective (95.9%)** | imperfective (88.1%) | imperfective (86.1%) |
| | Verb mode | indicative (80.4%) | indicative (86.3%) | indicative (92.7%) | indicative (87.8%) |
| | Type of verb process | material (50%)<br>relational(17%)<br>mental (12%) | material (56%)<br>relational(19%)<br>mental (11%) | **material (39%)**<br>**relational (37%)**<br>mental (13%) | **material (33%)**<br>**mental (20%)**<br>relational (18%) |
| | Kinesis | **dynamic (92%)** | dynamic (75.3%) | dynamic (68.8%) | dynamic (71.6%) |
| | Realis/irrealis | realis (78.8%) | **realis (91.8%)** | realis (83.5.8%) | **realis (75%)** |
| | Affirmation | affirmative (92.2%) | affirmative (93.2%) | affirmative (89.9%) | affirmative (86.1%) |
| Participant-related parameters | No. of participants | 2+ participants (98.6%) | 2+ participants (95.9%) | **2+ participants (57.8%)** | **2+ participants (50.2%)** |
| | Agent | **defocused (85.5%) + inanimate *others*** | defocused | **animate human (100%)** | **animate human (100%)** |
| | Patient | inanimate (79.6%) | inanimate (95.7%) | **inanimate (66.2%)** | inanimate (82.7%) |
| | Individuation P | non individuated (58.1%) | **non individuated (94.2%)** | non individuated (56.1%) | non individuated (71.3%) |
| | Affectedness P | affectedness (97.2%) | affectedness (94.5%) | affectedness (95.4%) | affectedness 100% |
| Context-related parameters | Spatio-temporal restriction | presence (16.6%) | **presence (28.77%)** | presence (22%) | presence (11.7%) |
| | Prepositional phrase | presence (56.7%) | **presence (71.2%)** | **presence (17.4%)** | presence (20.5%) |

*Table 24. Overview of the distinctive features of French passive-like structures*

A look at Table 24 reveals the following characteristics. Let us first summarise the features related to the verb. When expressing impersonal content, the verb is mainly dynamic and used in present tense, imperfective aspect and indicative mode. It expresses material, relational and mental processes in different proportions according the structure. The verb encodes real events situations and is used in the affirmative. As displayed by Table 24, proportions within these characteristics vary. The periphrastic passive is the construction that resorts the least to present tenses while the *se*-construction relies almost exclusively on these tenses. Similarly, aspect

varies according to the structure as it is correlated with the parameter tense. Verb mode undergoes slightly less variation as the four structures use the indicative mode in more than 80% of the cases. As for the type of verb process, the three most frequent categories are the same in the four structures, though their distribution differs greatly and, as explained in 7.2.4, the subcategories of the types of processes vary. The high frequency of dynamic verbs seems to be a characteristic of the periphrastic passive that displays the highest proportion of such verbs. The *se*-construction is the construction that encodes the most the events and situation in the real world compared to the other three. Finally, the proportion of affirmative clauses does not vary greatly across the four passive-like structures. Let us now turn to the characteristics related to the participants.

The MCA revealed that participant-related parameters have a preponderant role in explaining the variation between the passive-like structures. Our analysis allows to explain how these parameters vary across the four constructions. The structures can be paired in two groups. On the one hand, the constructions based on an impersonal pronoun and, on the other the periphrastic passive and the *se*-construction. The two R-impersonals usually resort to one-participant clauses while this trend is not observed in the other group. While the agent is always an animate human in non-anaphoric third person plurals and man-impersonals, the agent is most of the time defocused in the *se*-construction and the periphrastic passive. When the agent does appear in the periphrastic passive it mainly belongs to the category inanimate *others*. As for the patients, the use of animate patients seems to be a distinctive feature of the non-anaphoric third person plural, setting it apart from the behaviour of the man-impersonal *on*. The individuation of patients is a parameter that, overall, varies greatly across the structures, with the *se*-construction containing the highest proportion of non-individuated patients and the non-anaphoric third person plural the lowest. Affectedness does not differ greatly from one structure to another.

Results show that the elements in the discourse fragments within which the passive-like constructions occur also play a key role. Spatio-temporal restrictions are much more present with the *se*-construction than with the other passive-like constructions. This construction is also the one to resort most to prepositional phrases used for agenthood management. This higher use of contextual clues with the *se*-construction may be explained by the fact that the agent is generally defocused. The presence of elements such as spatio-temporal restrictions and prepositional phrases helps to restrict the possible agents involved in the passive-like structure. When the agent is specified, for instance in the case of a non-specific pronoun in subject position or introduced by the typical agentive preposition *by*, contextual clues may appear as less essential to delimit the event or the situation.

The influence of the mode and the register on the parameters in French passive-like structure is substantial. First, the analysis reveals that within a parameter the influence of either the mode or the register or both can apply to a structure. Moreover, this influence can be noted for the periphrastic passive and not for the man-impersonal *on*, the other way around or affect both structures. This disparity in the impact of mode and register is especially present in verb-related parameters. For instance, the analysis reveals that tenses in man-impersonals are influenced by the mode whereas this influence is not observed in the periphrastic passive. Overall, as for verb-

parameters are concerned, results indicate that mode has a greater impact on the variation between the structures. The influence of register is rather subtler and punctual. There is, however, one exception to this observation. The analysis shows that the type of verb process suffers considerable influence of both mode and register in the distribution of the different types of processes across the data sets, the most varied and distinct set being the formal oral corpus. While mode and register only exercise a minor influence in some verb-related parameters, the picture is different as for participant- and context-related parameters. Regarding participants, the analysis indicates that the number of participants is greatly impacted by both mode and register, one-participant clauses being more frequent in oral productions and informal situations. The presence of an explicit agent and/or a patient as well as their nature undergo an important influence of mode and register. Overt agents, analysed in the light of the periphrastic passive, are more frequent in written data than oral data. Similarly, the use of a patient and the category it belongs to does not only depend on the type of passive-like construction but also on the register and the mode. Results indicate that the periphrastic passive, which combines more frequently with a patient than the man-impersonal *on*, uses much more animate human patients in informal oral data. Conversely, man-impersonal *on* displays more animate patients (human and non-human) in written data. Context-related parameters, the last group of features commented on in this investigation, also differ according to the degree of formality and the written/oral nature of the discourse. While the use of spatio-temporal restrictions receives much more influence from the register, the presence of a prepositional phrase is rather governed by the mode. A last comment has to be made regarding formal oral data which consist of parliamentary debates, a genre considered hybrid due to its written-to-be-spoken nature. In French, three parameters seem to be influenced by this specific type of data: the verb mode, the patient, and spatiotemporal restrictions. The analysis of these three variables shows that passive-like structures in formal oral data display similar behaviour to written data, be it formal or informal. This phenomenon has been especially observed for the periphrastic passive. However, the tendency for formal oral data to resemble written data remains limited in our study. The corpus composition may partly explain this finding since the priority has been given to spontaneous interventions. When an influence is noted it is, in most parameters, either governed by the mode of production or the degree of colloquialism, most of the time creating the two following pairs: formal oral/informal oral vs formal written/informal written or formal oral/formal written vs informal oral/informal written.

Before closing this section, I will briefly discuss the discursive functions of the four passive-like structures surveyed in French. As Neumann (2012: 192) highlights, "speakers may choose from the options provided by the langue system and talk about the same subject matter quite differently, depending on the situation in which they interact." To this statement, I would add that the choice also depends on the communicative intention of the speaker. The analysis in this chapter revealed that French speakers can use at least four options to diminish the salience of an agent. Let us consider example (241) extracted from the French data and (242-244), constructed examples including the other three options to talk about the same subject.

(241) *Le panais **se sème** de février à juin (…)* . (WIKI FR)
'Parsnips **are sown** from February to June (…).'

(242) ***On sème*** *le panais de février à juin (…).*
‘We **sow** parnsips from February to June (…).’

(243) *Le panais* ***est semé*** *[par X] de février à juin (…).*
‘Parnsips **are sown** [by X] from February to June (…).’

(244) ***Ils sèment*** *le panais de février à juin (…).*
‘They sow **parnsips** from February to June (…).’

In example (241), the *se*-construction, accompanied by a present tense and an imperfective form, two typical features of French *se*-construction as highlighted by this study, is used to express that generally parsnips are sown from February to June. The agent is absent, probably because its mention is superfluous and what matters most is the parsnip and the sowing season. Depending on the situation and his/her intention the speaker can opt for a man-impersonal *on*. Compared to (241), in (242) the parsnip is defocused and placed in object or comment position. Though the man-impersonal impedes the immediate recuperation of the agent, by choosing such an option, the speaker maintains the agentive force in topic position. In (243) the short periphrastic passive allows to place the parsnip in topic position, thus placing focus on it. Depending on the nature of the agent carrying out the sowing process and on the situation, the speaker has the possibility of including an overt agent introduced by an oblique complement *by*. Through this process, the short passive becomes a long one. As demonstrated in this study, the addition of such a device with the *se*-construction is very rare. The analysis of the last option, the non-anaphoric third person plural, resembles the one offered for the man-impersonal *on*. The main difference lays in the fact that a non-anaphoric third person plural automatically excludes the speaker from the interpretation of the passive-like structure. To sum up so far, the examination of the different syntactico-pragmatic parameters has allowed not only to provide an overview of the overall characteristics of passive-like structures but also to highlight a series of differences between four agent-defocussing mechanisms in French. This detailed account of the role played by verb-related, participant-related and context-related features in the selection of passive-like structures has helped to yield further insights into the French impersonal system. The next chapter aims at providing a similar account of the Spanish system to further complete the panorama of agent-defocussing mechanisms.

# 8. Passive-like structures in Spanish

This chapter is concerned with the analysis of four passive-like structures in Spanish: the periphrastic passive, the *se*-construction, the non anaphoric third person plural and the construction with *uno*. It shows that these structures, though they may share some proporties, display different relative frequencies and occupy a different position in the Spanish system.

| **Structure** | **Raw frequency** | **Relative frequency (pttw)** |
|---|---|---|
| **Periphrastic passive** | 325 | 15.56 |
| ***Se*-construction** | 968 | 46.36 |
| **Non-anaphoric third person** | 498 | 23.85 |
| ***Uno*** | 5 | 0.24 |

*Table 25. Raw and relative frequencies per ten thousand words (pttw) in Spanish data*

Before entering the core of the analysis, it is important to remind the reader that the three Spanish structures do no appear at the same frequency. Table 25 indicates that the *se*-construction is by far the most frequent structure in Spanish with a raw frequency of 968 occurrences, followed by the non-anaphoric third person plural (498 occurrences), the periphrastic passive (325 occurrences) and the construction with *uno* (5 occurrences). *Uno* is only included for informative purposes but, as this structure only accounts for a very restricted number of occurrences in the present set of data, I will not further comment on it in this section. This variation in the frequency of use explains the fact that, when the breakdown of the data in mode/register components is operated, some data sets gather a rather restricted number of occurrences. This is the case for formal written and formal oral data in non-anaphoric third person plurals, for which information has been left out of the graphs. Similarly, only 6 occurrences of the periphrastic in informal written data have been obtained. They have been included on the graphs but only for informative purposes given the higher frequency of this form in the other subsets.

This chapter is organized around five axes. The first part (8.1) introduces the results of the mutliple correspondences analyses. Then section 8.2 details the parameters related to the verb, section 8.3 those related to participants and section 8.4 those related to the context. The parameters examined in these three sections have been selected on the basis of the general MCA presented in Chapter 6. Since these parameters are the same for the three languages they will form a common ground for comparison. In addition to these parameters, parameters standing out in the language-specific MCA have been added. This is the case, in this chapter, for the parameter *presence of another passive-like structure*, whose analysis will be specific to the Spanish data. As the French and English MCA did not reveal the importance of this parameter in explaining the variation between the structures, the analysis of the presence of another passive-like structure will only be carried out for Spanish data. A brief comment on the use of *uno* is presented in 8.5. Finally, the last section (8.6) recapitulates the findings and aims at portraying the Spanish impersonal system.

## 8.1. Panoramic overview

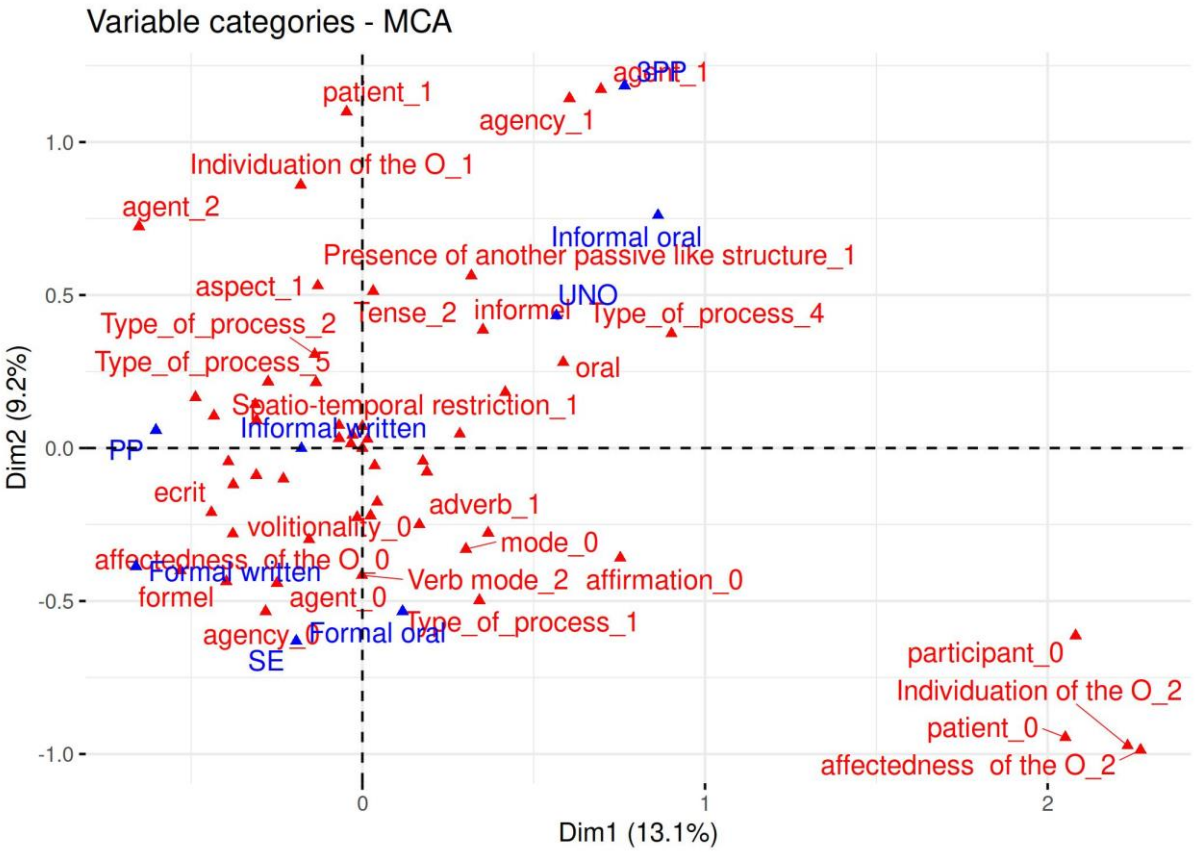



*Figure 18. Distribution of parameters in Spanish data*

Figure 18 shows the dispersion of the different variables (represented by red triangles) – i.e., parameters in the Spanish data, considering the different registers and modes as well as the four impersonal structures studied (in blue). If we look at the plot, one group of parameters clearly stand out. As can be seen, participants (category 0; one-participant clause), affectedness and individuation of the patient (category 2; individuated) and patients (category 0; absence of a patient) are isolated on the right-hand side of the figure. This means that these parameters are essential when explaining the variation between the impersonal structures in Spanish. Along the vertical axis, patient (category 1; presence of a patient) as well as agency (category 1; presence) and agent (category1; animate), situated at the top of the plot, are expected to be associated with the non-anaphoric third person plural and, to a smaller extent, with informal oral data. Individuation of the patient (category 1; individuated), agent (category 2: inanimate) and aspect (category 1; perfective), placed on the left-hand side of the graph are expected to be important parameters for the periphrastic passive. A link between these variables and informal written data is also expected since they are placed close together on the factor map. The dispersion also reveals an association between informal oral data and the presence of another passive-like structure, especially for the non-anaphoric third person plural. The other variables are rather situated near the origin of the plot, thus contributing slightly less to explaining the variation between the impersonal structures.

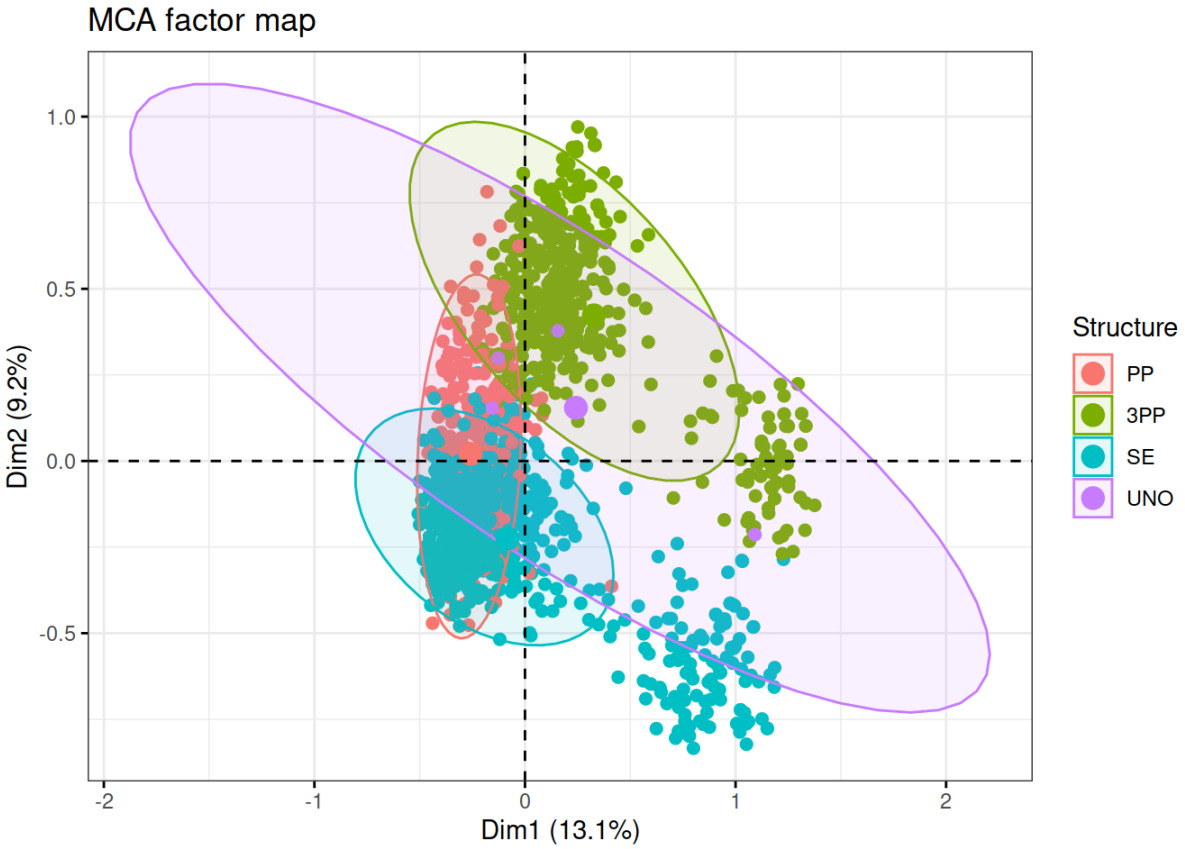



*Figure 19. Distribution of the four passive-like structures in Spanish*

Figure 19 provides us with information on the distribution of the different Spanish impersonal structures and how their characteristics possibly overlap. A series of interesting findings emerge from this Figure. The periphrastic passive (represented by red points on the image) and the *se*-construction partially overlap. This means that the two structures share several features. The projection of this figure on Figure 18 presenting the variables allows to say that this phenomenon is particularly present in written data and, among others, involves the parameter *mode* (category 2). Both structures also appear independently. Part of the occurrences of the *se*-construction are situated towards the bottom right-hand side of the graph while a series of occurrences of the periphrastic passive are situated towards the top left-hand side for the periphrastic passive, though rather close to the origin. When projected on the preceding figure, these areas correspond respectively to the group of parameters associated with participants (for the *se*-construction) and to spatio-temporal restrictions and types of processes (for the periphrastic passive). Figure 19 also reveals that the non-anaphoric third person plural shares a few variables with the periphrastic passive but none with the *se*-construction. Quite a distinct behaviour can thus be expected from the non-anaphoric third person plural. The plot also shows a stronger association between this impersonal structure and the parameters *type of process* and *presence of another passive-like structure.*

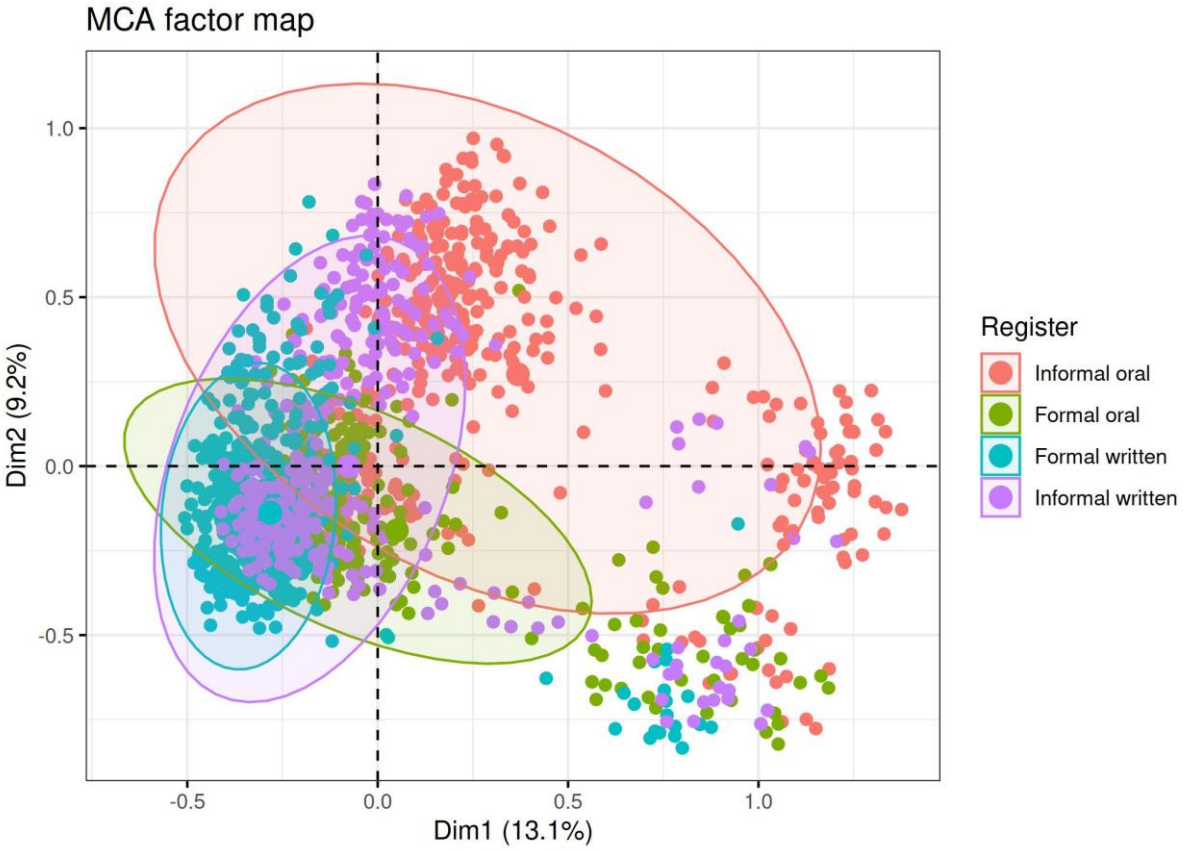



*Figure 20. Factor map for mode and register in Spanish*

Figure 20 visualises how the occurrences of the impersonal constructions are distributed according to the register and the mode. Taking into account Figure 19 with the distribution of the structures, we can observe that that the *se*-construction and the periphrastic passive will be particularly associated with written data as they develop alongside the vertical axis. Formal oral data, represented with green points, mainly contain *se*-constructions. Finally, informal oral data, situated at the bottom right of the figure are associated with non-anaphoric third person plurals.

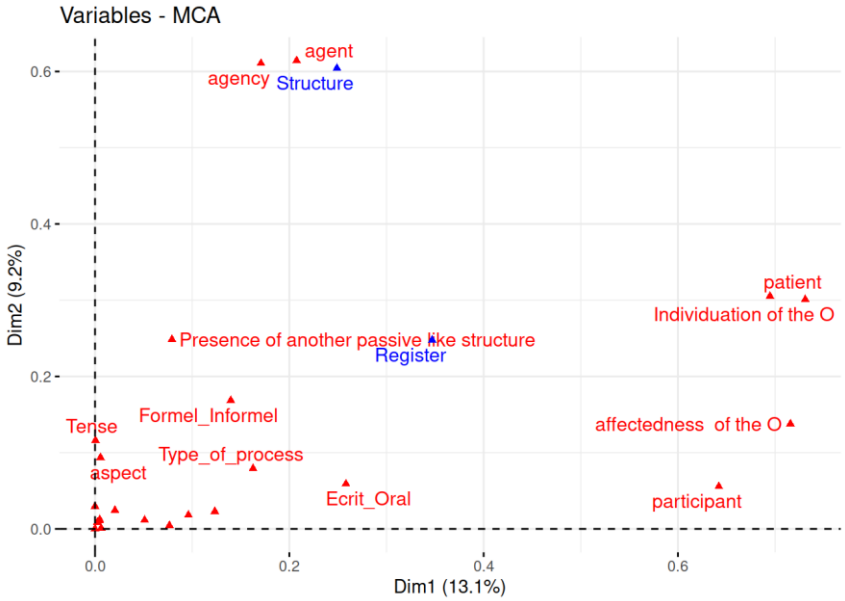



*Figure 21. Correlation(s) in Spanish data*

Figure 21 allows to see whether there is a possible correlation between some parameters within Spanish data. To be considered correlated, two (or more) parameters have to be situated close to each other on the graph and be positioned between 0.5 and 1. The figure shows that the pair *agent* – i.e., the nature of the agent, and *agency* i.e., the presence of an agent, at top of the plot, are likely to be correlated. A link between these two parameters and the type of structure can also be expected. The pair formed by the parameters *patient* and *individuation of the patient* can also be considered as correlated as they are situated at 0.7 alongside the horizontal axis.

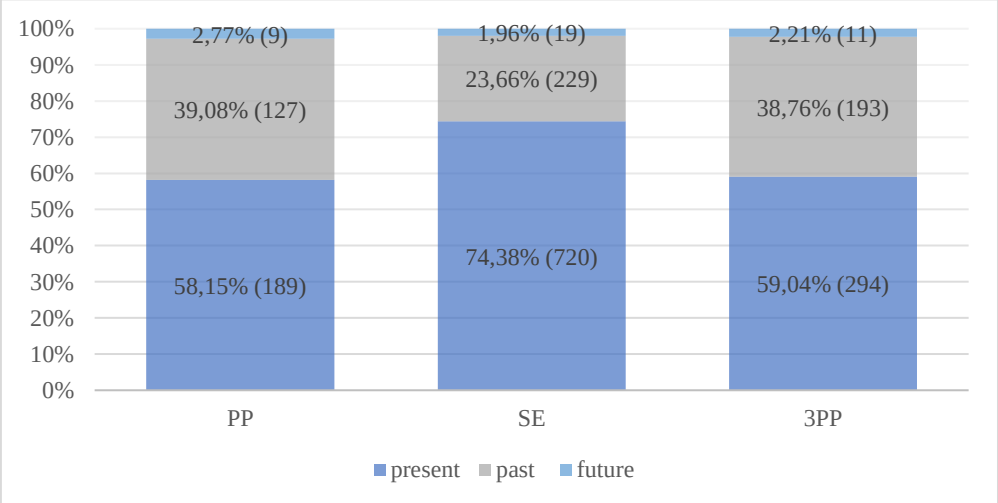
The visualisation of the Spanish data through MCA figures allows to see that, in Spanish, each of the three impersonal structures studied in this work seems to have its area of specialization and is associated with a particular register and mode, though the periphrastic passive and the *se*-construction share some features. A fine-grained description of the parameters will allow to provide more details on the specific features and the behaviour of each impersonal structure.

## 8.2. Verb-related parameters

This section describes the main contributing features related to the verb, obtained from the MCA: *tense* (8.2.1), *aspect* (8.2.2), *verb mode* (8.2.3), *type of verb process* (8.2.4), *kinesis* (8.2.5), dichotomy *realis/irrealis* (8.2.6) and *affirmation* (8.2.7).

### 8.2.1. Tense

The analysis of the parameter *tense*, divided into present, past and future tenses, yields the following results (see Graph 40).




*Graph 40. Distribution of present, past and future tenses in Spanish passive-like structures*

Present tenses predominate in the three structures and, more especially, in the *se*-construction as illustrated in (245) with the verb *se mueve* ('is moved'). Results indicate that this structure uses significantly more present tenses than the other two impersonal structures[32].

(245) *¿Qué tipo de Carga es la que **se mueve** mayormente vía aérea*? (YAHOO SP)
'What type of cargo **is** mostly **moved** by air?'

These results are in line with what was found for French data. And, as explained in Chapter 7, the high presence of present forms may be linked to the fact that the *se*-construction is frequently used to express general norms or generic situations. This is exemplified in example (245) where the speaker is asking about general information on the type of cargo generally moved by air.

The periphrastic passive and the non-anaphoric third person present very similar profiles as far as the distribution of the three categories of tenses is concerned. Both structures use present tenses in respectively 58.1% and 59% of the cases and resort much more to past tenses than the *se*-construction. Examples (246 and 247) illustrate past tenses in a periphrastic passive and a non-anaphoric third person plural construction. In both cases, the verb is used in *pretérito indefinido*, which expresses punctual or temporally delimited actions which started and finished in the past.

(246) *Según la información oficial es que la Luna ya **fue explorada** al 100 porciento (...)* (YAHOO SP).
'According to official information, the Moon **has already been** 100 percent **explored**.'

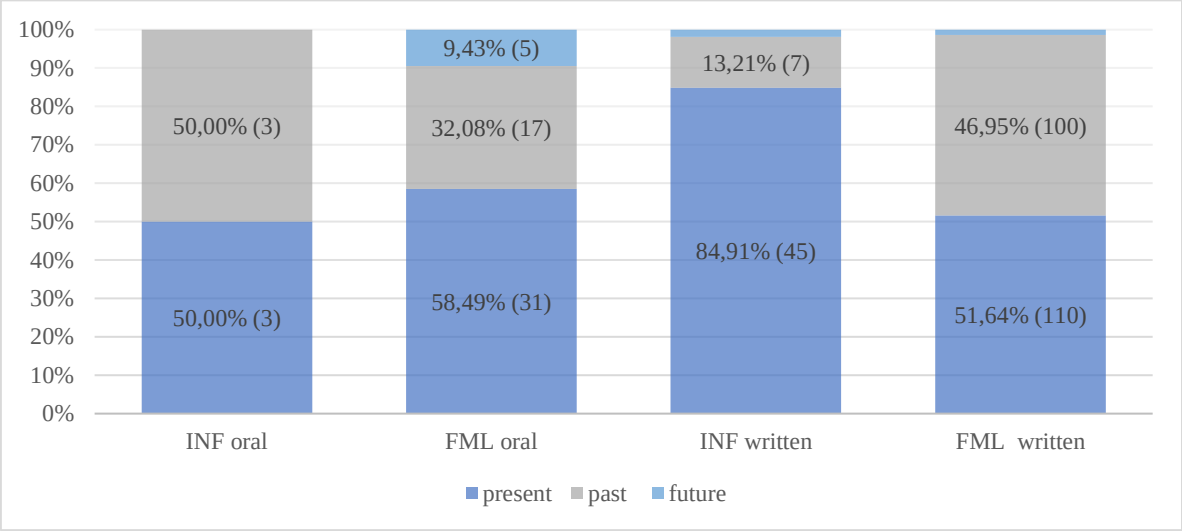
(247) *Al desarrollarse esta raza en Gales se perseguía conseguir a un perro que correteara a las gallinetas arreándolas de sus escondites en los arbustos. Su versatilidad creció y fue cuando **se dieron cuenta** de que este perro era feliz en diferentes condiciones y también era capaz de nadar.* (WIKI SP)
'When this breed was developed in Wales, the aim was to get a dog that would run the moorhens out of their hiding places in the bushes. Its versatility grew and **they realised** that this dog was happy in different conditions and was also able to swim.'

The MCA reveals that the confidence ellipses of the periphrastic passive and the non-anaphoric third person plural partially overlap. The similar behaviour of these two structures regarding *tense* appears as one of their shared characteristics. A last observation reveals that, overall, future tenses only totalize very few occurrences, regardless of the passive-like structure. This means that speakers only use defocussing mechanisms to express themselves on events and situations that are either anchored in the present or that occurred in the past. Situations considered as occurring in the future, thus situated more on the level of eventuality, either use other resources or the nuance of probability is expressed through other mechanisms such as

[32] Chi-square tests produce the following results for the pair *se* vs periphrastic passive : $\chi^2 = 30.6833$, $p < .05$ and $\chi^2 = 36.3028$, $p < .05$ for the pair *se* vs non-anaphoric third person plural.

auxiliaries, negation etc. This will be further examined in section 8.2.6, which addresses the issue of realis/irrealis anchorage.

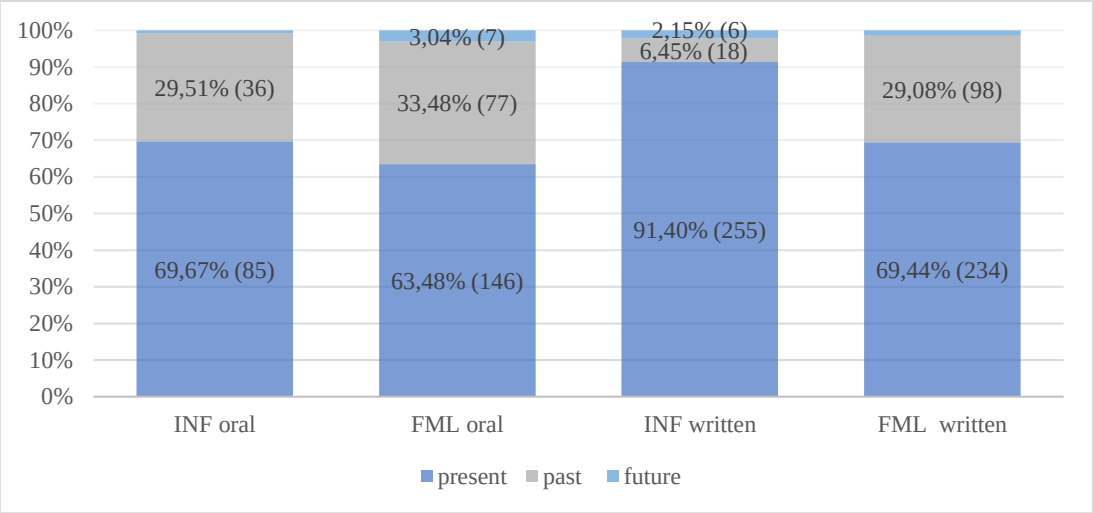
Breaking down the data makes it possible to see whether the mode and the register have a possible impact on the distribution of the categories of tense in the three passive-likes structures.




*Graph 41. Distribution of tenses in Spanish periphrastic passives*

Graph 41 shows that tenses are distributed differently according to the register and the mode. Informal written data resort much more to present tenses than formal written data ($\chi^2 = 18.4979$, $p < .05$). Given the small number of occurrences of periphrastic passives in the informal oral data set, it is difficult to say whether this tendency to resort to numerous occurrences in present tenses is also present in oral data or whether it is particular to informal written productions.

A look at formal oral and formal written data reveals that formal oral data contain far fewer past tenses (32.08%) than formal written data (46.9%). The only factor explaining this difference in percentages is the presence of future tenses in formal data, though in small proportion (9.43%; 5 occurrences), since the difference between the uses of present tenses in the two data sets is not significant. It can thus be suggested that, overall, the use of tenses in periphrastic passives is much more varied in formal oral data than in the other three combinations of mode and register.




*Graph 42. Distribution of tenses in Spanish se-constructions*

Let us now examine the distribution of tenses in the *se*-construction. A first look at Graph 42 seems to reveal an influence of the register, with a higher proportion of present tenses in informal data. This observation is, however, only valid for written data since a chi-square test indicates that there is no significant difference between informal and formal oral data ($\chi^2$ = 1.4223, p > .05). As was the case for the periphrastic passive, *se*-constructions in informal written data contain a relatively high proportion of present tenses (91.4%; 255 occurrences).

As for the non-anaphoric third person plural, no significant difference could be observed between informal oral and informal written data. They resort to present tenses in respectively 57.3% (205 occurrences) and 64.2% (86 occurrences), the rest of the occurrences being almost exclusively used in past tenses.

The influence of register and mode on tenses in passive-like structures can be summarised as follows. The distribution in oral data seems rather stable across formal and informal data. Conversely, in written data, there are far more present tenses in the informal data set. This was observed both for the periphrastic passive and the *se*-construction. This result corroborates Biber's claim that "involved productions", which broadly correspond to informal interactions, (usually associated with a higher degree of colloquialism) contain a higher proportion of present tense verbs (Biber 1995:142). It can thus be assumed that as far as *tense* is concerned, register has a stronger impact than mode and is, therefore, an essential factor to consider when analysing passive-like structures in the light of *tense*.

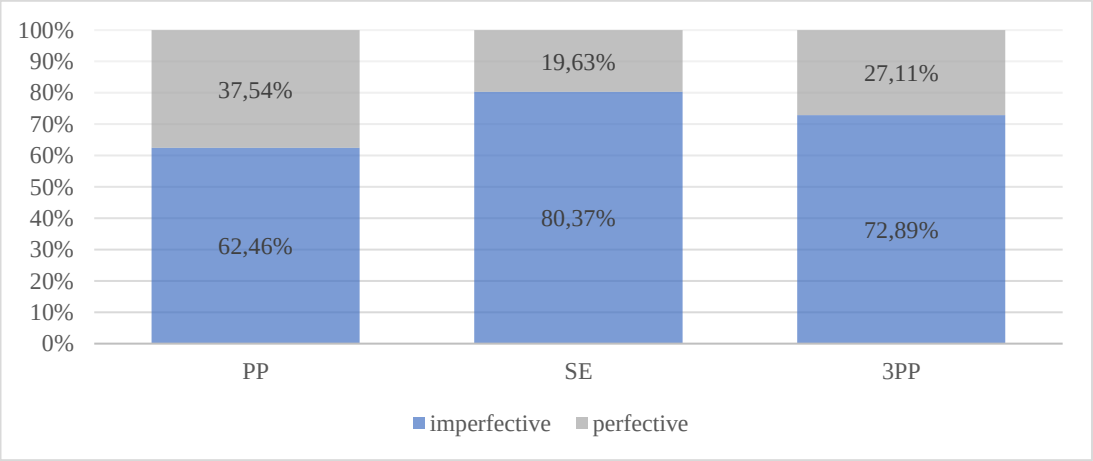
### 8.2.2. Aspect

Aspect and tense are two characteristics often discussed together when analysing syntactic phenomena. This section analyses the distribution of the imperfective and perfective aspect in Spanish passive-like structures as well as potential links with the results obtained for the parameter *tense*.

Compare (248) with (249), which presents a periphrastic passive with an imperfective (248) and a perfective aspect (249). While in (248) the process of knowing is still ongoing, in (249) the exploration can be considered as over.

(248) *El pomelo **es conocido** como toronja,al menos en mi pais.*(YAHOO SP)
'Grapefruit is known as toronja [grapefruit], at least in my country'

(249) *Según la información oficial es que la Luna ya **fue explorada** al 100 porciento (...).* (YAHOO SP)
'According to official information it is that the moon **ha**s already **been explored** 100 percent (…).'

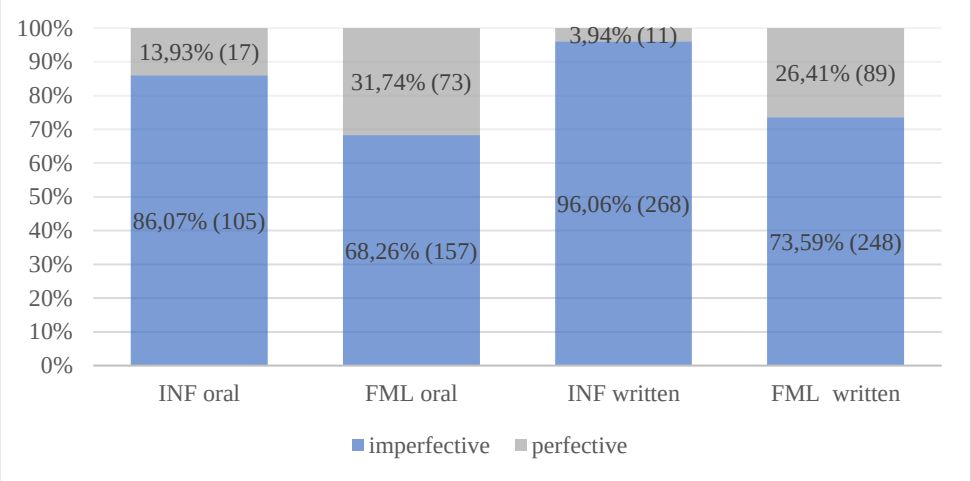



*Graph 43. Distribution of imperfective and perfective forms in Spanish passive-like structures (parameter aspect)*

As shown in Graph 43, the imperfective aspect (illustrated in 236) predominates in the three passive-like structures. However, the proportion varies considerably according the structure. The *se*-construction contains the highest proportion of verbs in the imperfective (80.4%), followed by the non-anaphoric third plural (72.9%) and the periphrastic passive (62.5%). This means that events and situations tend to be presented much more as ongoing, habitual or iterative in *se*-constructions and as completely over in periphrastic passives. The non-anaphoric third person plural occupies an intermediate position between the *se*-construction and the periphrastic passive. The preponderant use of imperfective forms in *se*-constructions compared to the periphrastic passive largely corroborates the results found by Fernández (2007) on Spanish. Since the imperfective aspect is the preferred aspect used in generic statements (Hacquard 2009:280), we can therefore argue that *se*-constructions are better candidates to express generalization while periphrastic passives would rather be employed in specific and episodic events.

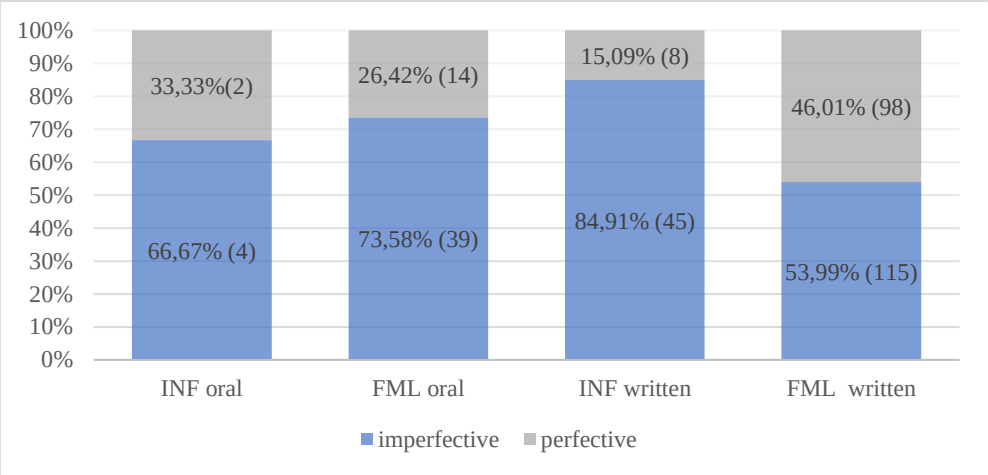
A comparison with Graph 40 presenting the distribution of tenses (see 8.2.1) shows that the high use of the imperfective aspect in the *se*-construction can be correlated with the large proportion of present tenses. A striking result is the fact that, while tenses are distributed rather similarly in periphrastic passives and non-anaphoric third person plurals, this similarity is not observed for *aspect*. The non-anaphoric third person plural uses significantly more the imperfective aspect than the periphrastic passive. An explanation would be a higher proportion of *imperfectivo* in non-anaphoric tenses, classified as a past tense with an imperfective aspect while, the periphrastic passive would resort more to past tenses considered as perfective such

as the *pretérito simple* ('simple past'). *Aspect* appears thus as a key parameter to distinguish the three structures in Spanish and can even be considered as more influential than *tense* in explaining the variation between the three passive-like structures.



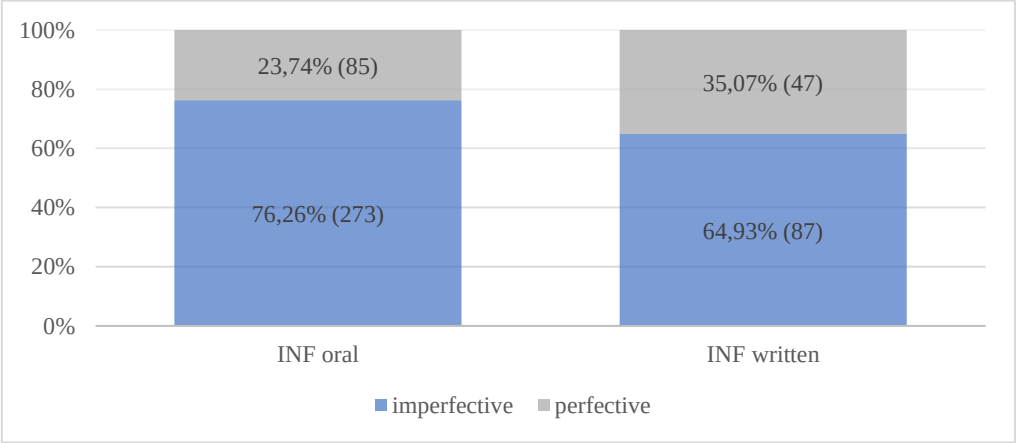

*Graph 44. Distribution of imperfective and perfective forms in the Spanish se-constructions*

As seen above, the *se*-construction is the passive-like structure that displays the highest use of forms with an imperfective aspect. Graph 44 reveals that register considerably impacts the distribution of imperfective aspect since its proportion is significantly higher in informal oral data compared to formal oral data ($\chi^2 = 13.2787$, $p < .05$) and in informal written data compared to formal written data ($\chi^2 = 56.6564$, $p < .05$). It can be assumed that the generalizing function of an imperfective aspect in the *se*-construction is rather specific to informal settings. In addition to register, mode is also an important factor explaining the variation of imperfective and perfective aspects in the *se*-construction. A closer look at Graph 44 shows that there is slightly more clauses in the imperfective aspect in written data than in oral data. This observation holds true in informal data ($\chi^2 = 13.0475$, $p < .05$) but not in formal data ($\chi^2 = 1.9027$, $p < .05$). It can be concluded that *se*-constructions receive a double influence, from the register and, to a smaller extent, from the mode though this observation is only valid for informal data.




*Graph 46. Distribution of imperfective and perfective forms in Spanish periphrastic passives*

The analysis of *aspect* identifies register as an important factor explaining variation in periphrastic passives. As shown in Graph 45 informal written data contain a high proportion of clauses in the imperfective (84.9%; 45 occurrences) while the use of the imperfective and perfective aspect is much more balanced in formal written data (54%; 115 occurrences and 46%; 98 occurrences, respectively). The difference between the frequency of use of imperfective periphrastic passives in formal and informal written data is significant for $\chi^2 = 16.922$, $p < .05$. Aspect seems to display a more stable behaviour across oral data although the number of occurrences in informal oral data does not allow to draw firm conclusions. A look at formal data reveals that formal oral data resort to more imperfective clauses than formal written ones ($\chi^2 = 6.6843$, $p < .05$).




*Graph 45. Distribution of imperfective and perfective forms in Spanish non-anaphoric third person plurals*

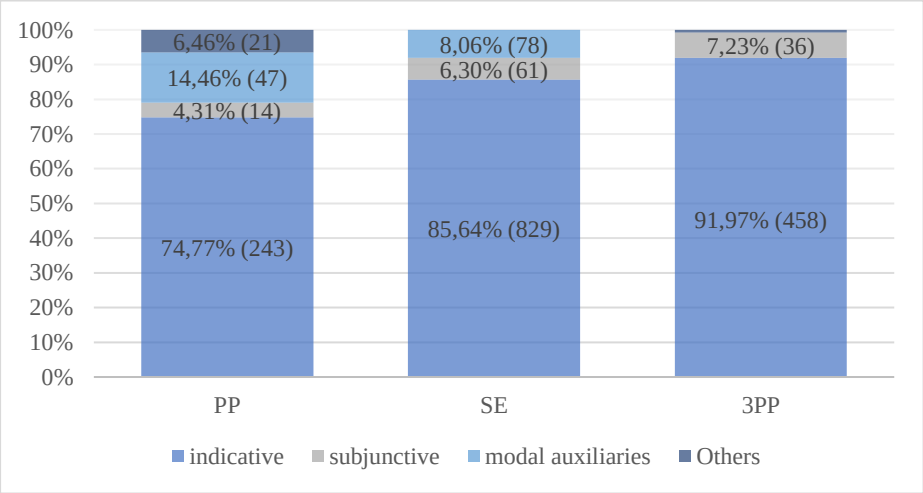
The picture is rather different as for the non-anaphoric third person plural (Graph 46). While for the *se*-construction, results indicate a significantly higher use of imperfective constructions in informal written data, the opposite can be observed in non-anaphoric third person plural. Informal oral data use significantly more imperfective clauses than informal written data ($\chi^2 = 6.3776$, $p < .05$).

In sum, the three passive-like structures display quite different behaviour regarding the parameter *aspect*. In line with what was found in the literature for Spanish, the *se*-construction uses more imperfective clauses than the periphrastic passive. This work allows to bring further

information as for aspect in passive-like structures since it shows that the non-anaphoric third person plural occupies an intermediate position between the other two structures. It also shows the influence of mode and register within each structure, informal written data appearing as better candidates for imperfective situations in *se*-constructions and periphrastic passives while in non-anaphoric third person plurals informal oral data are the ones with the highest percentage of imperfective clauses. The mode of production and the degree of formality are thus key variables to include in the analysis of *aspect.*

### 8.2.3. Verb mode

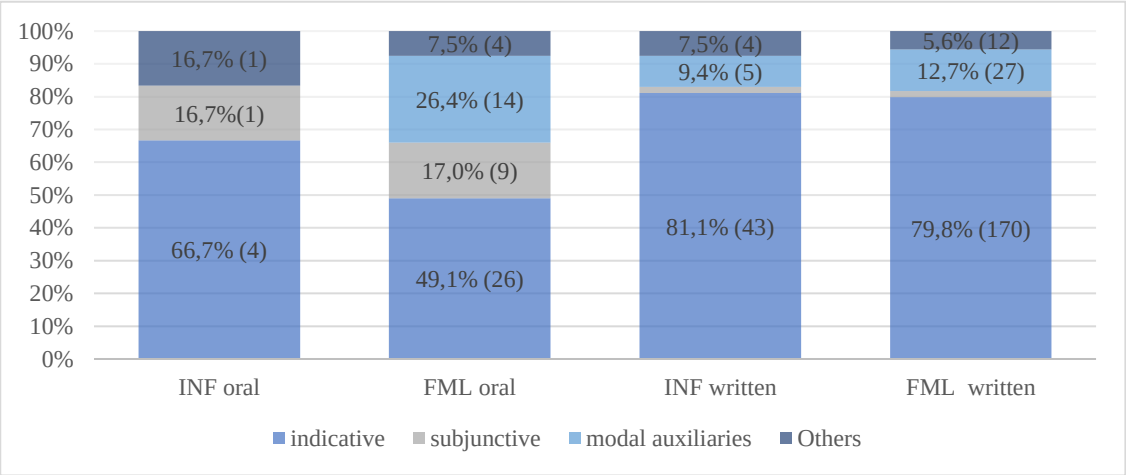
Verb mode is another syntactic parameter that may explain the variation between the passive-like structures. Verb modes are classified in four categories: indicative, subjunctive, modal auxiliaries and others.




*Graph 47. Distribution of verb modes in Spanish passive-like structures*

Graph 47 presents the proportion of the four categories of verb modes in the three passive-like structures. The results indicate that the indicative mode is clearly preferred, though in significantly different proportions. The non-anaphoric third person plural is the structure that resorts the most to it (92%; 458 occurrences) and the periphrastic passive the least with 74.8% (243 occurrences). The *se*-construction occupies an in-between position with 85.6% (829 occurrences) of indicative verbs. The fairly small proportion of indicative mode found in the periphrastic passive can be explained by the relatively high presence of modal auxiliaries and, to a smaller extent, to the use of the subjunctive and of occurrences classified as *others.* Whereas verbs in the subjunctive are found in the three passive-like structures in approximately the same proportions, the presence of the category *others* and the rather high proportion of modal auxiliaries seem specific to the periphrastic passive.

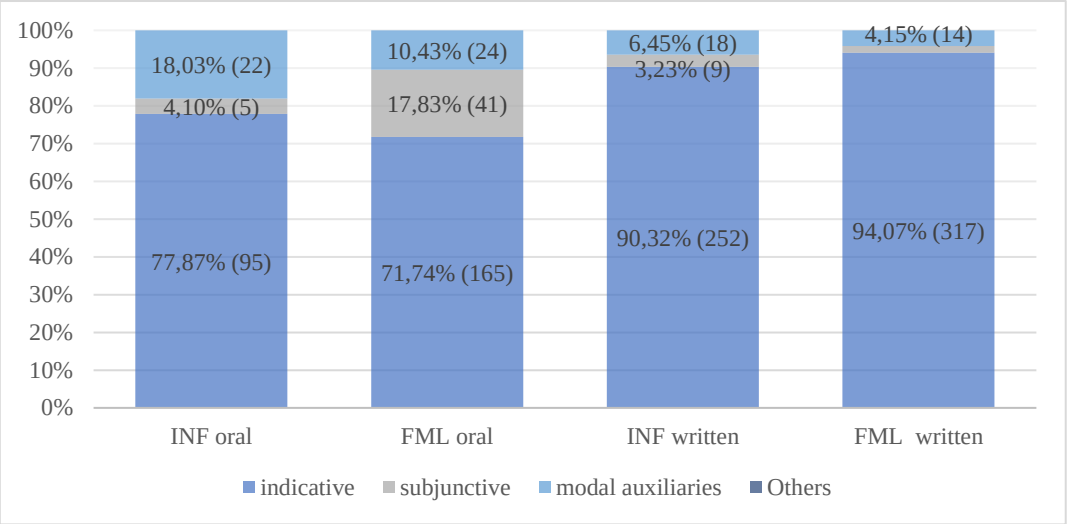
Biber (1995) lists modals as one of the linguistic characteristics likely to vary across registers. Based on these results, I hypothesize that the parameter *verb mode* varies across register and mode.




*Graph 48. Distribution of verb modes in Spanish periphrastic passives*

As presented above, verb modes are distributed differently across register and mode (Graph 48). A first distinction can be drawn between oral and written data. While informal and formal written data display a fairly similar behaviour, informal oral and, even more, formal oral data appear as strikingly different. In formal oral data, the indicative mode hardly reaches the 50% threshold, thus giving much more importance to the subjunctive, modal auxiliaries and the category *others*.

Yus (2001) reports that English written discourse is usually associated with a higher proportion of modal auxiliaries. The picture obtained in our Spanish data is much more contrasted depending on the specific structure. For the periphrastic passive, results indicate that both the informal and formal written data sets contain modal auxiliaries. Though they are absent in informal oral data, which would corroborate Yus (2001)'s findings, our formal oral data set includes 26.4% (14 occurrences) of elements classified as modal auxiliaries. The relatively high use of modal auxiliaries in formal oral data, which makes them share characteristics of written data, can be explained by the fact that formal oral data were extracted from parliamentary debates, a genre that may contain some features of written productions. Similar results were found by Biber (2006) for English regarding the use of modal verbs in spoken university registers, genres that are considered as having a high degree of formality.




*Graph 49. Distribution of verb modes in Spanish se-constructions*

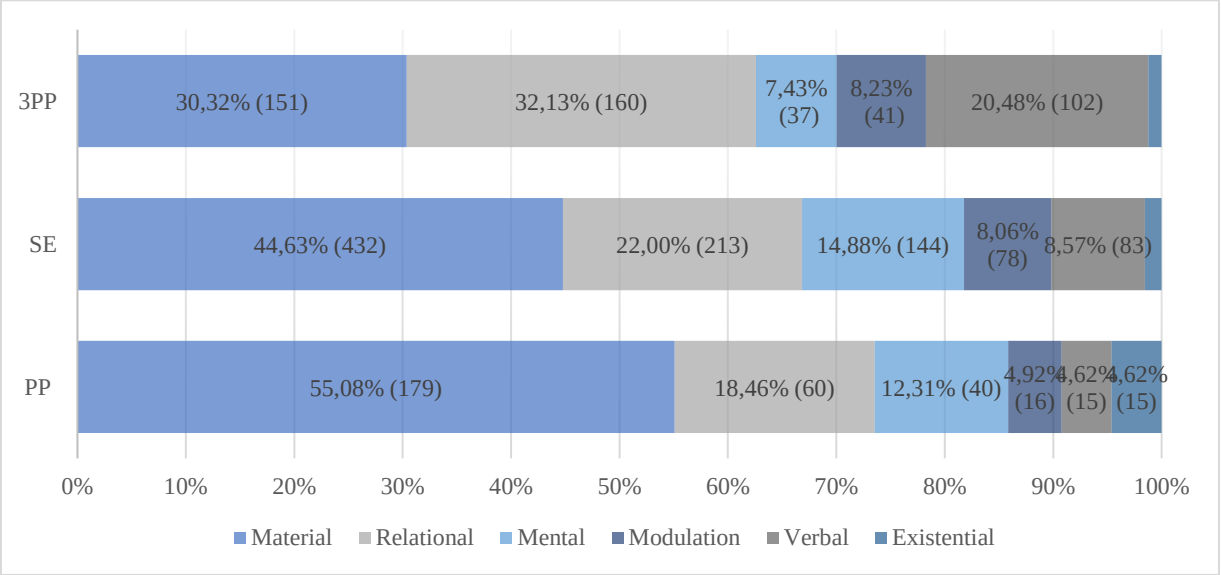
A look at verb modes in *se*-constructions reveals again that mode and register are at play when explaining variation (Graph 49). Written data (both formal and informal) contain almost exclusively occurrences of *se*-constructions in the indicative. This trend is not observed in oral data, much more varied as for verb modes. An interesting result is the relatively high proportion of subjunctives in formal oral data. Such proportion of subjunctive is observed in none of the other data sets. The subjunctive allows to express nuances such as doubts, possibility or irreality. In using such devices, the speaker involves himself/herself in a more cautious way than using the indicative mode, much more anchored in reality. This high presence of subjunctive in formal oral data may partly answer the higher need of cautiousness when taking part in parliamentary debates, as well as by a greater presence of complex clauses including subordinates, which are a syntactic context in which subjunctives often appear.

Concerning the non-anaphoric third person, results indicate that the parameter verb mode does not vary since both in formal and informal written data the indicative mode largely predominates (93.3%; 125 occurrences and 91.9%; 329 occurrences).

The distribution of verb modes is not a straightforward matter. On the one hand, it varies considerably according to the passive-like structure and, on the other, it is also crucial to consider the register within which the passive-like structure appears as well as the oral versus written nature of the discourse production.

### 8.2.4. Type of verb process

Spanish passive-like structures can select different types of verb process among the following ones: *mental* process, *relational* process, *material* process, *verbal* process, *existential* process and *modulational* process. Although, as pointed out in the literature, the periphrastic passive and the *se*-construction present a series of similarities, a detailed examination of the results coupled with statistical tests reveals that each of the three structures display a rather specific profile.




*Graph 50. Distribution of verb processes in Spanish passive-like structures*

A glance at Graph 50 reveals that, overall, two types of verb processes stand out: *material processes* and *relational processes*. The two categories are, nevertheless, used in relatively different proportions. Let us first examine *material process* since Graph 50 indicates that, except for the non-anaphoric third person plural, it is the most common type of verb process in Spanish passive-like structures. Material processes can be divided in four subcategories: *space*, *change*, *behaviour* and *others*. The periphrastic passive is the structure that resorts the most to material processes (55.1%; 179 occurrences). Out of the four categories, occurrences of material processes are more or less equally distributed across *space (*43 occurrences) (Example 250), *change* (61 occurrences) (Example 251) and *others* (52 occurrences) (Example 252). Verbs related to behaviour are much less represented (23 occurrences). The next three examples illustrate the category *space*, *change* and *others.* In (250) the cows suffer a change of localization – i.e., from the farm they are sent to the slaughterhouse. The subcategory *change* involves processes of creation, destruction or modification such as in (251), where the engines are created, designed by someone. The category *others* is rather heterogeneous and gathers processes such as *use*, *control*, *contact* or *activities.* The verb *utilizar* ('to use') in (252) illustrates a case of material process belonging to the subcategory *use.* Products are manipulated and employed for a specific purpose.

(250) *Son inseminadas una vez al año durante cuatro años, fecha tras la cual su producción de leche desciende y* ***son enviadas*** *al matadero al igual que pasa con sus hijos los terneros* (WIKI SP).
'They are inseminated once a year during four years, after which their milk production drops and they are sent to the slaughterhouse as are their calves.'

(251) *Estos motores* ***fueron diseñados*** *por el Prof. Dr. Ing. Karl Leist para aumentar considerablemente la potencia con un consumo reducido de combustible*. (WIKI SP)

‘These engines **were designed** by Prof. Dr. Ing. Karl Leist to increase considerably power with reduced fuel consumption.’

(252) *También, por supuesto, que haya esos 20 millones extra en la dotación a la leche porque, claro, no podría ser de otra manera: ampliamos el dinero porque ampliamos los productos que puedan* ***ser utilizados****.* (PROCEP SP)
‘Also, of course, that there is that extra 20 million in the milk envelope because, of course, it could not be otherwise: we increase the money because we increase the products that **can be used**.’

Material processes are also the most frequent verbs in *se*-constructions (44.6%; 432 occurrences). However, a statistical test indicates that these verbs are used significantly less in *se*-constructions compared to periphrastic passives ($\chi^2 = 10.6578$, $p < .05$). This means that *se*-constructions are involved in different semantic patterns than the periphrastic passive. While the periphrastic passive relies largely on material verbs which, most of the time can be linked to dynamic situations, this link is weaker with *se*-constructions.

When used with material processes two subcategories predominate in the *se*-construction: *space* (149 occurrences) as in (253) and *change* (152 occurrences) as in (254). In (253) the specimens are taken away from the creosoted and thus experience a change in their location. In (254) the verb *soluciona* (‘solved’) involves a change in the situation, where possible solutions are found to improve a situation.

(253) *Los especimenes de esta especie* ***se han removido*** *de la madera creosotada y el preservante se puede exprimir literalmente de sus cuerpos, con todo eso continúan atacando la madera.* (WIKI SP)
‘Specimens of this species **have been removed** from the creosoted wood and the preservative can literally be squeezed out of their bodies, yet they continue to attack the wood.’

(254) *Esto* ***no se soluciona*** *solamente hablando de pirómanos, como hicieron algunos de los diputados aquí hoy, o como hizo la ministra de Agricultura española, la señora García Tejerina.* (PROCEP SP)
‘This **cannot be solved** just by talking about arsonists, as some of the Members here today did, or as the Spanish Minister for Agriculture, Mrs García Tejerina, did.’

The categories *others* (109 occurrences) and *behaviour* (22 occurrences) are slightly less used. As far as the subcategories of material processes are concerned, it can be noted that the periphrastic passive and the *se*-construction display a similar profile.

The non-anaphoric third person plural is the passive-like structure that includes the smallest percentage of material processes (30.3%; 151 occurrences). Within this category, verbs belonging to the sub-category *space* predominate (70 occurrences). Example (255) illustrates a material verb classified as *space*, where calves are kept in a place away from the others animals.

(255) *los terneritos son aislados de sus madres apenas nacen para darle la leche a la industria y ustedes la consuman.....a los terneros machos, los **aislan** y los dejan sin comida ascinados hasta que mueran porque no sirven para nada o los venden para ser consumidos como carne blanda (YAHOO SP)*
'calves are isolated from their mothers as soon as they are born to give milk to the industry and you consume it..... male calves, they **isolate** them and leave them without food until they die because they are useless or are sold to be eaten as tender meat.'

Up till now, we have seen that material processes are the most frequent type of verb process in the periphrastic passive and the *se*-construction as well as an important category of verbs in the non-anaphoric third person plural. Let us now turn to relational processes, which appear as the most frequent category in non-anaphoric third persona plural and the second most represented one in the periphrastic passive and the *se*-construction. Relational verbs are either verbs of *possession* or *attribution.*

As for the non-anaphoric third person plural, relational verbs are slightly more frequent than material processes as they totalize 32.1% (160 occurrences) of the data. Results show that relational verbs in our data almost exclusively pertain to the category *possession* (147 occurrences out of 160). Verbs of *possession* as illustrated in (256), usually imply a transfer from an agentto a patient. In (256), the quinces experience a transfer from the sellers to the buyers, namely the clients on the market, which become the new owners of the fruits.

(256) *H2 [¿tie]nes membrillos? / aah es que el otro día*
*HI no / no- no- es que ni me= ni me acuerdo porq-*
*H2 [el otro día] **estaban vendiéndolos** en el mercado*
*HI pues / <enf>si quieres</enf> / en la época de los membrillos / recuérdamelo*
(ESLORA SP)
'H2 do you haves quince? / aah it's that the other day
HI no / no- no- no- it's that I don't even remember because-
H2 [the other day] **they were selling** them at the market.
HI well / <enf>if you want</enf> / at the time of the quinces / remind me.*'*

It can be suggested that, given this high proportion of verbs of possession, non-anaphoric third person plural forms are good candidates of two-participant clauses, as will be further developed in Section 8.3. In the periphrastic passive and the *se*-construction, relational verbs account for 18.5% and 22% of the data, with no significant difference between the two constructions ($\chi^2$ = 1.8333, p > .05). Within these constructions, verbs of possession and attribution are more or less equally distributed, though verbs of possession are slightly more used than verbs of attribution. These findings point toward the use of verbs of possession as a specific feature of the non-anaphoric third person plural.

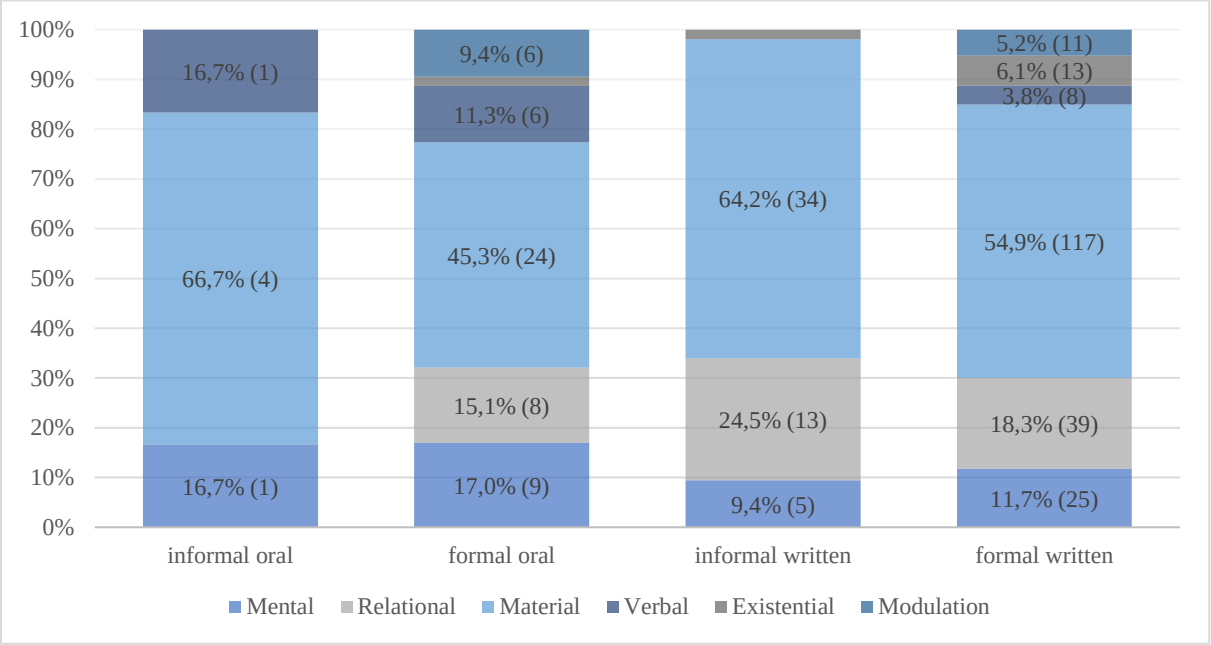
Another specificity of the non-anaphoric third personal plural is its large use of verbal processes (20.5%; 102 occurrences), a phenomenon absent from the other two passive-like structures. Within verbal processes, the subcategory *communication* largely predominates (100 out of 102 occurrences) (see example 257 with the verb *dicen* 'they tell').

(257) *entre otras cosas, en el curso te **dicen** a ke velocidad debes despegar y aterrizar el avion ke te asignaron, a cuantos grados debes girar los flaps segun el clima y la pista, o ke velocidad de crucero debes mantener. en un determinado viaje.. osea la informacion basica suerte! (YAHOO SP)*
'among other things, in the course **they tell** you at what speed you should take off and land the plane you have been assigned, how many degrees you should turn the flaps depending on the weather and the runway, or what cruising speed you should maintain on a given trip... that is the basic information, good luck!'

While verbal processes are characteristic of the non-anaphoric third person plural, mental verbs are rather used with the periphrastic passive and the *se*-construction (respectively 12.3%; 40 occurrences and 14.9%; 144 occurrences). Such proportion of mental verbs is not observed in the non-anaphoric third person plural and thus constitutes a considerable difference between the passive-like structures in Spanish. In both the periphrastic passive and the *se*-construction the subcategory *cognition* predominates.

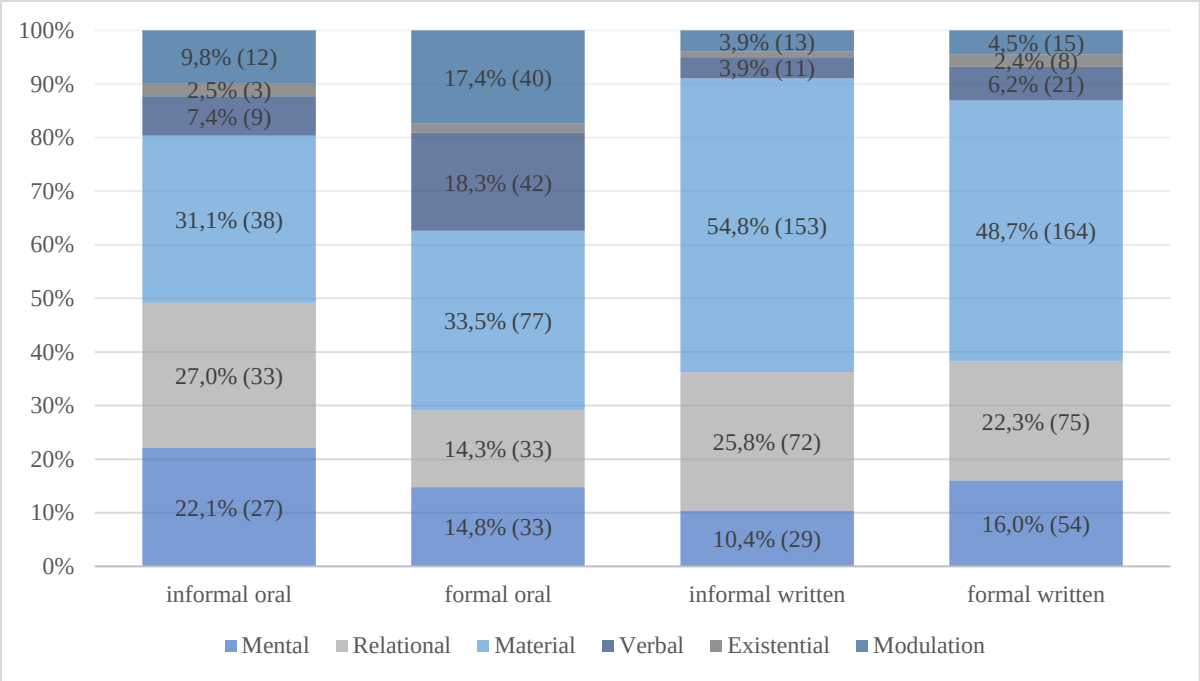
Modulation and existential types of verb processes are the least frequent in the three passive-like structures.

To sum up what was observed so far, it can be suggested that the periphrastic passive and the *se*-construction rely on the same types of verb processes but in slightly different proportions, especially regarding *material* processes. Conversely, the non-anaphoric third person plural differs considerably from the other two passive-like structures. While material processes constitute the major category in periphrastic passives and *se*-constructions, two main types of verb processes dominate in non-anaphoric third person plurals: *relational* and *material* processes. Moreover, it is the only structure to resort to such a high number of verbal processes. Interestingly, the analysis of the type/token ratio makes it possible to say that, even though the periphrastic passive and the *se*-construction appear to rely on similar types of verb processes, they display quite a different type/token ratio (see Chapter 6). Our findings point to a smaller lexical density in the *se*-construction. It is assumed that the higher frequency of the construction can explain this result since we hypothesised that the more common the linguistic phenomenon, the smaller lexical density it has. In what follows, I will complement these results with information on the impact of the mode and the register.




*Graph 51. Distribution of verb processes in Spanish periphrastic passives*

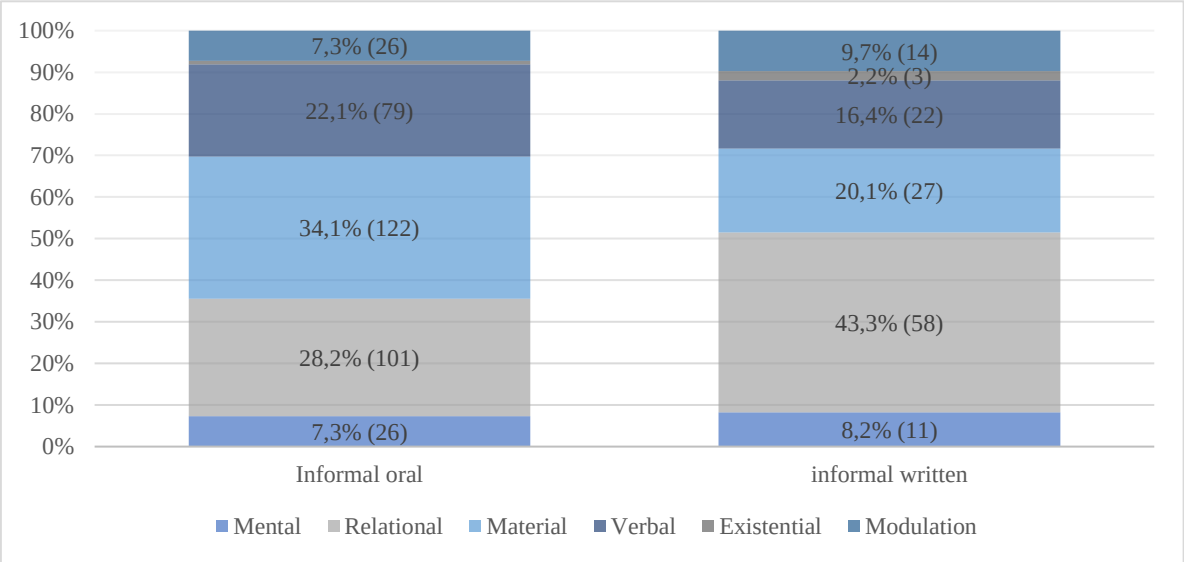
As indicated in Graph 51, the distribution of types of verb processes differs according to the mode and the register. Informal written data contain much more material processes than formal written and than formal oral data. In addition, informal written data is the data set that includes the smallest number of types of verb processes. Compared with informal written data, formal written data use more 'less represented' categories – i.e., verbal, existential, and modulation processes. This ample variety of types of verb processes, though in restricted proportions, in formal data maybe be associated with the fact that the formal written mode is usually syntactically more elaborate. It can also reflect a different perspective adopted by speakers to represent the world. While in informal written data speakers will rely on material verbs, including notions of *space*, *changes*, formal written data include more mental, verbal, existential and modulation processes, linked with more cognitive and abstract levels of representation of events and situations. Formal oral data seem to have its own features since this combination of mode/register can be characterized, in addition to material processes, by an important quantity of mental, verbal and modulation verbs. As few occurrences were obtained for informal oral data, no further observation can be made. As we will see, the picture is somehow different for the *se*-construction.




*Graph 52. Distribution of verb processes in Spanish se-constructions*

Graph 52, illustrating the feature *verb process* in the *se*-construction, reveals a rather similar profile for informal written and formal written data as the graph concerning the periphrastic passive suggesting that the two constructions have a similar use of types of verbs in written data. Statistical tests reveal that the main difference concerns the slightly higher proportion of mental verbs in formal written data ($\chi^2 = 4.149$, $p < .05$). This suggests a slight influence of the register within written data. The picture is far more contrasted as for oral data. While a glance at written data shows a preference for material processes, this is not the case in oral data. Oral data, be it formal or informal, are much more heterogeneous. This first observation leads us to consider mode as having a preponderant role in the distribution of the types of verb processes. Second, oral data undergo a strong influence of register. Informal data gather much more mental and relational verbs than formal data while these latter include a greater proportion of verbal and modulation processes. Informal and formal oral data both show their own preferences. This rather important use of verbal and modulation processes confirms what was noted in the periphrastic passive, whose more limited number of occurrences did not allow, at that stage, to offer strong conclusions as regards the use of these two types of processes.

As regards the influence of register on the distribution of the different types of processes in the non-anaphoric third person plural, Graph 53 suggests what follows.




*Graph 53. Distribution of verb processes in Spanish non-anaphoric third person plurals*

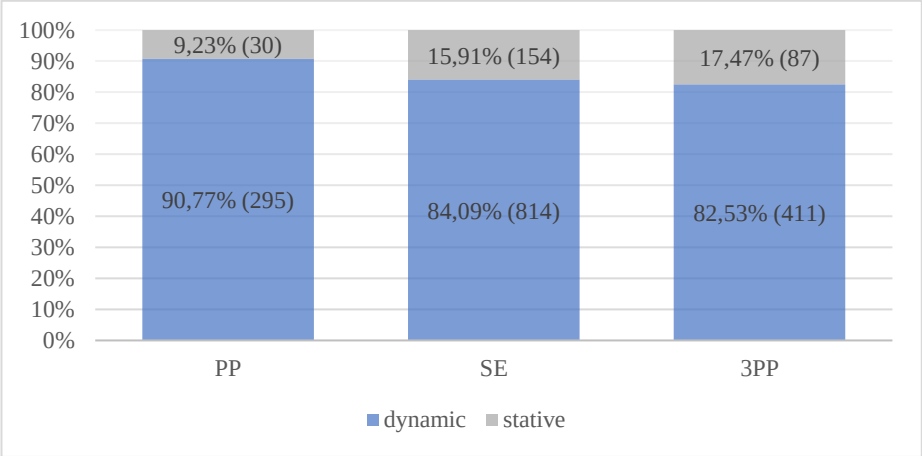
Informal oral and written data share a common background but differ on several other points. The main diverging points are the higher proportion of relational verbs in informal written data (43.3%; 58 occurrences compared to informal oral ones (28.2%; 101 occurrences) as well as the smaller percentage of material verbs in informal written (20.1%; 27 occurrences) compared to oral ones (34.1%; 122 occurrences). This rather small proportion of material verbs in informal written data was not found in the other two passive-like structures analysed previously. Graph 56 also shows a slightly higher use of verbal processes in informal oral (22.1%; 79 occurrences), however, this difference is not significant. This leads us to conclude that register influences the distribution of verb processes in non-anaphoric third person plurals, and more especially, impacts the proportions of material and relational verbs.

In sum, verb processes are distributed differently in the three passive-like structures though a series of similarities could be observed as for the periphrastic passive and the *se*-construction (see for instance the category of verb processes involved in the two constructions). The main differences can be summarised as follows. Material and relational processes represent the two major types of processes in these two structures, though the *se*-construction displays a slightly smaller number of material verbs and higher number of relational verbs than the periphrastic passive. The picture obtained for the non-anaphoric third person plural differs substantially from the results obtained for the other two passive-like structures. In non-anaphoric third person plurals, the proportion of material and relational verbs is balanced and verbal processes occupy a much more important space.

Differences in the influence of mode and register could also be observed. While an overall high proportion of material verbs could be observed in written data for the periphrastic passive and the *se*-construction, this was not the case for the non-anaphoric third person plural. Differences of register could also be identified, especially in *se*-constructions used in oral productions.

### 8.2.5. Kinesis

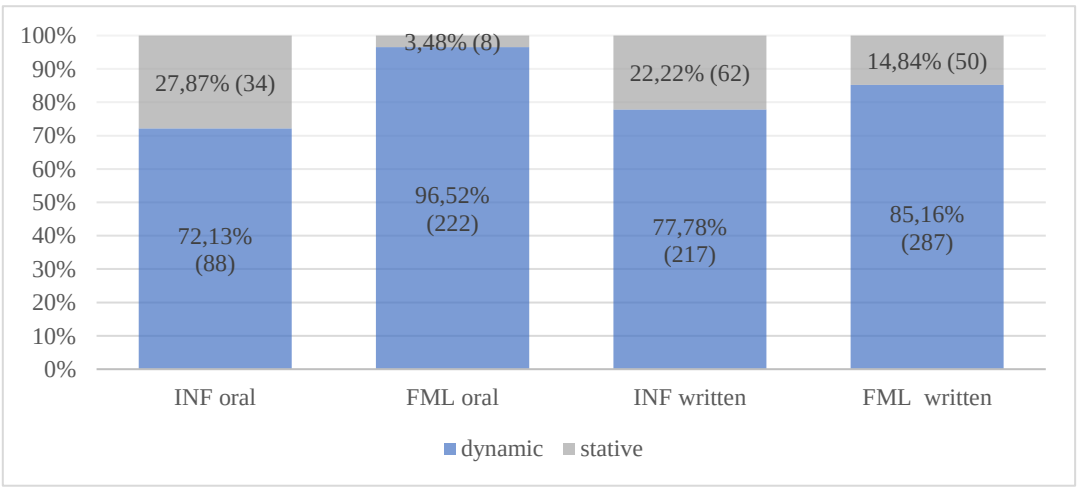
General MCA revealed that the dynamic versus stative nature of the verb may enter at play when explaining the variation between the different passive-like structures in the three languages. According to the MCA specific to Spanish, the parameter kinesis does not appear as a major contributing factor to explain the difference in Spanish however, it has been included in this part in order to be able to compare the results obtained for French with those from the other languages. For Spanish, our data suggest the following results (Graph 54).




*Graph 54. Distribution of dynamic and stative verbs in Spanish passive-like structures*

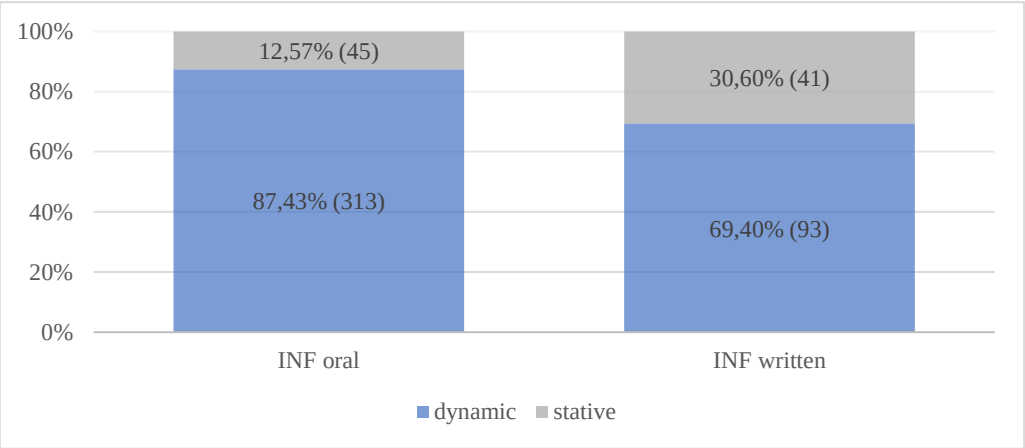
Results show then that in Spanish dynamic verbs largely predominate in passive-like constructions. The difference in the proportions is significant between the periphrastic passive and the *se*-construction ($\chi^2 = 8.8908$, $p < .05$) but not between the *se*-construction and the non-anaphoric third person plural. This high use of dynamic verbs means that passive-like structures are mainly used to defocus the agent in situations involving actions rather than states. While Collins (2012) in his study on grammatical variation in English suggested that the passive is typically used with a dynamic verb, our study allows to further extend his observation to Spanish and to French as already explained in Chapter 7.

Breaking down the data into the four combinations of mode and register reveals the following.




*Graph 55. Distribution of dynamic and stative verbs Spanish se-constructions*

In *se*-constructions there are substantially more dynamic verbs in formal data than in informal data (Graph 55). This would mean that, in informal interactions, speakers are more prone to exchange about inner states, conditions, emotion, attitude etc. No clear pattern emerged as for the influence of the mode.




*Graph 56. Distribution of dynamic and stative verbs in Spanish non-anaphoric third person plurals*

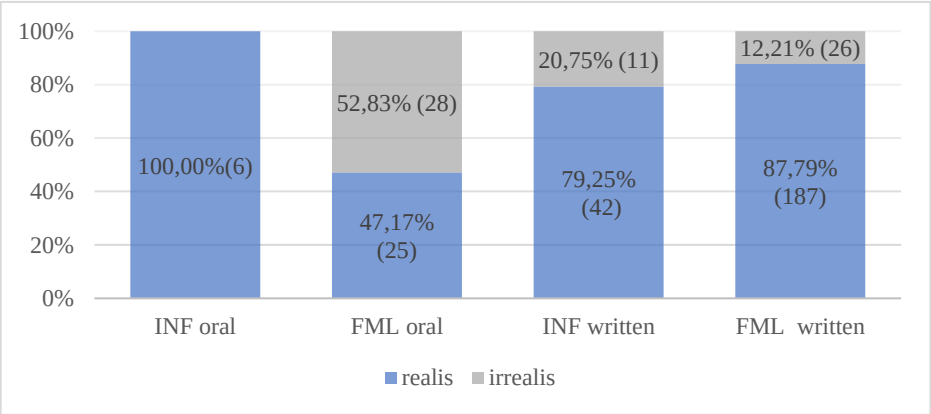
As for the non-anaphoric third person plural, results indicate that informal written data display much more stative verbs than informal oral ones (Graph 56). This result does not match the pattern obtained for the *se*-construction. A possible explanation is that while the non-anaphoric third person plural is typically used in oral contexts (72%, 360 occurrences out of 498) the use of the *se*-construction is less associated to a specific register or mode. As a consequence, it may be assumed that, when the non-anaphoric third person plural is used in another context than its typical one, it modifies its characteristics.

Finally, the periphrastic passive displays a stable behaviour across the four data sets with approximately an overall 10% of stative verbs.

In sum, a more detailed examination of the parameter *kinesis* indicates that the differences between the three passive-like structures are slightly more limited than what has been observed so far for the other verb-related parameters. This confirms that, as can be seen in the MCA, kinesis has a smaller impact in contributing to explain the variation between the three passive-like constructions. The more fine-grained analysis, nevertheless, revealed a distinct behaviour of according to mode and register for the *se*-construction and the non-anaphoric third person plural.

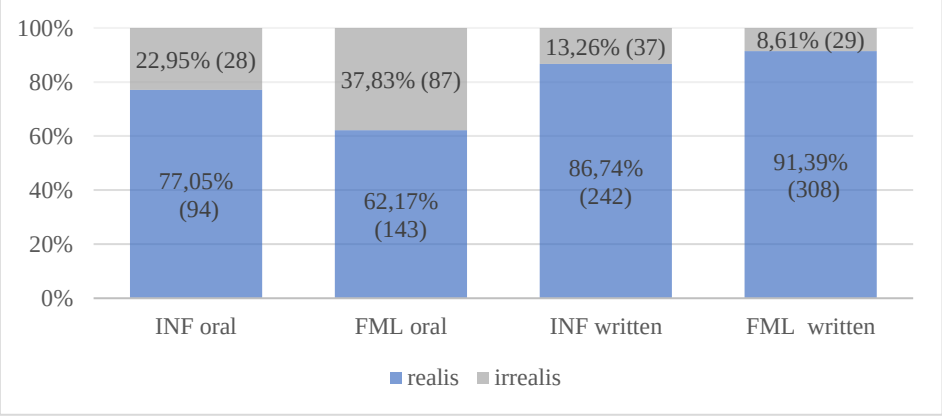
### 8.2.6. Realis/irrealis dichotomy

Speakers can either refer to events or situations as anchored in reality or indicate their non-actualisation, or, set differently, that what is referred to has not happened or does not hold true at the moment of speaking. Passive-like structure in Spanish are most often used with events and situations that are considered as real or actualized, or ongoing. No difference between the three structures could be found in our corpus since realis accounts for 80.92% (263 occurrences) in the periphrastic passive, 81.3% (787 occurrences) in the *se*-construction and 82.9% (413 occurrences) in the non-anaphoric third person plural. This suggest that the way to anchor events and situations is rather stable across passive-like structures. Nevertheless, as will be developed hereunder, the distribution of realis and irrealis clauses differs according to register and mode.




*Graph 57. Distribution of realis and irrealis clauses in Spanish periphrastic passives*

As shown in Graph 57 formal oral data tend to use far more irrealis clauses than formal written data. An influence of the mode is thus suspected but since no sufficient data were obtained for informal oral data it is premature to generalize this finding. A look at the written data show that informal data seem to resort to more realis clauses (79.2%) than formal data (87.8%). A chi-square test, however, indicates that this difference is not significant in our data ($\chi^2 = 2.5897$, $p > .05$). This leads us to conclude that the formal oral data set is the one contributing the most to the overall use of irrealis clauses in periphrastic passives. Let us now examine whether similar observations emerge for the *se*-construction.



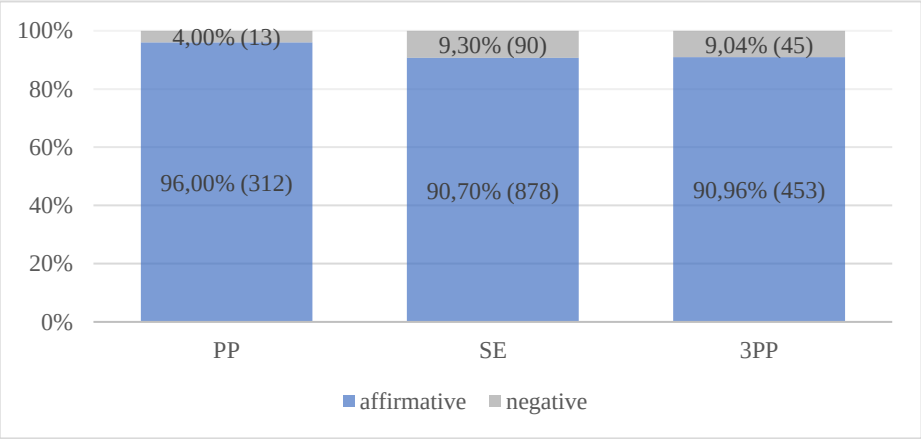

*Graph 58. Distribution of realis and irrealis clauses in Spanish se-constructions*

The impact on the distribution of realis and irrealis clauses in *se*-constructions seems to be twofold (Graph 58). First, data receive an influence of the mode since there is a significantly higher use of irrealis clauses in oral data than in written data[33]. Second, the degree of formality also contributes in explaining the variation but its impact seems much more limited since statistical tests indicate that formal data resort significantly more to irrealis clauses in oral mode whereas there is no significant difference between formal and informal written data. The distribution of realis and irrealis clauses in *se*-constructions is therefore largely impacted by the mode of production and, to a lesser extent, by the register, in oral data.

As for the non-anaphoric third person plural, the available data reveal that with informal oral data use more irrealis clauses (18.2%; 65 occurrences) than informal written ones (13.4%; 18 occurrences). Although the difference is not significant, maybe due to the smaller number of occurrences in informal written data, this result points toward the same direction that was found for the *se*-construction, i.e., higher use of irrealis clause in oral data.

### 8.2.7. Affirmation

Clauses can either be used in the affirmative to express the validity of an assertion while the negative expresses the opposite.




*Graph 59. Distribution of affirmative and negative clauses in Spanish passive-like structures*

A brief glance at Graph 59 reveals an overall predominance of affirmative clauses in the three passive-like structures. A chi-square test reveals, nevertheless, that this tendency is significantly more marked in the periphrastic passive than in the other two structures[34]. Our results also show that overall, negations are more present in oral data.

[33] The statistics are $\chi^2 = 5.8672$, $p < .05$ for informal oral vs informal written and $\chi^2 = 71.728$, $p < .05$ for formal oral vs formal written.
[34] The statistics are $\chi^2 = 9.3136$, $p < .05$ for the periphrastic passive vs the *se*-construction and $\chi^2 = 7.6141$, $p < .05$ for the periphrastic passive vs the non-anaphoric third person.

### 8.2.8. Other parameters

The present section aims at briefly presenting the results obtained for the two verb-related parameters considered as not salient by the MCA, namely *punctuality* and *volitionality*. The results concerning *punctuality* show that the use of durative verbs largely predominate in the three structures: 76.3 % (380 occurrences) in the non-anaphoric third person plural, 73,4% (709 occurrences) in the *se*-construction and 56.9% (185 occurrences) in the periphrastic passive. Examples (258) and (259) illustrate a durative and a punctual verb in a *se*-construction. The verb *atan* 'to tie up' in (258) can be classified as durative as the realization of the action (tying up trunks) requires different stages. The accomplishment of the event then implies a duration. Sentence (259) is constructed with a punctual verb, to integrate (*se integraran*). Compared to (258), in (259), the action of starting to run a new airline is instantaneous, whereas *se atan* in (258) is a discrete process developing in time.

(258) *si hay buena corriente de agua se sueltan los troncos con cuidado de que no se atasquen pero si hay poca corriente **se atan** haciendo balsas que se guían hasta donde haga falta.( WIKI ESP)*
'If there is a good water current, the trunks are released, taking care that they do not get stuck, but if there is little current, they **are tied up**, making rafts to guide them as far as necessary.'

(259) *Nuevas aerolíneas **se integraran** al aeropuerto, a comienzos de este año se integró la aerolínea española Iberia con vuelos a Madrid y a mediados de este año se integraran 5 aerolíneas más: Air Canada, China Airlines, Emirates, KLM y Spanair. (WIKI ESP)*
'New airlines **will be integrated** to the airport, at the beginning of this year the Spanish airline Iberia was integrated with flights to Madrid and in the middle of this year 5 more airlines will be integrated: Air Canada, China Airlines, Emirates, KLM and Spanair.'

As far as the parameter *volitionality* is concerned, passive-like structures appear as used almost exclusively used with volitional verbs since they account for 89.5% (866 occurrences) of the data in the *se*-construction, 90.15% (293 occurrences) in the periphrastic passive and 95.4% (475 occurrences) in the non-anaphoric third person plural.

(260) *H1 eso / en= la revolución iraní Pausa*
*H1 eehmm Pausa*
*H1 es lo que pone que <cita>**se asalta** la= la embajada de los Estados Unidos Pausa*
*H1 hay cinco personas que logran escapar Pausa*
*H1 y se= y se= se refugian en la= en la embajada de Canadá</cita>* (ESLORA SP)
'H1 eso / en= the Iranian revolution Pause
H1 eehmm Pause
H1 that's what it puts that <quote>the= US embassy **is assaulted** Pause
H1 there are five people who succeed in escaping Pause

H1 and they= and they= take refuge in the= Canadian embassy</cite>.'

Example (260) illustrate a volitional verb with a *se*-construction. The attack on the US embassy is an action performed purposefully by an animate entity.

### 8.2.9. Summary

To briefly summarise the main results yielded by this section, the following points can be highlighted. As for the distribution of tenses, the periphrastic passive and the non-anaphoric third person plural use a similar proportion of present tenses while the *se*-construction resorts to them in a higher proportion. The analysis has revealed that *tense* and *aspect* are two correlated parameters, a high use of present tenses goes hand in hand with imperfective aspect.

The type of verb process largely contributes in explaining the difference between the three passive-like structures. While the periphrastic passive and the *se*-construction display a rather similar profile, the non-anaphoric third person plural considerably differs in its preference of type of verb process.

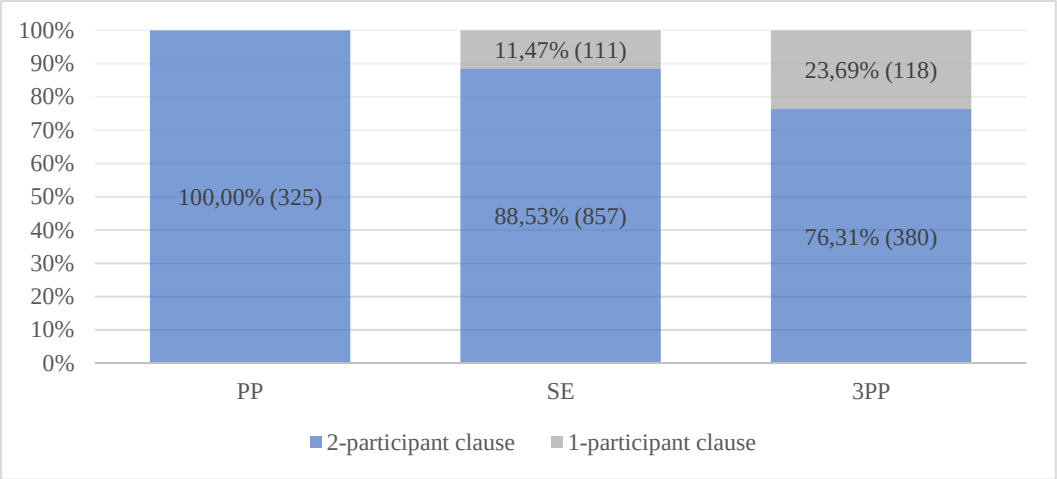
Another interesting result yielded by this detailed analysis of verb-related parameters is that, in some cases, the passive-like structures display an overall similar behaviour. Nevertheless, breaking down the data into mode and registers reveals numerous differences between the data sets. This phenomenon occurs, for instance, in the case of the dichotomy *realis/irrealis*. The distribution of the verb-related features appears thus as rather register and mode sensitive.

## 8.3. Participant-related parameters

This section complements the verb-parameters by examining in closer detail the participants occurring with the verb and develops as follows. First, I will discuss the number of participants used in the different passive-like structures. As we will see, some structures appear to be more constrained than others. Then, I will turn to the detailed analysis of the agent and the patient. It is assumed that the nature of these two elements may partly explain the variation between the passive-like structures. I will further bring additional information on the degree of individuation and affectedness of the patients. The last part of the section recapitulates the main findings regarding participants.

### 8.3.1. Number of participants

Clauses can imply one or several participants. This internal architecture may impact the variation between the different passive-like structures.




*Graph 60. Distribution of one- and two-participant clauses in Spanish passive-like structures*

Graph 60 shows the distribution of one- and two-participant clauses in the three Spanish constructions analysed in this study. As can be observed, the number of participants varies according to the passive-like structure. While periphrastic passives only consist in two-participant clauses (325 occurrences) such as in (249), the non-anaphoric third person plural allows more flexibility with 23.7% (118 occurrences) of one-participant clauses (Example 250). Compare (261) and (250). In (262) the clause consists in two participants: the Alca plane and the implicit agent which does the action of using the Alca. Example (262), conversely, implies only one participant – i.e., the persons who work during the semester (*trabajan).*

(261) *Debido a su alta capacidad de maniobra al volar a bajas velocidades, el Alca puede* ***ser utilizado*** *también contra helicópteros.* (WIKI SP)
'Due to its high manoeuvrability when flying at low speeds, the Alca **can** also **be used** against helicopters.'

(262) *H2 y allí en King's empezaban las clases el día nueve de enero (...) H2* ***trabajan*** *un montón durante el curso y luego aparte todos los exámenes juntos yo yo prefiero hacerlos por cachos porque Pausa (...).* (ESLORA SP)
'H2 and there at King's they started classes on the ninth of January (...) H2 **they work** a lot during the course and then apart from all the exams together I prefer to do them in chunks because Pausa (...).'

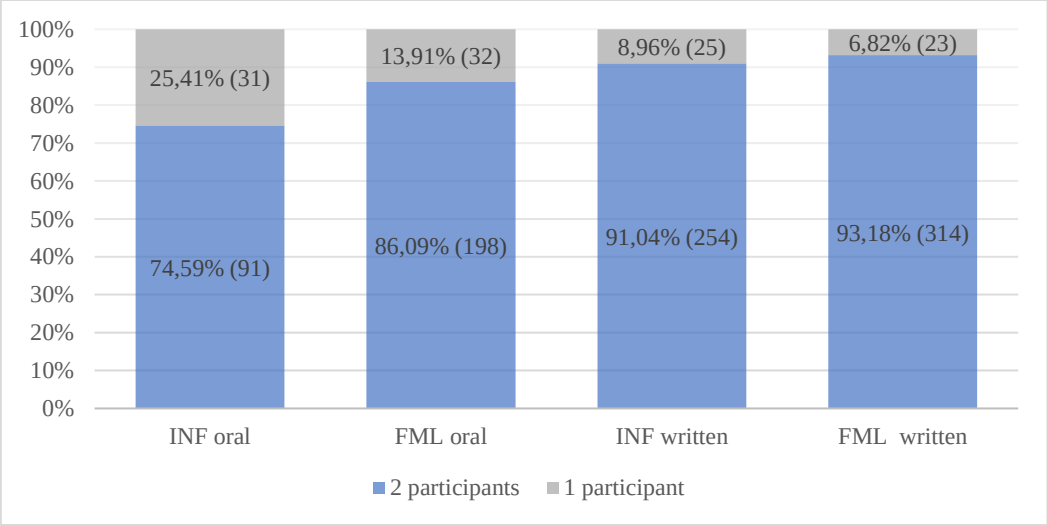
As explained in the analysis of the French passive-like structures, the specific architecture of the periphrastic passive which promotes a patient in the position usually occupied by the agent explains the exclusive use of two-participant clauses.

The *se*-construction occupies an intermediate position between the periphrastic passive and the non-anaphoric third person plural as it displays 11.5% (111 occurrences) of one-participant clauses. The syntactic properties of the *se*-constructions largely contribute in explaining this high use of two-participant clauses since one of their central purposes is to place the focus on a patient. They are thus good candidates to contain a patient in addition to the (defocused) agent. When *se*-constructions consist in one-participant clause, a brief qualitative review of the data indicates that they are built with a subordinate clause as in (263). As explained in the

methodological section (Chapter 5), subordinate clauses are not considered as participants in this study. Example (263) therefore is considered as a one-participant clause with the verb *se supone* ('is supposed') which only implies the participant performing the action of supposing.

(263) *Pero mi pregunta es, si ya tienen una licencia de piloto de aerolinea o comercial, (1)porque tienen que hacer el curso si* ***se supone*** *que ya saben pilotiar un avion?* (YAHOO SP)
'But my question is, if they already have an airline or commercial pilot's licence, (1) why do they have to do the course if **it is supposed** that they are already able to fly a plane?'

Let us now examine the possible influence of register and mode on the distribution of one- and two-participant clauses. As the periphrastic passive only contains two-participant clauses in our data the analysis of mode/register variation will be centred on the *se*-constructions and the non-anaphoric third person plural.

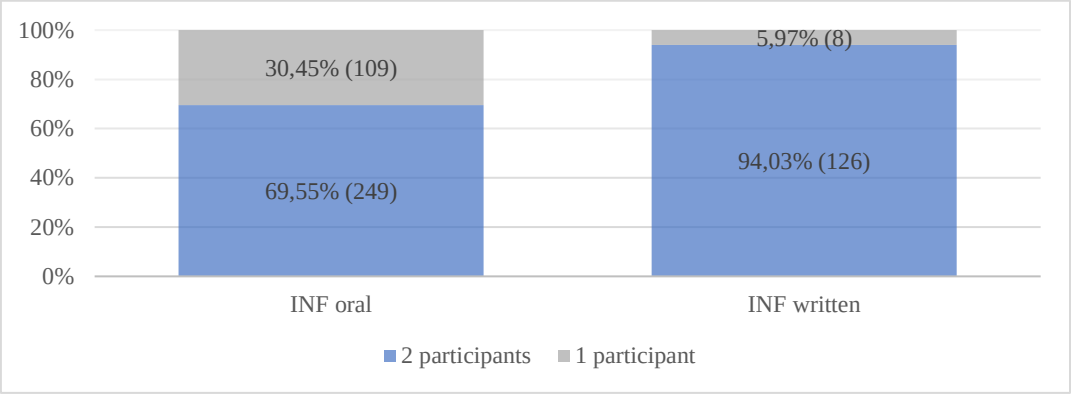



*Graph 61. Distribution of one- and two-participant clauses in Spanish se-constructions*

First, Graph 61 indicates that the distribution of one-participant clauses varies according to the register. There is a significantly higher use of two-participant clauses in informal oral data ($\chi^2$ = 7.1704, $p < .05$) compared to formal oral ones. Such difference between formal and informal data is, however, not observed in the written data set. Second, the number of one- and two-participant clauses also varies according to the mode. Results reveal a significant difference between the use of one-participant clauses in informal oral and informal written data ($\chi^2$ = 19.1159, $p < .05$). A similar difference can be observed between formal oral and formal written data, with a more frequent use of one-participant clauses in the oral data set ($\chi^2$ = 7.841, $p < .05$). A more qualitative overview of the data reveals that overall, one-participant clauses in oral data, be it formal or informal, imply mental and verbal processes such as in (264).

(264) *H1 ¿no le viste en el Twitter en pla=n? Pausa H1 <cita>preparando los exá=menes</cita> Pausa H1 no sé qué= en plan que* ***se notaba que*** *estaba estudiando (...).* (ESLORA SP)

‘H1 didn't you see him on Twitter like? Pause H1 <quote>preparing for exams</quote> Pause H1 I don't know what= like **you could tell** he was studying.’




*Graph 62. Distribution of one- and two-participant clauses in Spanish non-anaphoric third person plurals*

Results in Graph 62 confirm the trend found in the *se*-construction, that is a more frequent use of one-participant clauses. Nearly a third of the occurrences in informal oral data consists in one-participant clauses (30.4%; 109 occurrences). Similarly to the *se*-construction, the use of one-participant clause with non-anaphoric third person plurals is highly associated with verbal processes.

The number of participants appear to be largely constrained by the syntactic properties and internal architecture of the passive-like structures. While the periphrastic passives and the *se*-constructions are almost exclusively used with two-participant clauses, the non-anaphoric third person plural allows much more variation regarding the number of participants. The analysis also revealed that the variation between the passive-like structures is influenced by mode and register and that a link could be established between the type of verb process and the number of participants implied in the clause.

### 8.3.2. Agent

The participants analysed in the present work either play the semantic role of agent or patient. This section is devoted to the analysis of agent, which, as was the case for the French part of this work, is based only on the periphrastic passive. Indeed, the agent in the non-anaphoric third person plural is always a human entity and the agent in the *se*-construction is rarely overt. The periphrastic passive is thus the best candidate to experience variation as far as its agent is concerned.

Similarly to what was found in French data, Spanish periphrastic passives contain a large proportion of occurrences used without any overt agent – i.e., introduced by the oblique complement *por* (‘by’) (78.8%; 256 occurrences) as illustrated in (265).

(265) *Entonces, llegó el día en que uno de los cerdos **fue asesinado**: todavía tengo recuerdos vívidos de todo el proceso -incluyendo los gritos*. (WIKI SP)

‘Then, the day came when one of the pigs **was killed**: I still have vivid memories of the whole process - including the screams.’

What is of particular interest for us is when there is an overt agent in the periphrastic passive. I will first give the main results, regardless of register and mode variations and, in a second phase, supplement the findings with information on these two types of variations.

The figures in Table 26 allow to draw a first distinction between animate and inanimate agents.

| | **Agent** | **No. occurrences** | **%** |
|---|---|---|---|
| **Animate** | Human | 24 | 32.7% |
| | Non human | 8 | 11.52% |
| **Inanimate** | Organization | 14 | 16.59% |
| | Country/City | 3 | 3.23% |
| | Inanimate Other | 25 | 35.94% |
| | **Total** | 74 | 100% |

*Table 26. Classification of animate and inanimate agents in Spanish periphrastic passives*

Results reveal that the use of animate (45%; 32 occurrences) and inanimate agents (55%; 42 occurrences) is more or less balanced. Within animate agents, the category *human* predominates such as in (266) where the action is realized by the last vice president of the airline.

(266) *La nueva aerolínea **fue guiada por** el último vicepresidente y jefe de operaciones de Pan Am, Marty Shugrue, quien también ayudo en la creación del programa de viajero frecuente WorldPass (…).* (WIKI SP)
‘The new airline **was guided** by Pan Am's late vice president and chief operating officer, Marty Shugrue, who also helped create the WorldPass frequent flyer programme.’

Within inanimate agents, inanimate *others* is the most common type of agents, as illustrated in (267) with *por el fuego* (‘by fire’).

(267) *Pueblos y ciudades **fueron cercados** por el fuego.* (PROCEP SP)
‘Villages and towns **were encircled** by fire.’

(268) *El acuerdo, jurídicamente vinculante, deberá **ser ratificado** por el Parlamento.* (PROCEP SP)
‘The agreement, legally binding, will have **to be ratified** by the Parliament.’

Organizations, including cases such as the Parliament in (268), also represent a large part of inanimate agents.

In Chapter 7, the analysis revealed that the presence of an overt agent is either prompted by the nature of the agent (famous person, organization or institution) or by the problematic nature of the situation. In a similar way, these two contexts help to explain the presence of an overt agent in Spanish periphrastic passives. A qualitative analysis of the data reveals the presence of agents such as el *Parlamento* (‘the Parliament’), el *vicepresidente* (‘the vice president’), the New York Times, las *autoridades* (‘the authorities’). A brief review of the semantic of the verbs displays

instances of verbs such as *secuestrado* (‘kidnapped’), *agravado* (‘aggravated’), *empeorado* (‘worsen’) o *matado* (‘killed’), which describe a negative situation. These findings confirm thus what is usually suggested in the literature about the use of long (or non-basic) periphrastic passives. A main difference with French data is the fact that Spanish agents seem to be much more used to serve the textual cohesion of the discourse.

(269) *La mayoría de los hongos de la pudrición suave requieren de la adición de alimentos exógenos para causar el ataque substancial. Estos alimentos a menudo* ***son proporcionados*** *inadvertidamente* ***por*** *los fertilizantes en suelos agrícolas, restos de basura en torres de enfriamiento, y otras fuentes nutrientes.* (WIKI ESP)
‘The majority of fungi of soft rot requires the addition of exogenous nutrients to cause substantial attack. These nutrients **are** often inadvertently **provided by** fertilisers in agricultural soils, debris in cooling towers, and other nutrient sources.’

In (269), the use of a periphrastic passive allows to position the patient *alimentos exógenos* ‘exogenous nutrients’ of the preceding sentence in topic position of the new sentence. The textual cohesion is strengthened by the presence of the demonstrative pronoun *estos* ‘these’ which accompanies the patient of the periphrastic passive. The overt agent provides information on how these nutrients are provided.

Example (270) also illustrates a periphrastic passive serving a discourse organizing function but slightly different from example (269). While the topic of the excerpt in (270) is Linda McCartney, the use of the periphrastic passive in (270) allows to foreground one of her achievements, that is, the acquisition of the Heinz Company, while still mentioning her as agent, ensuring the cohesion of the discourse.

(270) *En 1995, Linda apareció en forma animada con su marido en Los Simpsons episodio "Lisa la vegetariana" , y el episodio, La "Papelera de Titanes" fue dedicado a su memoria, en 1998. El HJ Heinz Company* ***fue adquirido por*** *Linda McCartney Foods en marzo de 2000, y el Grupo de Hain Celestial, en 2007. McCartney fue una firme defensora de los derechos de los animales, y prestó su apoyo a muchas organizaciones como PETA (Personas para el Tratamiento Ético de los Animales), […].* (WIKI SP)
‘In 1995, Linda appeared in animated form with her husband in The Simpsons episode "Lisa the Vegetarian", and the episode, " The Trash of the Titans " was dedicated to her memory, in 1998. The HJ Heinz Company **was acquired by** Linda McCartney Foods in March 2000, and the Hain Heavenly Group in 2007. McCartney was a strong advocate for animal rights, and gave help to many organisations such as PETA (People for the Ethical Treatment of Animals), [...].’

This textual cohesive function may be regarded as a specificity of the Spanish periphrastic passive, and would help to explain the different contexts of use of the periphrastic passive and the *se*-construction, a passive-like structure in fierce competition with the canonical passive in Spanish. Indeed, though the *se*-construction allows to focus on the patient, the construction is rarely accompanied by an agent. The absence of an overt agent would result in a loss of information.

Since the presence of overt agents remains limited in our data, these observations are only first assumptions that may serve as bases for further investigation.

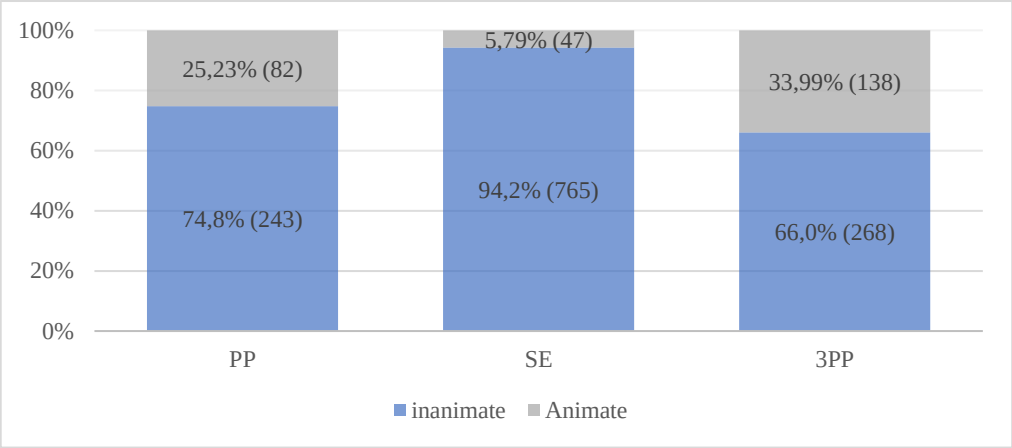
As far as register and mode variation is concerned, results indicate that 76% of the overt agents (55 occurrences out of the 74 occurrences) are used in formal oral data. This means that the other combinations of mode/register either provide clues as for the agent through other devices such as the use of spacebuilders or prepositional phrases or leave the agent entirely defocused. This part is detailed in the next section. Now that the panoramic overview of agents has been established, we will turn to the examination of patients.

### 8.3.3. Patient

As seen in 8.3.1, the three passive-like structures largely rely on two-participant clauses, including thus a patient. This leads us to consider patients as a key factor in explaining the variation between these structures. A first approach provides a general picture on the animate versus inanimate nature of the patients (see Graph 63). Examples (271) and (272) illustrate an inanimate patient used in a *se*-construction (*armamento* 'armaments') and an animate patient, the third person plural in the verb *son obligados* ('they are obliged') which represents the pilots, used with a periphrastic passive.

(271) *HI lo que importa es / <ruido=palmada/> sus fábricas funcionan / **se vende** armamamento sale la cri- el país de la crisis=*
*H3 sí / pero también mueren / también mueren= / ah / soldaditos americanos (ESLORA SP)*
'HI what matters is / <noise=palm/> their factories are working / armaments **are being sold** the cris thecountry escapes the crisis= H3 yes / but they also die / they also die= / ah / american soldiers'

(272) *Más del 70 % de la plantilla de pilotos de Ryanair está compuesta por falsos autónomos, sin derecho a cobrar durante ausencias por enfermedad, ni las vacaciones. No se respetan los tiempos de descanso y, para poder trabajar, **son obligados** a abrir su falsa empresa en Irlanda, beneficiando en la tributación a la compañía low cost.* (PROCEP SP)
'More than 70 % of Ryanair's pilot workforce is made up of bogus self-employed, without the right to be paid during sick leave or holidays. Rest periods are not respected and, in order to be able to work, they are forced to open their bogus company in Ireland, benefiting the low cost company in terms of taxation.'




*Graph 63. Animate versus inanimate patients in Spanish passive-like structures*

Results indicate a significantly different behaviour for the three structures (see Graph 63). The *se*-construction almost exclusively resorts to inanimate patients (94.2%; 765 occurrences), followed by the periphrastic passive (74.8%; 243 occurrences) and the non-anaphoric third person plural (66%; 268 occurrences). The nature of the patient can therefore be considered as a variational factor explaining the differences between the structure. In what follows, detailed information on the categories of patients is provided.

| Structure | Animate human | Animate non human | Organization | Country/City | Others |
|---|---|---|---|---|---|
| **Periphrastic passive** | 7.69% (25) | **17.54% (57)** | **6.77% (22)** | 0.31% (1) | 67.70% (218) |
| ***Se*-construction** | 1.48% (12) | 4.31% (35) | 0.37% (3) | 0.99% (8) | **92.86 %** (754) |
| **Non-anaphoric third person** | **27.83% (113)** | 6.16% (25) | - | 0.99% (4) | 65.02% (264) |

*Table 27. Classification of animate and inanimate patients in Spanish passive-like structures*

Table 27 indicates that each of the three passive-like structure has its own specificities as far the nature of the patient is concerned. The periphrastic passive displays two peculiarities : a high use of animate non human patients (17.5%; 57 occurrences) exemplified in (273) with *mascotas* ('pets') as well as an important number of organizations (6.8%; 22 occurrences), as Heinz Company in (274), which is hardly found in the other two structures.

(273) *muchisimas personas tienen mascotas de manera responsable,yo siempre he tenido mascotas(perros,gatos,peces,hansters,tor... todo) y la mayoria* ***han sido adoptadas****,ahora mismo tengo tres perros (...) (YAHOO SP)*
'many people have pets in a responsible way,I have always had pets(dogs,cats,fish,hamsters,tor... everything) and most of them been adopted,right now I have three dogs.'

(274) *El HJ Heinz Company **fue adquirido** por Linda McCartney Foods en marzo de 2000, y el Grupo de Hain Celestial, en 2007.* (WIKI SP)
'The HJ Heinz Company was acquired by Linda McCartney Foods in March 2000, and the Hain Celestial Group in 2007.'

The *se*-construction can be associated to an extensive use of inanimate patients *others* (92.9%; 754 occurrences). The soft drink INKA KOLA in example (275) illustrates the category inanimate *others.*

(275) *Para los abstemios, cristianos verdaderos y alcohófobos, la deliciosa INKA KOLA es un refresco que **no se consigue** en ninguna otra parte, excepto en el Ecuador y en partes de USA donde se produce y exporta respectivamente. (YAHOO SP)*
'For abstainers, true Christians and those who don't like alcohol, the delicious INKA KOLA is a soft drink that **cannot be obtained** anywhere else, except in Ecuador and parts of the USA where it is produced and exported respectively.'

This abundant presence of inanimate patients can be partly explained by the fact that the use of animate referents in *se*-constructions can lead to several interpretations: reflexive, reciprocal or passive (Cabañas Maya 2006:126). The choice of inanimate entities in *se*-constructions acting as passive-like structures makes it possible to avoid any ambiguous reading.

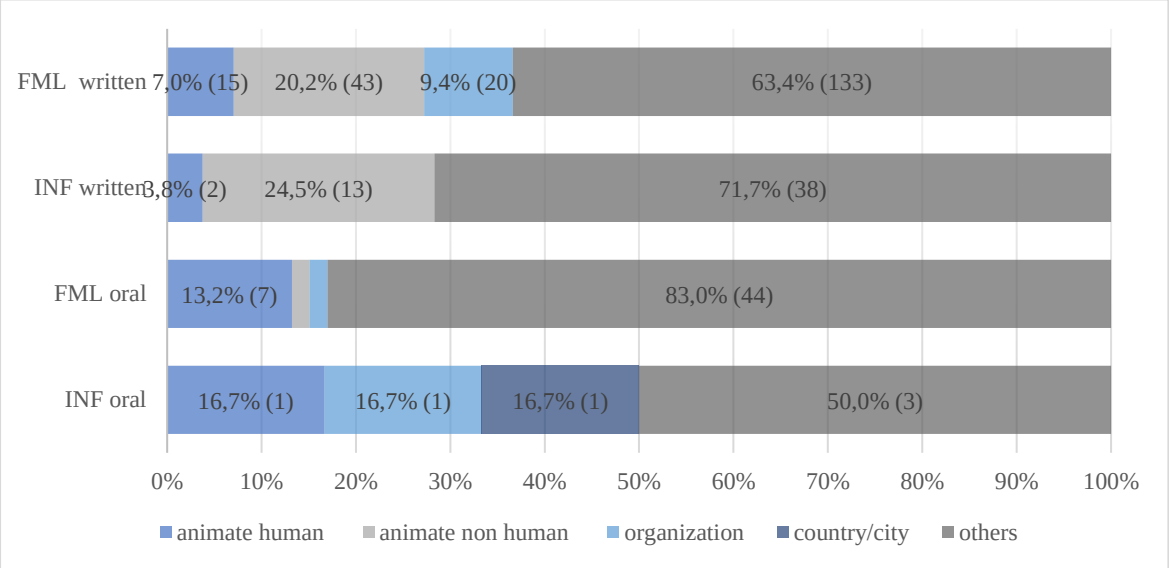
The non-anaphoric third person plural can be characterized by its numerous cases of animate human patients (27.8%; 113 occurrences) such as the personal pronoun *te* ('you') in (276).

(276) *mira acá en argentina por lo menos te **dicen** que tenes que cortar el pelo bien corto bien flaco porque no te aceptan tan gordos.* (YAHOO SP)
'look here in argentina at least **they tell** you that you have to cut your hair very short and very thin because they don't accept you so fat.'

It can be observed that all three passive-like constructions are in complementary distribution as regards the nature of the patient included in the structure.

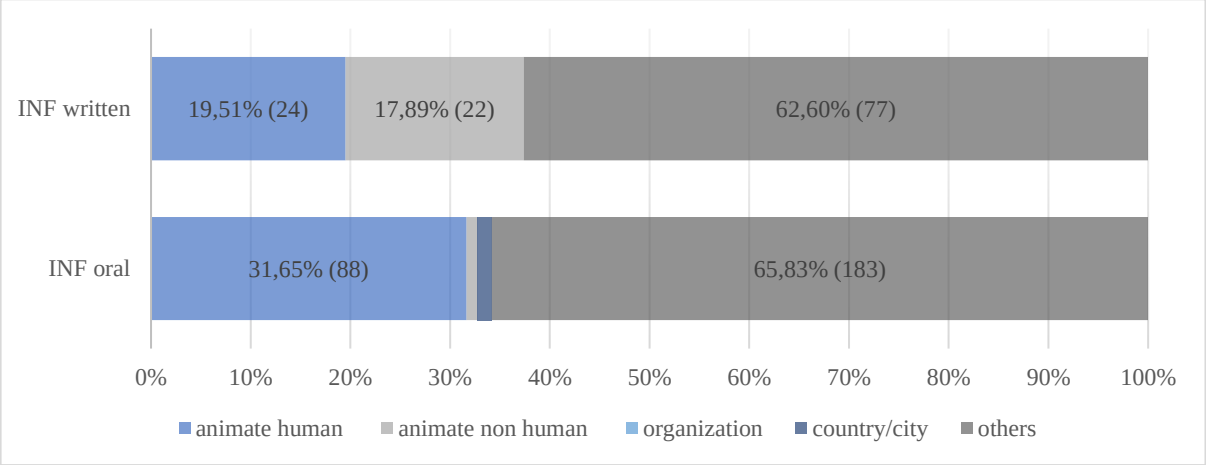
Breaking down these results according to the mode and the register components makes it possible to obtain further information regarding patients. First, let us examine the results obtained for the *se*-construction. The analysis indicates that 5.8% of the patients in *se*-constructions are classified as animate. A closer look at the data reveals that the large majority of these animate patients are used in formal written data (34 out of the 47). This may be explained by the fact that formal written data may be more prone to give additional information to the reader since there may be less shared background between the writer and the reader than in oral or informal situations, where the interlocutors typically share an interactional situation and the related contextual knowledge. Since inanimate *others* largely predominate in *se*-

constructions (92.8%) no other relevant information on the influence of mode and register could be obtained for this passive-like structure.




*Graph 64. Distribution of patients in Spanish periphrastic passives*

Though the figures are small, Graph 64 gives a more detailed overview of the nature of the patients in the four combinations of mode/register in the periphrastic passive. Particular attention is will be paid to non human animate agent and organizations as these two categories seem, as explained above in Table 27, to be characteristic of the periphrastic passive. Animate non human patients are mainly used in written data, be it formal (20.2%) or informal (24.5%). This result confirms what was found for the *se*-construction, that is, a high use of animate non human patients with formal written data. Since this type of patient is absent from the oral data sets, it can be assumed that mode impacts the distribution. As for the use of organizations, these patients are mainly gathered in formal written data (20 occurrences out of the 22).



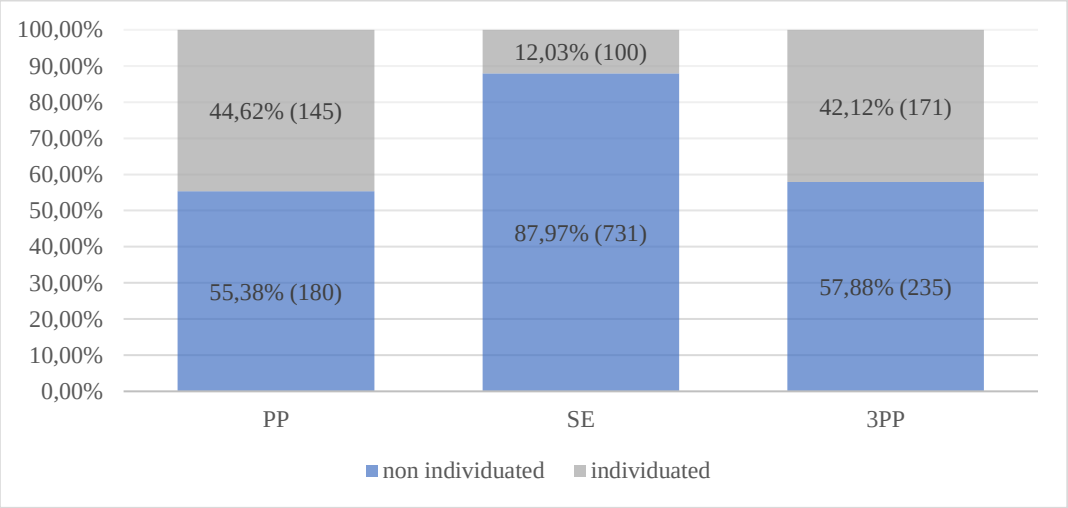

*Graph 65. Distribution of patients in Spanish non-anaphoric third person plurals*

Graph 65 supports the fact that animate non human patients are strongly associated with written data. For the non-anaphoric third person plural, our data only allow to confirm this observation for informal written data, given the structure's low frequency in formal data. The graph also

allows to note a significantly higher use of animate human in informal oral data compared to oral data ($\chi^2 = 63.6962$, $p < .05$).

### 8.3.4. Individuation and affectedness of the patient

This section provides further information on patients through a detailed examination of the parameters *individuation* and *affectedness*. Six features can be used to characterized patients (Hopper & Thompson 1980): proper name, animate, concrete, countable, singular, definite/referential.




*Graph 66. Distribution of individuated and non-individuated patients in Spanish passive-like structures*

As observed in Graph 66, the individuated and non-individuated patients are represented in different proportions according to the passive-like structure. The *se*-construction displays a relatively high use of non-individuated patients (88%; 731 occurrences). Example (277) illustrates a case of non individuated patient. The noun phrase *tortilla* meets with only three of the six criteria of *individuation* (concrete, countable and singular).

(277) *si se dobla la tortilla a la mitad y **se baña** con mole se llama enchilada o enmolada, si la masa cruda se rellena con frijol y se frie se vuelve una gorda o un bocol. (*YAHOO SP*)*
'if the tortilla is folded in half and **bathed** with mole it is called enchilada or enmolada, if the raw dough is filled with beans and fried it becomes a gorda or bocol.'

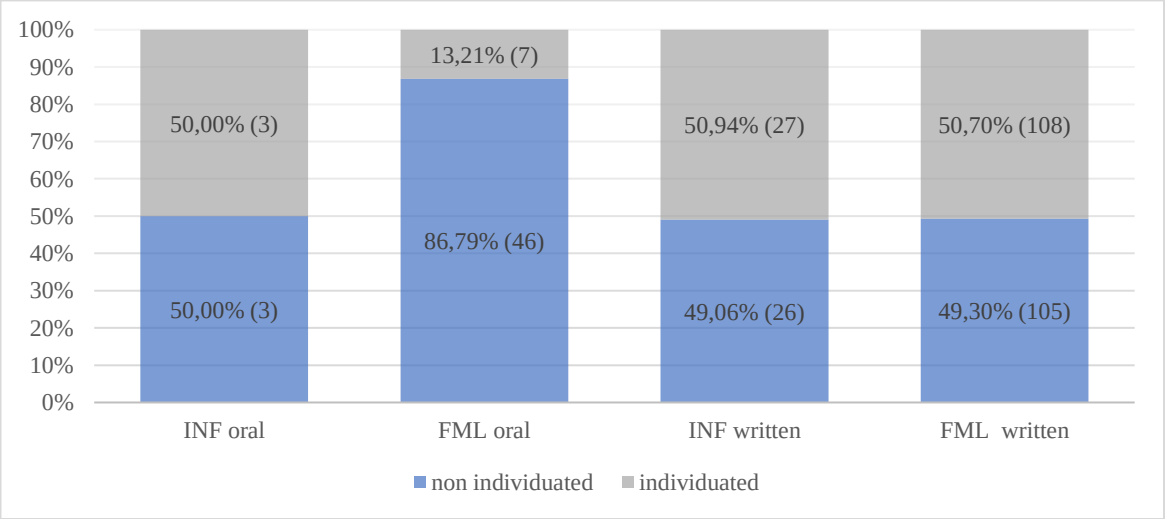
Interestingly, this use correlates with the high frequency of use of inanimate patients *others* presented in the previous section. Inanimate *others* are more likely to be classified as non individuated since, due to their nature, they are automatically deprived from the feature *animate*. Moreover, as already explained, this category of patients can either be concrete or abstract entities. Patients considered as abstract such as in (277) have a lower degree of individuation (Hopper & Thompson 1980).

(278) (…) *en cualquier caso, debemos aproximar nuestras posturas y llegar a una legislación que sea coherente, porque de nada sirve legislar si luego no* ***se puede aplicar*** *esta legislación o no puede comprobarse que se aplica.* (PROCEP SP)
'In any case, we must bring our positions closer together and reach legislation that is consistent, because there is no point in legislating if this legislation cannot be applied or cannot be verified that it is applied.'

(279) ***Se corta*** *la cebolla en cuadrados grandes y se frien hasta que queden transparentes* (…) (YAHOO SP)
'The onion **is cut** the onion into large squares and fried until translucent.'

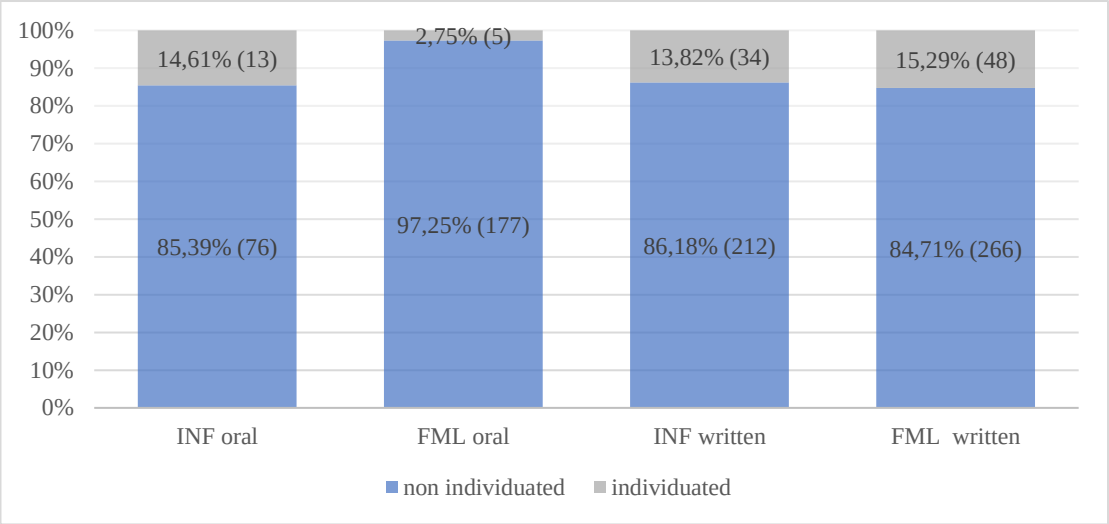
Legislation in (278) refers to a set of laws rather than a tangible entity. Its degree of individuation is thus lower than that of the onion in (279), which is a concrete element.

Compared to the *se*-construction, the periphrastic passive and the non-anaphoric third person plural contain a much more balanced distribution of individuated and non-individuated patients. A chi-square test reveals that there is no significant differences in the distribution of (non)-individuated patients between the two structures ($\chi^2 = 0.4587$, $p > .05$). This larger presence of individuated patients in periphrastic passives and non-anaphoric third person plural can be associated to the higher frequency of use of animate patients observed in 8.3.3.



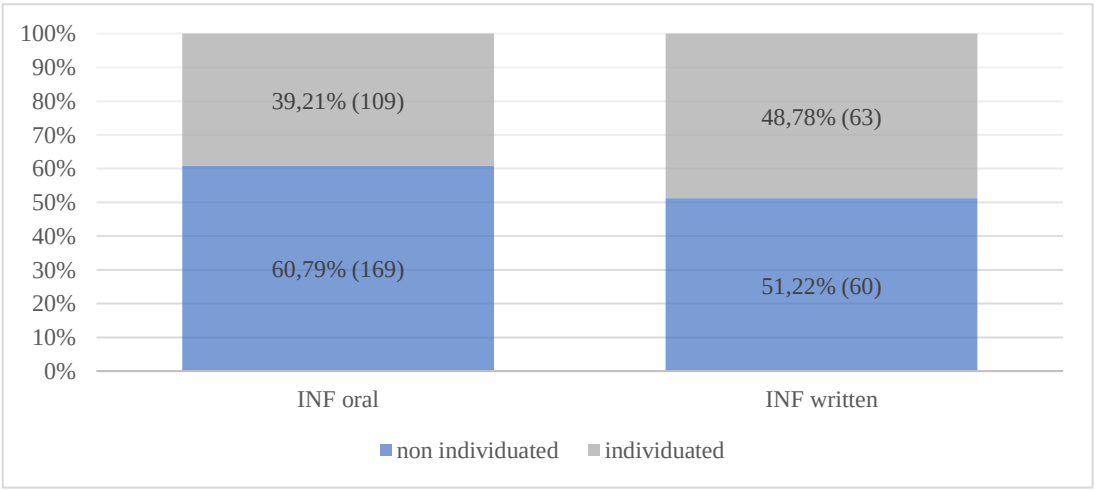

*Graph 67. Individuated and non individuated patients in Spanish periphrastic passives*

A more detailed look at the periphrastic passive reveals that, while the proportion of individuated and non-individuated is perfectly balanced in informal and formal written data, the use of non-individuated patients is much more frequent in formal oral data (see Graph 67). This suggests that the written data set does not undergo any influence of the register but the oral one does. This observation should, however, be considered cautiously as figures in formal oral data remain low.




*Graph 68. Individuated and non individuated patients in Spanish se-constructions*

The balanced profile between individuated and non-individuated patients found in the periphrastic passive is confirmed for the *se*-construction (see Graph 68). The tendency to resort to more non-individuated patients in formal oral data than in formal written is also observed in the *se*-construction. Interestingly, informal oral data resemble more informal written data than formal oral ones. An influence of the register can therefore be suspected. This finding did not emerge for the periphrastic passive as the number of occurrences of this construction is limited in informal oral data.




*Graph 69. Individuated and non individuated patients in Spanish non-anaphoric third person plurals*

In Graph 69, it can be observed that informal written data use significantly more individuated patients compared to informal oral data ($\chi^2 = 5.022$, $p < .05$). The mode of the production is thus assumed to impact the degree of individuation of patients. This higher use of individuated patients may be linked to the fact that, as seen previously, animate patients are more common in informal written data than in oral ones.

So far, we have seen that the *se*-construction is the passive-like structure to resort the most to non-individuated patients. The nature of the patients may explain this phenomenon since, as indicated in the MCA, the parameters patient and individuation are positively correlated. We

suggest that patients classified as inanimate *others*, which largely predominate in the *se*-construction, are the best candidates to be classified as non-individuated. This would explain the high use of non-individuated patients in the *se*-construction. The periphrastic passive and the non-anaphoric third person, which both display a more heterogeneous use as regards the nature of the patient, resort to much more individuated patients. Regarding the influence of mode and register, no significant influence of the register could be observed in the written data set. Conversely, the register appears as a key factor in the oral situations, with a formal oral data set that resembles more written data than informal oral ones.

The degree of *affectedness* is another feature that characterizes the patient. They can be either totally or partially affected. Since one of the main purposes of passive-like structures is, among others, to place the focus on the patient, it is expected that the latter will be totally affected. Our data largely confirm this assumption since the three passive-like structures include a patient totally affected in more than 98% of the cases. With such a high proportion of totally affected patients, no relevant information on mode and register could be obtained.

#### 8.3.5. Summary

The number of participants included in the passive-like structure greatly varies according to the construction. While the periphrastic passive only relies on two-participant clauses, the *se*-construction and the non-anaphoric third person plural are more flexible. These two constructions display, nevertheless, a clear preference for two-participant clauses.

As for agent, since non-anaphoric third person plurals are by definition constructed with human agents and *se*-constructions usually defocus their agents, the analysis has been centred on agents in periphrastic passives. Results showed that, overall, inanimate agents are slightly more frequent than animate ones. Human agents predominate in the animate category while *others* account for the largest part of inanimate agents. As far as patients are concerned, our findings made it possible to highlight the key role played by the patient in differentiating the three Spanish passive-like structures. The analysis revealed that the three structures are complementary, to some extent, regarding the nature of the patient. The periphrastic passive can be characterized by a high presence of animate non human patients, the *se*-construction by the almost exclusive use of inanimate *others* and the non-anaphoric third person plural by a high frequency of animate human patients. The following correlation could be observed between the nature of the patient and its degree of individuation : the two passive-like structures which display a high proportion of animate patients contain a larger number of individuated patients while the *se*-construction includes much more non-individuated patients.

The next section shows that, in addition to participant-related parameters, contextual factors also impact the variation in the use of the three passive-like structures in Spanish.

## 8.4. Context-related parameters

In order to capture the potential impact of the linguistic context, three elements will be scrutinized: the presence of a prepositional phrase, the use of spatio-temporal restrictions and the concomitant use of another passive-like construction alongside the studied structure. This last parameter is specific to the Spanish language analysis.

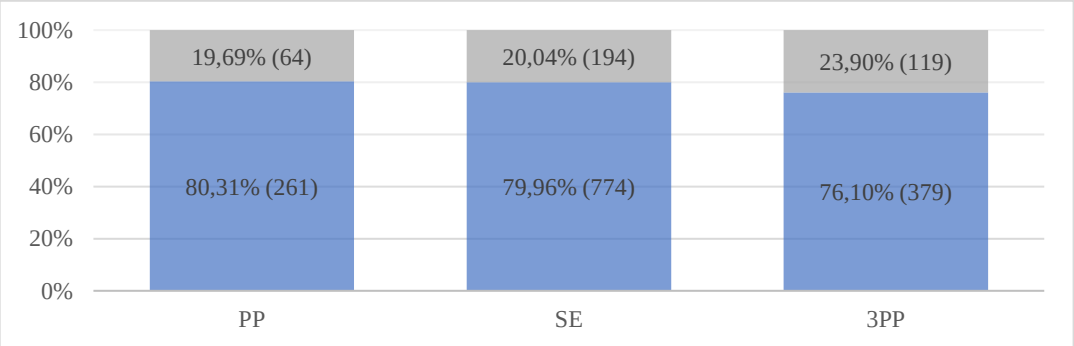
### 8.4.1. Presence of a spatiotemporal restriction

Spatio-temporal restrictions constitute a valuable resource to obtain information on the possible participants implied in the passive-like structures. They can either operate at the geographical level or at the temporal level as exemplified in (280) (*en la mayoría de los países* 'in most countries'). The addition of the restrictive element makes it possible to understand that the fact that the daily amount of DHA is not reached does not apply to all countries but only to a majority. In (281), the presence of the temporal restriction *En la actualidad y desde principios de la revolución industrial* ('Nowadays and since the beginning of the industrial revolution') indicates that the use of metals and plastics replacing wood describes the current situation and did not hold true in the past.

(280) *La cantidad diaria recomendada de DHA es de 220 mg diarios, un nivel que no* ***es alcanzado*** *en la mayoría de países, ni siquiera por las personas que no son vegetarianas.* (WIKI SP)
'The recommended daily amount of DHA is 220 mg per day, a level that **is not reached** in most countries, even by non-vegetarians.'

(281) *En la actualidad y desde principios de la revolución industrial muchos de los usos de la madera* ***han sido cubiertos*** *por metales o plásticos, sin embargo es un material apreciado por su belleza y por que puede reunir características que difícilmente se conjuntan en materiales artificiales*. (WIKI SP)
'Nowadays and since the beginning of the industrial revolution, many of the uses of wood **have been covered** by metals or plastics, however it is a material appreciated for its beauty and because it can combine characteristics that are difficult to find in artificial materials'.

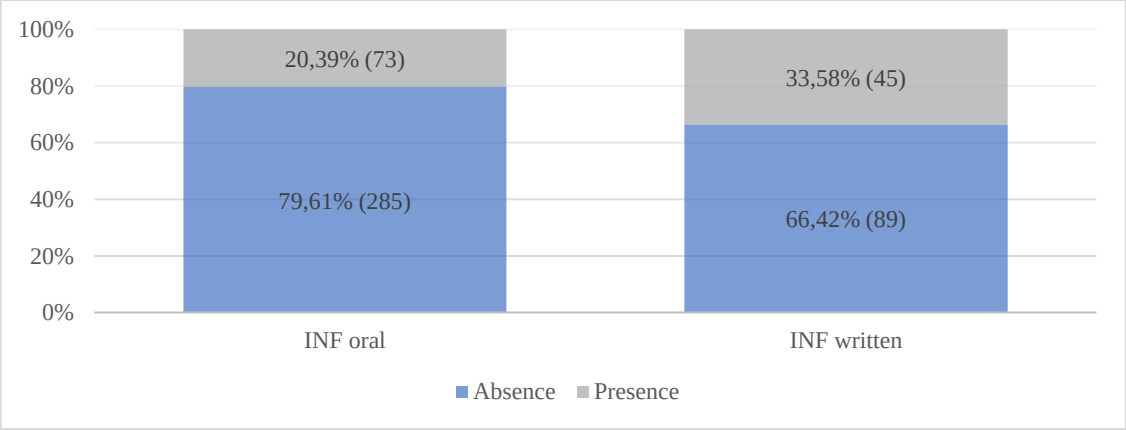
Graph 70 indicates that spatio-temporal restrictions are used in between 19.7% to 23.9% of the cases, depending on the passive-like structures.




*Graph 70. Use of spatiotemporal restrictions in Spanish*

Statistical chi-square tests reveal, however, that these slight frequency differences of spacebuilders are not significant. These contextual elements are one of the rare characteristics commonly shared by the three passive-like structures studied in this work. The more detailed examination of the use of spatio-temporal restrictions in the light of mode and register variation yields the following results.

Regarding the influence of the mode and the register in the three passive-like structures, our results indicate that these factors do not influence the distribution of the spatio-temporal restrictions in the periphrastic passive and the *se*-construction but they do impact their distribution in the non-anaphoric third person plural. A slightly more frequent use of spacebuilders with the periphrastic passive is suspected in written data compared to oral data, however, the limited number of occurrences obtained in this study for the periphrastic passive does not allow to confirm this observation. Even though much more frequent than the periphrastic passive in Spanish, our data for the *se*-construction do not point to any influence of the mode or the register.



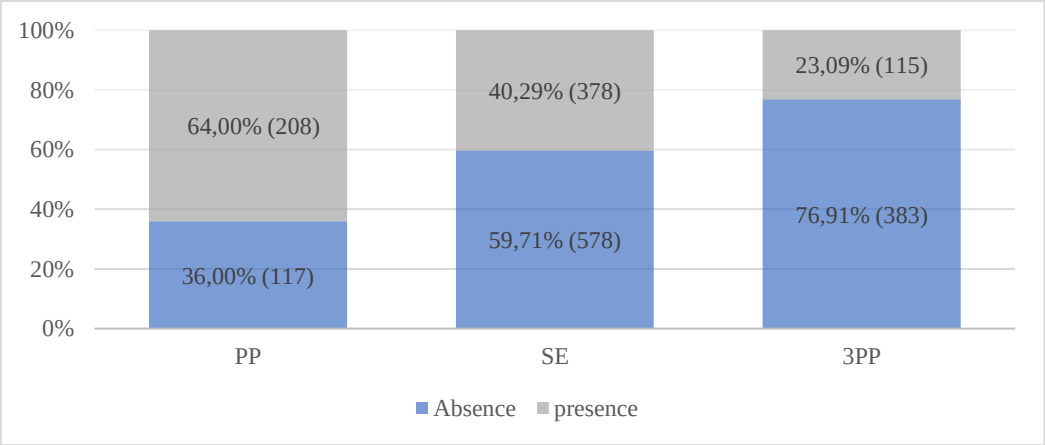

*Graph 71. Presence of a spatiotemporal restriction in Spanish non-anaphoric third person plurals*

The only significant difference as for the use of spatio-temporal restrictions was found in the non-anaphoric third person plural (Graph 71). Results indicate that these devices are significantly more used in informal written data than in informal oral data. This may be explained by the fact that, usually, written productions contain additional information as for spatio-temporal reference as the degree of common background between the speakers is less straightforward than during oral interactions. This result is in line with the findings yielded by the analysis of French passive-like structures.

It can be concluded that the use of spatio-temporal elements, which can guide the interpretation of the passive-like structure is stable across the three structures in Spanish and does not undergo a major influence of the mode and the register.

### 8.4.2. Presence of a prepositional phrase

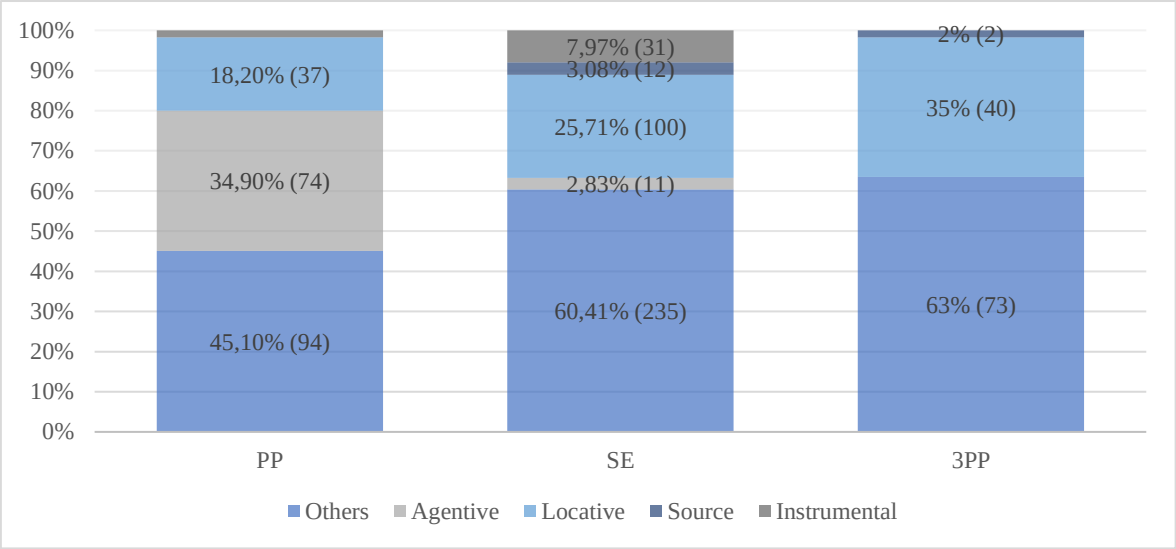
The prototypical oblique complement *by*, *por*, *par*, is not the only complement available to manage agenthood in passive-like structures. Delbecque (2014) shows that in Spanish, in addition to agentive preposition, source, locative and instrumental prepositional phrases are used to provide additional information on the defocused or under-elaborated agent. Before examining in more detail the role of each of these types of prepositional phrases, let us look at Graph 72 which offers the proportion of passive-like structures with and without prepositional phrases. It can be seen that the three constructions resort to them in considerably different proportions.




*Graph 72. Use of prepositional phrases in Spanish passive-like structures*

As indicated in Graph 72, the periphrastic passive is the structure to resort the most to prepositional phrases (64%; 208 occurrences), followed by the *se*- construction (40.3%; 378 occurrences) and the non anaphoric-third person plural (23.1%; 115 occurrences). This greater presence of prepositional phrases in periphrastic passives and *se*-constructions leads us to consider them as a factor participating in the syntactic complexity of the two passive-like structures. As will be detailed in what follows, when accompanied by a prepositional phrase, its category greatly varies from one passive-like structure to another[35].

[35] Locative prepositional phrases and spatiotemporal restrictions overlap in 22 (out of 37) cases for the periphrastic passives, 66 (out of 100) cases for the *se*-constructions and 33 (out of 40) cases for the non-anaphoric third person plurals.




*Graph 73. Distribution prepositional phrases in Spanish passive-like structures*

Graph 73 indicates that the most frequent preposition is the category *others*, which includes all the prepositional phrases that do not serve for agenthood management. Example (282) illustrates this category. The periphrastic passive accompanied by the preposition *para* ('for') does not bring any additional information as to who the agent could be.

(282) *El Rolls-Royce Merlin, de dimensiones aproximadas a las del Allison V-1710, y cuya producción alcanzó enormes proporciones en 1.941,* ***fue seleccionado*** *para el P-40, y durante ese año un P-40D de serie, se equipó experimentalmente con un Merlin 28, adoptando la denominación de XP-40F (Hawk 87D). (WIKI SP)*
'The Rolls-Royce Merlin, whose dimensions approximate those of the Allison V-1710, and whose production reached enormous proportions in 1941, **was selected** for the P-40, and during that year a production P-40D was experimentally fitted with a Merlin 28, adopting the designation XP-40F (Hawk 87D)'.

The non-anaphoric third person plural is the passive-like structure that contains the highest proportion of such prepositional phrase (63%; 73 occurrences). The majority of the other prepositional phrases found with the non-anaphoric third person plural pertains to the category locative (35%; 40 occurrences) exemplified in (283). This result shows that, although the non-anaphoric third person plural construction has an overt agent, it frequently appears alongside a locative device, which provides the interlocutor with further information on which are the possible entities included in the agent.

(283) *Bueno yo a mis perros los coonsidero parte de mi familia, aclaro que no ocupan en mi corazon el lugar que ocupa mi hijo, pero son parte importante en mi vida, bueno no contemplemplo mi vida sin tener a mi lado un peludo amigo. Eso me* ***enseñaron*** *en casa, una casa sin perrito no es un hogar completo* (YAHOO SP)
'Well, I consider my dogs to be part of my family, I must clarify that they do not occupy in my heart the place that my son occupies, but they are an important part of my life, well, I do not consider my life without having a furry friend by

my side. That's what I **was taught** at home, a house without a dog is not a complete home.'

Though in a smaller proportion, the locative preposition is also the second most frequent category in *se*-constructions, after the category *others* (example 284).

(284) *no son los unicos platillos que **se consumen** en la gastronomia mexicana estan por ejemplo la carne a la tampiqueña, cochinita pibil o pescado a la veracruzana,* (…). (YAHOO SP)
'they are not the only dishes **consumed** in Mexican gastronomy, for example, carne a la tampiqueña, cochinita pibil or pescado a la veracruzana, (…).'

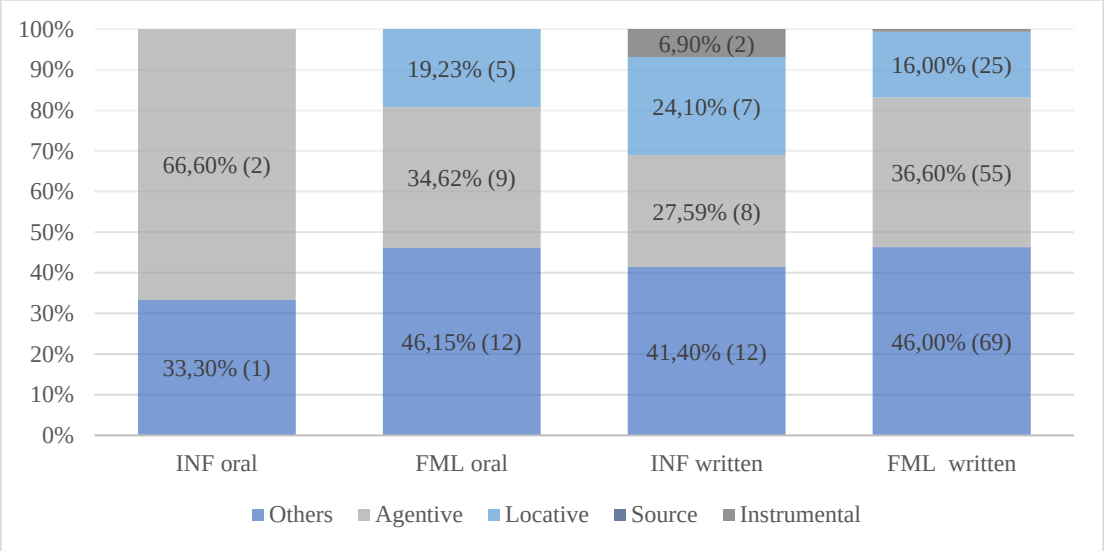
A specificity of the *se*-construction is the use of instrumental prepositions, which is absent or hardly found in the other two passive-like structures. An instrumental preposition is exemplified in (285), viz. sodium sulphide helps in the treatment of kraft.

(285) *Proceso de Kraft . **Se trata** con solución de sulfuro sódico e hidróxido sódico en relación 1:3 durante 2-6 h a temperaturas de 160 -170 ºC.* (WIKI SP)
'Kraft process . It **is treated** with sodium sulphide and sodium hydroxide solution in a 1:3 ratio for 2-6 h at temperatures of 160 -170 ºC.'

Unlike the *se*-construction and the non-anaphoric third person plural, the second most frequent prepositional phrase in the periphrastic passive is not the locative preposition but the agentive one. This result was expected since it is the typical proposition to introduce an oblique complement in periphrastic passives. By-phrases constitute 34.9% (74 occurrences) of the prepositions found with this passive-like structure (see example 286).

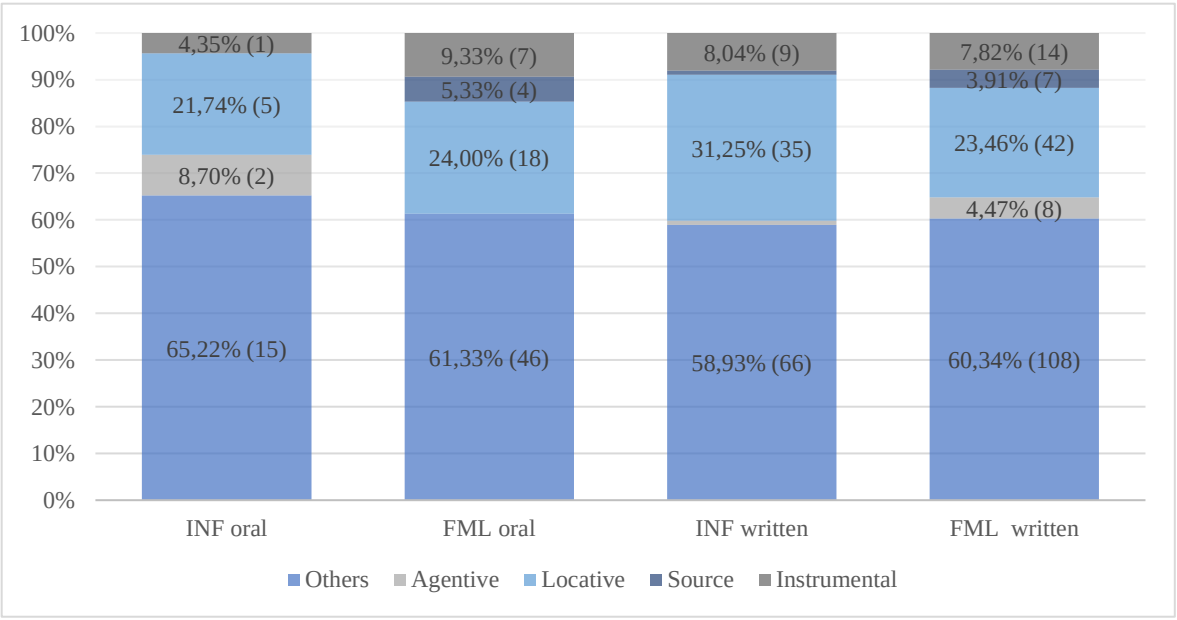
(286) *La reina **es protegida** por los soldados y es fortalecida y alimentada por las obreras, que también construyen el nido y causan el daño a la madera*. (WIKI SP)
'The queen **is protected** by the soldiers and is strengthened and fed by the workers, who also build the nest and cause the damage to the wood.'

The breakdown of the data into mode and register components yields the following findings.




*Graph 74. Distribution of prepositional phrases in Spanish periphrastic passives*

As appears in Graph 74, very few periphrastic passives are accompanied by prepositional phrases in oral productions and in informal written texts. Formal written data is the combination of mode/register that contains the highest number of occurrences of periphrastic passives with prepositional phrases. It can be observed that two categories predominate: *agentive* (36.6%) and *others (*46%*).* The category *locative* gathers the remaining occurrences of prepositional phrases. Due to the relatively low number of figures, no general conclusion can be suggested as for the influence of mode and register on the use of prepositional phrases with periphrastic constructions.




*Graph 75. Distribution of prepositional phrases in Spanish se-constructions*

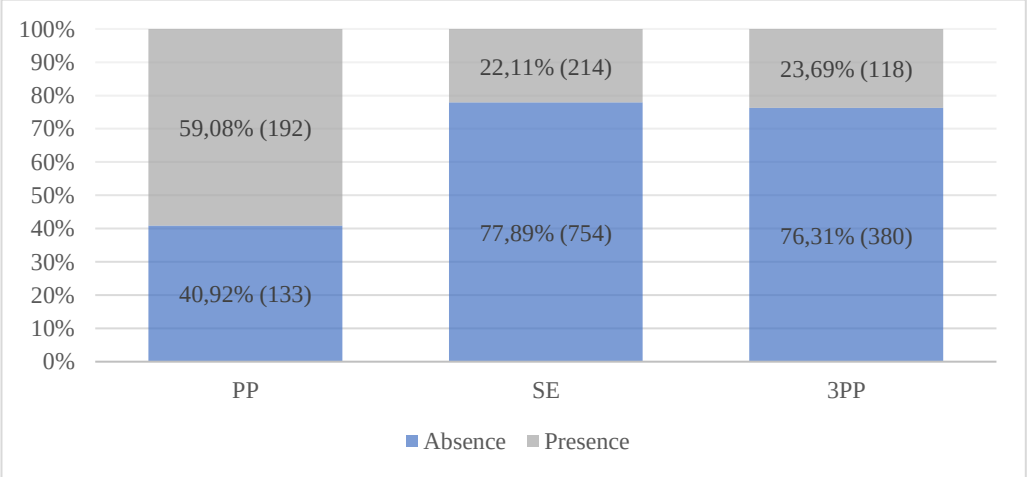
As shown in Graph 75 and confirmed by the calculation of chi-square tests, there is no significant difference as for the distribution of the five categories of prepositional phrases in *se*-constructions. This leads us to assume that register and mode have no major impact on the

category of prepositional phrases in *se*-construction. Similarly, the results obtained for the non-anaphoric third person plural did not show significant differences between informal oral and informal written data, the two combinations of mode/register available for this passive-like structure.

In summary, the three passive-like structures vary according to two factors: the presence/absence of a prepositional phrase in the clause from the one hand and the category of prepositional phrase from the other. While the use of agentive oblique complements is specific to the periphrastic passive, the use of locative prepositions appears as a strategy to provide additional information on the agent in *se*-constructions and non-anaphoric third person plurals. The *se*-construction can, in addition, be characterized by its use of instrumental prepositions. Finally, no influence from mode and register was observed for the distribution of prepositional phrases in these three Spanish passive-like structures.

### 8.4.3. Presence of another passive-like structure

The language-specific MCA plot revealed that, in our Spanish data set, the use of another passive-like structure in the linguistic context is a distinctive feature between the three passive-like structures. Graph 76 indicates that this use of another passive-like structure is specific of the periphrastic passive.



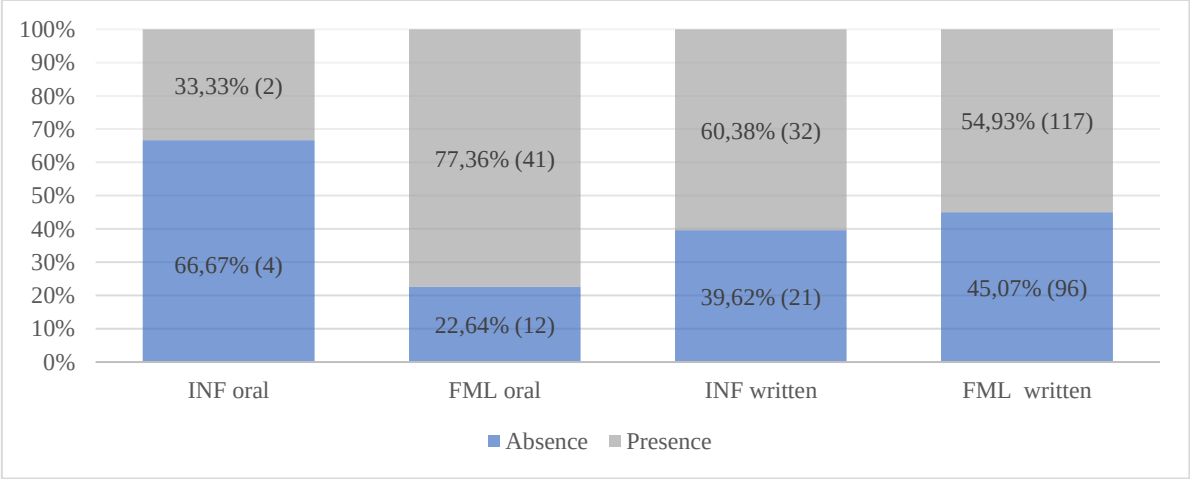

*Graph 76. Presence of another passive-like structure with Spanish passive-like structures*

(287) *Al igual que el lobo ibérico, fue nombrada por Cabrera en 1907, pero en este caso siempre ha habido muchas dudas de su autenticidad, pues* ***fue descrito*** *a partir de varios animales que estaban cautivos en Murcia y no se han observado en estado salvaje. (WIKI SP)*
'Like the Iberian wolf, it was named by Cabrera in 1907, but in this case there have always been many doubts as to its authenticity, so it **was described** from several animals that were captive in Murcia and have not been observed in the wild.'

In (287) the periphrastic passive *fue descrito* ('was described') occurs in the same linguistic context as *fue nombrada* ('was named'), another periphrastic passive, which is, in this case, accompanied by the agentive prepositional complement *por Cabrera* ('by Cabrera). The presence of this other passive-like structure with an overt agent may provide clues on which agents are engaged in the process of *fue descrito.* A third passive-like structure *no se han observado* ('have not been observed') can also bring complementary information on the topic addressed in the interaction.

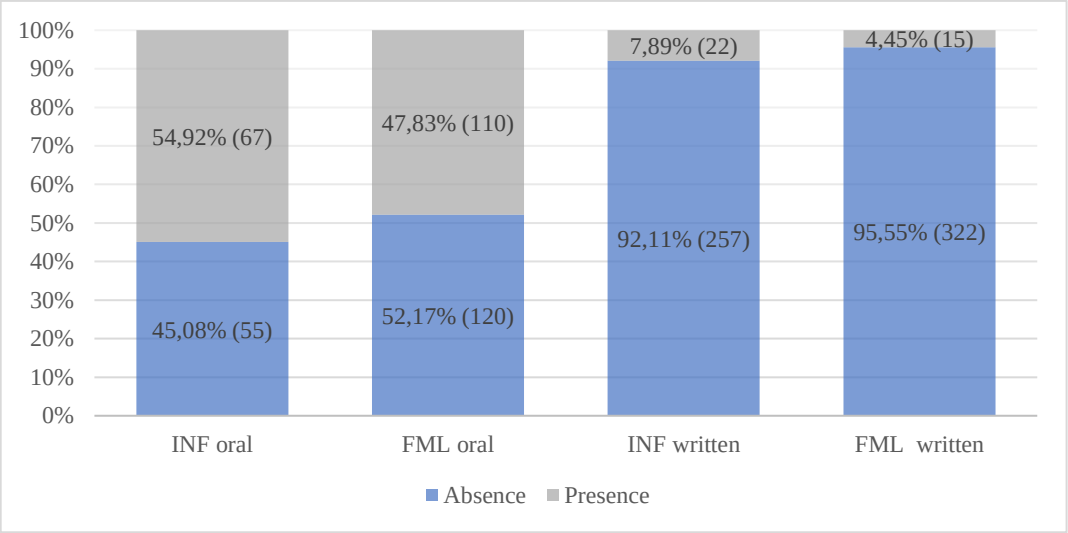
In addition to providing potential further information, the presence of another passive-like structure in the linguistic context can also play the opposite role, viz. to strengthen the impersonal character of the discourse. This case is exemplified in (288) with the double use of a periphrastic passive in the same sentence.

(288) *Son inseminadas una vez al año durante cuatro años, fecha tras la cual su producción de leche desciende y* ***son enviadas*** *al matadero al igual que pasa con sus hijos los terneros. (*WIKI SP)
'They are inseminated once a year for four years, after which their milk production drops and they **are sent** to slaughter as are their calves.'




*Graph 77. Presence of another passive-like structure with Spanish periphrastic passives*

Two important facts emerged from Graph 77. A stable behaviour across written data sets ($\chi^2$ = 0.5112, $p < .05$) can be observed. Compared with written data, a higher percentage of use of another passive-like structure in the near linguistic context can be noted in formal oral data. A chi-square test shows that there is a significant difference between formal oral (77.4%) and formal written (54.9%) data in this respect. An influence of the mode can be thus suspected for the periphrastic passive.




*Graph 78. Presence of another passive-like structure with Spanish se-constructions*

This influence of the mode on the presence of another passive-like structure found for the periphrastic passive is confirmed in the *se*-construction (Graph 78). The difference between written and oral data is significant both in the informal data set ($\chi^2 = 108.7337$, $p < .05$) and in the formal data set ($\chi^2 = 149.654$, $p < .05$). Conversely, there is no influence of the register on the presence of another passive-like structure.

The non-anaphoric third person plural displays a similar trend as results indicate that the presence of another passive-like structure is significantly higher in informal oral data than informal written data ($\chi^2 = 34.8753$, $p < .05$).

To briefly summarise, the presence of another passive-like structure seems to be specific of the periphrastic passive where it appears in 59% of the cases. The detailed analysis of the results also allowed to show that mode plays a preponderant role in the distribution of these additional passive-like structures, with oral contexts being much more likely to contain such linguistic elements. Conversely, the degree of formality does not appear as a key factor regarding this parameter.

### 8.4.4. Other parameters

The language-specific MCA graphs revealed that, in Spanish, the presence of another passive like-structure, the use of prepositional phrases as well as the presence of spatiotemporal restrictions are key factors in the comparison of Spanish passive-like structures. This section summarises briefly the results obtained for the other context-related parameters included in our work, namely the presence of a person reference in the linguistic context, the use of an adverb and the type of content. Though these parameters are not considered as salient by the statistical tests, they can provide some additional information on the behaviour of the studied mechanisms. Let us start with the use of a person reference in the linguistic context as illustrated in (289), where the *se*-construction co-occurs with two references to the second person singular (imperative *usa* 'use' and indicative present *puedes* 'you can'). Through the presence of these

two elements, it can be assumed that the interlocutor is highly involved in the passive-like structure *se aprende (*'you learn').

(289) *Hola usa el flight simulator x es una gran simulador y **se aprende** bastante, ademas puedes jugar con radiocontroladores (...).* (YAHOO SP)
'Hello use the flight simulator x is a great simulator and **you learn** a lot, also you can play with radio controllers (...)*'*

Results suggest that the three Spanish constructions include a person reference in the clause in rather different proportions (15.1% for the periphrastic passive, 35% for the *se*-construction and 50% for the non-anaphoric third person plural). The three structures share the fact that, when they use a person reference, it tends to be either a third person singular or plural. When used with the periphrastic passive and the *se*-construction, this reference tends to provide information on the agent whereas when a person reference occurs with the non-anaphoric third person plural it is assumed that it rather brings information on the interpretation of the patient.

(290) *Normalmente **se cree** que la coloración del pelaje del lobo sirve como mecanismo de camuflaje.* (WIKI SP).
'It is generally believed that the wolf's coat colouring serves as a camouflage mechanism'

The presence of an adverb, as illustrated in (290) with *normalmente* ('generally'), can help, among others, to generalize a situation. However, the analysis revealed that the three passive-like constructions hardly resort to this type of device to provide supplementary information. Only 11.2% of adverbs are found for the non-anaphoric third person plural, 13.2% for the periphrastic passive and 17% for the *se*-constructions. The third context-related parameter, namely the type of content, which can be considered as neutral, restrictive or generalizing, displays little variation between the three constructions as the structures almost exclusively appear in neutral contexts, that is, not containing any elements to guide towards either a specific or a general interpretation.

### 8.4.5. Summary

As for the parameters in relation with the linguistic context, the analysis has revealed the following results. The use of spatio-temporal restrictions, included in this Chapter to ensure a common basis of comparison between the three languages, reveals a stable behaviour across the three structures and the mode/register components in Spanish. The presence of a prepositional phrase and another passive-like structure in the linguistic display much more variation. The analysis shows that these two contextual elements are especially used with the periphrastic passive. An additional difference could be observed as for the type of prepositional phrases. While the most typical one in periphrastic passives is the preposition *por*, the other two structures rely on locative propositions.

## 8.5. Numeral-based *uno*

Before turning to the interim conclusions of this chapter, it is worth briefly commenting on the numeral-based *uno,* a construction frequently associated with and suggested as translational equivalent to the French man-impersonal *on.* We put forward that these two structures, besides having a different etymological root, cannot be considered as equivalents. A first considerable difference between the two is their frequency. While, as explained in Chapter 7, the man-impersonal *on* is one of the most frequent defocussing devices in French (995 occurrences with a relative frequency of 48.83 pttw), such claim cannot be made for the numeral-based *uno* since only 5 occurrences (relative frequency of 0.24 pttw) of the phenomenon were found in our data. Only absolute frequencies are commented in this section, since a very low number of occurrences of the construction was collected. I will first start by describing the verb-related parameters and the types of process, and supplement the description with information on participant-related parameters.

In our data, the construction with *uno,* occurs with mental, relational, material and modulation verbs. Each time only one occurrence of the phenomenon was found in the set of data. One case of *uno* with another semantic role than agent was found (see example 291)

(291) *En el budismo, respecto a matar a otros seres, Buda rechazó incluir el vegetarianismo estricto para los monjes aunque prohibió matar ningún animal o comer de un animal muerto explícitamente para* ***uno mismo*** *(WIKI SP).*
'In Buddhism, regarding killing other beings, Buddha refused to include strict vegetarianism for monks although he forbade killing any animal or eating a dead animal explicitly for **oneself.**'

The four verbs involved in the constructions with *uno* are dynamic, two of them occur in a durative situation and other two in a punctual one. Regarding the aspect, three cases of *uno* are found with a perfective form and only one with an imperfective form. All the occurrences appear in a realis and affirmative clause (see example 292).

(292) *La raíz latina que significa "pez" es pisci y del "pollo" es pulli. Pesce-pollotarianismo (o chickifishitarian) es un neologismo peyorativo que significa que* ***uno incluye*** *pollo y pez como productos no carnicos, pero pescetarianismo y pollo-vegetarianismo son entidades separados. (WIKI SP)*
'The Latin root for "fish" is pisci and for "chicken" is pulli. Pesce-pollotarianism (or chickifishitarianism) is a pejorative neologism meaning that **one includes** chicken and fish as non-meat products, but pescetarianism and chicken-vegetarianism are separate entities'. (WIKI SP)

When it comes to participants, *uno* appears more frequently in two-participant clauses. The action is always performed by an animate human agent. There is a patient in three out of the four clauses. In the present data, the patient pertains to the category 'inanimate others' and is always totally affected.

## 8.6. Portraying the Spanish system

Table 28 recapitulates the findings obtained for the Spanish passive-like structures. The major characteristics of each structure, selected on the basis of chi-square tests and significant differences, have been highlighted in bold.

| | | Periphrastic passive | *Se*-construction | Non-anaphoric third person plural |
|---|---|---|---|---|
| Verb-related parameters | Tense | present (58.2%) | **present (74.4%)** | present (59%) |
| | Aspect | **imperfective (62.5%)** | imperfective (80.4%) | imperfective (72.9%) |
| | Mode | indicative (74.8%) | indicative (85.6%) | **indicative (92%)** |
| | Type of verb process | material (55%)<br>relational(18.5%)<br>mental (12.3%) | material (45%)<br>relational(22%)<br>mental (15%) | **relational (32%)**<br>material (30%)<br>**verbal** (20%) |
| | Kinesis | **dynamic (90.8%)** | dynamic (84.1%) | dynamic (82.5%) |
| | Realis/irrealis | realis (80.9%) | realis (81.3%) | realis (82.9%) |
| | Affirmation | affirmative (92.2%) | affirmative (93.2%) | affirmative (89.9%) |
| Participant-related parameters | No. of participants | 2+ participants (100%) | 2+ participants (88.5%) | **2+ participants (76.3%)** |
| | Agent | **defocused (78.8%) + inanimate *others*/animate human** | **defocused (100%)** | **animate human (100%)** |
| | Patient | inanimate (74.8%) | **inanimate (94.2%)** | inanimate (66%) |
| | Individuation P | non individuated (55.4%) | **non individuated (88%)** | non individuated (57.9%) |
| | Affectedness P | affectedness (99.1%) | affectedness (100%) | affectedness (95.5%) |
| Context-related parameters | Spatio-temporal restriction | presence (19.7%) | presence (20%) | presence (23.9%) |
| | Prepositional phrase | **presence (64%)** | presence (40.3%) | presence (23.1%) |
| | Other passive-like structure | **presence (59.1%)** | presence (22.1%) | presence (23.7%) |

*Table 28. Overview of the distinctive features of Spanish passive-like structures*

The first observation to be made is that, as far as verb-related parameters are concerned, the periphrastic passive, the *se*-construction and non-anaphoric third person plural have their specificities. The *se*-construction can be characterized by a higher use of present tenses compared to the other two structures. A low proportion of imperfective aspect is specific to the periphrastic passive while the *se*-construction is the structure that resorts the most to it. As for verb mode, the non-anaphoric third person plural is the structure that uses the most the indicative modality. This passive-like structure is also the one that, compared to the *se*-construction and the periphrastic passive, differs greatly as far as the type of verb process is concerned. Verbs in the non-anaphoric third person plural consist in relational processes and material processes in similar proportion as well as in verbal processes. Conversely, verbal processes are hardly used in the other two passive-like structures. A last interesting difference is the high use of dynamic verbs in periphrastic passives. Two parameters display a stable behaviour across the three passive-like structures: dichotomy *realis/irrealis* and *affirmation*. However, as will be reminded afterwards, their distribution across mode and register differs according to the construction. Let us now focus on the features related to the participants.

The MCA has revealed that participant-related parameters, and more precisely those related to the number of participants and the patients, are a set of parameters contributing greatly to explaining the variation between the structures. Table 28 indicates indeed that the number of participants varies considerably according to the structure, the non-anaphoric third person plural resorting the least to two-participant clauses. As for the nature of patients, the *se*-construction is the construction to resort to the highest number of inanimate patients, which seems to correlate with the high number of non-individuated patients. The degree of affectedness of the patient does not seem to vary across the three passive-like structures. As for the agent, it is always an animate human in non-anaphoric third person plurals and is most of the time defocused in the *se*-construction and the periphrastic passive. When the agent does appear in the periphrastic passive it belongs mainly to either the category animate human or inanimate *others*.

The analysis of the last set of parameters, those related to the linguistic context, has revealed the following results. While the use of the spatio-temporal restrictions does not vary across the three passive-like structures, the use of prepositional phrases and other passive-like structures in the linguistic context differs according to the structure. The periphrastic appears as the structure to resort most to these two elements. While the use of the agentive oblique complement is specific to the periphrastic passive, the use of locative prepositions appears as a strategy to provide additional information on the possible agent in *se*-constructions and non-anaphoric third person plurals.

The influence of the mode and the register on the structures depends on the parameter. It could be observed that for some parameters such as *tense*, *aspect*, *verb mode* or dichotomy *realis*/*irrealis,* register and mode are important parameters to take into account. Their influence has appeared as slightly less marked in other parameters, especially those related to the linguistic context.

So far, the examination of the different syntactico-pragmatic parameters has allowed to draw a clear portrait of three agent-defocussing mechanisms in Spanish as far as verb-related,

participant-related and context-related features are concerned. In order to compare the three languages on a common basis, the next chapter aims at providing a similar account of the English system to further complete the overall panorama of agent-defocussing mechanisms.

# 9. Passive-like structures in English

This last language-specific analytical chapter presents the results obtained for the English part of the data. Four passive-like structures have initially been included in this study: the periphrastic passive, the non-anaphoric third person plural, the *get*-passive and the construction based on the numeral *one*. As shown in Table 29, the relative frequency of two of them, the *get*-passive and the construction based on the numeral one, does not allow to carry out a reliable quantitative analysis.

| Structure | Raw frequency | Relative frequency (pttw) |
|---|---|---|
| **Periphrastic passive** | 1415 | 64.5 |
| **Non-anaphoric third person plural** | 186 | 8.48 |
| ***Get*-passive** | 49 | 2.2 |
| ***One*** | 9 | 0.41 |

*Table 29. Raw and relative frequencies per ten thousand words (pttw) in English data*

Since the *get*-passive displays a lower relative frequency compared to the periphrastic passive and the non-anaphoric third person plural; and as only 49 occurrences were obtained in our corpora, the main results are presented in a short separate section at the end of this chapter and only tentative conclusions will be suggested. As for the construction based on the numeral *one*, which displays an even more restricted number of occurrences, the results will be briefly addressed at the end of this chapter. Two passive-like structures will thus be analysed in detail in this chapter: the periphrastic passive and the non-anaphoric third person plural

The chapter is structured as follows. Starting with general findings yielded by the Multiple Correspondence Analyses (Section 9.1), it then focuses on verb-related, participant-related and context-related parameters (Sections 9.2 – 9.4). Section 9.5 summarises the results obtained for the *get*-passive, considered as a possible alternative to the periphrastic passive. Section 9.6 offers a qualitative overview of the numeral-based *one*. The chapter closes with a general portrait of the passive-like structures in English (9.7).

## 9.1. Panoramic overview

This section details the results of the Multiple Correspondence Analyses. It presents four different graphs: a biplot containing the distribution of the different parameters, a factor map visualising the distribution of the passive-like structures, a factor map patterning the influence of mode and register, and a graph indicating possible correlations between the parameters. These four types of graphs allow to capture a first overall picture of how the different parameters, structures and mode/register combine. Though they will not be commented on in this section, or very briefly in the case of the *get*-passive, the passive-like construction based on *one* and the *get*-passive are included on the graphs for informative purposes.

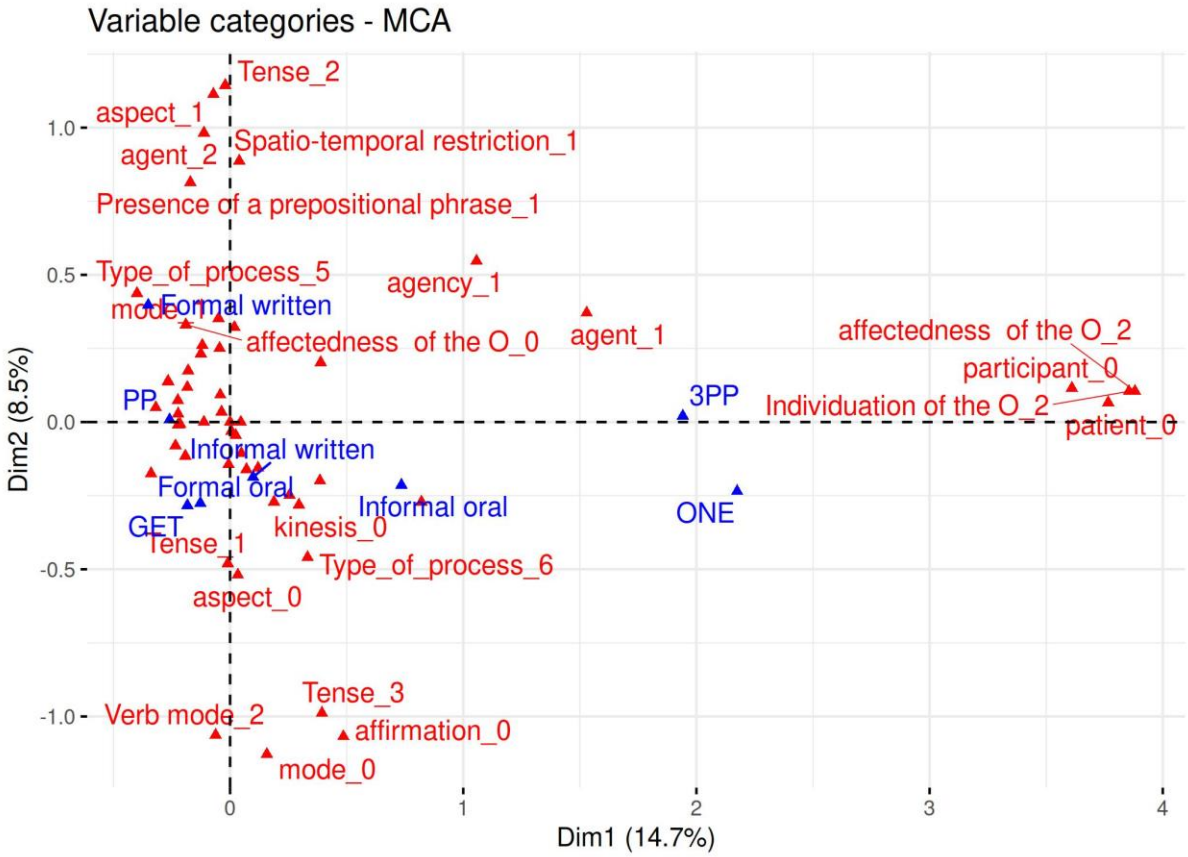



*Figure 22. Distribution of parameters in English data*

This figure visualises the distribution of the different parameters (represented in red on the graph) and their possible affinities. The parameters that contribute the most to explain the variation between the different passive-like structures are situated on the extremes of the graph. The following information can be obtained from Figure 22. If we first start with the horizontal axis, it can be observed that a group of parameters, in close relation with patients, stands out. The number of *participant* (category 0; one-participant clause), *patients* (category 0; absence of a patient) and *affectedness* and *individuation* of the patient (category 2; individuated) appear as highly associated and contributing factors. As indicated in Figure 22, this group of variables is specific to the non-anaphoric third person plural and to *one*, though data are scarce for that passive-like structure. Alongside the Y-axis, we can observe that several parameters, linked to the verb, the participants and the linguistic context, are likely to vary considerably. The position

of *aspect* (category 1; perfective) and *tense* (category 2; past) at the top of the plot suggest that these two verb-related parameters will largely impact the distribution of the structures. Two parameters related to the linguistic contexts, the presence of a spatiotemporal restriction and of a prepositional phrase, will also play a preponderant role. A last valuable information is the presence of the variable *agent* (category 2; presence). All these parameters show greater affinities with formal register and with periphrastic passive. At the bottom of the Y-axis, four parameters appear as important elements in the variation: Verb mode (category 2; past), dichotomy *realis/irrealis* (named *mode* in lowercase on the graph) (category 0; *irrealis*), affirmation (category 0; negative clause) and tense (category 3; future). As will be detailed in this chapter all these parameters are somehow linked to the way speakers encode reality.

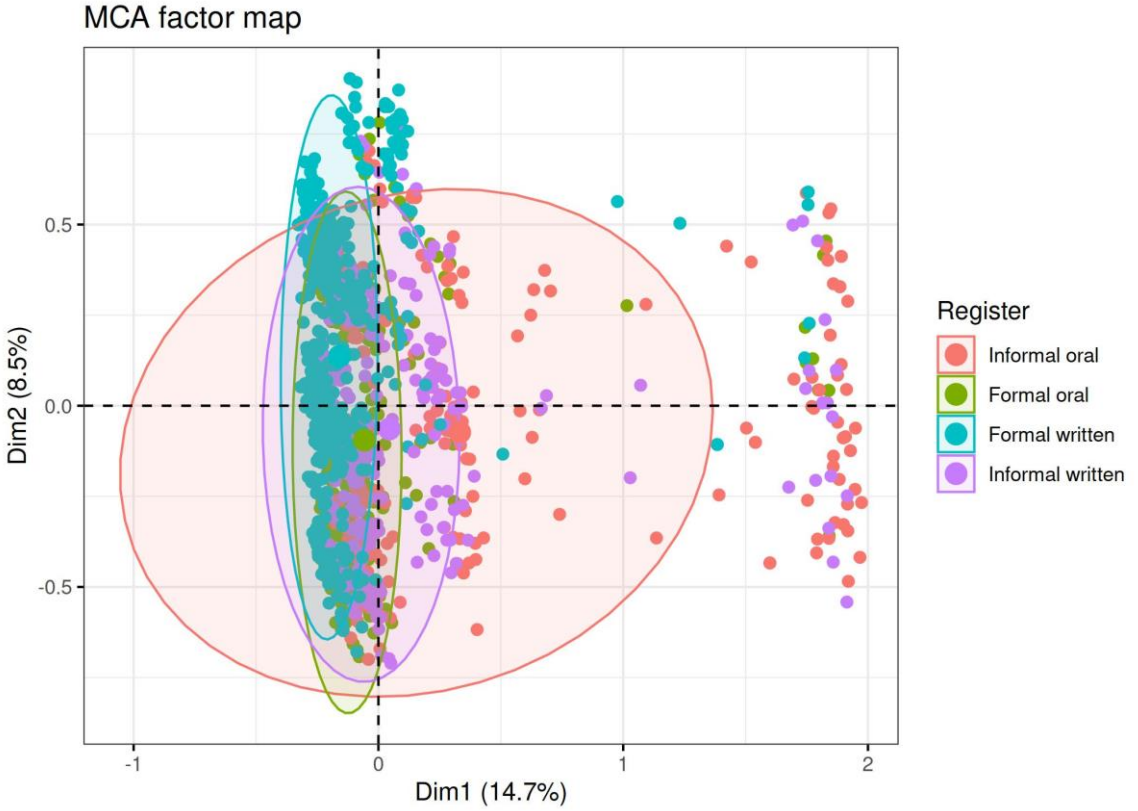



*Figure 23. Factor map for mode and register in Spanish*

Figure 23 allows to visualise all the individuals, namely all the occurrences of passive-like structures in formal/informal registers and oral/written modes. Concentration ellipses are added around each group to better delimit the different areas of each register/mode combination. It can be observed that while formal oral, formal written and informal written data occupy a similar position on the graph, the confidence ellipsis of informal oral data differs greatly. This means that informal data are more likely to present a different behaviour compared to the other three sets. It is interesting to note that the ellipsis of formal oral data resembles more the ones of written data than the one of informal oral data. Therefore, we can expect formal oral data to share a large series of characteristics with the written mode. A last additional information provided by Figure 23 is the specific association of the parameters placed at the top of Y-axis – i.e., *tense*, *aspect* and *spatio-temporal restrictions* with the formal written register. Particular

attention will be devoted to these points when analysing the behaviour of the two passive-like structures in English.

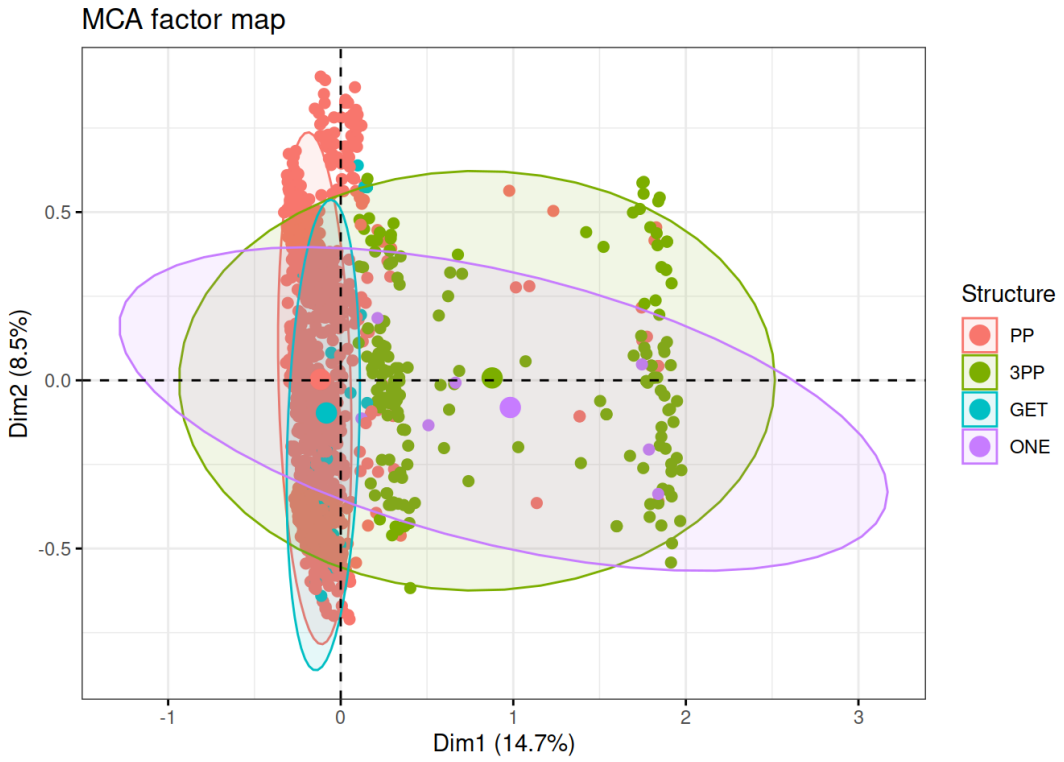



*Figure 24. Distribution of the four passive-like structures in English*

Figure 24 reveals that while the characteristics of the periphrastic passive develop alongside de Y-axis (red ellipsis), those of the non-anaphoric third person plural (green ellipsis) rather extend alongside the horizontal axis. Both passive-like structures, nevertheless, do overlap on some areas of the plot, which means that the two structures will share a series of features. It is expected that the parameters at the top of the vertical axis will be specific of the periphrastic passive and those at the extreme of the horizontal axis will best describe the non-anaphoric third person plural. Though the number of occurrences obtained for the *get*-passive and *one* is more limited, Figure 24 indicates that the *get*-passive largely overlaps with the periphrastic passive whereas the ellipsis of *one* is more comparable to the one of the non-anaphoric third person plural. Two pairs of passive-like structures emerge, therefore, from Figure 24: on the one hand the periphrastic passive and the *get*-passive and on the other, the non-anaphoric third person plural and *one*.

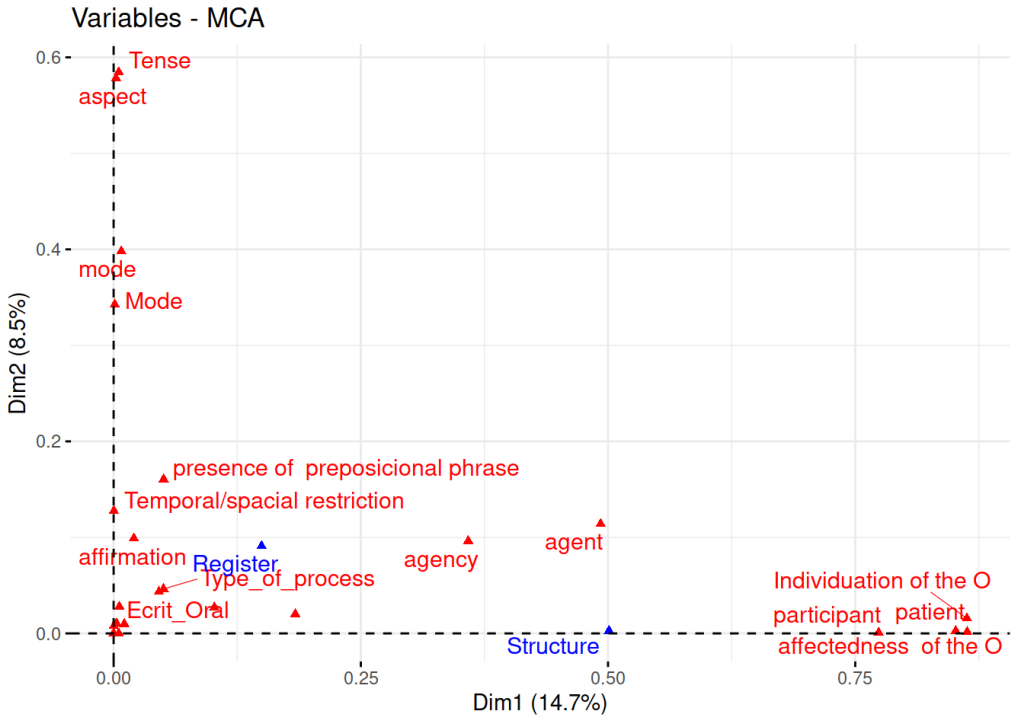



*Figure 25. Correlation(s) in English data*

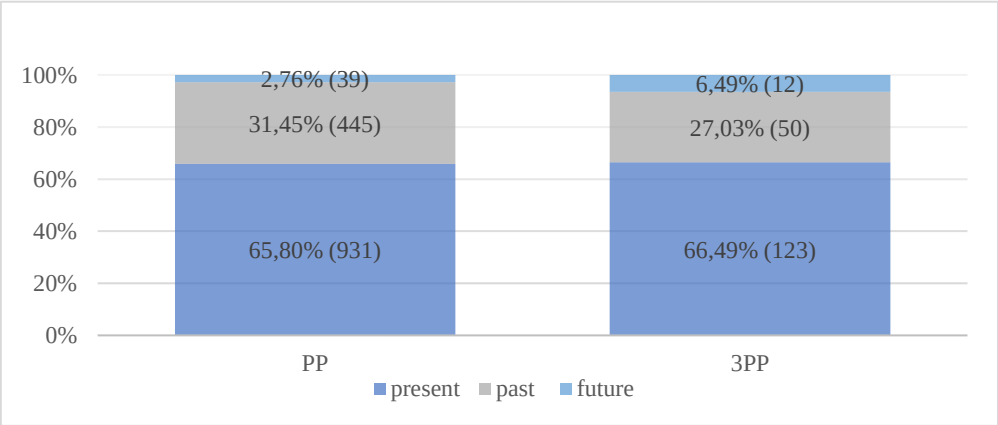
Figure 25 allows to see whether potential correlations between parameters or pairs of parameters can be established. To be considered as correlated, parameters have to comply with two criteria: be situated between 0.5 and 1 and be positioned close together on the graph. Two different groups of parameters meet these two requirements. On the vertical axis, *tense* and *aspect*, situated at 0.6, constitute a first group of correlated parameters. On the horizontal axis, *patient*, *individuation of patient* and *affectedness of the patient* form another series of correlated variables. Particular attention will be given to these two sets of parameters to see how they correlate.

## 9.2. Verb-related parameters

This section examines seven parameters related to the verb and selected according to the general MCA: *tense* (9.2.1), *aspect* (9.2.2), *verb mode* (9.2.3), *type of verb process* (9.2.4), *kinesis* (9.2.5), dichotomy *realis/irrealis* (9.2.6) and *affirmation* (9.2.7). The results are complemented with information of less contributing verb-related parameters in 9.2.8. The section closes on a summary of the findings in 9.2.9.

### 9.2.1. Tense

As already seen in the two previous analytical chapters, tenses have been divided into three main categories: present, past and future. In the periphrastic passive and the non-anaphoric third person plural, tenses are distributed in the following manner (see Graph 79):




*Graph 79. Distribution of present, past and future tenses in English passive-like structures*

Present tenses account for approximately 66% of the data, followed by past tenses (between 27% and 31%) and future tenses (between 2.8% and 6.5%). Examples (293) and (294) illustrate present tenses, the most common category, in a periphrastic passive and a non-anaphoric third person plural.

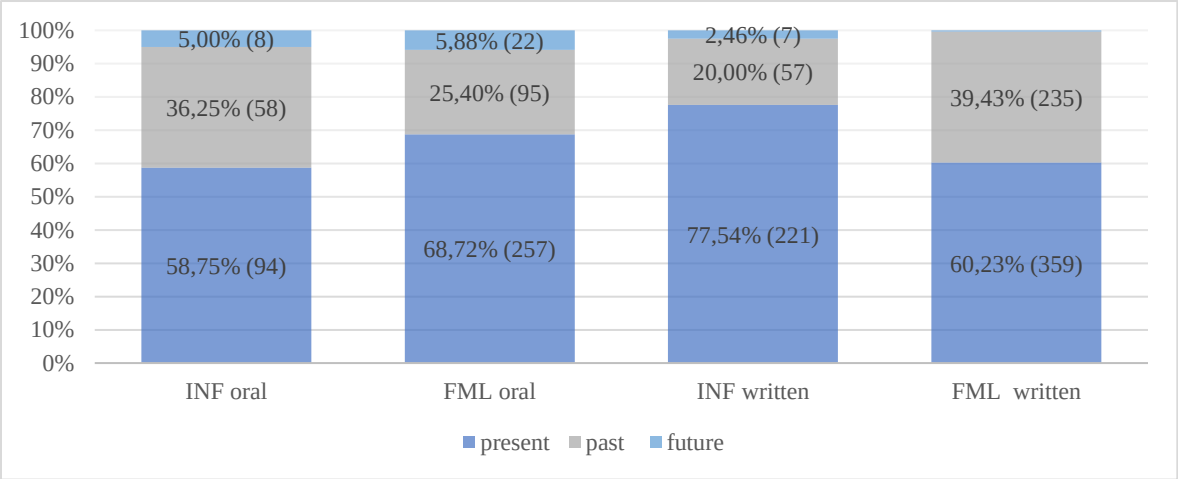
(293) ***It is used*** *as a working dog for tasks including detection of odors such as explosives , accelerants ( for arson investigation ) , and narcotics ; tracking of humans for suspect apprehension in police work ; and search and rescue* . (WIKI EN)

(294) *You will need a pilot's licence and then find a museum where* ***they still fly*** *Spitfires*. (YAHOO EN)

Chi-square tests reveal that the differences between the proportions of tenses in the two passive-like structures is only significant for the category *future* ($\chi^2 = 7.3776$ $p < .05$): the non-anaphoric third person plural resorts more frequently to future tenses (illustrated in 295) than the periphrastic passive.

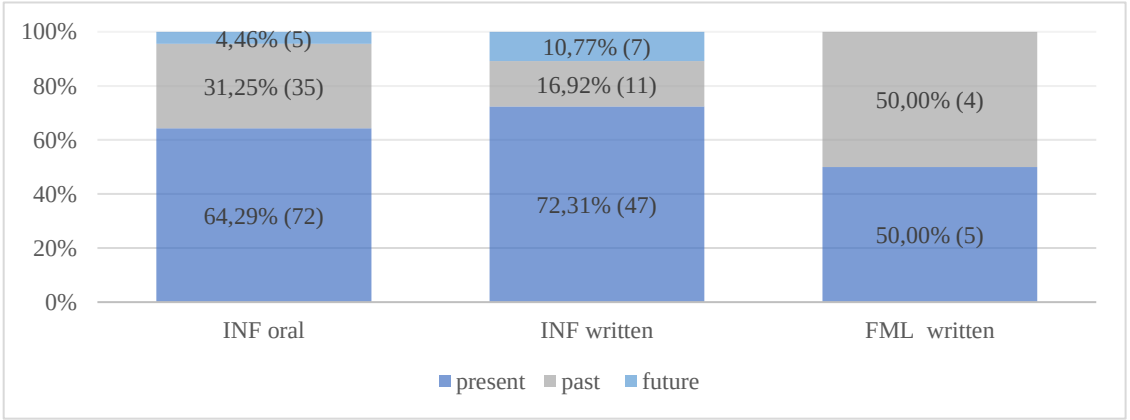
(295) *"S0037" trans="overlap">but you are right it is rained all day today</u>*
*<"S0115">Cos cos it is it seems that the volume is there you I mean</u>*
*<"S0037">yeah</u>*
*<"S0115">somehow anyway</u>*
*<"S0037">but I bet* ***they will still water*** *the grass at King's and all that they will just pay it will not they?</u>* (BNC EN)

In addition to this small difference for the use of future tenses, the distribution of the parameter *tense* across the two passive-like constructions remains stable. Let us now examine whether the mode and the register impact the use of tenses with these structures in a different way.




*Graph 80. Distribution of tenses in English periphrastic passives*

Graph 80 indicates that tenses in periphrastic passives receive an influence from both the register and the mode. A look at oral data reveals that formal situations include significantly more present tenses than informal ones ($\chi^2 = 4.9416$. $p < .05$). This smaller proportion of present tenses in informal oral data is compensated by a higher use of past tenses. The frequency of use of future tenses is stable across the two registers. Unexpectedly, the opposite pattern is found in the written data set. Informal written data use significantly more present tenses than formal written ones ($\chi^2 = 25.1088$. $p < .05$). The graph shows that in informal data, written data use much more present tenses whereas in formal data, present tenses predominate in oral data. It can be concluded that the distribution of tenses varies considerably according to the degree of formality and the mode of production. However, no clear overall pattern emerged as each of the data set appears has having its own characteristics.




*Graph 81. Distribution of tenses in English non-anaphoric third person plurals*

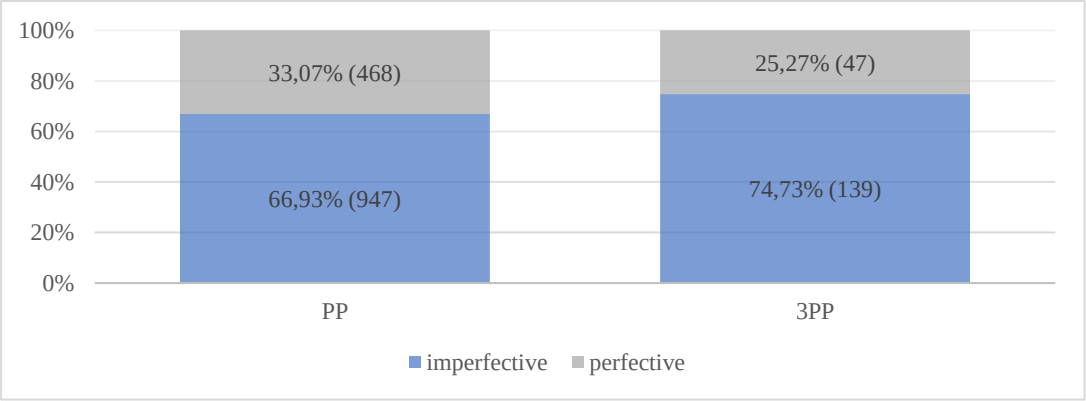
As shown in Graph 81, a higher use of present tenses can be observed in informal written data compared to informal oral ones. Though the difference is not significant ($\chi^2 = 1.2014$. $p > .05$), the results concur with those obtained for the periphrastic passive. Compared to informal oral data, the informal written set can be characterized by a restricted use of past tenses and a greater presence of future tenses. As pointed out by Biber (1995), tense is one of the factors usually considered when examining differences between registers. This analysis shows that, as for the

use of passive-like structures in English, a relatively high proportion of present tenses characterises the informal written data.

To briefly summarise, tenses in periphrastic passives and non-anaphoric third person plurals are distributed rather similarly, with a predominance of present tenses. However, breaking down the results into mode and register components reveals that the parameter *tense* is highly sensitive. Each mode/register combinations appears to be having its own characteristics: a higher use of present tenses in informal written data and a marked presence of past tenses in informal oral ones.

### 9.2.2. Aspect

*Aspect* is another syntactic characteristic that may come into play in establishing the profile of the passive-like structures. In a few words, situations and events are classified according to the way they are presented. They can either be seen as accomplished, viewed from the endpoint (perfective) or as denoting incompletion (imperfective).




*Graph 82. Distribution of imperfective and perfective forms in English passive-like structures (parameter aspect)*

Graph 82 indicates that the imperfective aspect predominates both in the periphrastic passive (66.9%; 947 occurrences) (296) and in the non-anaphoric third person plural (74.7%; 139 occurrences) (283). In (296) it is stated that the origins of the dog breed are not known, there is no specific endpoint in time, the situation is still ongoing and reflects incompletion. Similarly, the non-anaphoric third person in (297) expresses an ongoing situation. It is a current fact that China possesses a large quantity of bamboo trees.

(296) *Its exact origins **are not known** but it appears that some possibly spaniel and setter-type dogs, retriever-type dogs , and farm collie may have gone into the mix (WIKI EN)*

(297) *How have chopsticks help china? how have they solve problems </question> (…) <answer from>{S} </answer from> <answer> **they have** a lot of bamboo which is cheaper than copper.they are billions imagine if most of them using metal cutlery's. </answer> (YAHOO EN)*

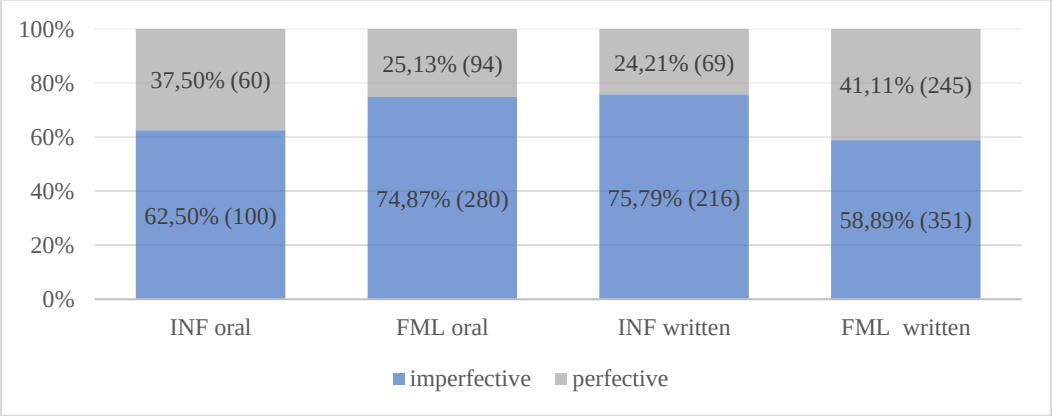
The calculation of a chi-square test shows that the English periphrastic passive uses significantly more perfective forms than the non-anaphoric third person plural ($\chi^2 = 4.59$. $p < .05$). Example (298) illustrates a case of periphrastic passive used with a perfective aspect. Unlike (296) and (297) where the situation has no specific endpoint, the action to remove the Spitfire plane in (298) took place years ago and the process is not on-going anymore but considered as accomplished.

(298) *<answer>{S}c The Spitfire **was retired** from the USAF more than 50 years ago. </answer> (YAHOO EN)*

As can be seen in the MCA graph, *tense* largely correlates with the parameter *aspect.* Results show that the proportion of past tenses, usually associated with the perfective aspect, is higher in the periphrastic passive than in the non-anaphoric third person plural. Present and future tenses are thus assumed to correlate with an imperfective reading while past tenses correlate with a perfective one.

In their study on impersonal passives in Swedish and Dutch, two Germanic languages, Johansson and Rawoens (2019) report that this passive construction is usually used with atelic verbs or verb phrases. A similar trend is observed in our data since the imperfective aspect is dominant in both passive-like constructions. The researchers specify that, when used with a telic verb, impersonal passives tend to occur with an adverbial modifier. As very few occurrences of the English impersonal passive, constructed with the expletive *it*, were found in our data, the correlation between a perfective aspect and an adverbial modifier could not be verified. However, results seem to show that, in our data coded as periphrastic passives, when used in a situation considered as perfective, 25.8% of the occurrences are accompanied by a spatio-temporal modifier while this figure falls to 8% in imperfective situations. This finding goes in the same direction as what is reported by Johansson and Rawoens (2019) on the use of telic verbs and adverbial modifiers with impersonal passives.

Interesting results appear when examining the influence of the mode and the register on the distribution of the parameter *aspect*.



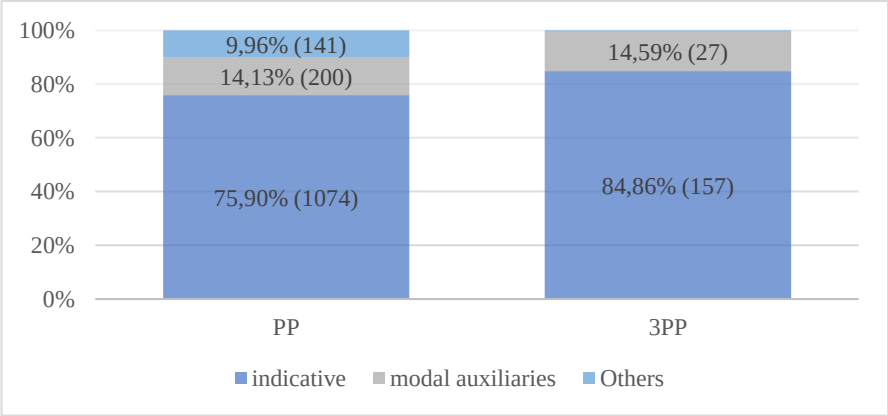

*Graph 83. Distribution of imperfective and perfective forms in English periphrastic passives*

For the periphrastic passive, Graph 83 indicates that there is a major influence of both the register and the mode, though the imperfective aspect prevails in all datasets. While in the oral data set, formal data resort to significantly more imperfective forms than informal ones, the

reverse pattern can be observed in written data[36]. As for the impact of mode, it can be observed that written data contain significantly more imperfective forms in informal contexts while, in formal situations, oral data totalize the highest number of imperfective forms[37]. Rather unexpectedly, the pattern found in formal oral data resembles the one of informal written data whereas informal oral data resemble formal written ones in this respect. This significantly higher use of imperfective forms in informal written data (84.6%; 55 occurrences) compared to informal oral data (70.8%; 80 occurrences) is also found in the non-anaphoric third person plural. This leads us to assume that, in addition to differences in proportions between periphrastic passives and non-anaphoric third person plurals, *aspect* also undergoes a strong impact of the degree of formality and the mode of production. Though largely correlated with the parameter *tense*, the behaviour of *aspect* appears as much more heterogeneous depending on the structures and the combination mode/register. These results confer the parameter *aspect* with an essential role in explaining the variation between the passive-like structures.

### 9.2.3. Verb mode

*Verb mode* relates to the way the verb expresses the event or the situation. In this study, four categories have been created: indicative, subjunctive (absent from English data), modal auxiliaries and *others.*




*Graph 84. Distribution of verb modes in English passive-like structures*

A first relevant finding is the absence of the subjunctive. Neither the periphrastic passive nor the non-anaphoric third person plural resort to that verb mode. This may be attributed to the fact that subjunctive forms are hardly used in English compared with the Romance languages included in this study. Results seem to indicate that nuances such as probability, possibility or eventuality are rather expressed through modal auxiliaries.

As suggested by Graph 84 the distribution of verb modes in periphrastic passives clearly differs from the one observed in the non-anaphoric third person. The non-anaphoric third person uses

[36] Significant for the pair: formal oral vs informal oral ($\chi^2 = 8.3504$. $p < .05$) and for the pair: formal oral vs informal written ($\chi^2 = 8.7915$. $p < .05$).
[37] Significant for the pair: informal oral vs informal written ($\chi^2 = 23.9975$. $p < .05$) and for the pair: formal oral vs formal written ($\chi^2 = 25.7912$. $p < .05$).

significantly more indicative forms than the periphrastic passive ($\chi^2$ = 8.1636. p < .05). Examples (299) and (300) illustrate cases of indicative mode in the two passive-like structures.

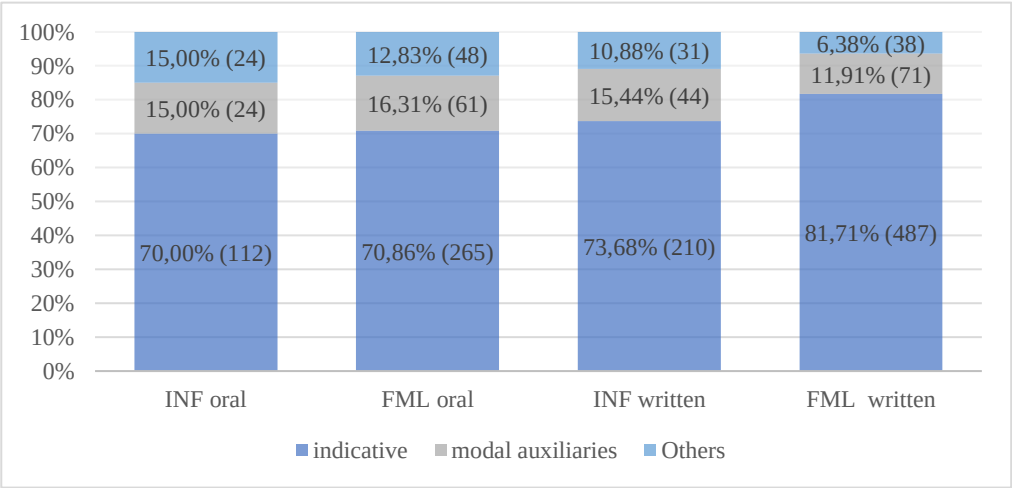
(299) *The breed **was** originally **known** as the Little River Duck Dog or the Yarmouth Toller* . (WIKI EN)

(300) *Dangerous dog breeds?{S} **Do they not have** any Blue Heelers or Jack Russell Terriers in NSW? http://au.news.yahoo.com/a/-/latest/5124...{S} What do you think?{S} Aside from the statistics, do you believe these particular breeds are the most dangerous?{S}* (YAHOO EN)

Since the proportion of modal auxiliaries appears to be stable across the two structures, the higher use of indicative forms in the non-anaphoric third person plural may be explained by the fact that the periphrastic passive includes 9.96% (141 occurrences) of forms classified as *others*, which are mainly infinitives, *as* exemplified in (301) with *to be retrieved*. A look at Graph 87 shows that these forms are absent from the non-anaphoric third person plural.

(301) *Tollers are named for their ability to entice or lure waterfowl within gunshot range , called tolling .{S} The hunter stays hidden in a blind and sends the dog out to romp and play near the water , usually by tossing a ball or stick **to be retrieved**.* (WIKI EN)

When the breakdown of the data through mode and register components is operated, the following results emerge.




*Graph 85. Distribution of verb modes in English periphrastic passives*

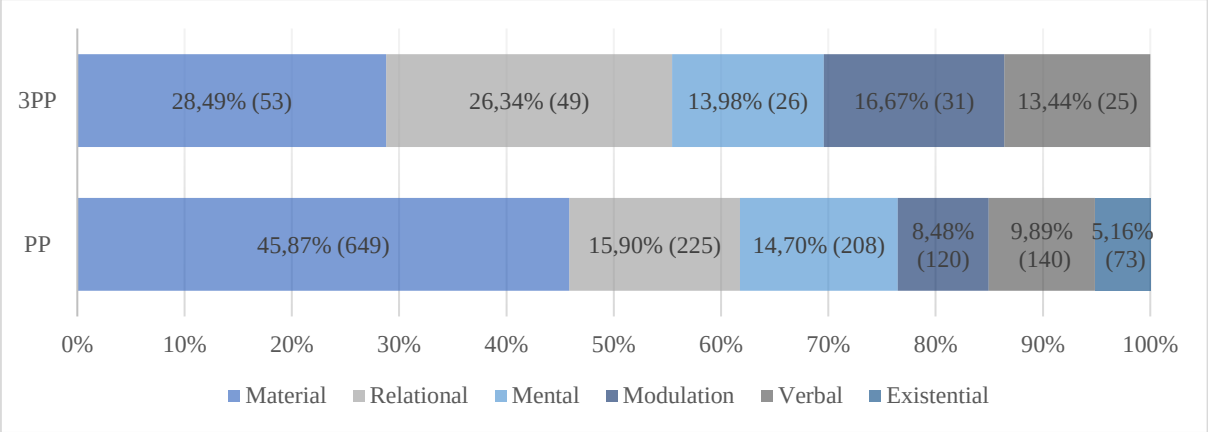
For the periphrastic passive, while the degree of formality does not impact oral data sets, Graph 85 shows that it influences the distribution of the parameter *verb mode* in written data. A chi-square test reveals that formal written data use significantly more indicative forms than informal written data ($\chi^2$ = 7.5187. p < .05). The comparison of the formal written data set with the formal oral one, reveals that written data resort to a higher proportion of indicative forms ($\chi^2$ = 15.5431. p < .05). Conversely, oral data and informal productions do not display any significant differences.

The analysis of the non-anaphoric third person plural reveals that there is no significant difference depending on the register or the mode.

Therefore, it can be suggested that the impact of mode and register on the distribution of *verb mode* is relatively restricted since the mode of production is only affecting formal data and the degree of formality only impacts written data in periphrastic passives. On the other hand, we see that the type of passive-like structure is an important element to consider as the periphrastic passive and the non-anaphoric third person plural display a different behaviour regarding their use of verb mode forms.

### 9.2.4. Type of verb process

As far as the type of verb process is concerned, the periphrastic passive and the non-anaphoric third person plural display a different picture.




*Graph 86. Distribution of verb processes in English passive-like structures*

Graph 86 reveals that the most common type of verb occurring with periphrastic passives are verbs related to *material* processes (45.9%; 649 occurrences). This phenomenon is not observed in the non-anaphoric third person plural since there is no predominant type of verb as is the case for the periphrastic passive. Nevertheless, in the non-anaphoric third person plural, two categories, in relatively similar proportions, totalize more than half of the data: *material* processes (28.5%; 53 occurrences) and *relational* processes (26.3%; 49 occurrences). Relational processes are also present in periphrastic passives although in smaller numbers (15.9%; 225 occurrences). Let us first examine these two types of verb processes.

A closer look at the periphrastic passive shows that, within the category *material*, verbs mainly express *changes* (17.4%; 249 occurrences) (illustrated in 302), denote *space* (13.8%; 195 occurrences) (exemplified in 303) or fall in the category *others* (11.2%; 159 occurrences)

(302) *Jets **could be made** to run on gasoline but kerosene is used because they burn so much fuel that gasoline would be cost prohibitive </answer> (YAHOO EN)*

(303) *There will not be any "flying cars," in the classic sense of a flying vehicle that is as simple to operate as an automobile and that **can be parked** in--and flown*

*from— the typical driveway to provide aerial transportation for you and your family. (YAHOO EN)*

The category *others* includes verb processes denoting control, uses or activities. The following examples respectively illustrate a periphrastic passive using a verb of control (304) and a verb of use (305).

(304) *Can you learn to fly a glider in the UK if you have Diabetes?{S} I have type 1 diabetes that* ***is controlled*** *using insulin (YAHOO EN)*

(305) *<answer>{S} It could be literally anything... have they ever done full blood-work on him?{S} Tested everything that needs* ***to be tested****?{S} Sounds a little like how my dog acted when she had a liver problem,(...). (YAHOO EN)*

A glance at material verbs used in non-anaphoric third person plurals reveals that they mainly belong to the domain of *space* (12.4%; 23 occurrences out of the total data) (Example 306).

(306) *It is a fantastic ambition but one that only a lottery winner may be able to satisfy.{S} You will need a pilot's licence and then find a museum where* ***they still fly*** *Spitfires. (YAHOO EN)*

In (306), the process of flying involves a move through the air, thus a change in the spatial location. The other types of *material* processes in non-anaphoric third person plurals totalize a relatively restricted number of occurrences and will not be further commented on.

Turning to *relational* processes (divided in verbs of *possession* and verbs of *attribution*), which appear as one of the two most frequent categories in non-anaphoric third person plurals and the second most represented one in the periphrastic passive, the analysis yields the following results. In periphrastic passives, verbs are rather equally distributed between verbs of *attribution* (114 occurrences) (307) and verbs of *possession* (111 occurrences) (308). In (307), the term *rigidity* is assigned to the capacity of not being easily moved or changed. In (308), a pilot is given the possibility of working in the non-flying area. This person possesses a new right he/she did not have before.

(307) *There are several gyroscopes (we just call them gyros) that give different information.{S} They work so well because they have what* ***is called*** *rigidity in space, meaning they resist movement.(YAHOO EN)*

(308) *B your training route will reflect your ability - only the cream become Fighter pilots{S} C you* ***will be offered*** *another non-flying trade, but you will be free to resign your commission should you wish to do so.{S} The RAF do not like this, so to minimise the chances of it happening they only choose the best at the recruitment stage. (YAHOO EN)*

(309) *<"S0245"><unclear/> I will see if it will let me do it again <unclear/> oh shit yeah</u>*
*<"S0246">is it back in stock?</u>*

*<"S0245">it probably was the last one it is gone now</u>*
*<"S0246">has it?</u>*
*<"S0245">yeah</u> <"S0246">ah well the worst thing you can do is you can order it and then* ***they just have to refund*** *your money</u>*
*<"S0245">mm I will see how much it is <unclear>hold on I will see</unclear> </u> <"S0244">I fucking hate sentinels they scare the shit out of me</u* (BNC EN)

In the non-anaphoric third person plural, verbs of *possession* (20.4% 38 occurrences), as illustrated in (309) with the verb *to refund*, predominate over verbs of attribution (5.9%; 11 occurrences).

Let us now examine the other types of verb processes. In the non-anaphoric third person plural, they are nearly equally distributed between the categories *modulation* (16.7%; 31 occurrences) (*help* in 310) , *mental* (14%; 26 occurrences) (*to know* in 311) and *verbal* (13.4%; 25 occurrences) (*tell* in 312) .

(310) *is it simply a flat deck 8 or so inches off the ground? chances are the deck alone would be enough to kill the grass, but i'd spray anyway. goto the store you are buying the lumber from.* ***they will help*** *you figure out how much you will need. it will depend on how long the boards are that they sell. (YAHOO EN)*

(311) *What would you do if you decide not to use a flight you have booked? ?{S} Would you go to the trouble of cancelling it or would you just not turn up?{S} I will not be getting any money back on it anyway </question> (…) <answer> as you said you wont get any cash back so why not do the decent thing and cancel that way t****hey will know*** *you are not coming and if they had a standby passenger they could get a seat earlier go on do the decent thing ! (YAHOO EN)*

(312) *But the same approval can be given to civilian flights. </answer> <asker's comment> thanks, ive always wondered why,* ***they dont tell*** *you that in fschool </asker's comment> (YAHOO EN)*

As for the periphrastic passive, the other categories of verb processes are distributed as follows: *mental* (14.7%; 208 occurrences), *verbal* (9.9%; 140 occurrences), *modulation* (8.5%; 120 occurrences) and *existential* (5.2%; 73 occurrences). A comparison with the non-anaphoric third person plural reveals that both passive-like structures uses *mental* verbs in a similar proportion (approximately 14%). It can also be observed that *verbal* and *modulation* types of processes are less frequent in periphrastic passives than in non-anaphoric third person plurals.

Several researchers (see Viberg 2010 or Johansson & Rawoens 2019) state that verbal communication verbs (for instance *to tell* or *to announce*) constitute a prominent type of verb in impersonal passives in Germanic languages. Since our English data contain very few cases of impersonal passives, this result could not be confirmed. This tendency to include an important amount of verbal types of process could not be generalized to the periphrastic passive.

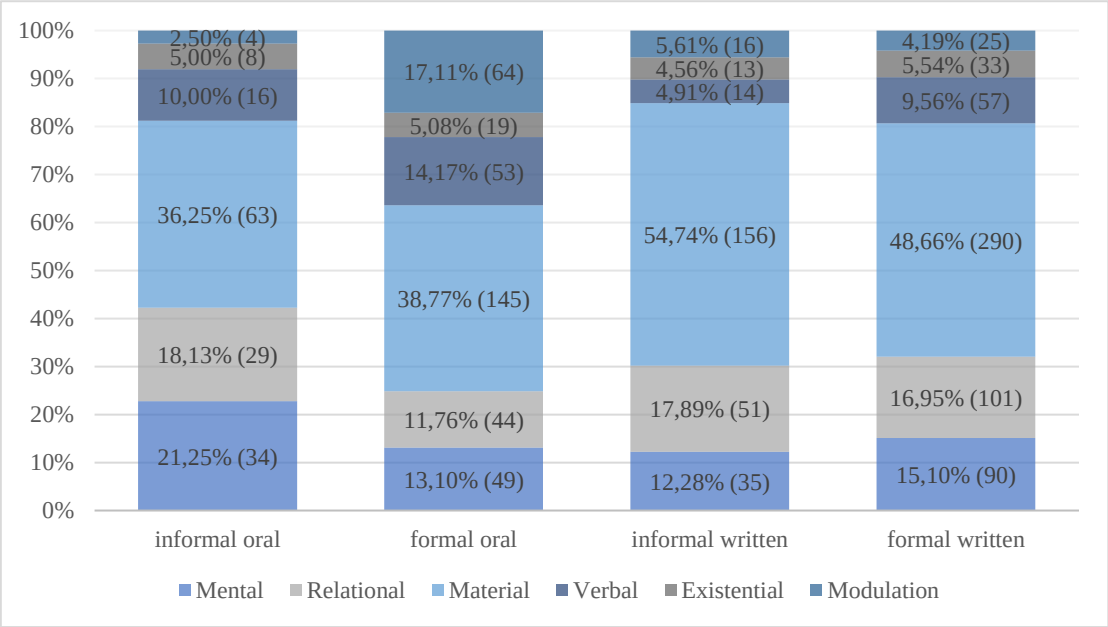
Our study indicates indeed that, when expressing a verbal semantic meaning, speakers tend to opt for a non-anaphoric third person plural rather than a periphrastic passive.

One final relevant finding yielded by this analysis is the presence of *existential* verbs in periphrastic passives (as in 313), which are absent in non-anaphoric third person plurals. In (313) two friends are talking about the fact that tigers or lions should not be killed. In the process of killing, an animate being experience a change of state affecting its existence.

(313) *<"S0439">this is so weird is not it though that we not not that I am no not judging I would've done pretty much the same but it is weird that we think a tiger or a lion* ***should not be killed*** *and then we are quite happy to kill spiders and bugs</u> <"S0441">I know</u>* (BNC EN)

By way of interim conclusion, it can be suggested that the periphrastic passive shows a preference to be used with material processes while, in the non-anaphoric third person plural, material and relational processes appear to be having a similar frequency. Mental, verbal and modulation, which are rather equally distributed, are the other three categories used with the non-anaphoric third person plural. In the periphrastic passive, mental verbs are slightly more frequent than verbal and modulation processes. Finally, the exclusive presence of existential verbs in periphrastic passive can be considered as a defining characteristic of this structure.

When zooming in on the influence of mode and register, Graph 87 indicates the following results for the periphrastic passive.



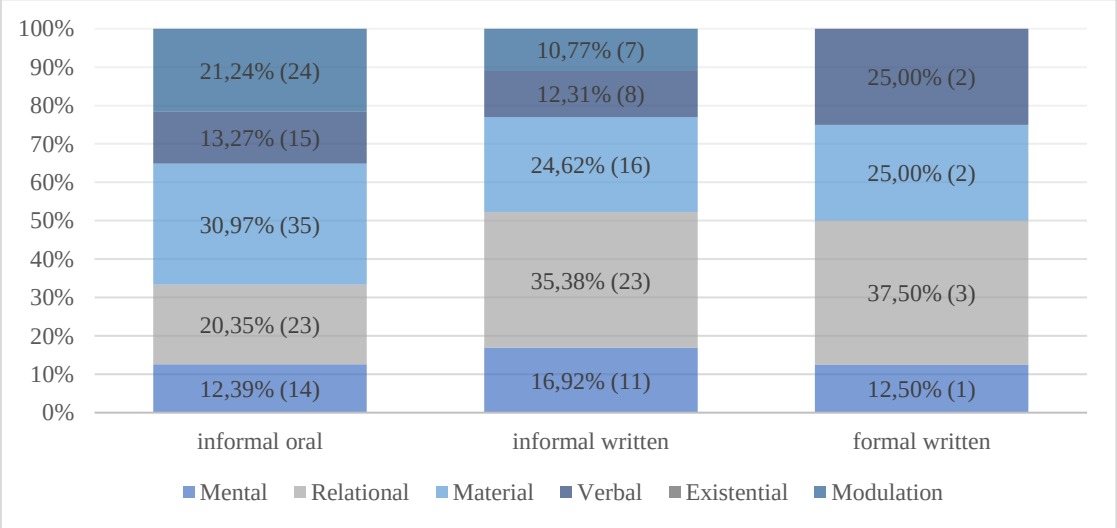

*Graph 87. Distribution of verb processes in English periphrastic passives*

Though only to a limited extent, the degree of formality plays a role in the distribution both in written and in oral data. In written data, results indicate that formal and informal data sets only differ as to their proportion of verb processes. A chi-square test reveals that formal written data use significantly more verbal processes (9.6%) than informal written ones (4.9%) ($\chi^2 = 5.63$. $p < .05$). The other categories of type of process are distributed in rather similar proportions in

both data sets. It can therefore be assumed that the register has a limited impact on the distribution of the type of verb in written data. At first glance, oral data seem to be more heterogeneous. Chi-square tests reveal however that only two categories differ significantly between informal and formal data: mental ($\chi^2 = 5.6679$. $p < .05$) and modulation ($\chi^2 = 21.5315$. $p < .05$) processes. Overall, the analysis suggests that register does not impact the main categories such as *material* processes but rather affects less represented types of verb processes such as *mental* or *modulation* verbs.

The examination of the distribution of types of verb processes reveals that both in the formal and in the informal data sets, there are more *material* processes in written data than in oral ones[38]. This smaller use of material processes in oral data seems to go hand in hand with the greater presence of mental and verbal processes. A last interesting difference to highlight regards the use of modulation verbs, which is significantly higher in formal oral data ($\chi^2 = 19.3979$. $p < .05$) compared to formal written ones. All those observations suggest that there is an influence of the mode both on larger categories such as *material* processes and on less represented ones such as *verbal* and *modulations* verbs.

It can be concluded that, for the periphrastic passive, register and mode are essential factors to consider when analysing the distribution of the type of verb process.




*Graph 88. Distribution of verb processes in English non-anaphoric third person plurals*

The data available for the non-anaphoric third person plural only allow us to examine the impact of the mode in informal data (Graph 88). Only one significant difference could be observed between the two data sets: informal written data use significantly more relational verbs than informal oral ones ($\chi^2 = 4.8645$. $p < .05$). This phenomenon was not observed in the periphrastic passive. It can thus be assumed that a high use of relational verbs in informal written is a specific feature of the non-anaphoric third person plural. Other differences could be noted when looking at Graph 89, such as a higher use of *modulation* or *mental* verbs in informal data, however chi-

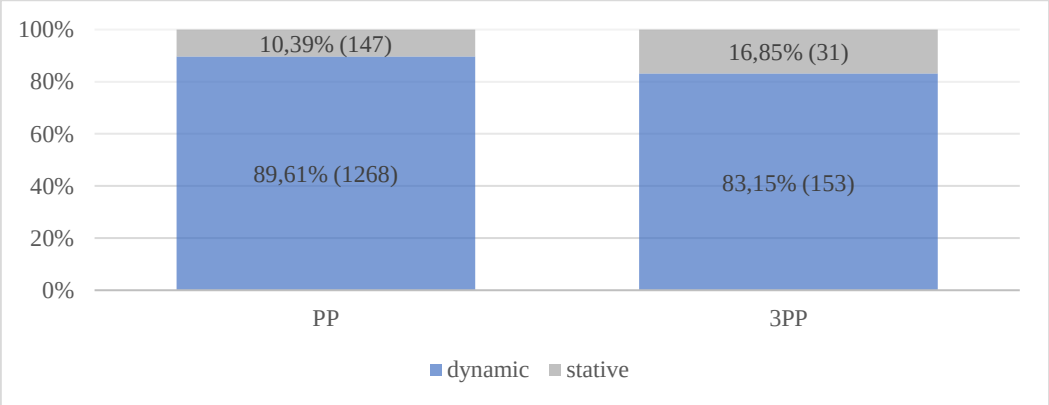
[38] Significantly different for informal oral vs informal written data ($\chi^2 = 9.6752$. $p < .05$) and for formal oral vs formal written data ($\chi^2 = 9.0831$. $p < .05$).

square tests indicate that there are not significant. This may be explained by the relatively small number of occurrences obtained for the periphrastic passive.

In sum, verb processes are distributed differently in the periphrastic passive and in the non-anaphoric third person plural. While in the periphrastic passive, *material* processes account for approximately half of the data, the different types of processes are much more balanced in the non-anaphoric third person. An additional point on which the two passive-like structures differ is the use of existential verbs, which only appear in the periphrastic passive. It could also be observed that mode and register affect the distribution of verb processes in both structures. These differences cannot be attributed to a difference in the type/token ratio as our results indicate that both structures have a similar lexical density (see Chapter 6). All these results make the parameter *type of process* especially important when explaining the variation between passive-like structures in English.

### 9.2.5. Kinesis

Verbs can be classified as either dynamic or stative. Periphrastic passives and non-anaphoric third person plurals show a clear preference to be used with dynamic verbs over stative ones (see Graph 89).




*Graph 89. Distribution of dynamic and stative verbs in English passive-like structures (parameter kinesis)*

(314) *The Toller **was bred** to retrieve from icy waters and must have a water-repellent double coat of medium length and softness , and a soft dense undercoat (WIKI EN*)

In (314), the action of breeding a dog requires a dynamic process. A more detailed examination of the data reveals that, though dynamic verbs predominate in both passive-like structures, there is a significantly higher proportion of stative verbs in the non-anaphoric third person plural (16.8%; 31 occurrences) ($\chi^2 = 6.8668$. $p < .05$).

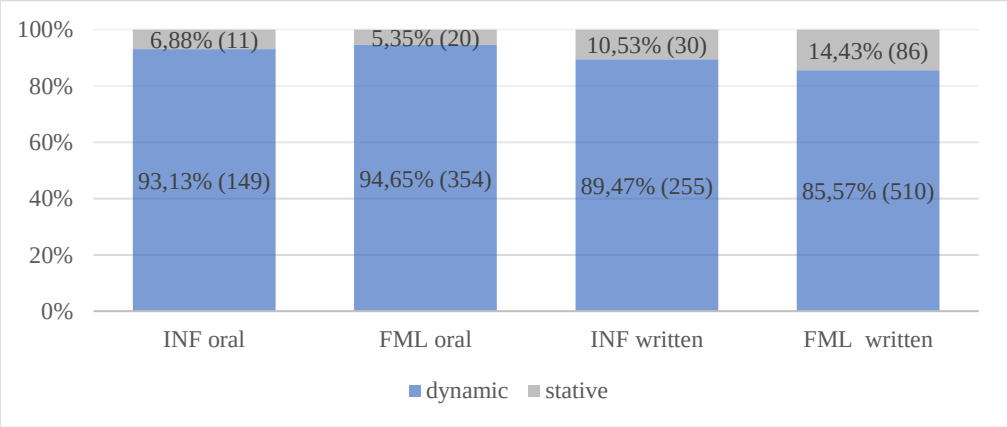
(315) *What would you do if you decide not to use a flight you have booked? ?{S} Would you go to the trouble of cancelling it or would you just not turn up?{S} I will not be getting any money back on it anyway </question> (…)*

*<answer> as you said you wont get any cash back so why not do the decent thing and cancel that way t**hey will know** you are not coming and if they had a standby passenger they could get a seat earlier go on do the decent thing !* (YAHOO EN)

(316) *More energy,slimmer body,hardly and bad breath or gas problems.Plus less parasites in your system.Your skin glows more.No cellulite!No stub burn belly fat.Threes lots of things you just all around feel better.Plus have you seen what **they inject** into chickens that you eat.And cows carry so much you can catch. (*YAHOO EN)

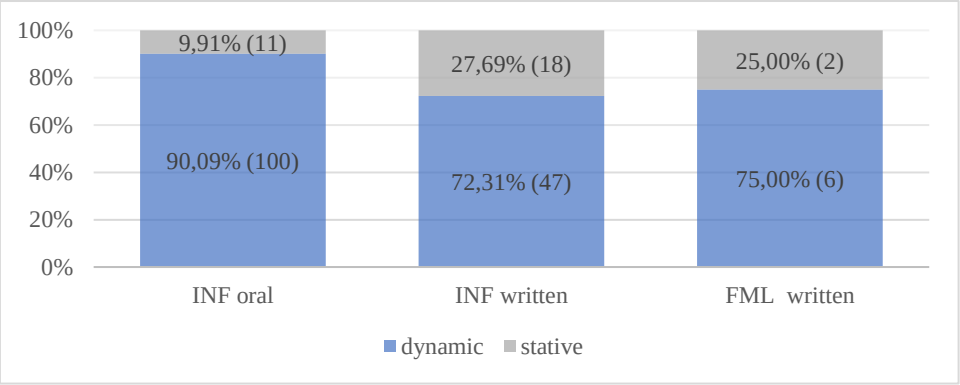
Compare (315) and (316), which illustrate the use of stative and dynamic verbs in the non-anaphoric third person plural. In (315) the verb *know* describes a state as it relates to the fact of possessing information about something. Conversely, *inject* in (316) implies dynamicity in the process of putting a substance into chickens.

Let us examine whether mode and register influence the distribution of dynamic and stative verbs in English passive-like structures.




*Graph 90. Distribution of dynamic and stative verbs in English periphrastic passives*

As can be observed from Graph 90, the degree of formality does not influence the distribution of dynamic and stative verbs in the periphrastic passive. In informal and formal oral data sets, the proportion of dynamic verbs accounts for 93.1% and 94.6%. Similarly, the degree of formality does not impact the distribution of written data since results show no significant difference between informal written (89.5%) and formal written data (85.6%). The only significant difference is found in formal data: the calculation of a chi-test shows that formal oral data use much more dynamic verbs than formal written data ($\chi^2 = 19.4727$. $p < .05$).




*Graph 91. Distribution of dynamic and stative verbs in English non-anaphoric third person plurals*

A glance at the distribution of dynamic verbs in the non-anaphoric third person plural (Graph 91) reveals that the frequency of dynamic verbs is significantly higher in informal oral data than in informal written ones ($\chi^2 = 9.4192$. $p < .05$). This result confirms the tendency to a higher use of dynamic verbs in informal oral data in the periphrastic passive, which was however not significant for that structure. This would mean that speakers tend to express themselves on matters usually linked to thoughts and opinions more in written contexts than oral situations, which leads us to consider that the use of stative verbs in passive-like structures is a specific feature of written data.

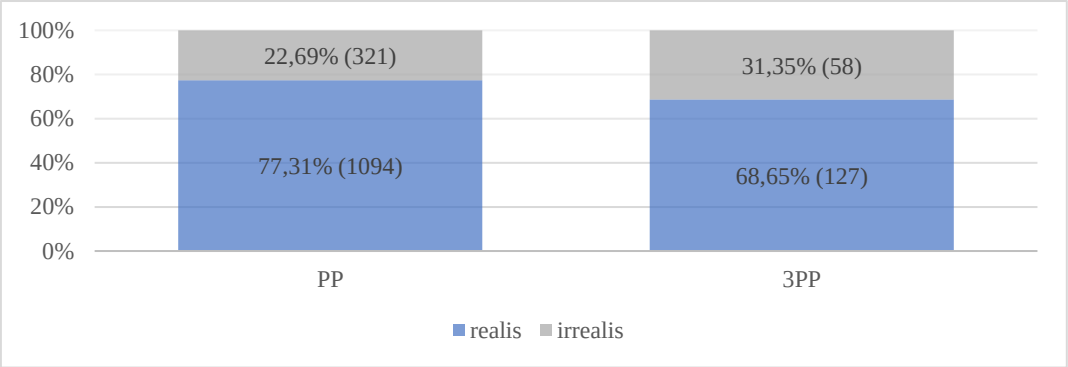
The influence of register and mode on the distribution of dynamic and stative verbs is therefore slightly more limited compared, for instance, with what was observed for the parameter *aspect* as it affects only some of the combinations mode/register in the two passive-like structures. Still, the analysis provides us with relevant information on the behaviour of *kinesis* across passive-like structures and the speaker's different practices according to the mode of production.

### 9.2.6. Realis/irrealis dichotomy

Content can either be actualized (or anchored in reality) or not, thus happening in the non-real (contingent) world. In this study, the following elements have been considered as irrealis-inducing: the use of a negative particle, the use of future and conditional tenses and the use of the subjunctive and modal indicators. Compare (317) and (318). Example (317) illustrates a situation anchored in the real world. It is a fact that the Spitfires planes are, at the time of the speech production, owned by a specific company. Conversely in (318), the verb *contaminate* is accompanied by the modal auxiliary *may* which confers a meaning of possibility to the clause. Drinking water may or may not be contaminated, thus expressing that both possibilities should be considered and that there is no certainty.

(317) *There are VERY few air worthy Spitfires left. (b) Most* ***are owned*** *by Preservation Groups.* (YAHOO EN)

(318) *Plants from the ground that are not washed properly may contain B12 from bacteria in the soil, often from faeces ; drinking water* ***may also be contaminated*** *with B12-producing bacteria (…).* (WIKI EN)



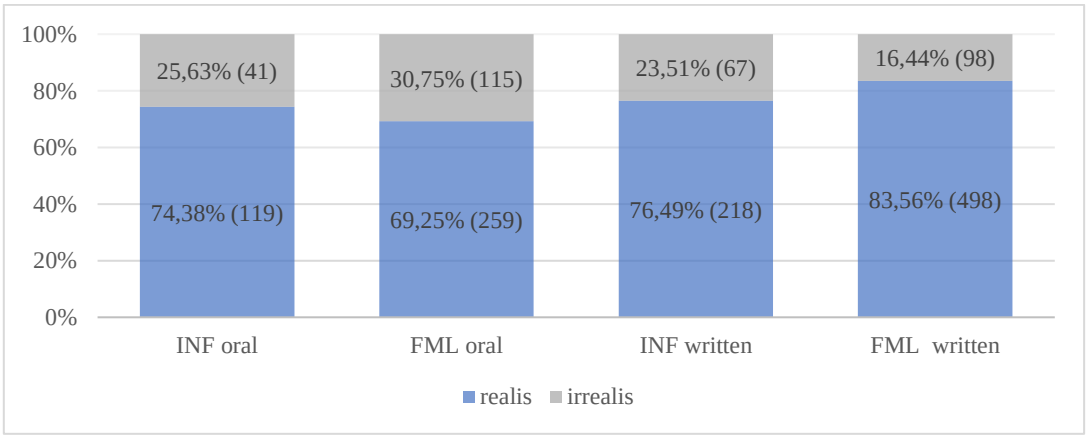

*Graph 92. Distribution of (ir)realis clauses in English passive-like structures*

The periphrastic passive and the non-anaphoric third person plural display a different behaviour as for their way to anchor information in reality. As indicated by Graph 92, though realis clauses predominate in both passive-like structures, the non-anaphoric third person plural includes much more irrealis content than the periphrastic passive ($\chi^2 = 6.797$. $p < .05$).

(319) *Military does have to follow the rule unless they are approved to go faster.{S} But the same approval can be given to civilian flights. </answer> <asker's comment> thanks, ive always wondered why,* ***they dont tell*** *you that in fschool </asker's comment>* (YAHOO EN)

(320) *<answer> it is a little disk that they put under the skin on the dogs neck so if he gets lost or stolen* ***they can scan*** *his neck and find out where he belongs. the disk has ur info on it so he can be returned. </answer>* (YAHOO EN)

Examples (319) and (320) illustrate two different ways to encode irrealis in the non-anaphoric third person plural. In (319) the action of telling did not occur and, following Hopper & Thompson's taxonomy, is therefore not considered as real. In (320) the irrealis mode is encoded through the use of the modal *can*.




*Graph 93. Distribution of realis and irrealis clauses in English periphrastic passives*

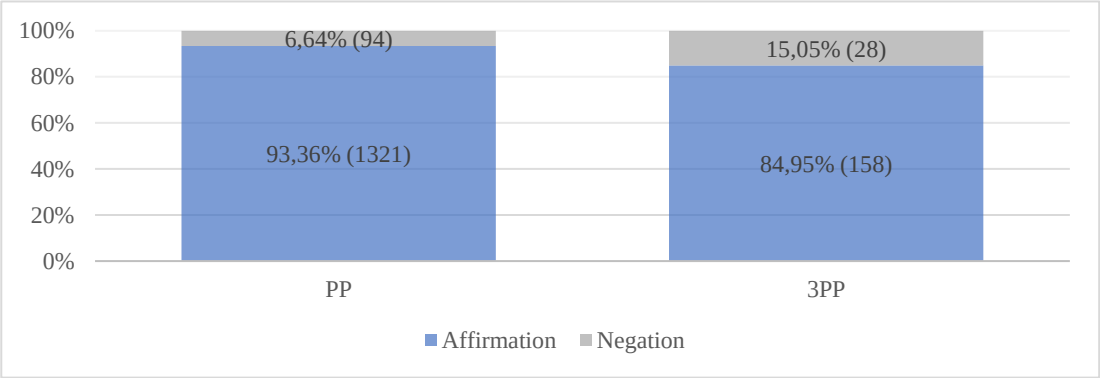
Breaking down the data reveals that, for the periphrastic passive (Graph 93), there is significantly more realis clauses in formal written (83.6%; 498) data than informal written ones (76.5%; 218) ($\chi^2 = 6.324$. $p < .05$). Such difference between informal and formal data is not observed in the oral data set. It can therefore be suggested that the influence of the degree of formality only operates in written productions. Regarding the impact of mode, results indicate

that while there is no difference in informal data sets, formal written data use significantly more realis forms than formal oral data ( $\chi^2 = 28.2021$. $p < .05$). As for the non-anaphoric third person plural, no influence of the mode or the register can be noted regarding the way to anchor events and situations.

This suggests that the impact of mode and register is relatively restricted since these two factors only affect some data sets.

### 9.2.7. Affirmation

The parameter *affirmation* has a direct relation with the notion of realis that has just been discussed. Affirmative sentences are indeed better candidates for real clauses.




*Graph 94. Distribution of affirmative and negative clauses in English passive-like structures*

As shown in Graph 94, the periphrastic passive and the non-anaphoric third person plural are mainly used in affirmative sentences such as exemplified in (321) where the speaker sets out a fact in an affirmative manner, namely that cows are inseminated.

(321) *Dairy cows **are** artificially **inseminated** every year*. (YAHOO EN)

Even though affirmative clauses predominate in the two passive-like structures, results indicate that the non-anaphoric third person plural resorts far more to negative forms than the periphrastic passive ($\chi^2 = 16.5193$. $p < .05$). Chi-square tests reveal that there is no impact of the mode and the register, be it for the periphrastic passive or the non-anaphoric third person plural.

### 9.2.8. Other parameters

This section offers a brief summary of the results obtained for verb-related parameters considered as less salient by the MCA when explaining the variation between the passive-like structures in English: *volitionality* and *punctuality*. The analysis revealed that volitional verbs largely predominate in both the periphrastic passive (89.4%; 1265 occurrences) and the non-anaphoric third person plural (91.9%; 170 occurrences). Example (322) illustrates a periphrastic

passive used with a volitional verb. The decision of allowing military flights to go faster is made on purpose. An entity deliberately decides to give this right to these flights.

(322) *Military does have to follow the rule unless **they are approved** to go faster.{S} But the same approval can be given to civilian flights*.(YAHOO EN)

As for the parameter *punctuality*, results indicate that, in both passive-like structures, approximately two third of the occurrences (70.2%; 993 occurrences for the periphrastic passive and 76%; 140 occurrences for the non-anaphoric third person plural) appear with non punctual verbs such as in (323). The action of preparing the peppers requires a certain period of time and several steps are needed to accomplish the activity.

(323) *As you fry, be careful not to crowd the pan.{S} Remove the peppers from the pan with tongs, drain on paper towels, and repeat until all peppers are fried.{S} Serve immediately, accompanied by salsa, if desired.{S} Serves 6.{S} Note: The peppers **can be prepared** up to the point of being dredged in cornmeal and fried up to two days in advance and refrigerated, tightly covered with plastic wrap.* (YAHOO EN)

### 9.2.9. Summary

In sum, the following characteristics can be highlighted for verb-related parameters in periphrastic passives and non-anaphoric third person plurals.

As regards *tense* and *aspect*, the non-anaphoric third person plural uses significantly fewer past tenses and more future tenses compared to the periphrastic passive. This result appears to be in direct relation with the high proportion of imperfective forms observed in the non-anaphoric third person plural. Conversely, the periphrastic passive displays a higher proportion of past tenses, which positively correlates with a higher use of the perfective aspect. The *verb mode* is another feature that distinguishes the two passive-like structures. While the non-anaphoric third person plural rather relies on the indicative, the periphrastic passive appears as much more heterogeneous, including forms such as infinitives or gerunds. As for the dynamicity of the two passive-like structures, results indicate that non-anaphoric third person plurals are more likely to contain stative verbs than periphrastic passives. Differences were also noted in the way the two passive-like constructions anchor events and situations in the real world since the non-anaphoric third person plural resorts to much more irrealis clauses than the periphrastic passive. This may be party linked to a higher use of negative clauses found in the non-anaphoric third person plural. Another important difference between the two structures regards the type of verb they include. While material verbs predominate in the periphrastic passive, the distribution of the type of verb processes is far more heterogeneous in the non-anaphoric third person, with an ample use of relational, material, mental, verbal and modulation types of processes.

A last point that is worth mentioning is the role played by mode and register in the distribution of the verb-related parameters. The parameters *tense*, *aspect* and *type of verb process* appear as

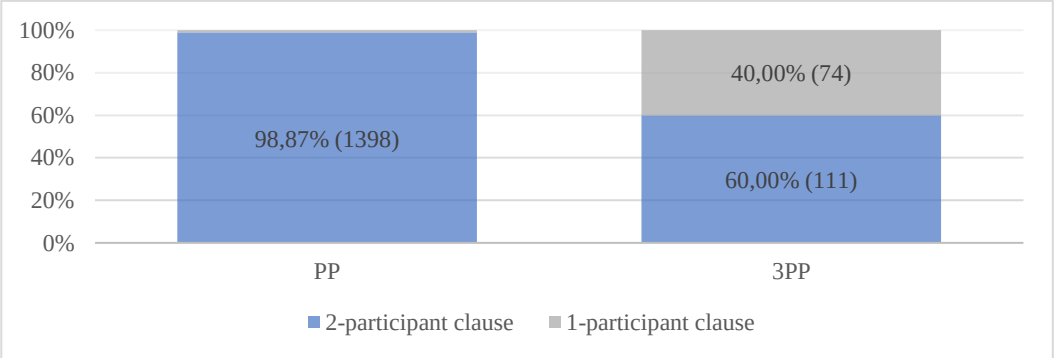
highly sensitive to the mode of production and the degree of formality. As for the other features, their impact is either limited to one or two data sets, or absent.

## 9.3. Participant-related parameters

This section addresses a series of parameters related to the participants included in the passive-like structures. It first starts with a detailed examination of the number of participants involved in the clause (9.3.1), then it analyses the presence and the nature of the agent and the patient (9.3.2 -9.3.3). It follows with a focus on the degree of individuation (9.3.4) and affectedness of the patients (9.3.5) and closes on a summary of the results (9.3.6).

### 9.3.1. Number of participants

Every clause can imply one or more participants. The periphrastic passive seems to occur almost exclusively with two or more participants (98.9%) while the distribution between one- and two-participant clauses is much more balanced in the non-anaphoric third person plural (see Graph 95).




*Graph 95. Distribution of one- and two-participant clauses in English passive-like structures*

The pronoun *most* and the proper noun *Preservation Groups* in example (324) illustrate a clause with two explicit participants in a periphrastic passive. When the periphrastic passive occurs in a one-participant clause, it usually appears with a subordinate clause such as *it was found* in (325).

(324) *There are VERY few air worthy Spitfires left. (b) Most* ***are owned by*** *Preservation Groups. (YAHOO EN)*

(325) *It reduces both volume and height growth as well as eventually killing the trees and causing them to be more susceptible to windthrow and other diseases and insects . In the Southeastern US* ***it was found*** *that as many as 30% of trees can* [unreadable]*@ @ @ @ @ @ @ @ @ irregulare ' ' (WIKI EN)*

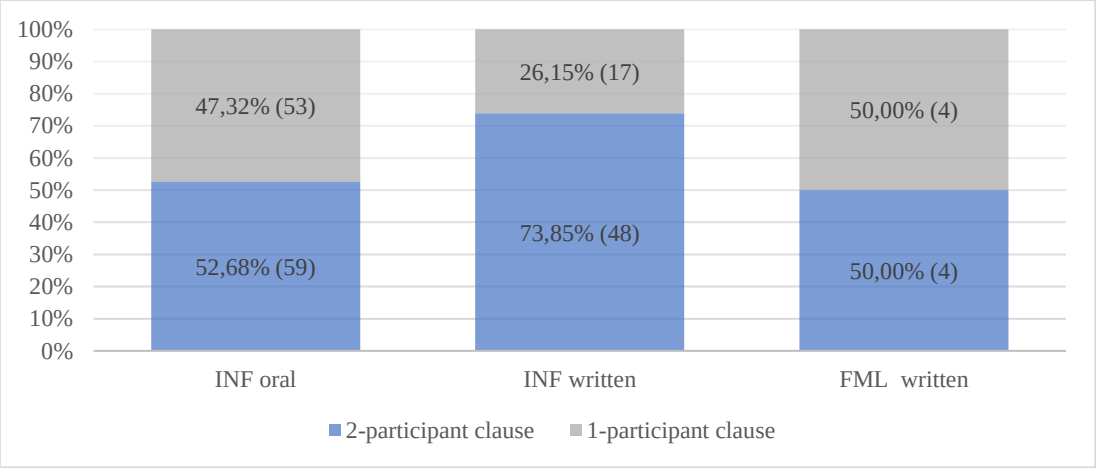
The non-anaphoric third person construction always contains an animate agent in subject position. This passive-like construction therefore implies at least one explicit participant.

Further investigation reveals, nevertheless, that as far as the number of participants is concerned, a majority of non-anaphoric third person constructions occurs in clauses with two participants (60%; 111 occurrences) such as in Example (326), containing the pronoun *they* in subject position and *you* in object position. Example (327) is considered as including only one participant, namely the subject *they*. Such types of clauses are less frequent as they only occur in 40% of the cases (74 occurrences).

(326) *But the same approval can be given to civilian flights. </answer> <asker's comment> thanks, ive always wondered why,* ***they dont tell you*** *that in fschool </asker's comment>* (YAHOO EN)

(327) *hrees lots of things you just all around feel better.Plus have you seen what they inject into chickens that you eat.And cows carry so much you can catch.And how about you pour vinegar on raw pork and see all the bugs you eat.Yeah* ***they say*** *the heat kills it but the dead body are in you now for other parasites to form off of.And you are one less human helping kill helpless animals!* (YAHOO EN)

Due to the fact that periphrastic passives almost exclusively resort to two-participant clauses, there is no possible variation regarding the mode and the register. As for the non-anaphoric third person plural, data suggest the following pattern (see Graph 96).




*Graph 96. Distribution of one- and two-participant clauses in English non-anaphoric third person plurals*

There are significantly more two-participant clauses in informal oral data than in informal written data ($\chi^2$ = 11.8106. $p < .05$). This result confirms again the lower transitivity of conversational data suggested by Hopper & Thompson (2006). A brief qualitative review of the data reveals that in most cases, in addition to the human agent, the non-anaphoric third person is constructed with a second pronoun, as illustrated above in (326) with *you*. This would mean that in oral situations, speakers tend to include more frequently other interlocutors in their discourse than in written data.

In sum, the number of participants not only varies according to the passive-like structure but also depends on the mode of production, in the case of the non-anaphoric third person plural.

### 9.3.2. Agent

Since the non-anaphoric third person construction always contains an animate human as agent, the analysis of the agent will be centred on the periphrastic passive, expected to display greater variation regarding the nature of the agent. Results indicate that the absence of an overt agent in the periphrastic passive constitutes the predominant type of periphrastic constructions in English. In our data set, the presence of an overt agent, introduced by the prototypical *by*-phrase, accounts for 12.8% (181 occurrences) of the total data included in the study. This set of 181 occurrences will serve as the basis for the analysis of the parameter *agent*. I will first comment on the main results (included in Table 30), regardless of register and mode variations and, in a second phase, information on these two types of variations will be incorporated.

| | **Agent** | **Occurrences** | **%** |
|---|---|---|---|
| **Animate** | Human | 62 | 34.2% |
| | Non human | 2 | 1.1% |
| **Inanimate** | Organization | 31 | 17.1% |
| | Country/City | 3 | 1.6% |
| | Inanimate Other | 78 | 43% |
| | Uncodifiable | 5 | 2.7% |
| | **Total** | 181 | 100% |

*Table 30. Classification of animate and inanimate agents in English*

A first distinction can be operated between animate and inanimate agents. As reported in Table 30, 62 occurrences of the agents are animate entities and 112 occurrences are classified as inanimate. In the final set of data, 5 occurrences were not codifiable due to the the fact that part of the discourse was unclear or missing such as in (328). However, the presence of the prepositional phrase *by* clearly indicates the speaker's intention to include an agent introduced by a prototypical agentive complement.

(328) <"S0261">kind of different is not it?</u> <"S0262">er yeah and actually **was directed** by a you know I think <trunc> (BNC EN)

A look at the figures in Table 30 reveals that human agents predominate in the animate category and agents classified as *others* totalize the highest proportion. Examples (329) and (330) illustrate the category of animate human and inanimate *others*.

(329) *Thyroid problems **have been identified** by American breeders as a priority issue , together with epilepsy and hip dysplasia* (WIKI EN).

(330) *Physically, the colour of water **is affected** by factors like the light source, absorption and scattering of light, as well as suspended materials in the water.* (YAHOO EN)

Animate humans are exemplified in (329) where the agents performing the action of identifying the problems are people who breed dogs as their job. Example (330) illustrates the category

inanimate *others.* Factors like *light* and *suspended materials in the water*, which would be expected and considered to be inanimate entities, impact the color of water. Organizations appear as the third most frequent category of agents (17.1%; 31 occurrences) and include agents such as *the American Dietetic Association, the Australian National Health and Medical Research Council* and *Dietitians of Canada* illustrated in (331). Organizations such the abovementioned ones can involve humans and can thus be considered as playing an agentive role.

(331) *Well-planned vegan diets appear to offer protection against certain degenerative conditions , including heart disease , and* ***are regarded*** *as appropriate for all stages of the life-cycle by the American Dietetic Association, the Australian National Health and Medical Research Council , and Dietitians of Canada.* (WIKI EN).

A more detailed approach of the data, considering mode and register variations, shows that 81.2% of the overt agents are included in formal written data (101 occurrences) and in formal oral data (46 occurrences). Given this predominance of overt agents in formal data, an influence of the register can be observed on the use of overt agents in periphrastic passive. Conversely, the mode of production does not seem to affect the nature of the agents as no significant difference could be noted between oral and written data sets.

As shown in the literature and confirmed for French and Spanish data, the presence of an overt agent is often observed in a problematic situation or when the agent is a well-known person, occupies an important position in society such as a president or is an organisation. The English data confirm these findings. A quick scan of the expressed agents reveals that, although present, the negative effect triggered by the semantic of a verb seems to be a less frequent explanation for the presence of the agent compared to French and Spanish data. When the situation is considered problematic, the adversity is often created by the context rather than by the meaning of the verb as illustrated in (332).

(332) *Mr President, the simple answer is that I think the speaker is wrong.{S} It is a question of whether or not these groups are helped.{S} I do not believe that they* ***are helped by*** *this extra proposal.{S} I do not think it helps at all..{S}* (PROCEP EN)

Example (332) is extracted from a parliamentary debate on the inclusion of an additional proposal in a food information consumer plan. In (332), the problematic effect is created by the context in which the periphrastic passive occurs such as through clauses such as *I think that the speaker is wrong* or *I do not believe that*. The use of the agent complement maintains the focus on the fact that the speaker believes that groups are not helped, while providing information on the fact that they are not helped by the specific proposal made by a fellow member of parliament.

At first sight, the use of verbs with negative semantics appears slightly more restricted in English than in Spanish and French. This phenomenon may be explained by the fact that English possesses a *get*-passive mechanism that, as widely shown in the literature and as will be developed in 9.5, is frequently associated with contexts seen as problematic from the speaker's

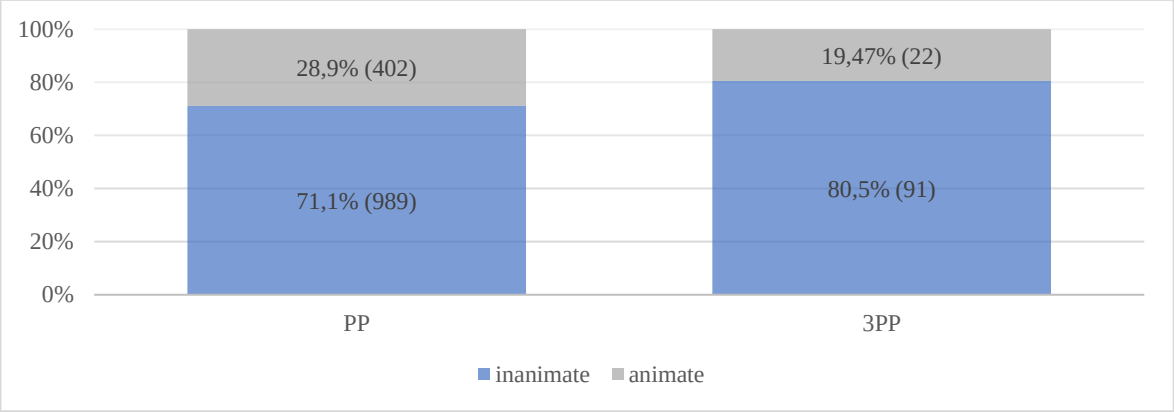
perspective. These first intuitions must however be considered as avenues for further reflection and have to be confirmed through a larger quantitative study.

### 9.3.3. Patient

This section intends to compare the nature of the patients included in the two passive-like structures. As seen in 9.3.1, while the periphrastic passive occurs almost exclusively in two-participant clauses, the presence of a second participant, in addition to the subject pronoun, in the non-anaphoric third person plural is much more flexible. The analysis developed hereunder is therefore based on the complete set of periphrastic passives and the 111 occurrences of non-anaphoric third person plural consisting of a two-participant clause. As a first approach Graph 97 suggests a general picture on the animate versus inanimate nature of the patients. Example (333) and (334) respectively illustrate this distinction with *trees* as inanimate patient and *farmers* as animate patient.

(333) *The **trees were cut down** during the Neolithic, and onwards both to provide fuel and to produce open grazing for neolithic herdmen's sheep. (YAHOO EN))*

(334) *It was meant to be simple because it was designed at home and people would understand what they were at and we ended up with mayhem on it and it was anything but simple; but the biggest problem that we had with it is that when farmers then came to us and **were told** that they had to do x, y and z, when we brought them to the Commission, the Commission told them they could do nothing about it.* (PROCEP EN)




*Graph 97. Distribution of animate versus inanimate patients in English passive-like structures*

Results indicate that inanimate patients predominate in the two passive-like structures (Graph 97). Within inanimate patients, the category *others* largely prevails. Patients of that type can be concrete entities such as for example *credit* (335) or even more abstract concepts such as *tight controls* in (336).

(335) *<question>{S} What would you do if you decide not to use a flight you have booked? ?{S} Would you go to the trouble of cancelling it or would you just not*

*turn up?{S} I will not be getting any money back on it anyway </question> (…) <answer> i would call in advance* ***they should give*** *you a credit that is good for one year (YAHOO EN)*

(336) *You especially want the fuel to not have water suspended in it, as it could freeze in the fuel, so* tight controls *in manufacture and distribution* ***are needed****. (YAHOO EN)*

The findings reveal that the periphrastic passive contains significantly more animate patients than the non-anaphoric third person plural ($\chi^2 = 4.5918$. $p < .05$).

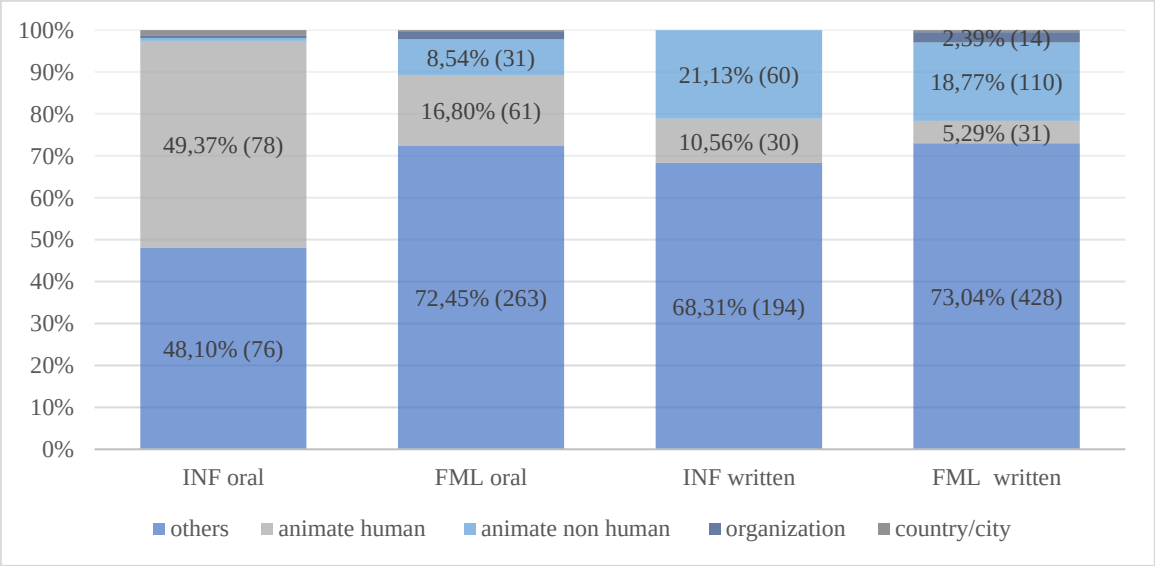
Table 31 makes it possible to have a close look at the nature of the patients.

| Structure | Animate human | Animate non human | Organization | Country/City | Others |
|---|---|---|---|---|---|
| **Periphrastic passive** | 14.4% (200) | **14.5% (202)** | 1.6% (22) | 0.4% (6) | 69.1% (961) |
| **Non-anaphoric third person plural** | 18.6% (21) | 0.9% (1) | - | 0.9% (1) | 79.6% (90) |

*Table 31. Classification of animate and inanimate patients in English passive-like structures*

The analysis reveals that, while the use of animate humans is stable across the two passive-like structures, the periphrastic passive resorts far more to animate non human patients (14.5%; 202 occurrences) (such as *Alaunts* dogs in 337) than the non-anaphoric third person plural (0.9%; 1 occurrence). The presence of animate non human patients in the periphrastic passive can thus be considered as one of the defining features of this passive-like structure.

(337) *In France , Alaunts* ***were separated*** *into three main categories , based on physical appearance and the duties they performed* (WIKI EN)




*Graph 98. Distribution of patients in English periphrastic passives*

Graph 98 reports the distribution of the different types of patients in the periphrastic passive. The behaviour of the informal oral data set considerably differs from the other three mode/register combinations. While in the other three data sets inanimate *others* prevail, in informal oral data patients are equally distributed between animate *others* and animate humans. This high percentage of human patients can be considered as a defining characteristic of informal oral data. Let us focus on the impact of the oral vs written nature of the data. A comparison between informal oral and informal written data shows that while animate humans are specific of the informal oral data set, an important proportion of animate non human patients (21.1%; 60 occurrences) combined with a small proportion of animate human patient seems to characterize informal written data. This higher proportion of animate human patients in oral data is confirmed in the formal data set[39]. A last interesting result to highlight is the fact that both in the oral and in the written data sets, there are significantly more animate human patients in informal data.

While the impact of mode and register appears to be a key element in the distribution of patients in the periphrastic passive, no influence could be observed on the non-anaphoric third person plural.

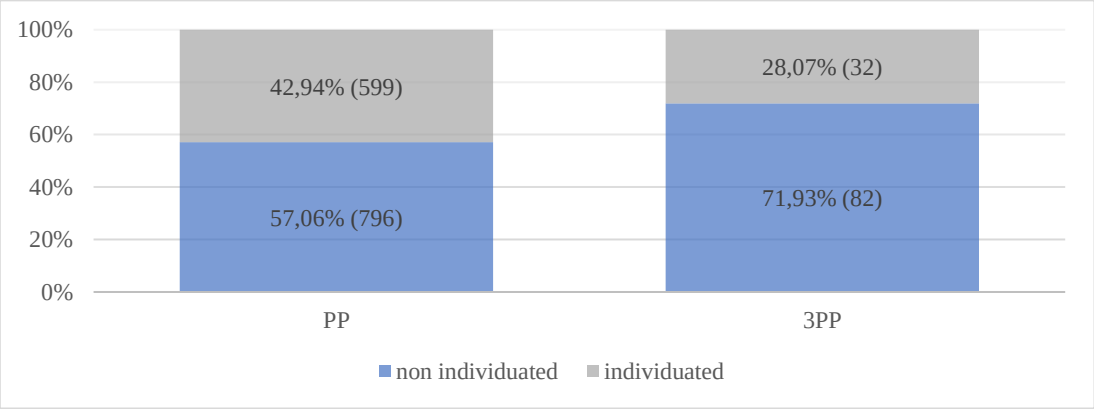
### 9.3.4. Individuation of the patient

Patients are classified as either individuated or non-individuated. Example (338) includes a patient considered as individuated. Indeed, the personal pronoun *you* possesses the following characteristics: it is human, it refers to a concrete entity, it is singular, countable and definite. It therefore fulfils five of the six characteristics of an individuated patient highlighted by Hopper & Thompson (1980). In (339) *Spitfires*, which are fighter aircrafts, fall in the category of non-individuated participants. They are non-human plural concrete entities. The patient is countable and, in this specific situation, is used in a non-referential manner. As *Spitfires* only totalize two characteristics of highly individuated objects (countable and concrete entities), they are considered non individuated.

(338) *goto the store you are buying the lumber from.* ***they will help*** *you figure out how much you will need. it will depend on how long the boards are that they sell. (YAHOO EN)*

[39] A chi-square test shows that the difference between formal oral and formal written data is significant for $\chi^2 = 33.9451$ $p < .05$.

(339) *You will need a pilot's licence and then find a museum where t**hey still fly Spitfires**.{S} Unless you already have a several hundred hours of flying experience I doubt that you stand any chance at all of them letting you fly their aeroplane. (YAHOO EN)*

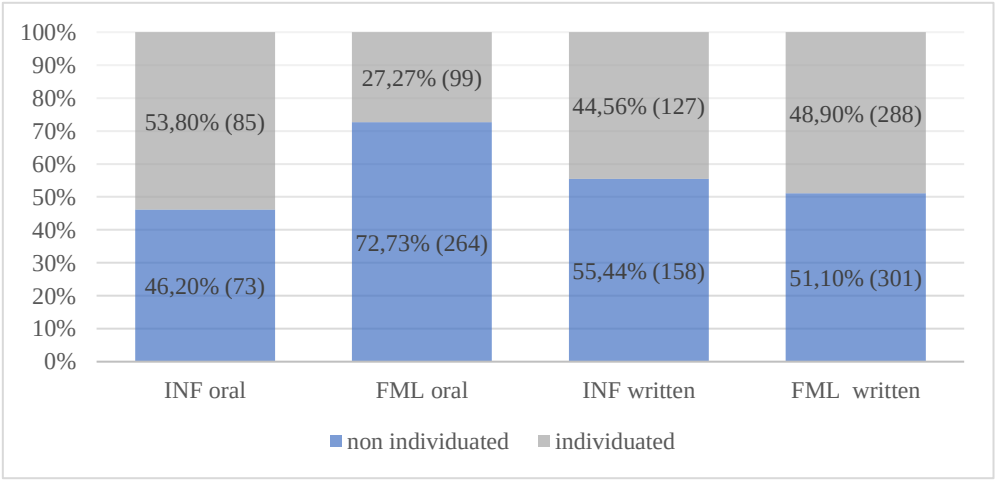



*Graph 99. Distribution of (non-)individuated patient in English passive-like structures*

Graph 99 indicates that periphrastic passives contain significantly more individuated patients than the non-anaphoric third person plural ($\chi^2 = 9.5764$. $p < .05$). Example (340) illustrates a periphrastic passive containing an individuated patient.

(340) *Saint Genevieve , the Patron Saint of Paris , **is mentioned** as having observed a vegetarian diet - but as an act of physical austerity , rather than out of concern for animals . (WIKI EN)*

In (340), *Saint Genevieve* can be considered as highly individuated as the patient fulfills the criteria proper, human, concrete, countable and refers to a specific person. Animate patients are better candidates to be classified as individuated. As found in 9.3.3, the periphrastic passive tends to include much more animate patients than the non-anaphoric third person plural, which would be a possible explanation of the larger presence of individuated patients in the periphrastic passive. This assumption is reinforced by the results offered in Figure 25 in Section 9.1, which indicate a correlation between the parameters *patient* and *individuation.*




*Graph 100. Distribution of (non-)individuated patient in English periphrastic passives*

The examination of (non)-individuated patients in mode and register indicates the following results for the periphrastic passive (Graph 100). In oral data, the informal data set contains significantly more individuated patients ($\chi^2 = 33.9044$. $p < .05$). It can be suggested that these individuated patients are animate entities as we saw that highly individuated patients correlate with animate patients. Such a higher use of individuated patients in informal data is not observed in the written data set. In written data the proportion of individuated patients does not differ according to the degree of formality of the data[40]. Another interesting result suggested by this graph is the greater presence of non-individuated patients in formal oral data compared to formal written ones. This difference is not observed in informal data sets ($\chi^2 = 3.4748$. $p > .05$). In summary, while the behaviour of written data is stable across the two registers, oral data display much more variation, a marked presence of individuated patients being specific to informal oral data and a high use of non-individuated patients characterizing formal oral data.

In the non-anaphoric third person plural, no significant difference could be observed between informal oral and informal written data. This confirms what was found for the periphrastic passive. The distribution of the (non)-individuated patients is thus affected by the mode of production and, though only in oral data, by the degree of formality of the situation.

### 9.3.5. Affectedness of the patient

Patients can either be totally or only partly affected by the event or the situation. Compare (341) where the patient is totally affected, the whole *eggs* are used, and (342), where the patient is not totally affected since only a *great part of* and not the entire blame is put on the former president.

(341) *Eggs **are used** in recipes as thickeners and binders ; the protein in eggs thickens when heated and binds the other ingredients together . (WIKI EN)*

(342) *Like in the Air Maldives bankruptcy , a great part of the blame **is put** on former president Maumoon Abdul Gayoom 's wish of having a Maldivian prestige international airline at any cost , regardless of the company 's shady financial operations . (WIKI EN)*

Results indicate that, in English, the periphrastic passive and the non-anaphoric third person plural almost exclusively contain totally affected patients (99.4%; 1386 occurrences and 99.5%; 114 occurrences, respectively).

[40] No significant difference was found between informal written and formal written data ($\chi^2 = 1.4474$. $p > .05$).

### 9.3.6. Summary

This section has explored the notion of *participant* in periphrastic passives and non-anaphoric third person plurals by analysing the following parameters: the number of participants included in the event or the situation, the agent, the patient and the individuation and affectedness of the patient.

Findings suggest that the two passive-like constructions greatly differ regarding these factors. First, while the periphrastic passive almost exclusively relies on two-participant clause, the non-anaphoric third person plural appears to be far less constrained regarding the number of participants included, with a more balanced profile between one- and two-participant clauses. Second, the agent behaves differently within the two structures. While the non-anaphoric third person plural consists of an explicit animate human agent, the periphrastic passive tends to defocus its agent. Third, the presence and the nature of the patient are two other distinctive features. The smaller proportion of patients is explained by the higher use of one-participant clauses in non-anaphoric third person plural. The degree of individuation of the patients has been shown to correlate with its nature. The degree of affectedness of the patient appears as the only parameter that does not vary across the two passive-like structures. Last, the written versus spoken nature of the data, as well as the degree of formality were proved to be key variables in explaining the distribution of the different participant-related parameters.

## 9.4. Context-related parameters

According to the results yielded by the general Multiple Correspondence Analyses, two context-related parameters are considered as salient in the study of passive-like structures: the use of a spatiotemporal restriction (presented in 9.4.1) and the presence of a prepositional phrase (analysed in 9.4.2). The last part of the section briefly summarises the findings obtained for the other context-related parameters (9.4.3).

### 9.4.1. Presence of a spatiotemporal restriction

Results indicate that the periphrastic passive and the non-anaphoric third person plural do not diverge[41] as for the proportion of spatiotemporal restrictions included in the linguistic context (13.9%; 197 occurrences and 18.3%; 34 occurrences, respectively). In (343) the presence of *during the Neolithic* allows to restrict the range of people involved in the process of cutting trees. Similarly, in Example (344) the adverb *today,* placed at the beginning of the sentence, restricts the people classified as vegetarians. While these two examples illustrate temporal

[41] The chi-square indicates no significant difference between the periphrastic passive and the non-anaphoric third person plural with $\chi^2 = 2.5279$. $p > .05$.

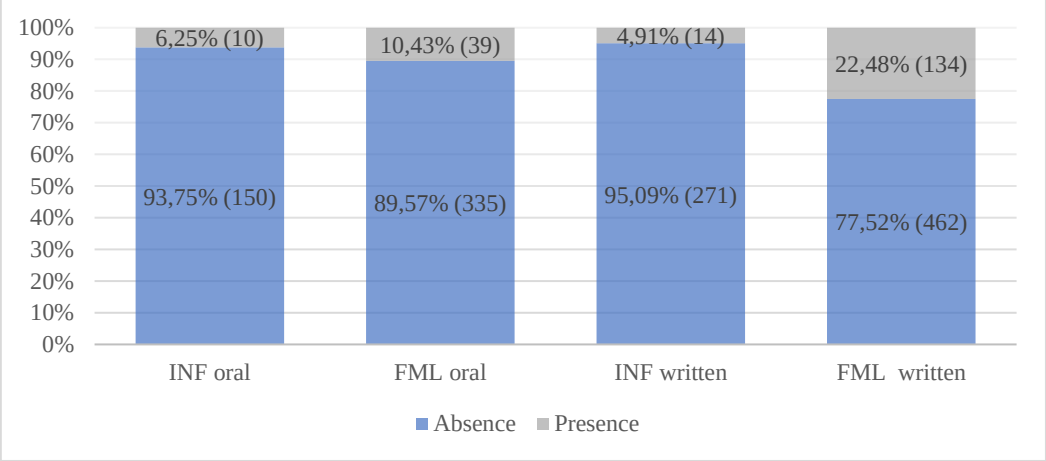
restrictions, the mention of the country *Romania* in (345) allows to spatially reduce the scope of the possible agents involved in the action.

(343) *The **trees were cut down** during the Neolithic, and onwards both to provide fuel and to produce open grazing for neolithic herdmen's sheep.* (YAHOO EN)

(344) *Today Indian vegetarians , primarily lacto-vegetarians , **are estimated** to make up more than 70 percent of the world 's vegetarians .* (WIKI EN)

(345) *Now we move on to another subject, and that is wild stray dogs in Romania, which **are** apparently **being shot** and killed one way or another.* (PROCEP EN)

However, differences emerge as for the distribution of the spatiotemporal restrictions in the four combinations of register/mode data sets (see Graph 101).




*Graph 101. Distribution of spatiotemporal restrictions in English periphrastic passives*

The analysis shows that the formal written set is the one to resort the most to spatiotemporal restrictions and a chi-square test confirms a significant difference between formal oral and formal written data ($\chi^2 = 22.7903$. $p < .05$). This phenomenon was already observed in French and Spanish data. This frequent use of spatiotemporal elements seems to be a specific characteristic of the periphrastic in formal written data.

As for the non-anaphoric third person plural, no difference could be observed between informal oral and informal written data as in both sets, the passive-like structure appears with a restriction in respectively 13.3% and 12.5% of the occurrences.

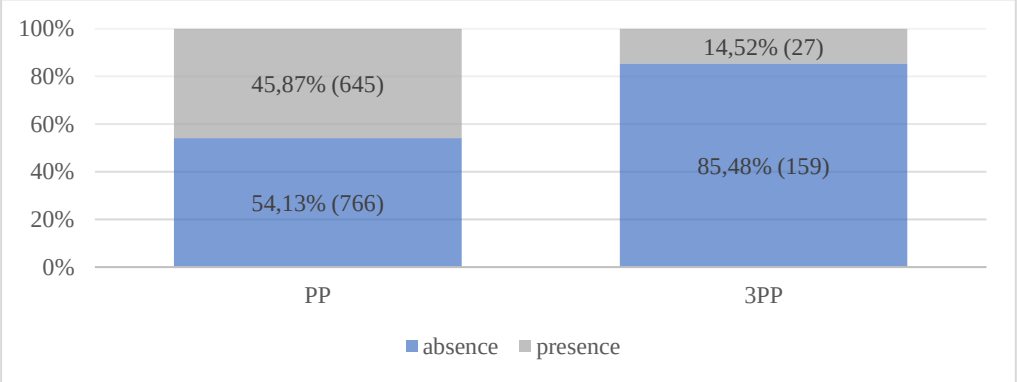
It can thus be concluded that the spatiotemporal restrictions display a rather stable behaviour across the passive-like structures, except in the periphrastic passive where they appear as a distinctive feature of formal written data. And, as already explained in Chapter 7 and pointed out by Steiner (2017), spatiotemporal restrictions are more frequently employed in written productions.

### 9.4.2. Presence of a prepositional phrase

The presence of a prepositional phrase may bring additional information on the participants involved in the passive-like structure. This analysis adopts the taxonomy developed by Delbecque for Spanish (2014), which distinguishes four oblique complements to manage agenthood, and applies it to English periphrastic passives and non-anaphoric third person plurals. Before examining the different categories of prepositional phrases, let us have a look at the presence of this item in the two passive-like constructions. Examples (346) and (347) illustrate a case of passive-like structure with (*by)* and without a prepositional phrase.

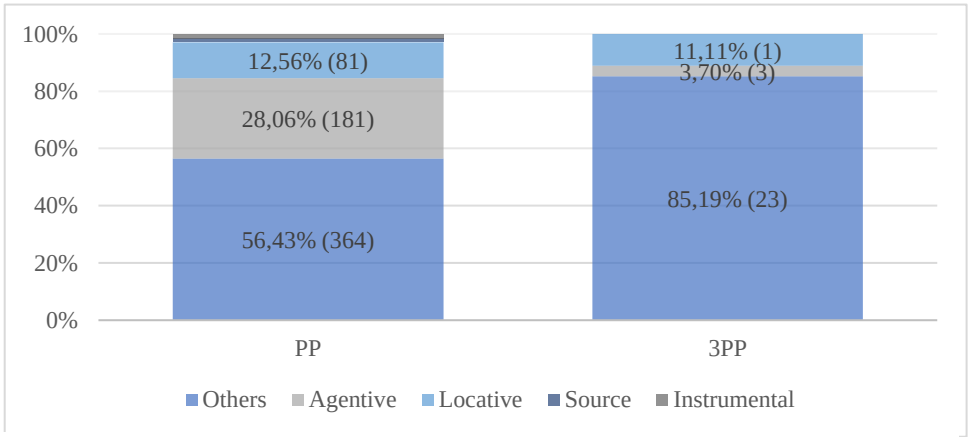
(346) *Pollen **is** released and **carried by** the wind to female cones . (*WIKI EN)

(347) *I have green carnations could somebody tell me **have** these flowers **been dyed**? or can there be grown like that.? </question>* (YAHOO EN)




*Graph 102. Presence of a prepositional phrase in English passive-like structures*

As can be seen in Graph 102, the periphrastic passive contains a higher percentage of prepositional phrases (45.9%; 645 occurrences) than the non-anaphoric third person plural (14.5%; 27 occurrences). A more detailed approach reveals the following distribution of the different categories of prepositional phrases[42] (see Graph 103).




*Graph 103. Distribution of prepositional phrases in English passive-like structures*

[42] Locative prepositional phrases and spatio-temporal restrictions overlap in 33 (out of 81) cases for the periphrastic passives and 3 (out of 3) cases for the non-anaphoric third person plurals.

The non-anaphoric third person plural, which hardly resorts to prepositional phrases, almost exclusively contains the type *others* (85.2%; 23 occurrences) such as illustrated in (348).

(348) *<"S0556">but I live more in <anon type="place"/> than you do and I have got free travel</u> <"S0405">how?</u>*
*<"S0402">how?</u> <"S0402">oh you moved innit and then you kept the Oyster cheeky</u> (...)*
*<"S0556">mine was a sixteen plus and I have been going to <anon type="place"/> so where did I move from?</u>*
*<"S0405">you just you just kept the same address</u>*
*<"S0556">what?</u>*
*<"S0405">you just use the same address as when you</u>*
*<"S0556" trans="overlap">yeah but when you order an Oyster it comes to your home address* ***they deliver*** *it to you</u>*
*<"S0405">oh you just get them in the post</u>* (BNC)

In (348), a group of people is speaking about how to get an Oyster card, which is a card allowing you to travel on most transport services in London. The non-anaphoric third person *they deliver* is used with the preposition *to*, which in is classified as *others* as it does not bring further information as for agenthood management.

While it is the main category in the non-anaphoric third person plural (85.2%), the type of preposition *others* accounts for a smaller proportion (56.4%;364 occurrences) of the data in the periphrastic passive. The typical agentive preposition *by* (as in 349 which introduces the agent *fruit-eating birds*) constitutes the second main group of prepositional phrases. The high use of the preposition *by* with the periphrastic passive was expected since it is the most usual device to introduce a demoted agent in this type of structure.

(349) *In the families Podocarpaceae , Cephalotaxaceae , Taxaceae , and one Cupressaceae genus ( ' ' Juniperus ' ' ) , the scales are soft , fleshy , sweet and brightly colored , and* ***are eaten*** *by fruit-eating birds , which then pass the seeds in their droppings .* (WIKI EN)

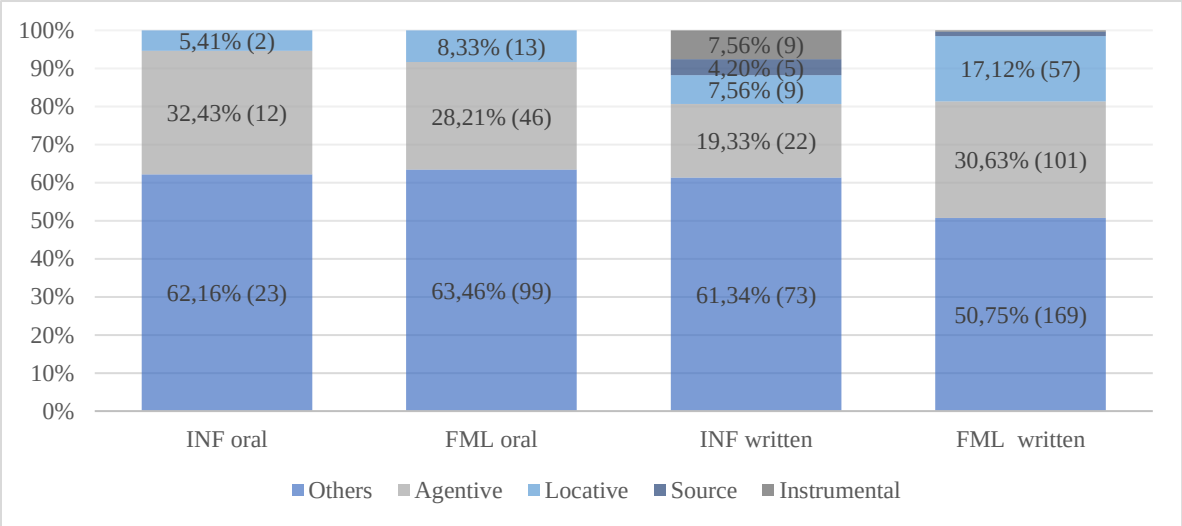
It is worth mentioning that locative prepositions, though less frequent than *agentive* and *others* oblique complements, still gather 12.6% (81 occurrences) of the data. Locative prepositions bring information on the settings within which the passive-like structure happens.

(350) All Member States should sign up now and implement common standards of justice so these gangs know **they will be caught and punished** anywhere in the EU (PROCEP EN)

In (350), the gangs will *be caught and punished* in the whole European Union. The use of the prepositional complement *in the EU* can be considered as a metonymy for member states and individuals working for the institution and ensuring respect of laws. The use of this locative complement allows to restrict the range of possible agents involved in the process of catching and punishing the gangs.

Source and instrumental oblique complements are hardly used with the periphrastic passive and absent from the non-anaphoric third person plural.

The breakdown of the data obtained for the periphrastic passive reveals that mode and register affect the type of preposition used in each of the four data set. First, propositional phrases are rather used in written data (452 occurrences) than in oral data (193 occurrences). It should also be mentioned that 51.6% (333 occurrences) of all the prepositional phrases found in periphrastic passives are used in formal written data.




*Graph 104. Distribution prepositional phrases in English periphrastic passives*

While, as can be seen in Graph 104, prepositions are distributed similarly in informal and formal oral data, the two written data sets display much more differences. Informal written data is the most heterogenous data set since, though in limited proportions, it is the only one containing the five types of prepositional phrases. Compared to informal written data, formal written data include significantly more agentive and locative prepositional phrases[43]. The high proportion of these two prepositional phrases goes hand in hand with the restricted use of prepositions classified as *others*. This leads us to consider formal written data as the data set that resorts the most to agenthood management devices. In sum, while the behaviour of oral data is rather stable in the periphrastic passive, each set of written data displays its own specificities. As the non-anaphoric third person plural hardly resorts to prepositional phrases, the influence of mode and register could not be examined.

### 9.4.3. Other parameters

This section summarises the results obtained for the four parameters considered by the MCA as contributing less to explaining the variation between the passive-like structures: presence of another passive-like structure in the linguistic context, use of an adverb, presence of a person reference in the near context and type of content. The use of another passive-like structure in

[43] Chi-square tests reveal a significant difference between formal and informal written data in the use of agentive ($\chi^2 = 5.5982$. $p < .05$) and locative prepositions ($\chi^2 = 6.4177$. $p < .05$).

the same sentence or in the near-context of the passive-like structure (illustrated in 351 with *was granted* and *are employed*) appears to be rather similar in the periphrastic passive and the non-anaphoric third person plural (70.7% and 79%, respectively).

(351) *However, a serious problem arose because the Committee on Employment and Social Affairs* ***was granted*** *exclusive competence on the Social Agenda under Rule 54.{S} Unbelievably their text infers that professional pilots on European airlines* ***are employed*** *on illegal, bogus contracts, thereby putting passengers in danger*. (PROCEP EN)

In the same vein, no important differences were found as for the use of an adverb. Both passive-like structures appear to be accompanied by an adverb in respectively 17% and 10.8% of the cases, as exemplified in (352) with the adverb *generally*.

(352) *The following breeds* ***are*** *generally* ***described*** *as giant breeds : * alsatian dog * Brazilian mastiff Presa canario *argentine dogo * Alaskan Malamute * American Akita * Anatolian Shepherd * Bernese Mountain Dog * Black Russian Terrier * Boerboel * Borzoi * Bullmastiff (…).* (WIKI EN).

As far as the type of content is concerned, both passive-like structures show a preference to be used in a neutral context, that is, not containing any indicators that could guide towards either a restrictive interpretation of the defocussing mechanism (such as personal experiences) or, conversely towards a more general or universal lecture, for instance by including adverbs favouring generalisation.

The last parameter related to the context is the presence of a person reference in the linguistic context of the passive-like structure. Results indicate that while the non-anaphoric third person plural tends to appear with such elements in its context, the periphrastic passive would rather appear without any reference.

Example (353) illustrates a case of periphrastic passive with a reference to a third person singular, namely *the dog*. The presence of the common noun *the dog* helps in understanding who has to accomplish the action of the periphrastic passive *to be retrieved*.

(353) *The hunter stays hidden in a blind and sends the dog out to romp and play near the water , usually by tossing a ball or stick* ***to be retrieved*** . (WIKI EN)

Though some limited differences can appear in percentages between the periphrastic passive and the non-anaphoric third person plural, the calculation of the MCA reveals that they do not contribute considerably to explain the different behaviour of the English passive-like structures.

## 9.5. *Get*-passive, an alternative to periphrastic passive?

The *get*-passive is considered as a possible alternative to the periphrastic passive in English, though more restricted and usually associated to adversative contexts. In Chapter 6 we already observed that the *get*-passive is far less used than the periphrastic passive. This finding largely

concurs with what is generally reported in the literature. As the figures are quite low (49 occurrences), the general description of the results for this structure considers the occurrences altogether and does not distinguish between the two modes and registers. The analysis has however revealed that the *get*-passives in our data are mainly found in oral informal productions (65,3%) and, to a lesser extent, in written informal data (24,5%). Formal data hardly resort to *get*-passives for defocussing purposes. In the literature, the *get*-passive is often recognized as a possible alternative to the French *se faire* (Raineri 2010). The present study could not provide more information on this issue as *se faire* constructions are absent from our set of data.

### 9.5.1. Verb-related parameters

As far as verb-related parameters are concerned, results indicate that the *get*-passive forms found in our data are mainly used with present tenses (71.43%; 35 occurrences) and, in a smaller proportion, past tenses (26.53%; 13 occurrences). Future tenses hardly occur with this passive-like structure. The parameter *tense* largely correlates with the imperfective aspect (73.37%; 36 occurrences). These two variables are exemplified in (354).

(354) *S0555">cos you know what? in Pakistan they have a blasphemy law</u> (...) <"S0555">erm so like if they do something which is blasphemous you* ***get charged*** *and people just use it because if there is someone that they do not like</u>* (BNC EN)

The analysis of *verb mode*, in direct relation with the previous two parameters shows that *get*-passives are usually used in the indicative (71.43%; 35 occurrences). In approximately 20% (10 occurrences) of the cases, the clues indicating *verb mode* have been classified as *others*. This category mainly contains gerunds such as (355).

(355) *<"S0555">yeah they just show the videos of where people are dying</u> <"S0402">oh yeah and some people* ***getting shot*** *in the face</u>* (BNC EN)

All the occurrences of *get*-passives are found to be used with dynamic verbs such as in (356) where the process indicates an action, in this case to hurt someone by hitting them. This exclusive use of dynamic verbs appears as one of the defining features of the *get-passive.* Let us remember that the periphrastic passive contains 89.6% of dynamic verbs, thus allowing more flexibility regarding the parameter *kinesis*.

(356) *<S0556" trans="overlap"><unclear/> a woman* ***getting beaten up*** *she got beaten to death because she apparently burnt burnt the</u>* (BNC EN)

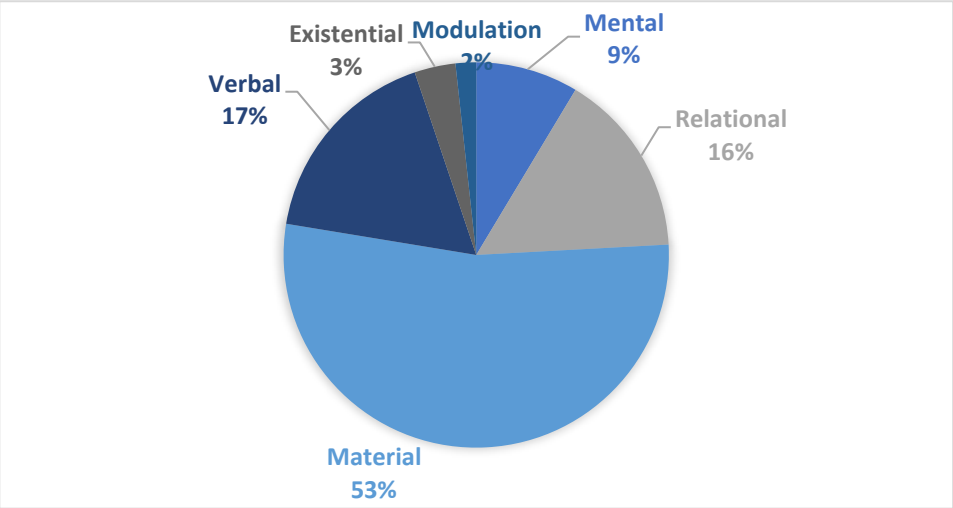
As for the dichotomy realis/irrealis, *get*-passives rather encode real events (73.47%; 36 occurrences) than irreal ones. Examples (357) and (358) respectively illustrate these two options to represent the situations or actions.

(357) *Mr President, some 35 years ago I got a job as a history teacher.{S} I was a terrible teacher.{S} Terrible.{S} Fortunately* ***I got elected*** *to Liverpool City*

*Council and I found an excuse for getting out of the classroom, which was good for me and good for my students*. (PROCEP EN)

(358) *<"S0572">it **could get changed** for a week well cos cos apparently England have that vote now so like well England have the control to do that</u>*
*<"S0571" trans="overlap">what?</u>*
*<"S0572">cos that was part of the EU the EU made sure that people got their bins cleaned every two weeks I searched it up</u>* (BNC EN)

The large proportion of realis clauses can be linked with the high proportion of affirmative clauses found with get-*passives* (83.67%; 41 occurrences).




*Graph 105. Distribution of types of verb processes in the get-passive*

The types of verb processes included in *get*-passives are distributed in the following way. As shown in Graph 105, half of the *get*-passives (53%) in our data describes material processes such as (359).

(359) *<"S0571"><> really?</u>*
*<"S0572">it could **get changed** for a week well cos cos apparently England have that vote now so like well England have the control to do that</u>*
*<"S0571" trans="overlap">what?</u>* (BNC)

Nearly equally represented, the second and third most frequent categories of verbs are verbal (17%) and relational processes (16%).

Following the results obtained in our data, the prototypical verbal features of a *get*-passive construction would be a construction with a present tense, an indicative mode, imperfective aspect and describing a dynamic situation, rendered through a material type of verb. The situation would be described in the affirmative way and encoded in the real world.

### 9.5.2. Participant-related parameters

*Get*-passives are always built with at least two participants, an agent (expressed or not) and a patient. The agent is usually defocused (83.67%; 41 occurrences) or, when it is expressed, is an animal such as in (360), where the agent is a dog.

(360) *My mother **got bitten by** a small fluffy mutt (Had to go to the hospital) and never got asked what type of dog bit her so I do not trust any bite statistics*. (YAHOO EN)

In our data, inanimate entities never endorse the role of agent in *get*-passives. Regarding patients, they are mainly human (Example 360) (79.59%; 39 occurrences) and in a much more restricted number of cases (18.37%; 9 occurrences), inanimate *others* such as in (361).

(361) *{S} SO! i looked up vegan videos who i was hoping to have some advice on youtube, and i found videos like '10 reasons meateaters are stupid' and stuff like that.{S} I watched a few and these were their pointers of most{S} 1. you have rotting flesh in you (fact : food does not rot inside you it **gets digested**){S} 2. meat eaters are violent cause of it{S} 3. livestock is fed on antibiotics which makes you sick* (…) (YAHOO EN)

All patients are totally affected and approximately three quarters (77.55%; 38 occurrences) of them can be considered as individuated: meeting four out of the six criteria established by Hopper & Thompson (1980) as illustrated in (362). The personal pronoun *I* complies with the following criteria: definite, animate, singular and concrete.

(362) *Mr President, some 35 years ago I got a job as a history teacher.{S} I was a terrible teacher.{S} Terrible.{S} Fortunately **I got elected** to Liverpool City Council and I found an excuse for getting out of the classroom, which was good for me and good for my students.{S} (PROCEP EN)*

The relatively high number of individuated patients can be linked to the high proportion of human patients, which, since they automatically meet the criterion *animate*, are good candidates to meet at least four of the criteria established by Hopper & Thompson (1980).

### 9.5.3. Context-related parameters

*Get*-passives appear as structures functioning with few elements in the linguistic context that could guide the interpretation of any possible agent. This relative absence of contextual elements helps to lay a greater focus on the process the patient undergoes. These constructions usually appear without a person reference in the context (83.67%) and very few adverbial complements (10.20%) or spatiotemporal restrictions (4.08%). Similarly, the general co-text in which they occur is considered as neuter as, most of the time (95.92%), there are no indicators towards a restrictive or generic/indefinite interpretation. As for the use of prepositional phrases, approximately half of the *get*-passives found in our data appear alone (44.90%). When there is

a prepositional phrase, two categories emerge. The first option is the use of the prototypical agentive preposition *by* (16.33%; 8 occurrences) as in (360) where the agent *fluffy mutt* is introduced with the agentive preposition. The second option is the category *others* (38.78%; 19 occurrences), thus not serving for agenthood management, as in (363) where the *get killed* is followed by the preposition *for*.

(363) *<"S0570" trans="overlap">do you know the sign of not being able to admit it is the sign that you are an addict?</u> <"S0571"><></u> <"S0517">how do you know?</u> <"S0572">I could lay into all three of you</u> <"S0571" trans="overlap"><unclear/></u> <"S0572">and* ***get killed*** *for it but when you two lay into me it is alright</u> <"S0517">we are not laying into you <anon type="name" nameType="m"/></u>* (BNC EN)

The only element in the context that could help interpret the *get*-passive and its relation to participants is the presence of another passive-like structure (48.98%; 24 occurrences). Example (364) illustrates a *get*-passive (*got asked)* co-occurring with another defocussing device, in this case another *get*-passive, *got bitten*.

(364) *{S} My mother got bitten by a small fluffy mutt (Had to go to the hospital) and never* ***got asked*** *what type of dog bit her so I do not trust any bite statistics.{S} And that does not even take into consideration how popular both breeds are. </answer>* (YAHOO EN)

This restricted presence of contextual elements guiding the interpretation of a possible agent could be linked to the fact that *get*-passives are specific to situations considered as problematic from the speaker's perspective (Carter &McCarthy 1999). In such problematic situations, a high degree of defocussing seems logical.

The *get*-passive shows a strong preference to appear in informal oral data (65.3%; 32 occurrences) and, to a lesser extent, in informal written data (24.5%; 12 occurrences). Some of the features observed below can be explained by this tendency to appear almost exclusively in the informal register. For instance, we assume that the absence of contextual elements such as adverbial complements or spatiotemporal restrictions can be linked to the fact that, in informal data, interlocutors are more likely to have a higher degree of shared knowledge on the matter discussed. The addition of elements to specify the settings may thus appear as superfluous. Biber (1995) also points out that the temporal organization of the discourse (*tense* and *aspect*) is highly sensitive to register. Results indicate that present tenses and imperfective forms are more frequent in the *get*-passive than in the periphrastic passive, a passive-like structure usually associated with more formal registers and written productions. A large proportion of present *tenses* and imperfective forms would therefore be a distinctive element between the two constructions, which largely correlates with the degree of formality and the mode of production of the data.

## 9.6. Numeral-based *one*

Before drawing the portrait of the English passive-like structures, it is worth briefly commenting on the behaviour of the numeral-based *one*. Weiner & Labov explain that even though *one* is presented as a major equivalent of the man-impersonal *on* (in informal data), the frequency of *one* appears as far more restricted, since in their study they did not obtain any occurrences of the construction (1983:35). Our results confirm the rarity of the phenomenon since only nine examples of *one* were obtained in our data. The low frequency of the structure in our data corroborates Gast's (2015) findings. The author adds that in English speakers would rather opt for other strategies to defocus the agent such as the passive or the non-anaphoric third person plural (2015:9).

The tentative observations made in this section, based on absolute frequencies, should therefore be considered cautiously as figures remain low and tendencies are only valid for this study.

Unlike the man-impersonal *on*, the numeral-based *one* can, in addition to its subject role, perform a number of other syntactic functions such as direct and indirect object complement or prepositional complement (see for instance Example (365)).

(365) *Part of maturity is being able to recognize that **one's desires** may not be obtainable..{S} (YAHOO EN)*

Nine occurrences of *one* were obtained in the data set used in this investigation, two of which occupy another position than the subject one. Hopper & Thompson's parameters therefore do not apply as the construction is not directly co-occurring with a verb. Therefore these two occurrences have not been considered in the present discussion.

The analysis revealed that the verbs used with the construction *one* in our present data are equally distributed between dynamic (3 occurrences) and stative verbs (4 occurrences). They tend to appear in durative situations rather than in punctual ones. The verbs present in the clause with *one* are only used in the imperfective aspect (7 occurrences). Example (366) illustrates a case of a dynamic verb used in a durative situation with an imperfective form.

(366) *We do not always know best in the west!{S} Not all civilisations eat meat, lots of Indians do not and I am sure there are hundreds others if **one reads** a little about the world....I have had excellent vegetarian cuisine in Thailand for example. (YAHOO EN)*

As for the parameter *volitionality*, the impersonal strategy *one* is as frequently used with a volitional (4 occurrences) than with a non-volitional verb (3 occurrences). The verbs preferentially occur in an affirmative clause (6 occurrences and tend to encode real events (5 occurrences). As far as types of processes are concerned, the verb either falls in the category of mental (3 occurrences), material processes (3 occurrences) and relational processes (1 occurrence).

As for participant-related parameters, *one* appears in one-participant and two-participant clauses (4 occurrences and 3 occurrences, respectively). The agent is always an animate human

entity. Finally, when there is a patient, it falls in the category of inanimate *others* such as in (367) where *skyr* is a type of yogurt and the patient is always individuated and totally affected.

(367) *By the way, how does **one** make skyr completely from scratch without existing skyr?* (YAHOO EN)

Due to the scarcity of the data, the defocussing mechanism based on *one* will not be further commented upon.

## 9.7. Portraying the English system

This final section recapitulates the main findings regarding verb-related, participant-related and context-related parameters in the periphrastic passive and the non-anaphoric third person plural in order to establish a detailed description of these two passive-like constructions. Data regarding the features of the *get*-passive have been added in grey scale as only 49 occurrences of this passive-like structure have been obtained. However, though only tentative conclusions can be made, these data may bring some relevant information to complete the panorama of passive-like structures in English. The distinctive features, based on significant differences, have been highlighted in bold.

| | | Periphrastic passive | Non-anaphoric third person plural | *Get*- passive |
|---|---|---|---|---|
| Verb-related parameters | Tense | present (65.8%) | present (66.5%) | present (71.4%) |
| | Aspect | imperfective (66.9%) | **imperfective (74.7%)** | imperfective (73.5%) |
| | Verb mode | indicative (75.9%) | **indicative (84.9%)** | indicative (71.4%) |
| | Type of verb process | **material (45.9%)** | material (28.5%) | material (53%) |
| | | relational(15.9%) | relational (26.3%) | verbal (17%) |
| | | mental (14.7%) | **modulation (16.7%)** | relational (16%) |
| | Kinesis | **dynamic (89.6%)** | dynamic (83.1%) | dynamic (100%) |
| | Realis/irrealis | **realis (77.3%)** | realis (68.6%) | realis (73.5%) |
| | Affirmation | **affirmative (93.4%)** | affirmative (84.9%) | affirmative (83.7%) |

| | | | | |
|---|---|---|---|---|
| Participant-related parameters | No. of participants | **2+ participants (98.9%)** | 2+ participants (60%) | 2+ participants (100%) |
| | Agent | **defocused (85.8%) + inanimate *others*/animate human** | **animate human (100%)** | defocused (83.7%) |
| | Patient | inanimate (71.1%) | **inanimate (80.5%)** | animate human (79.6%) |
| | Individuation P | non individuated (57.1%) | **non individuated (71.9%)** | individuated (77.5%) |
| | Affectedness P | affectedness (99.4%) | affectedness (99.5%) | affectedness (100%) |
| Context-related parameters | Spatio-temporal restriction | presence (13.9%) | presence (18.3%) | presence (4.1%) |
| | Prepositional phrase | **presence (45.9%)** | presence (14.5%) | presence (55.1%) |

*Table 32. Overview of the distinctive features of English passive-like structures*

As indicated in Table 32, the behaviour of the periphrastic passive and the non-anaphoric third person plural diverges for all verb-related parameters, except for *tense* where they both resort to present forms in similar proportion. The non-anaphoric third person plural can be characterized by a high use of imperfective forms and the indicative mode. Though the imperfective aspect and the indicative mode remain the most common features in the periphrastic passive, they appear in smaller proportions if compared with the non-anaphoric third person plural. The type of verb process is another parameter that distinguishes the two passive-like structures. While the periphrastic passive largely relies on the material processes, the distribution of the different verb processes is much more balanced in the non-anaphoric third person plural. A specific feature of this passive-like structure is the high proportion of modulation processes, whereas the presence of existential processes appears as a defining feature of the periphrastic passive, since they are absent from the non-anaphoric third person plural. A last observation regards the way the two passive-like constructions anchor their content in the real world. Table 32 shows that the periphrastic passive displays a higher percentage of realis and affirmative clauses, which means that this construction tends to encode more frequently its content in the real world than the non-anaphoric third person.

The number of participants included in the passive-like structure and their nature are other factors that allow to explain the variation between the periphrastic passive and the non-anaphoric third person plural. First, the number of participants appears as much more constrained in the periphrastic passive, which appears almost exclusively in two-participant clauses. Conversely, the non-anaphoric third person plural enjoys a larger flexibility with a more balanced profile between one- and two-participant clauses. The nature of the participants

included in the constructions also differs. While, due to its syntactic nature, the non-anaphoric third person plural always consists in an explicit human agent, the periphrastic passive mainly defocuses its agent, which can be animate or inanimate. As for the patients, inanimate patients predominate in both passive-like constructions. Nevertheless they appear in significantly different propositions, the non-anaphoric third person plural being the one to resort to the highest number of inanimate patients. This result contributes largely in explaining the higher use of non-individuated patients in the non-anaphoric third person plural since, as we have seen, the degree of individuation is positively correlated with the nature of the patients. Finally, the parameter *affectedness* remains stable across the two passive-like structures.

Let us now recapitulate the findings obtained for the two context-related parameters. Results indicate that the periphrastic passive and the non-anaphoric third person plural resort to spatiotemporal restrictions in similar proportions. Conversely, the presence of a prepositional phrase in the passive-like construction is a key factor contributing to the variation of the two constructions since the periphrastic passive is much more likely to occur accompanied by a prepositional phrase than the non-anaphoric third person plural. Agentive and locative prepositions are the two main oblique complements responsible for agenthood management in the periphrastic passive.

A last finding uncovered by this study is the importance of mode and register in the distribution of the data. The spoken versus written nature of the data as well as the degree of formality are crucial factors affecting the distribution of the data in three verb-related parameters: *tense, aspect* and *type of process.* While we observed that overall the distribution of the three different categories of tenses does no vary considerably between the periphrastic passive and the non-anaphoric third person plural, we saw that data are highly sensitive to mode and register. A similar phenomenon could be observed for the parameter *aspect,* positively correlated with the parameter *tense.* In a nutshell, the analysis revealed that present tenses and imperfective forms are closely linked to formal written data. As for the type of process, results have shown that, while material processes dominate in written data with the periphrastic passive, the non-anaphoric third person plural rather relies on relational verbs in written data. This difference may be linked to the higher use of the periphrastic passive in written productions while the non-anaphoric third person plural is rather associated with more informal and oral contexts. Such different contexts of use of the two passive-like constructions may explain the distinct profile as the type of verb process. We could also observe that written data sets are likely to be more homogeneous in the distribution of types of verb processes compared to oral data sets. Finally, the mode and the register exert a more restricted influence on the parameters *verb mode, kinesis, realis/irrealis* and *affirmation*.

Participant-related parameters are also largely affected by the mode and the register of the data within which the passive-like construction occurs. Findings showed that the use of two-participant clauses is a highly associated with written data. This may be explained by a higher degree of syntactic elaboration, which is usually reported to characterize written data. As for the use of agents, the analysis has indicated that the great majority of overt agents in the periphrastic passive are found in formal written data. As pointed out by Biber (1995), formal written data possess a high degree of explicitness, which could explain the more frequent

presence of overt agents in this type of production. The nature of patients and its degree of individuation are also largely impacted by the mode and the register, especially in the case of the periphrastic passive.

The last set of parameters, related to the context, also displays considerable differences regarding the nature of the data and the degree of formality. Our study indicates that spatiotemporal restrictions are far more used in formal written data, which would correspond to the higher need of explicitness of this type of context. Regarding the presence of a prepositional phrase, its presence in oral data sets is rather stable but in written data, a greater variety of prepositions is observed in informal written data.

To sum up, both similarities and dissimilarities between the two English passive-like structures have been uncovered. When examining specific parameters, the same value is often dominant in both constructions. However, it is regularly significantly more dominant in one construction than in another. Nevertheless, as discussed above, passive-like structures display much more dissimilarities, leading to the establishment of a distinct profile for each of the periphrastic passive and the non-anaphoric third person plural. Moreover, this analysis has proved that the spoken versus written nature of the data as well as the degree of formality play an essential role in the distribution of the data. They are thus important factors to consider when analysing syntactic-pragmatic parameters.

Hereafter, I briefly comment on the distinction between the periphrastic passive and the *get*-passive which are often presented as possible alternatives in English. Nevertheless, these two structures are highly register-sensitive. The periphrastic passive is generally considered as a formal device and is undoubtedly more present in formal and written data. The *get*-passive is usually presented as characteristic of oral productions and, more specifically, of informal oral situations (Huddleston 1971, Leech & Svartvik 1975, Anderwald 2018, Rühlemann 2007). This section suggests a tentative comparison of the two passive-like structures and examines the extent to which they can be considered as alternatives. Since the number of occurrences obtained for the *get*-passive is relatively low (49 occurrences), this section adopts a qualitative approach, only suggesting main tendencies.

Our results suggest that the parameters *tense* and *aspect* vary slightly between the periphrastic and the *get*-passive, with a higher use of present tenses and imperfective forms in the *get*-passive but a larger set of data would be required to confirm this tendency. No difference could be identified concerning the parameter *verb mode*. The proportion of realis and affirmative clauses, two parameters which provide information on the way the passive-like structures encode events and situations in reality, is again similar in the two constructions. The type of verb process is the parameter that is suspected to vary the most since results tend to indicate that the *get*-passive displays much more material and verbal processes than the periphrastic passive in informal data. The slightly more frequent use of dynamic verbs observed in *get*-passive could be correlated to the higher frequency of material processes since this type of process includes verbs involving physical actions such as activity, control, use, etc.

The number of participants involved in the clause is very similar in the two passive-like structures, which both display a strong preference for two-participant clauses. The *get*-passive

tends to use an overt agent slightly more often than the periphrastic passive. The preference of the *get*-passive for animate agents is one of its distinctive features since it was observed that the periphrastic passive rather includes inanimate entities as overt agents. A similar tendency, though less marked, can be observed for patients. Participants seem thus to be rather different in periphrastic passives and *get*-passives.

As for contextual elements, it can be suggested that the two structures have their own specificities. The larger presence of spatiotemporal restrictions seems to be characteristic of the periphrastic passive while the presence of prepositional phrases is a distinctive feature of the *get*-passive. In sum, it can be suggested that, though the two passive-like constructions share a series of features, they also possess their own distinctive properties mainly situated on the level of the participants and the semantic of the verb.

# 10. A contrastive view of the data

This final analytical chapter contrasts and compares the results obtained in the preceding three chapters devoted to the description of French, Spanish and English passive-like structures. The present chapter contains two parts. The first one contrasts the different passive-like constructions in the three languages involved while the second part focuses on the role played by the degree of formality and the mode of production in the distribution of our data.

## 10.1. Contrasting passive-like structures in French, Spanish and English

The first part of this chapter aims at examining to which extent the passive-like structures included in this study diverge or, on the contrary, present similar behaviour. In (10.1.1) a comparison of the canonical periphrastic passive in the three languages is offered, followed by a similar section on the non-anaphoric third person plural (10.1.2). The (dis)similarities of the *se*-construction in French and in Spanish are addressed in 10.1.3. Section 10.1.4 examines to which extent the *se*-construction and the periphrastic passive can be considered as complementary constructions in Spanish and in French. Finally, in 10.1.5 we intend to determine the position of the French man-impersonal *on* on the general continuum of passive-like structures.

### 10.1.1. The periphrastic passive

The objective of this section is to compare the periphrastic passive in French, Spanish, and English. Though the construction is available in the three languages, differences both regarding their frequency of use and the distribution of their respective characteristics can be observed. Table 33 offers a reminder of the frequency of use of the periphrastic passive (raw and per ten thousand words), already discussed more extensively in Chapter 6.

| | **English** | | **Spanish** | | **French** | |
|---|---|---|---|---|---|---|
| | raw | pttw | raw | pttw | raw | pttw |
| **Periphrastic passive** | 1415 | 64.5 | 325 | 15.56 | 857 | 42.05 |

*Table 33. Absolute and relative frequencies of the periphrastic passive in English, French and Spanish*

Table 33 indicates that English is the language that relies the most on that passive-like structure whereas Spanish shows a very restricted use of it. These results are in line with what is usually reported in the literature on the frequency of the periphrastic passive (see Siewierska 1984). The high use of periphrastic passives in English may partly be explained by the fact that English seems to possess a smaller range of passive-like structures than the other two languages. Orts (2018:809) reports that Spanish relies on a much wider range of agent-defocussing devices than English. Siewierska (1984: 230) also puts forward the rigid word order of English and the possibility of passivizing indirect and prepositional objects as factors explaining the higher use of the periphrastic passive in English.

The more restricted use of the periphrastic passive in Spanish concurs with what is usually reported in the literature, that is, a tendency for Spanish to prefer other passive-like structures than the periphrastic passive (Quesada 1998, Mendikoetxea 1999, Cabañas Maya 2006). French occupies an in-between position between English and French. Such frequency differences already suggest that formally similar structures may be used differently.

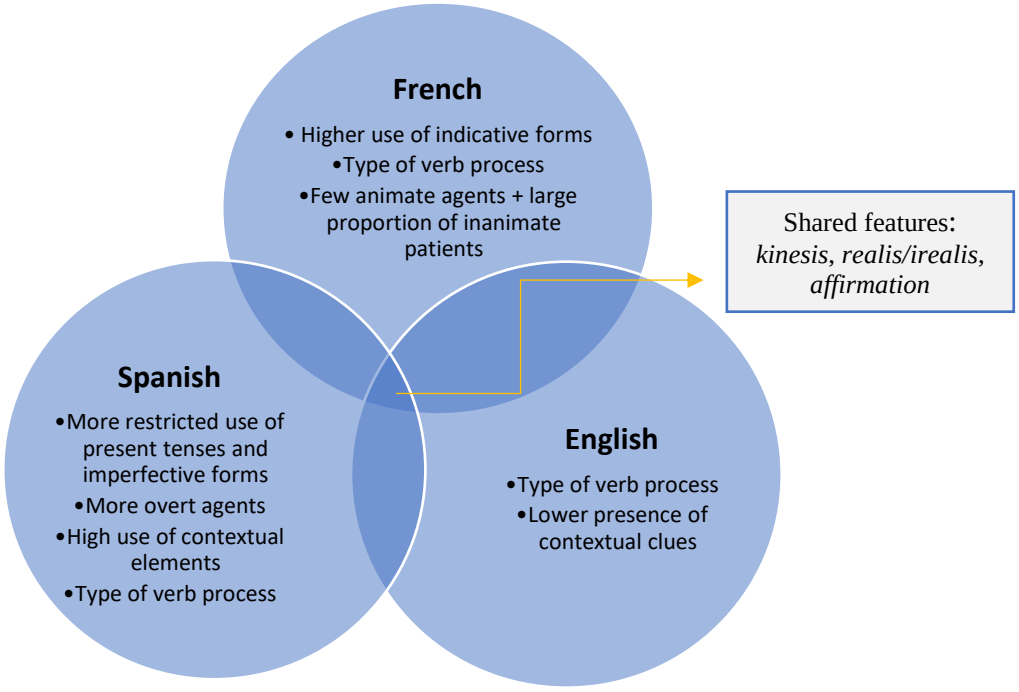



*Figure 26. Summary of the specific characteristics of the periphrastic passives in English, French and Spanish*

Figure 26 visualises the distinctive and the shared features of the periphrastic passive across the three languages, each represented by a circle. The features have been selected on the basis of chi-square tests. For the reader's convenience, these chi-squares can be found in Appendix 6.

A look at Figure 26 reveals that the three circles overlap in the centre. This common area represents the shared characteristics between the three languages. As for verb-related parameters, the three languages concur on the following features: *kinesis*, *realis/irealis*, *affirmation*. MCA graphs revealed that English and French share more characteristics between them than with Spanish. This phenomenon is represented in Figure 26 with a slightly larger

overlapping area between French and English circles. More specifically, these two languages have a similar relation to *tense* and *aspect* and defocus their agent in comparable proportions.

Let us now focus on the distinctive characteristics of the three periphrastic passives. As for verb-related parameters, the Spanish periphrastic passive is characterized by a relatively low number of present tenses compared to French and English. Since the proportion of future tenses is similar in the three languages, it can thus be concluded that the more restricted proportion of present tenses in Spanish goes hand in hand with a greater presence of past tenses compared to the other two languages. Moreover, this more restricted presence of present tenses correlates with a smaller proportion of imperfective forms. Spanish periphrastic passives are indeed those to display the smallest proportion of imperfective forms. This finding corroborates Weber's claim that Spanish periphrastic passive tends to use more perfective tenses than the French one (2014:307). It has to be mentioned, however, that the difference is only significant for the pair Spanish versus French. As for the *verb mode*, results indicate that while the English and Spanish passive periphrastic display a very similar proportion of indicative forms, the French periphrastic passive stands out with a significantly higher proportion. The type of verb process included in the periphrastic passive is another element that distinguishes the three languages. Although the three main categories of verbs are the same in English, French and Spanish, their proportion, and more especially the proportion of the most frequent category – i.e., material processes, varies. Spanish periphrastic passive is the one that resorts the most to them (55%), followed by French which occupies an intermediate position (50%) and then English (45.9%). The proportion of less represented types of verb processes also varies but not in a significant way. It is worth mentioning that, though in restricted proportion in some cases, all types of verb processes are included in the periphrastic passive. As it will be seen later on, this is not the case for the non-anaphoric third person plural. Let us now move on to participant-related parameters.

The main differences between the three languages affect the presence of an overt agent and its nature. The Spanish periphrastic passive contains significantly more overt agents (22.7%) than its French and English counterparts (14.5% and 12.8% respectively). This would reflect a higher degree of explicitness in Spanish periphrastic passive compared to the other two languages. Alternatively, it can be assumed that the Spanish periphrastic passive is the typical structure when overt agents are expressed whereas, when the agent is entirely defocused by non-elaboration, Spanish tends to use another passive-like structure, for instance the *se*-construction. Another interesting result yielded by this study regards the nature of the agent. While the proportion of animate versus inanimate agents is more or less equally distributed in English and Spanish, French displays a much smaller proportion of animate agents. A possible explanation to this phenomenon would be the presence of a man-impersonal construction in French. While, in our study, the Spanish and the English system only have the non-anaphoric third person plural consisting in an animate human as agent, French possesses a second option, the man-impersonal *on*. The possibility for French speakers to use this additional passive-like structure may partly explain the restricted number of animate agents in the periphrastic passive. Similarly, the nature of the patients also displays some differences according to the language. Results indicate that the French periphrastic passive is the passive-like structure to resort the most to inanimate patients while the English structure is the one to resort the least to them, the Spanish periphrastic occupying an in-between position. Rather unexpectedly, this high use of

inanimate patients does not correlate with a higher use of non-individuated patients as we could observe at some points in the preceding analytical chapters. The proportion of non-individuated patients is similar across the three languages. This means that French uses more inanimate individuated patients than English and then Spanish though the difference is not significant in our data, displaying a slightly different relationship as for its patients. It can therefore be assumed that the general nature of the participants included in French periphrastic passive is rather inanimate while the English and Spanish contain much more animate entities.

The exploration of the context-related parameters suggests that Spanish periphrastic passives rely heavily on the use of contextual elements to provide further information on the possible participants included in the defocussing mechanism. The Spanish construction is indeed the one that includes the largest proportion of prepositional phrases and spatio-temporal restrictions whereas English is the language that includes them the least. Within prepositional phrases, results indicate that English and French rely on oblique complements serving agenthood management purposes in relatively similar proportion. Conversely Spanish contains far more typical agentive prepositional phrases. In addition to these two parameters, the MCA graph specific to the Spanish language has indicated that the presence of another passive-like structure in the linguistic context is a salient feature only in Spanish. The analysis has revealed that such device is present in nearly 60% of the cases. This frequent use of the three contextual factors mentioned above contributes to the explicit character of the Spanish periphrastic passive already highlighted with its higher use of overt agents. While Spanish periphrastic passives display a relatively high degree of explicitness, it can be suggested that its English counterpart is placed on the extreme opposite of the continuum, thus with a tendency to totally defocus the agent. French seems to occupy an in-between position between Spanish and English. These differences reflect the fact that the three periphrastic passives do not function similarly and display different relative frequencies. As we explained, the English periphrastic passive is the most frequent structure (64 occurrences pttw), followed by French (42.05 occurrences pttw) and then the Spanish one (15.56 occurrences pttw).

To recapitulate, the main specificities of the periphrastic in English, French and Spanish can be summarised as follows. In Spanish, the periphrastic passive can be characterized by a higher use of past tenses and perfective forms, a marked presence of explicit agents and an especially dominant presence of contextual elements. The French periphrastic passive can be defined by its higher use indicative forms and restricted presence of animate agents. It also displays a larger proportion of inanimate patients in comparison to the other two languages. A lower presence of contextual clues is the central distinctive feature of the English periphrastic passive. Another interesting finding shows that each of the three periphrastic passives displays a particular relation regarding the distribution of the type of process. The specific pragmatic-syntactic behaviour found for the English, the French and the Spanish periphrastic passive partly explains the important difference in their frequency of use. And lastly, the comparison of the three languages allowed us to confirm that differences between passives across languages are best reported in terms of frequency (see Xiao et al. 2006 or Fredrikson 2016). Indeed, the present study highlighted that what distinguishes the three languages is not the nature of the parameter (e.g., presence of dynamic verbs or imperfective forms) but rather the frequency of use of the

parameter. As detailed later in the following point (10.1.2), this also holds true for the non-anaphoric third person plural.

### 10.1.2. The non-anaphoric third person plural

This section is devoted to the non-anaphoric third person plural, the other passive-like structure available in the three languages. Table 34 indicates that while English and French display a rather similar relative frequency, Spanish uses considerably more the non-anaphoric third person plural.

| | **English** | | **Spanish** | | **French** | |
|---|---|---|---|---|---|---|
| | raw | pttw | raw | pttw | raw | pttw |
| **Non-anaphoric third person plural** | 186 | 8.48 | 498 | 23.85 | 109 | 5.34 |

*Table 34. Absolute and relative frequencies of the non-anaphoric third person plural in English, French and Spanish*

Besides, though this structure is specific to informal oral data in the three languages, it could be observed that English, and to a smaller extent French, contain an important number of non-anaphoric third person plural in informal written data. This phenomenon is not observed in Spanish. Given these distinct preferences as for the register and the mode, maximal differences in the features could be expected between the Spanish construction on the one hand and the English and French ones on the other. However, as we will see below, this is not the case. Results indicate that English tends to stand alone while French and Spanish share a slightly wider set of features. It is also worth mentioning that the relatively high proportion of the non-anaphoric third person plural in Spanish and its more restricted use in English and French seem to go hand in hand with a lower proportion of the periphrastic passive in Spanish and a frequent use of that construction in English and French. The two passive-like structures then seem to be complementary.

The non-anaphoric third person plural in English, French and Spanish share a series of features, as represented by the overlap of the three circles in Figure 27. Except for the parameter *affirmation*, no verb-related parameters are common to the three languages. Similarly, as for participant-related parameters, they only share the fact that in the three languages the agent is animate human and that the patient is usually totally affected. The three languages also resort to spatiotemporal restrictions in similar proportions. In sum, despite their apparent similar surface structure, the three passive-like constructions only share a relatively limited number of characteristics. The main differences, summarised in Figure 27, are discussed hereafter.

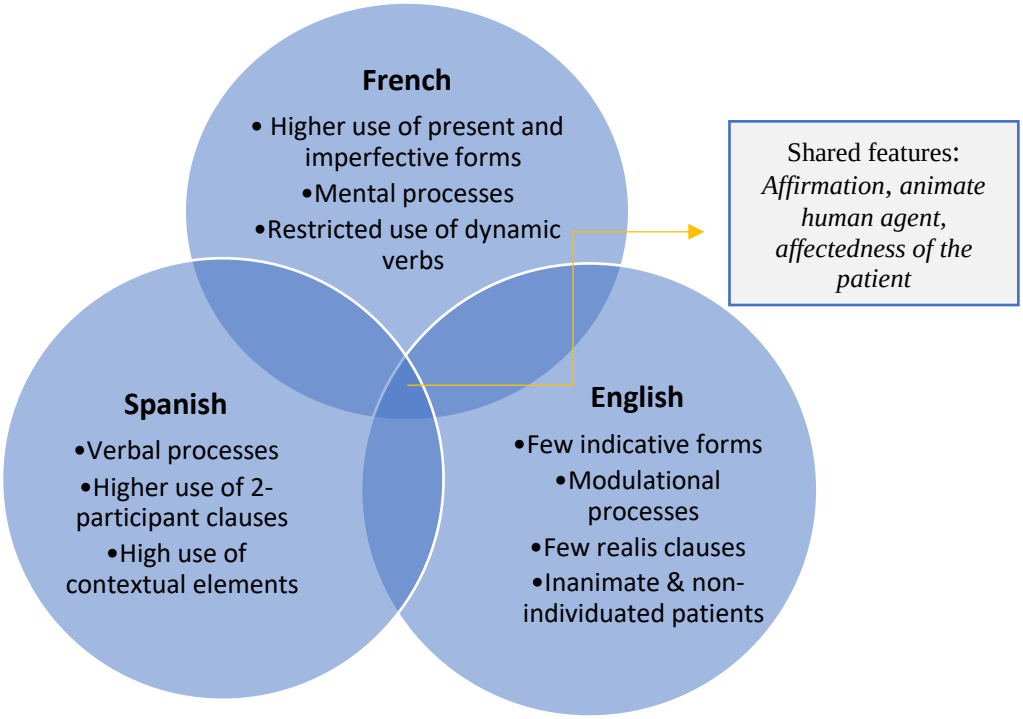



*Figure 27. Summary of the specific characteristics of the non-anaphoric third person plurals in English, French and Spanish*

Our analysis has revealed that differences could be identified in verb-related, participant-related and context-related parameters. Let us open the discussion with verb-related parameters. First, a sharp contrast can be observed as far as *tense* and *aspect* are concerned. The French non-anaphoric third person plural includes significantly more present and imperfective forms than its English and Spanish counterparts. This means that the French passive-like structure is more often used to describe on-going and habitual events or situations occurring in the present than the other two languages. Besides, French contains a significantly higher use of stative verbs compared to English and Spanish. This finding reinforces what has just been explained as stative verbs usually reflect states of feelings, thoughts or attitudes. The high use of mental types of verb processes is another specific feature of the French non-anaphoric third person plural, which can be directly related to the features commented so far. Mental verbs are indeed good candidates to be classified as stative as they express the speaker's inner thoughts which are usually more anchored in the present. It can thus be suggested that the non-anaphoric third person plural in French is rather oriented towards more on-going and habitual events and situations, inner experiences and the expression of processes of consciousness in situations anchored in the present. In comparison, English and Spanish non-anaphoric third person plurals contain a higher proportion of perfective forms and dynamic verbs, thus reflecting actions and processes considered as accomplished. A second interesting result that emerges from our analysis is the more restricted presence of indicative forms in English in favour of the category *modal auxiliaries.* As already explained, modal auxiliaries are generally good candidates to express nuances of probability and possibility. Put differently, they can reflect situations considered as not actualized. This result is supported by the high presence of irrealis clauses in

the English non-anaphoric construction. A third and last important finding this study has put forward regards the different use of types of verb processes in the three languages. While material and relational verbs remain the two most frequent categories in the three languages, the proportion of the other types of verb process varies considerably. Whereas, as already commented above, mental verbs constitute the third most frequent category of verb processes in French, verbal processes occupy this third category in Spanish and modulation verbs in English. This wide diversity as far as the type of verb is concerned means that the non-anaphoric third person plural serves different purposes according to the language. As a last comment, we would add that existential verbs are hardly used in non-anaphoric third person plurals and are even entirely absent in the case of the English and the Spanish construction.

The detailed examination of participant-related parameters has allowed to put into light two key findings. First, the Spanish non-anaphoric third person plural is the structure to resort the most to two-participant clauses. As observed in the previous chapters describing the results, when at least two participants are included in non-anaphoric third person plurals, they often include a personal pronoun. This suggests that the Spanish non-anaphoric third person plural is more concerned with providing information as for which participants are included in the clause. This may be linked with the especially high proportion of verbal processes since, as pointed out by Halliday & Matthiessen (2014), verbal processes may include three further participants in addition to the sayer: a receiver, a target and a verbiage. A second characteristic uncovered in this analysis is the particular affinity of the English non-anaphoric third plural with inanimate patients and, highly correlated to this result, an important proportion of non-individuated patients. The English non-anaphoric third person plural is indeed the structure that displays the highest proportion of inanimate patients classified as non-individuated. This means that when the patient is an animate entity English prefers to rely either on the periphrastic passive or on another syntactic structure to involve its animate patient.

The English non-anaphoric third person plural also differs from the French and the Spanish one in its relatively restricted number of elements in the linguistic context which could guide towards the interpretation of possible participants. Indeed, the English construction shows a relatively small number of prepositional phrases. Conversely, the Spanish non-anaphoric third person plural is the one to resort to the highest proportion of prepositional phrases, and more especially to a high proportion of locative complements, which, though they are not the prototypical agentive ones, can serve for agenthood management. In addition to a high use of prepositional phrases, Spanish, in particular, is also characterized by the presence of another passive-like structure in the context, which, as can be seen in the language-specific MCA graph, appears as a contributing factor in explaining the variation in Spanish data.

To recapitulate, in addition to differences in their overall frequency, especially marked for Spanish, the three non-anaphoric third person plurals display rather distinct behaviour, which leads us to conclude that considering them as perfect translational equivalents may result in inaccurate usages.

### 10.1.3. Se-constructions in French and in Spanish

This section aims to compare the construction based on *se*, available both in French and in Spanish. As reported in the literature and confirmed in Table 35, the frequency of this passive-like structure varies considerably according to the language. While *se* is considered as a major defocussing mechanism in Spanish (see Mendikoetxea 1999, Cabañas Maya 2006 or Laslop & Diaz 2010), its use is much more restricted in French.

| | **Spanish** | | **French** | |
|---|---|---|---|---|
| | raw | pttw | raw | pttw |
| ***Se*-constructions** | 968 | 46.36 | 73 | 3.58 |

*Table 35. Absolute and relative frequencies of* se-construction in *French and Spanish*

As developed more extensively in Chapter 2, Lamiroy (1993) highlights three points on which French and Spanish *se*-constructions differ: (i) the preference of the French *se* to appear with inanimate nominal clauses in the third person while with the Spanish *se* human nominal clauses would be accepted; (ii) the limited use of the French *se* to direct transitive verbs while the Spanish *se*-construction also allows indirect transitive and intransitive verbs; (iii) the greater presence of aspectual restrictions in French which would block the use of *se* with perfect tenses, a phenomenon not observed in Spanish.

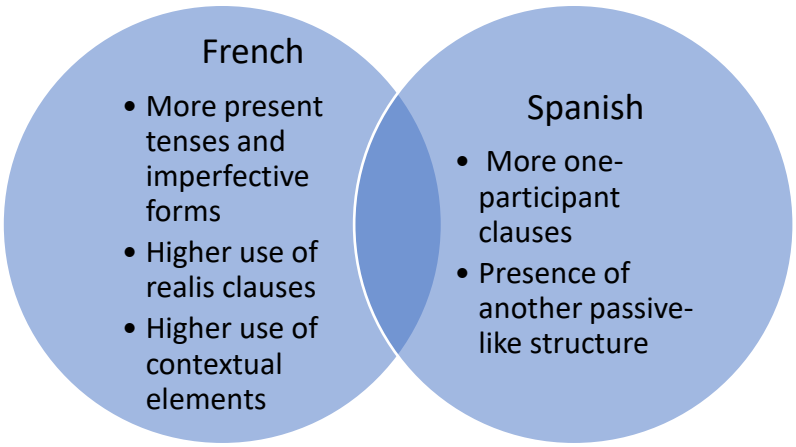



*Figure 28.Summary of the specific characteristics of the se-construction in French and Spanish*

Let us discuss these three aspects in light of the results yielded by this study, illustrated in Figure 28. The first point concerns the nature of the patients included in the *se*-constructions. Our findings reveal that, though the French *se* includes slightly more inanimate patients than its Spanish counterpart, the difference between the two languages is not significant. While, as pointed out by Lamiroy (1993) the Spanish *se* allows human nominal clauses, our results seem to indicate that both languages tend to almost exclusively rely on inanimate entities. The second point highlighted by Lamiroy (1993) is the large preference for the French *se* to appear with transitive verbs. This is largely confirmed by our results since they indicate that, while the

French *se*-construction almost exclusively appears in a two-participant clause, - i.e., including at least an object, its Spanish counterpart allows much more flexibility as for the number of participants. A one-participant clause would correspond to the use of *se* with intransitive verbs. As far as aspectual restrictions, our results concur again with the observations offered by Lamiroy (1993) since this study shows that French *se* occur almost exclusively with present tenses and imperfective forms. In comparison, Spanish includes much more past tenses and perfective forms.

In addition to the features pointed out by Lamiroy (1993), this study allows to bring additional information as for the differences between the French and the Spanish *se*-constructions. The first additional difference concerns the way the two structures encode events and situations in the world. Results indicate that in French, nearly all the occurrences of *se* are used in realis clauses, thus anchored in the real world, while in Spanish a higher proportion of irreal clauses is observed. Moreover, the French *se* tends to include much more spatio-temporal restrictions and prepositional phrases, elements in the linguistic context which help to identify possible participants.

#### 10.1.4. Se-constructions and periphrastic passives, complementary or overlapping structures?

This section contrasts the *se*-construction and the periphrastic passive in Spanish and, in a second step, in French, thus comparing the two languages that present these two strategies. Since the French *se*-construction displays a lower frequency than its Spanish counterpart, it is interesting to examine whether their respective relationship with the periphrastic passive is similar or not.

The literature often indicates that Spanish prefers the pronominal construction over the periphrastic passive (see Mendikoetxea 1999, Cabañas Maya 2006, Laslop & Diaz 2010). This higher use of the *se*-construction is confirmed by the relative frequencies observed in our study: 46.36 occurrences per ten thousand words for the *se*-construction and 15.56 for the periphrastic passive. As explained in Chapter 2, this preponderant use of the *se*-construction is usually explained by a preference for an active construction over a passive one. However, we assume that other factors may enter into play to explain this difference in frequency. To our knowledge, very few studies so far have adopted a corpus-based approach to document the syntactico-pragmatic behaviour of the Spanish periphrastic passive versus the *se*-construction, except for Fernández (2007), who offers a discursive-oriented analysis of these two constructions in the framework of her study on the Spanish passive voice. Our results make it possible to draw a more fine-grained picture of the distinct syntactico-pragmatic profile of the two Spanish passive-like constructions, whose main differences are visually represented in Figure 29.

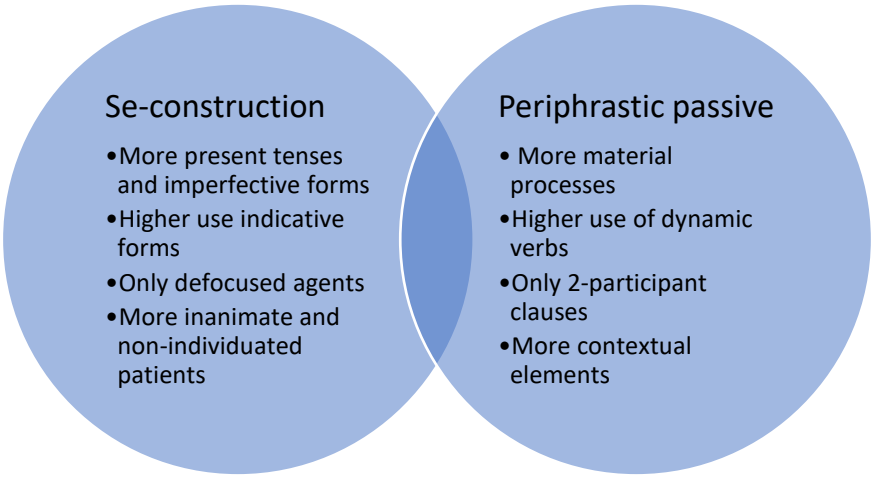



*Figure 29. Summary of the specific characteristics of the se-construction and the periphrastic passive in Spanish*

As for verb-related parameters, the periphrastic passive and the *se*-construction differ on five points. First, the analysis has revealed that the se-construction uses significantly more present tenses than the periphrastic passive. This important use of present tenses correlates with a large use of imperfective forms. This result confirms Espinoza (1997)'s claim that *aspect* is a main distinctive parameter between the two passive-like structures and a key feature in the choice of one structure over another. Our work provides further insights on this topic since, through a corpus-based study, it shows that imperfective forms are dominant in *se*-constructions (80.4% against 62.5% in periphrastic passives). In other words, when Spanish speakers need to express themselves on events or situations considered as not accomplished or still ongoing, they would rather opt for a *se*-construction than a periphrastic passive. In addition to this high use of present and imperfective forms, our results also indicate a significantly higher use of the indicative verb mode in *se*-constructions. The use of present tenses, imperfective forms and indicative verb modes appears thus as closely linked. Another finding regarding verb-parameters in this study is the particular association between the periphrastic passive and a considerably high use of dynamic verbs compared to the *se*-construction. This higher use of dynamic verbs in periphrastic passives can be linked to the marked presence of material types of verb processes, which constitutes the last difference between *se*-constructions and periphrastic passives. Though the differences are not significant, the *se*-construction tends to include slightly more relational and mental verbs than the periphrastic passive. This phenomenon can partly contribute to explaining the higher use of imperfective forms since mental and relational verbs are better candidates to express states or to describe ongoing or habitual situations. As we have seen until now, the two-passive-like structures are rather distinct constructions. However, they share two verb-related features: the presence of realis clauses and the use of affirmative constructions, two highly correlated parameters.

A sharp contrast between the two passive-like structures could be observed as for participant-related parameters. First, the periphrastic passive only occurs in two-participant clauses while the *se*-construction allows much more flexibility as to the number of participants included in the clause. Our results indicate that, though it is not the preferred pattern, the *se*-construction can be used intransitively while the periphrastic passive appears as more constrained since it does not display this syntactic pattern in our data. This study also points the difference in the way the two passive-like structures defocus their agent. While the *se*-construction only includes

cases of totally defocused agents, the periphrastic passive includes an explicit agent in nearly one occurrence out of four. When the agent is explicitly used in the periphrastic passive, results indicate that the distribution is fairly balanced between animate humans and inanimate *others*. The nature of the patient is another aspect that distinguishes the periphrastic passive from the *se*-construction. The latter contains significantly more inanimate patients, which largely correlates with a considerably high use of non-individuated patients. These results point towards an overall larger presence of animate entities in the periphrastic passive, be it as agent or as patient.

The two passive-like constructions also behave differently as for their use of contextual elements. While they both resort to spatio-temporal restrictions in similar proportions, the periphrastic passive resorts to significantly more prepositional phrases and other passive-like structures in its linguistic context. The results also indicate that, when a prepositional complement is used in a passive-like structure, the periphrastic passive includes much more complements which serve agenthood management than the *se*-construction. These results combined with the fact that the periphrastic passive contains an important proportion of overt agents confer this passive-like structure with a higher degree of explicitness concerning the participants included in the clause.

The periphrastic passive and the *se*-construction display, therefore a fairly different profile in Spanish. We can thus assume that the preference for active voice over the passive option is not the only reason explaining the difference of frequency observed between the two constructions. This study has uncovered an ample range of syntactico-pragmatic features which distinguish these two passive-like structures. Although they can, in certain specific contexts, be used as alternatives in Spanish, the *se*-construction and the periphrastic passive have their own specificities.

The *se*-construction and the periphrastic passive are two passive-like structures which are also available in French. Although the French *se* is far less productive than its Spanish counterpart, it is interesting to briefly examine whether a similar pattern to the one obtained for Spanish can be identified for French. Similarly to Spanish, the French *se*-construction contains much more present and imperfective forms than the French periphrastic passive. The higher presence of material processes in the Spanish periphrastic passive compared to the *se*-construction, is not found in French since both constructions display a fairly similar pattern as for the distribution of the type of verb process. The significantly higher use of dynamic verbs in the periphrastic passive is also observed in French. A difference between the two languages is the higher use of realis clause in the French *se*-construction compared to the periphrastic passive, a pattern which is not observed in Spanish. The marked contrast between the two Spanish structures as for the presence and the nature of the participants is also found in French. Finally, the considerable difference in the use of the contextual elements in the linguistic context between the *se*-construction and the periphrastic passive is also observed in French. The preference for an active option rather than a passive one is also frequently suggested to operate in French through the preponderant use of man-impersonal *on* (Lamiroy 1993). This issue is discussed in the next section.

### 10.1.5. Man-impersonal *on*, standing alone?

Multiple Correspondence Analyses revealed that the French man-impersonal *on* impacts considerably the distribution of the data. One central question was thus whether this passive-like structure stands apart on the continuum of passive-like structures or whether it shares some characteristics with other passive-like structures in French on the one hand but also with other constructions in English and in Spanish on the other. Based on the literature, three options are explored. Since Gaatone (1998) notes that *man*-impersonal and periphrastic passives may convey (partially) similar meaning and be used as alternatives, the first part of this section will be devoted to the comparison of these two passive-like structures in French. As pointed out by Siewierska (2011) man-impersonal *on* and non-anaphoric third person plural pertain to the category of R-impersonal (pronouns reduced in referentiality) which can lead to a large degree of overlap between such structures. The second type of comparison carried out in this section will thus involve the man-impersonal *on* and the non-anaphoric third person plural in French. Complementary information will be briefly provided on *on* versus the non-anaphoric third person in Spanish and in English. Finally, particular attention will be devoted to the relation between the French man-impersonal and the Spanish *se*-construction since literature often suggests the *se*-construction as a translational equivalent for the French *on*, depending on the morphosyntactic and pragmatic intentions of the speaker (Yao 2016).

Let us first briefly recapitulate the findings regarding the relative frequency of *on* and compare them in light of the three other three passive-like structures with whom the literature reports possible affinities.

| | | **English** | | **Spanish** | | **French** | |
|---|---|---|---|---|---|---|---|
| | | raw | pttw | raw | pttw | raw | pttw |
| **Structure** | 3PP | 186 | 8.48 | 498 | 23.85 | 109 | 5.34 |
| | ON | _ | _ | _ | _ | 995 | **48.83** |
| | PP | 1415 | 64.5 | 325 | 15.56 | 857 | **42.05** |
| | SE | _ | _ | 968 | **46.36** | 73 | 3.58 |

*Table 36. Absolute and relative frequencies for the non-anaphoric third person plural, the man-impersonal on, the periphrastic passive and the se-construction*

Table 36 indicates that *on* has a fairly high relative frequency (48.83 occurrences per ten thousand words), which ranks this construction as the second most common passive-like structure among those included in this investigation, after the English periphrastic passive. Interestingly, *on* presents a similar relative frequency as the periphrastic passive in French and the *se*-construction in Spanish, two of the constructions the literature suggested as possible alternatives. The other R-impersonal, the non-anaphoric third person plural, displays much lower frequencies.

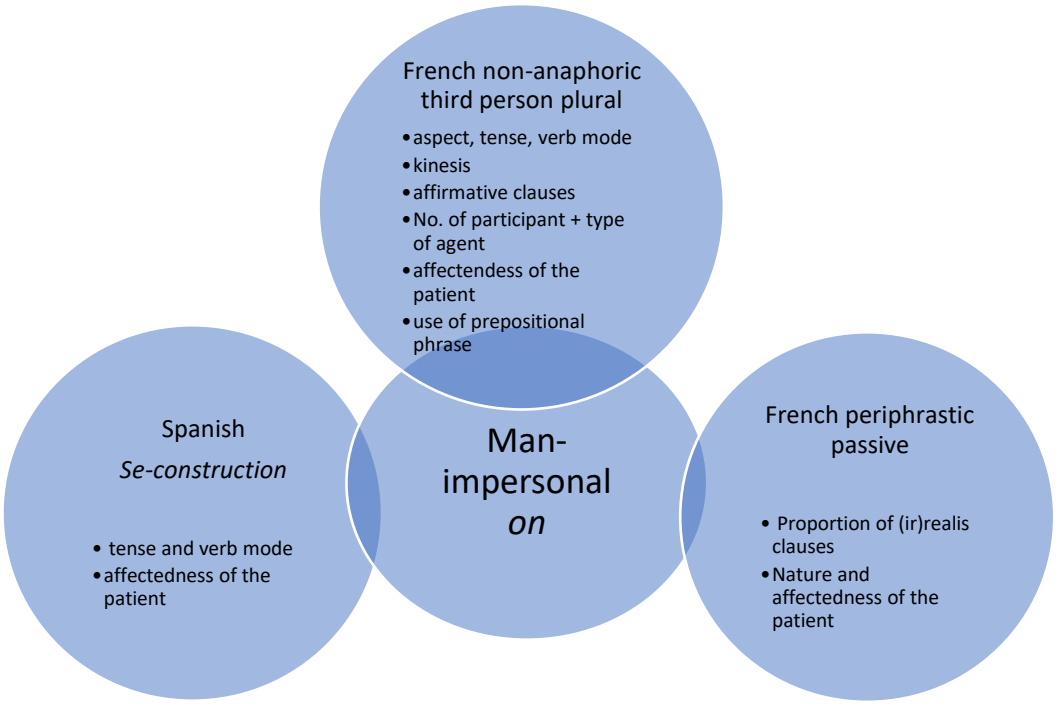



*Figure 30. Summary of the shared characteristics between on versus the periphrastic passive, the non-anaphoric third person plural and the se-construction*

Figure 30 summarises the common characteristics between the man-impersonal *on* and the other three structures. The comparison of the French periphrastic passive and the man-impersonal *on* reveals that these two passive-like structures considerably differ. The analysis of *tense* and *aspect*, two correlated parameters, reveals that *on* includes a higher proportion of present tenses and imperfective forms. The man-impersonal construction appears more frequently in the indicative mode than the periphrastic passive. It can thus be assumed that when speakers view a present event or situation with an internal perspective, they would rather opt for a man-impersonal *on.* Another aspect on which the two passive-like structures diverge is the type of verb and its dynamicity. Periphrastic passives use much more material types of processes and dynamic verbs than man-impersonals. *On* relies on larger proportions of mental and relational verbs. These two types of verbs are more likely to express state and denote relationship between a series of entities. This high presence of mental and relational verbs contributes to explain the smaller proportion of dynamic verbs used with the man-impersonal *on.* The number of participants included in the clause and their nature are additional distinctive features. While the periphrastic passive almost exclusively appears in two-participant clauses, with, in most cases, a defocused agent, the man-impersonal *on* is constructed with an animate human agent and only occurs in a two-participant clause in 50.2% of the cases. The nature of the patients involved in the passive-like structure also differs as the analysis revealed that man-impersonal *on* tends to use much more non-individuated entities than the periphrastic passive. A last difference between the two constructions concerns their use of contextual elements. Results indicated that the periphrastic passive includes a higher number of elements in the linguistic context that would help in the interpretation of the possible participants involved in the defocussing process.

This tendency is especially marked for the presence of a prepositional phrase which provides information through the use of agentive and locative complements. When the man-impersonal *on* uses a prepositional phrase serving agenthood management it is mainly a locative complement. Agentive prepositions, the typical device to express agentivity, is absent from this passive-like construction. In sum, when comparing the periphrastic passive with the man-impersonal *on,* we observe that they share very few syntactico-pragmatic features: the proportion of realis clauses and the nature and degree of affectedness of the patients. Though, as pointed out in the literature, they can be used as semantic alternatives to a certain extent, their overall preferred syntactic features diverge.

Whereas the man-impersonal *on* and the periphrastic passive greatly differ as to their representation of time and aspect, *on* appears as much closer to the non-anaphoric third person plural. As for features related to the verbal dimension, *on* and the non-anaphoric third person plural only diverge on two points: the types of verbs they include and the way they encode situations and events in the world. As for the type of verb, *on* includes much more mental verbs than the non-anaphoric third person plural while this latter relies on a larger proportion of relational verbs. The man-impersonal *on* appears thus as a better candidate to express cognitive processes and phenomena of perception and affection. Regarding the way the two passive-like structures encode the content in the real worlds, results indicate that man-impersonal *on* tends to resort to fewer clauses considered as anchored in the reality. A possible explanation could be the higher proportion of mental verbs. Since mental verbs involve the expression of the speaker's affection, attitude or perception, a less assertive approach as for the content of the clause may be expected. As for participant-related parameters, the two passive-like structures display a similar number of two-participant clauses. This result can be partly explained by the fact that they are both constructed with an overt animate human agent. They are, nevertheless, dissimilar as the nature of the patient is concerned. *On* contains much more inanimate and non-individuated patients than the non-anaphoric third person plural, which means that when animate entities are involved, the non-anaphoric third person plural tends to be preferred. A last difference between the two constructions is the higher presence of spatio-temporal restriction with the non-anaphoric third person plural.

Both man-impersonal *on* and the non-anaphoric third person plural are R-impersonals, namely constructions with a notional subject lacking referential properties (Malchukov & Siewierska 2011). This common property can explain their higher number of shared syntactic features if compared with the periphrastic passive, another prototypical agent-defocussing mechanism. This analysis has made it possible to uncover on which aspects these two R-impersonal diverge, namely: the type of verb process, the way they actualize events and situation and the nature of the patients. The context within which they appear also slightly differs as the non-anaphoric third person plural tends to provide more contextual information than the man-impersonal *on.* We suggest that the divergence in the use of contextual elements may partly be explained by the different usage patterns of two R-impersonals regarding the expression of referentiality. Viberg (2010) points out that, in Swedish, man-constructions frequently serve for the expression of genericity whereas the primary function of the non-anaphoric third person plural would be the expression of vagueness. It can be assumed that the expression of general references makes use of fewer contextual elements than the expression of vagueness.

Let us now compare the man-impersonal *on* to the Spanish *se*-construction, two structures sometimes suggested as translational equivalents. As for verb-related parameters, both structures present similar proportions of present tenses and indicative forms. However, they diverge on the percentage of imperfective forms, which is higher in the man-impersonal *on*. This may be due to the fact that the French passive-like structure includes more past tenses identified as *imparfait*, which are classified as imperfective forms. The man-impersonal *on* also contains a significantly higher proportion of stative verbs. This may be attributed to the fact that this passive-like structure shows a greater affinity with mental verbs than the *se*-construction, which includes much more material verbs. A last point related to verb-parameters which distinguishes both structures is their relation to realis/irrealis clauses. Clauses containing a man-impersonal *on* are less often anchored in the real world than constructions with *se.* The higher presence of negative clauses with man-impersonal *on* partly explains this difference. As for participant-related features, the two passive-like structures differ considerably. The man-impersonal *on* includes fewer patients than the *se*-constructions. Also, when *on* contains a patient, it is more often an inanimate and non-individuated one than those found in the *se*-construction. Contextual elements co-occurring with the passive-like structure is the last feature that allows to distinguish both constructions. The analysis has revealed that the Spanish construction resorts much more to those elements than the French *on.* The difference is especially marked regarding the use of prepositional phrases, considerably more present in *se*-constructions. It can be concluded that, though the literature suggests the two passive-like structures as translational equivalents, they can only be regarded as partially equivalent since our analysis reveals a series of differences between the two constructions. The main differences concern the following parameters: the aspectual dimension, the stative versus dynamic nature of the verbal argument, the dichotomy *realis/irrealis,* the presence, the nature and the degree of individuation of a *patient,* and, lastly, the use of contextual elements

In sum, we can suggest that the behaviour of the man-impersonal *on* comes closer to the one of the other R-impersonal as far as syntactico-pragmatic features are concerned. The comparison with the French periphrastic passive and the Spanish *se*-construction, often commented as possible alternatives to the man-impersonal *on*, indicates that regarding their core features, these structures differ considerably. However, the interpretation of which interaction participants (speaker, hearer or a third entity) are included in these defocussing mechanisms goes beyond the scope of this study. Siewierska & Malchukov (2011)'s crosslinguistic work or Rasson 2017's study on Spanish would constitute good bases to examine the various Spanish forms included from this perspective. On the Swedish man-imperonal Viberg (2010:140) indicates that, when translated in English, syntactic shifts or clause reductions or restructurations may be preferred to another impersonal pronoun such as *they.* The choice of a non-anaphoric third person plural to translate a French man impersonal *on* into English is also noted by Salkoff (1999). Interestingly, Viberg notes that when the Swedish *man* is translated into French, the French man-impersonal *on* covers only 57% of the translations (2010:141). This highlights again that there is no one-to-one mapping between two or more languages, each of them possessing its own system and specificities.

## 10.2. Mode and register, key factors in contrastive studies

The oral versus written nature of the data as well as the degree of formality of the production are elements that have long been neglected in contrastive studies. And, as pointed out, by Dieck (2016) for Spanish, information on the distribution of passive-like structures according to these two elements is often lacking. By including corpus-based data from informal and formal contexts as well as from oral and written situations, this study aims to bridge this gap and highlighting the importance of including these variables when carrying out contrastive studies in linguistics. The three language-specific analytical chapters have already shown the importance of these two factors for the distribution of the data. The present section suggests an in-depth look at the influence of mode and register from two perspectives. The first part focuses on some groups of structures that have been presented in the literature as prone to mode and register variation (10.2.1): (i) the use of the non-anaphoric third person and the periphrastic passive as alternatives in oral informal data (Weiner & Labov 1983); (ii) the relation between the periphrastic passive and the *get*-passive; (iii) the comparison between *uno, one* and *on,* often suggested as equivalents. The second part intends to determine which parameters are most affected by the mode and the register (10.2.2).

### 10.2.1. Structures and variation

#### *10.2.1.1. The non-anaphoric third person plural, an informal periphrastic passive?*

It seems widely acknowledged that the periphrastic passive serves to fulfil a series of requirements of formal discourse (Jespersen 1924, Weiner & Labov 1983, Siewierska 1984 Yamamoto 1984). Its dominant use in formal registers, compared to informal ones, is, therefore, of no surprise. However, as pointed out by Weiner and Labov (1983) passives are also quite common in colloquial spoken English. In informal oral situations, the authors highlight the fact that periphrastic passives and non-anaphoric third person plural can be used as possible alternatives. They add that usually, discussions on the English passive do not suggest any syntactic options when passives appear without any explicit agent. Yet, Weiner and Labov consider non-anaphoric third person plurals as a major active alternant of agentless passives in spontaneous speech (1983:36). The chief focus of this section is to determine to which extent the non-anaphoric third person plural can be considered as an oral informal alternative to the periphrastic passive. First, the situation in English will be discussed, since it is the language for which the literature has claimed that the non-anaphoric third person plural can be an informal periphrastic passive. Second, a comparison will be made with French and Spanish.

| | Non-anaphoric third person plural | | Periphrastic passive | |
|---|---|---|---|---|
| | raw | pttw | raw | pttw |
| **Informal written** | 65 | 13.54 | 285 | 59.38 |
| **Formal written** | 8 | 1.73 | 596 | 128.65 |
| **Informal oral** | 113 | 13.3 | 160 | 18.88 |
| **Formal oral** | 0 | 0 | 374 | 92.79 |

*Table 37. Distribution of non-anaphoric third person plurals and periphrastic passives in English*

Table 37 offers an overview of the distribution of non-anaphoric third person plurals and periphrastic passives in English according to the mode and the register. Our corpus-based data corroborate the idea that though the periphrastic passive is usually characteristic of formal and, more especially of formal written situations, the structure is not absent from informal and spoken interactions. Conversely, the non-anaphoric third person shows a marked tendency to appear in informal situations, be it written or oral since both datasets include a similar relative frequency. It is noticeable that in the informal oral dataset the non-anaphoric third person plural and the periphrastic passive obtain a fairly similar frequency (respectively 13.3 and 18.9 occurrences per ten thousand words). Let us examine these two sub-sets of data in more details in order to pinpoint potential similarities as for their syntactico-pragmatic behaviour in informal oral data.

Overall, it can be said that the non-anaphoric third person plural and the periphrastic passive share a larger set of features concerning verb-related parameters, in informal oral data than what was observed in their general behaviour, that is, regardless of the mode of production and the degree of formality. A closer analysis of the behaviour of the non-anaphoric third person and the periphrastic passive in informal oral data reveals a series of interesting findings which point toward a fairly similar profile for these two passive-like structures.

Our results indicate that, in informal data, the non-anaphoric third person plural and the periphrastic passive do not vary as for the parameter *tense* and *aspect* in informal oral data[44]. This was not the case in the general analysis, which does not consider register and mode variation. Similarly, the parameter *kinesis* does not display any differences between the periphrastic passive and the non-anaphoric third person plural when they are used in informal oral data. The analysis reveals, nevertheless, that both passive-like structures do vary in informal written. The stable proportion of dynamic verbs across the periphrastic passive and the non-anaphoric third person plural can thus be considered a specific characteristic of informal oral data. The way the two passive-like constructions encodes events and situations in the reality is another parameter that does not vary. This absence of substantial variation was also found in the informal written set of data[45], thus it is rather assumed that this parameter has a stable behaviour across informal data in general. As for the type of verb process, while the general distribution of verbs in the periphrastic passive and the non-anaphoric third person

[44] Chi-square tests indicate no significant difference for *tense* $\chi^2 = 0.8489$. $p > .05$) nor for *aspect* ($\chi^2 = 2.0295$. $p > .05$).

[45] No significant difference could be found for realis clause in informal oral data ($\chi^2 = 1.3791$. $p > .05$) nor for informal written data ($\chi^2 = 2.9828$. $p > .05$).

plural displays a fairly different profile, a closer look at informal data shows that they are relatively similar as for the type of verb included in informal data. The only difference is the strong presence of modulation verbs in the non-anaphoric third person plural, a phenomenon which is not observed in the periphrastic passive. This resemblance between the two passive-like structures as for the type of verb process is not observed in informal written data.

Although the periphrastic passive and the non-anaphoric third person plural share a wider range of verb-related features in informal oral data than in the other datasets, they do display two main differences. The first one concerns the verb mode. The non-anaphoric third person plural contains a significantly higher proportion of indicative forms compared to the periphrastic passive in informal oral data ($\chi^2 = 10.4368$. $p < .05$). Such a difference is not found in informal written data where both structures use the indicative verb mode in similar proportions[46]. The parameter *verb mode* can therefore be considered a distinctive feature between the non-anaphoric third person plural and the periphrastic passive as for their behaviour in informal spoken context. The second parameter which differs from one structure to the other is the use of affirmative clauses, largely dominant in periphrastic passives. This result may be explained by the fact that, regardless of the mode and the register, the periphrastic passive resorts to much more affirmative clauses (93.4%) than the non-anaphoric third person plural, which allows less space for variation.

With no surprise, their behaviour as for participant-related parameters is concerned varies considerably more since the non-anaphoric third person plural always implies an overt human agent, though reduced in referentiality, whereas the periphrastic passive allows much more flexibility regarding the presence and the nature of the agent as extensively commented in Chapter 9. This difference in the type and presence of the agent in these two constructions corroborates Viberg's findings (2010) which indicate that non-anaphoric third person plural impersonals are more referentially restricted than agentless passives (2010:139). Patients constitute thus the only element which can offer a relevant basis for the type of comparison carried out in this section. Though in different proportions, our data show that the informal oral dataset is the one that contains the largest quantity of animate human patients in the two passive-like structures. This result constitutes an additional similarity between the non-anaphoric third person plural and the periphrastic passive. As for the degree of individuation and affectedness of the patients, the analysis indicate a significantly higher use of individuated in the periphrastic passive ($\chi^2 = 6.1108$ $p < .05$) and a stable use of affected patients across the two structures. However, the phenomenon was already observed in the general analysis of the two structures. The strong presence of similarities uncovered for verb-related parameters is slightly less marked as for participant-related parameters.

Let us now comment context-related parameters, which, so far in contrastive studies have not been often considered when studying syntactic phenomena. Results indicate that, as far as the use of spatio-temporal restrictions is concerned, the behaviour of the non-anaphoric third person and the periphrastic passive in informal oral data diverges. The non-anaphoric third person plural tends to use significantly more restrictions than the periphrastic passive ($\chi^2 = 3.9281$. $p <$

[46] The chi-square test does not reveal any significant difference between the two sets of data ($\chi^2 = 0.637$. $p > .05$).

.05). This phenomenon can be counterbalanced by the fact that the periphrastic passive relies considerably more on prepositional phrases than the non-anaphoric third person plural. It can be hypothesised that both structures do yield further information in the linguistic context as for the potential participant included in the structures but their manner to realize it diverges.

In sum, our study has revealed that as for verb-related parameters the periphrastic passive and the non-anaphoric third person plural show a greater degree of similarity in informal oral data than what was observed in the general analysis of the English passive-like structure in Chapter 9. Overall, the participant-related parameters selected in this study display slightly more variation, mainly due to the nature and the presence of the agent. An additional element uncovered by this study is the fact that both passive-like structures include a large quantity of contextual elements but they differ on the strategy they use since the periphrastic passive rather relies on prepositional phrases whereas the non-anaphoric third person plural on spatio-temporal restrictions. It seems safe to say that, despite the differences about the participants, our results largely corroborate what is reported in the literature on the possible use of the non-anaphoric third person plural and the periphrastic passive as alternatives in informal oral data. This analysis has then shown a hitherto understudied phenomenon, namely that both passive-likes structures frequently use contextual elements in informal oral data though relying on different mechanisms. All these findings reinforce the importance to consider the mode and the register in the analysis of linguistic phenomena such as passive-like constructions.

The possibility of using the non-anaphoric third person plural as an alternative of the periphrastic passive in informal oral data appears as less straightforward in Spanish and in French. Our data reveal that, while the Spanish non-anaphoric third person plural obtains a relative frequency of 46.64 occurrences per ten thousand words in informal oral data, the periphrastic passive only displays a frequency of 0.78 occurrences per ten thousand words. This limited presence of periphrastic passives in informal data impedes a comparative analysis as the one carried out on English data. In French, the non-anaphoric third person plural is not a common passive-like mechanism and the periphrastic passive is relatively absent from informal oral data compared to the other combinations of modes and registers. The especially high use of the man-impersonal *on* in informal interactions in French is a contributing factor in the more restricted presence of these two passive-like structures.

#### *10.2.1.2. Periphrastic passives and get-passives*

Interesting findings regarding the influence of the mode and the register can also be obtained when examining the periphrastic passive and the *get*-passive, two structures often presented as possible alternatives in English. As pointed out in the literature, the *get*-passive is highly register-sensitive, appearing mainly in informal oral productions (Huddleston 1971, Leech & Svartvik 1975, Rühlemann 2007, Anderwald 2018). The comparison of the *get*-passive and the periphrastic passive in informal oral data reveals that the two passive-like structures share a higher number of features than when they are considered regardless of the mode and the register (see Chapter 9). The passive-like structures tend to have a more similar use of verb-related features. Nevertheless, they still differ regarding the presence and the nature of the agent. This

may be attributed to the fact that *get*-passives are considered semantically marked since the literature reports that they mainly occur in contexts or situations considered as problematic or adversative by the speaker (Carter &McCarthy 1999). Even though the periphrastic passive can occur in informal oral productions, this structure does not convey such a stance, which would partly explain why the two passive-like structures cannot be considered as interchangeable structures, even when used in informal oral situations.

#### *10.2.1.3. On, uno* and *one*

*On* is often considered as an equivalent to the numeral-based *one* in English and *uno* in Spanish. Even though in some contexts *on* and the numeral-based constructions can have a similar reading, they cannot be employed as alternatives without being aware of their frequency of use and their sensitivity to register and mode. As shown in this study, the man-impersonal is the most frequent passive-like structure in French (48.83 occurrences pttw). Conversely, numeral-based structures are scarce: 0.24 occurrences pttw in Spanish and 0.41 occurrences pttw in English. Malamud (2012) highlights that, in English, when speakers have to select an agent-defocussing strategy including a human and definite instigator, they tend to prefer the non-anaphoric third person plural over *one*, which can partly explain the low frequency of *one*.

In addition to this considerable difference in their frequency, the two types of constructions are associated with distinct registers and modes. *On* appears mainly in informal oral data (82.9 occurrences pttw) and, to a smaller extent, in informal written data (47.13 occurrences pttw). Our data reveal that all occurrences of the numeral-based constructions are found either in formal and informal written data. The absence of *uno* and *one* in oral interactions and their use in formal situations constitute two major differences compared to the French man-impersonal *on*. An additional distinction between the two types of structures regards their semantic role. While, in our data, *on* is almost exclusively used as instigator, data reveal that *uno* and *one* are much more prone to fulfil other semantic functions. This brief comparison of the two types of mechanisms highlights, again, the importance of raising the awareness of the register and the mode sensitivity of grammatical phenomena.

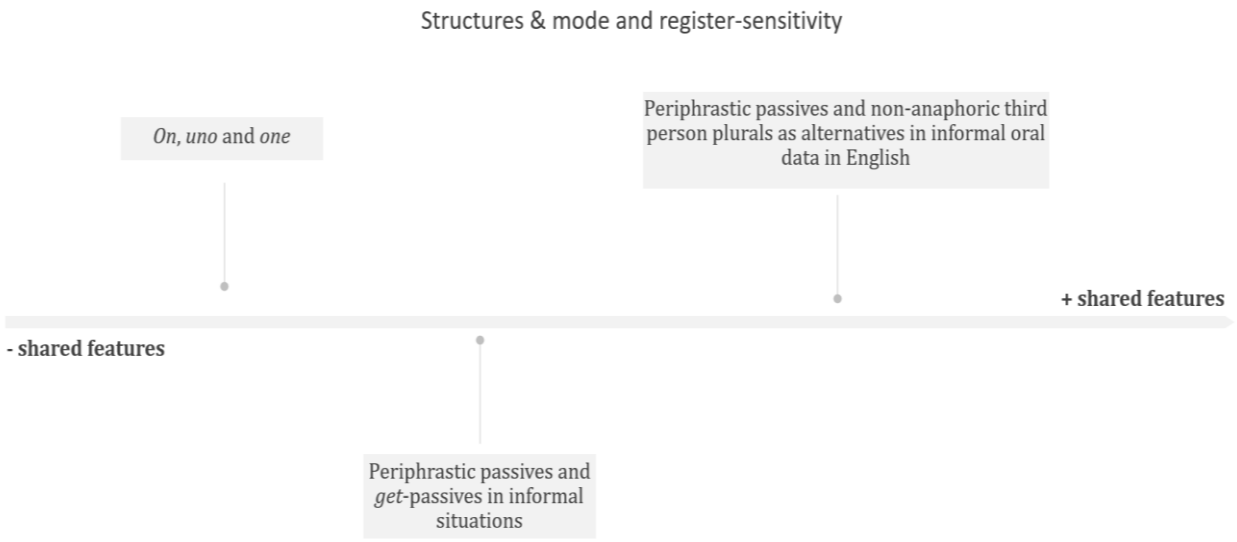



*Figure 31. Influence of mode and register on three sets of passive-like structures*

Figure 31 summarises the main findings regarding the mode and register-sensitivity of the passive-like structures. The focus of the analysis was placed on pairs or groups of structures usually described in the literature as depending on the register. The study revealed that when periphrastic passives and non-anaphoric third person plurals are used in English informal oral data, they share a higher number of features. This was not observed for the other two languages. A similar phenomenon, though less marked, occurs for the periphrastic passive and the *get*-passive. Though *on*, *uno* and *one* are sometimes presented as equivalents, the analysis revealed that, in addition to their different frequency of use, they tend to occur in highly different contexts and can rarely be considered as equivalents. These findings stress the importance of taking into account the degree of formality and the oral versus written nature of the data when carrying out linguistic studies.

### 10.2.2. Parameters and variation

Svartvik (1966) postulates that the degree of colloquialism of the situation is the key factor impacting the frequency of use of the passive in English rather than the mode, be it written or oral. In this section, we intend to evaluate to which extent this assumption can apply to our data and expand the scope of the analysis by intending to determine which syntactico-pragmatic parameters are the most impacted by mode and register variation. Due to the restricted number of occurrences obtained for the French *se*-construction and the French non-anaphoric third person plural, these two constructions are not included in the present discussion. The distinct areas where mode and register exert considerable influence on the data have been determined based on significant differences yielded by chi-square tests.

Let us first start with verb-related parameters. The analysis has revealed that *tense* undergoes a fairly important influence from mode and register, though to a different extent according to the language and the structure. In English, the distribution of the tenses in the periphrastic passive is both influenced by the register and the mode, with a more common use of present tenses in

informal written data. The picture appears as slightly different for Spanish since the periphrastic passive seems to be more impacted by the degree of formality than the mode of production. However, the distribution of data in Spanish and the English periphrastic passive converges on the fact that, overall, more present tenses are found in informal data. The Spanish *se*-construction undergoes a similar influence of the mode and the register as the Spanish periphrastic passive. The analysis has revealed that the French periphrastic passive displays a fairly stable behaviour across the four combinations of mode and register, thus identifying this French structure as the least impacted one as far as the distribution of the parameter *tense* is concerned. The French man-impersonal *on*, unlike the French periphrastic passive, is much more influenced by mode and register, with a higher frequency of past tenses in oral data, especially in informal oral data, than in written data. As far as the non-anaphoric third person plural is concerned, since this construction is rather specific to informal oral contexts and, in the case of French, slightly less frequent than other passive-like constructions, its sensitivity to mode and register is more arduous to detect and no clear pattern emerged as for the parameter *tense*.

The picture obtained for *aspect*, a parameter the analysis revealed as largely correlated to *tense* is substantially similar. In English, the periphrastic passive shows considerable influence of the register and the mode, an influence that is also observed in the non-anaphoric third person and which appears more clearly than was the case for the parameter *tense.* In French, similarly to what was found for *tense*, the behaviour of *aspect* is considered as stable in the periphrastic passive, with no influence of the mode or the register. While *aspect* does not vary greatly across the French periphrastic passive, the picture captured for the construction with *on* is relatively different. In the man-impersonal, *aspect* is particularly affected by the mode since our results indicate that oral data are better candidates for perfective forms compared to written data, which mainly include imperfective forms. The influence of the degree of formality can also be observed since a difference in the proportion of perfective forms can be observed between informal and formal oral data. Such influence was, nevertheless, not noted in written data. In Spanish, our results indicate an influence of both the mode and register within each structure, informal written data appearing as better candidates for imperfective readings in *se*-constructions and periphrastic passives while in non-anaphoric third person plurals, the informal oral dataset is the one including the highest percentage of imperfective clauses. The mode of production and the degree of formality are thus key variables to include in the analysis of *aspect.*

Interestingly, as for the impact of mode and register on the parameter *verb mode* is concerned, the three languages undergo a relatively similar influence, which was not the case for *tense* and *aspect.* This means that each of the three languages possesses a different way to organize time and aspectual features in the passive-like constructions while the *verb mode* appears as a much more stable feature across languages. In the parameter *verb mode,* the distribution of the indicative, subjunctive, modal auxiliaries and items classified as *others* (such as infinitives and gerunds) varies according to the register, the mode and the passive-like structure. In English periphrastic passives, *verb mode* is largely influenced by the mode of production of the data. In addition, an influence of the degree of formality, though only in oral data, could also be noted. The non-anaphoric third person plural, in contrary, does not display any differences. In French,

a similar influence is observed for the periphrastic passive and the man-impersonal *on*, with a marked difference between formal oral and informal data. The distribution of the four categories of *verb* modes is again greatly impacted by both the degree of colloquialism and the written versus spoken nature of the production in the periphrastic passive and the *se*-construction. Similarly to what was observed for English, while the periphrastic passive is largely affected by the mode and the register, the Spanish non-anaphoric third person plural is not impacted by the two external factors. The Spanish *se*-construction displays a rather similar behaviour to the one observed for the periphrastic passive.

The type of verb process is another parameter that is considerably influenced by the degree of formality and the mode of production of the data. In English periphrastic passives both factors play a key role in the distribution of the types of processes. The effect of the degree of formality is especially marked in oral data with, for instance, a marked presence of modulation verbs in formal oral data. A similar influence is noted for the French periphrastic passive though the impact of the register in oral data appear as slightly less important than what was observed in English, giving more prominence to the influence of the mode. In Spanish, the distribution of the types of processes is again affected by both the register and the mode, with a high proportion of material processes in written data and a more varied profile in oral data. A similar behaviour can be noted for the Spanish *se*-construction. In man-impersonals *on*, formal oral data especially stand out since their profile differs considerably compared to the other three mode and register combinations as for the types of verb processes included. The analysis also revealed that in French, the periphrastic passive and the man impersonal contain much more verbal processes in formal oral data than in the other three combinations of register and mode. In Spanish this trend is also noted for the *se*-construction but not for the periphrastic passive. This may be due to the fact that the periphrastic passive is a far less productive construction in Spanish. Lastly, regarding English, a larger set of data would be welcomed to confirm this observation but results also point toward a higher use of verbal processes in formal oral data. Once again, this highlights the need to consider factors as register and mode when carrying out contrastive linguistic analyses.

Compared to the parameters commented so far, a series of verb-related features undergo a much more limited effect of the mode and the register. The influence of the mode and the register on the two parameters related to the way the passive-like structures actualize events and situations, namely the dichotomy realis/irrealis and the use of affirmative clauses, appears as much more restricted than what emerged for *tenses, aspect, verb mode* etc. The impact of these two factors is not totally absent though, but is rather punctual, and when the case, only affects some sets of data or one passive-like structure. Similarly, the parameter *kinesis* is fairly stable across our data as it only affects some of the combinations mode/register or one specific construction in one language.

In a similar vein, participant-related parameters are also greatly impacted by the mode and the register within which the passive-like construction occurs. I will first start by commenting the impact of the mode and the register on the number of participants included in the clause, then on the use and degree of animacy of the agents and the patients and, finally, on the degree of individuation and affectedness of the patients.

In the three languages, results indicate that the use of two-participant clauses is highly associated with written data while one-participant clauses are more frequent in oral productions and informal situations. We put forward that the higher degree of syntactic elaboration, which is usually reported to characterize written data, could explain this finding.

In English, French, and Spanish, explicit agents, analysed in the light of the periphrastic passive, are specific to written data, and more precisely formal written data. The nature of patients and its degree of individuation are also largely impacted by the mode and the register in the three languages, especially in the case of the periphrastic passive. In French, findings show that the periphrastic passive, which includes more frequently a patient than the man-impersonal *on*, uses considerably more animate human patients in informal oral data. Conversely, the man-impersonal *on* contains more animate patients (human and non human) in the written datasets. In Spanish, animate patients seem to be rather associated with written data both in the periphrastic passive and the *se*-construction, thus resembling the profile of the French man-impersonal *on* rather than the French and English periphrastic passives.

As for the degree of individuation and affectedness of patients, our analysis has revealed that in Spanish and English, no significant influence of the register could be observed in the written data set. Conversely, the register appears as a key factor in the oral situations, with a formal oral data set that resembles more written than informal oral data. Unlike what has been found for English and Spanish, patients in French passive-like structures appear to be impacted both by the mode and the register. Finally, our findings indicate that the degree of affectedness of the patients does not vary.

In sum, it can be suggested that participant-related parameters are highly influenced by the mode and the register, with a fairly similar pattern of influences in the three languages, except for the parameter *individuation* in French.

The picture captured for context-related parameters, the last group of features analysed in this work, is rather contrasted. While in Spanish, the influence of the mode and the register is slightly less strong then observed for verb-related and participant-related parameters, they do impact the distribution of the features in English and French. In English, our findings reveal that spatio-temporal restrictions are far more used in formal written data, with an overall influence of the register, also observed in French. The limited presence of spatio-temporal restrictions in oral data may be explained by the constraints of oral communication. As pointed out by Kunz et al. (2017), oral communication usually implies reduced information density due to constraints on working memory capacity. The use of spatio-temporal devices would be considered as an additional superfluous load for the memory. Furthemore, in oral situations, speakers tend to share a higher common knowledge background where clarifications on location and time information may be unnecessary. Lastly, this study revealed that the presence of a prepositional phrase is rather governed by the mode in French, while in English the degree of colloquialism especially affects the distribution of the parameter in written data.

Figure 32 summarizes the parameters most prone to variation according to mode and register.

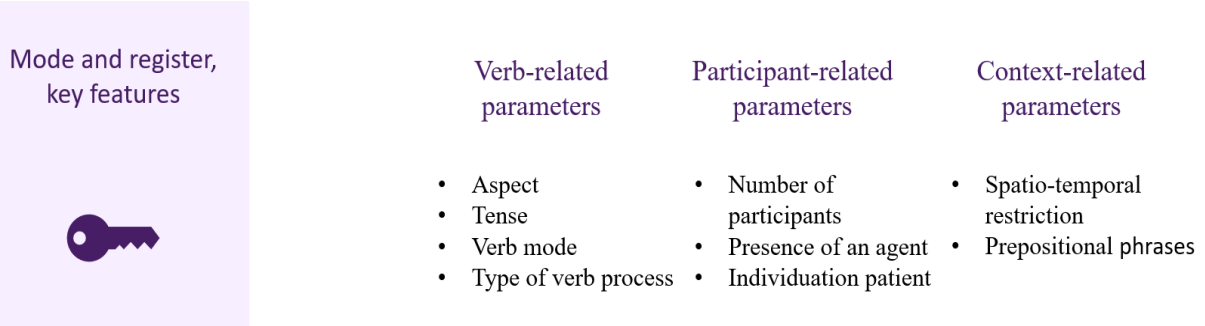



*Figure 32. Parameters most prone to variation according to mode and register*

In summary, our analysis has revealed that, for a series of parameters, oral data vary much more according to the degree of formality than written data (see for instance the parameter *aspect* or the type of verb process). This result can be explained by an overall higher degree of standardization of written productions. Formal oral data display a rather specific behaviour since, as discussed previously, they share characteristics of oral and of formal and informal written datasets. Moreover, it has been observed that, in a series of particular cases, formal oral data possess a distinct profile, not resembling any of the other three mode/register combinations. In addition to the influence of the register, this study has also highlighted the importance of the spoken versus written nature of the data. Overall, it suggests that both the register and the mode are essential factors to consider when studying linguistic phenomena, with a different influence depending on the passive-like structure, the language and the parameter under study. It has been observed that, for instance, the parameters in the non-anaphoric third person undergo slightly less influence of the register and the mode, which can be explained by the fact that the non-anaphoric construction is a highly register-sensitive structure *per se*. Conversely, the periphrastic passive is more prone to register and mode variation since its four subsets often appear as relatively different. Similarly, some parameters such as *tense* or *aspect* vary much more than others. In conclusion, as demonstrated in this section, a mere black or white approach on the matter would fail to identify some subtleties in the variation of the linguistic phenomena compared. These findings point towards the importance of having a prior reflection about this issue when carrying out contrastive corpus-based studies and of considering mode and register variations when compiling the data.

# 11. Conclusion

In embarking on this investigation I set out to explore six passive-like structures in English, French and Spanish: the periphrastic passive, the *get*-passive, the *se*-construction, the non-anaphoric third person plural, the man-impersonal *on*, and the numeral-based *uno* and *one*. Since the *get*-passive, *uno* and *one* proved to be fairly infrequent phenomena, the present work placed greater focus on the other four types of passive-like structures. These constructions were analysed along three different axes: (i) from the perspective of their main distinctive syntactico-pragmatic features; (ii) from a language-specific focus; and (iii) in light of *mode* and *register* variations. More precisely, I strived to answer the following questions:

- What are the main distinctive grammatical features of the different passive-like structures under study? Do these structures appear in different linguistic environments?
- How do language specificities impact the selection of the passive-like constructions?
- To which extent do mode and register affect the selection of a passive-like construction?

To answer the first questions, I developed an annotation protocol based on Hopper & Thompson's (1980) transitivity parameters. The present investigation supplements their work by considering Malchukov's (2006) proposal, which recasts Hopper & Thompson's transitivity scale and puts greater emphasis on the relationship between the different parameters. In addition to these parameters, the notions of *tense, verb mode,* and the *type of verb process* complete our analysis. Though *tense* and *verb mode* are usually not considered to be key parameters in studies on transitivity, these parameters are somehow connected to the parameter *aspect* and we believe that their examination refines the analysis of the aspectual-temporal dimension of passive-like structures. Moreover, and as confirmed in our study, *tense* and *aspect* are two parameters presented as highly correlated. The exploration of the *type of verb process* allows to add an additional layer to our investigation by obtaining lexical information on verbs involved in passive-like structures, a dimension which does not appear in Hopper & Thompson's approach (1980). As largely highlighted in the literature, when studying passive-like structures, it is crucial to consider contextual factors (e.g. Fløttum et al. 2007, Gjesdal 2008, Landragin & Tanguy 2014, Gast 2015, Yao 2016, De Cock 2020). Therefore, in addition to parameters related to transitivity, a series of contextual elements were included in this study.
The combination of parameters and contextual factors allows to capture the best possible picture of the distinctive features characterizing the passive-like structures under study.

### Answering the research questions

As clear from the analysis performed in the present work, pinpointing the main features of the passive-like structures is a task that cannot be accomplished without considering the language, the mode and the register. These factors appear as much more intertwined than expected. When

drawing a prototypical portrait of a structure such as the periphrastic passive, leaving aside the language specificities of English, French and Spanish would result in an erroneous picture of the phenomenon. Indeed, even though the surface structure of the three passive-like constructions is similar, our analysis reveals that, in addition to a different overall relative frequency, they also differ as for their syntactico-pragmatic features. The first two research questions, addressing the issue of the distinctive characteristics of the passive-like structures as well as the importance of the language are therefore answered simultaneously by summarizing the behaviour of the passive-like structures in each language in a first phase. Then, in a second phase, the focus is placed on the comparison of the structures available in two or three languages and on the connection between structures available in only one language such as the man-impersonal *on* and the passive-like structures in other languages. The third question, which addresses the influence of the mode and the register, is answered through both a focus on the parameters and a confrontation of our results with what is usually reported in the literature.

Starting with the passive-like structures available in French, namely the periphrastic passive, the *se*-construction, the non-anaphoric third person plural and the man-impersonal *on*, the results tend to show that, when expressing impersonal content, the same category of the parameter is usually dominant in all the passive-like structures under study. The verb is mainly dynamic, uses a present tense which largely correlates with the imperfective aspect and the indicative verb mode. The four passive-like structures show a preference for encoding real events and situations and largely rely on affirmative clauses. Nevertheless, the proportion of the dominant value varies from one structure to another. While the *se*-construction is almost exclusively formed with present tenses, the periphrastic passive displays a much wider flexibility as far as tenses are concerned. The two R-impersonals occupy an in-between position, with a percentage of present tenses of respectively 77.1% for the non-anaphoric third person plural and 72.9% for the man-impersonal *on.* In French, the parameter *tense* positively correlates with the *aspect*, where a high presence of present forms is associated with a high use of imperfective forms.

The analysis of the type of verb process, which provides this study with an additional lexical dimension, revealed that the four passive-like structures rely on the same three most frequent categories: *material*, *relational* and *mental* verbs. Nevertheless, the proportion of these categories greatly varies. While the periphrastic passive and the *se*-construction display a fairly similar profile with material verbs representing more than half of the data, this proportion of material verbs falls in man-impersonal *on* and the non-anaphoric third person plural. The man-impersonal *on* is characterized by a larger use of *mental* and *modulation* verbs whereas the distinctive feature of the non-anaphoric third person plural is the high frequency of relational verbs. A further characteristic this work uncovered is the specific link between realis clauses and the *se*-construction. This type of construction relies on realis clauses significantly more than the other three passive-like structures. It was also observed that R-impersonal constructions have a fairly similar profile when it comes to verb-related parameters, except for the parameter *realis/irrealis* dichotomy.

As for the participant-related parameters, the analysis revealed that two groups can be formed. The periphrastic passive and the *se*-construction can be grouped together as they share the following properties: (i) almost exclusively two-participant clauses; (ii) mainly appear without

any overt agent, when expressed in the periphrastic passive, the agent is mostly classified as inanimate *others* ; (iii) high presence of patients, mainly classified as inanimate *others*. This tendency is especially marked in the *se*-construction. The man-impersonal *on* and the non-anaphoric third plural form the second group of passive-like structures. Compared to the periphrastic passive and the *se*-constructions, they include a larger number of one-participant clauses. In both the man-impersonal *on* and the non-anaphoric third person plural, the agent is an animate human entity. This represents another major distinction with the other group of passive-like structures. As for patients, a considerably high proportion of animate patients was found in the non-anaphoric third person plural, setting it apart from the behaviour of the man-impersonal *on*. The degree of individuation of the patients is a parameter that, overall, varies considerably across the structures and seems to be associated with the animate versus inanimate nature of the patients. A possible explanation of the division of the four structures into two groups can be found in the specific syntactic architecture of the periphrastic passive and the *se*-construction, which places the patient in focus position.

This division of the four passive-like structures in two groups does not seem to apply in the case of context-related parameters. The behaviour of the *se*-construction clearly stands out, with an important presence of spatio-temporal restrictions and a high use of prepositional phrases with agentive force, more especially locative devices. We show that the higher presence of contextual elements with the *se*-construction may be due to the fact that, in *se*-constructions, the agent is almost always defocused. The presence of prepositional phrases and spatio-temporal builders can be considered as strategies to restrict the range of possible agents involved in this passive-like structure.

These findings made it possible to highlight the distinctive features for each French passive-like structure. Figure 33 visualizes these features which are further illustrated in examples 368 to 371.

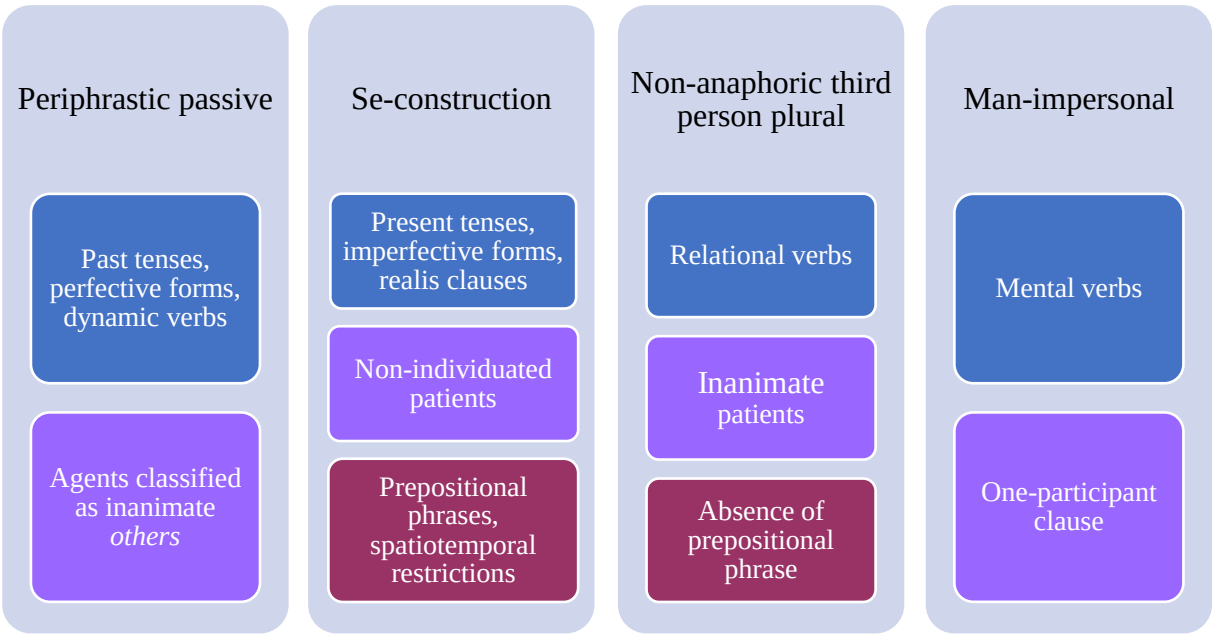



*Figure 33. Summary of the distinctive features for French passive-like constructions (in blue verb-related parameters, in violet participant-related parameters, and in burgundy context-related parameters).*

Examples 368 to 371 can then be considered as prototypical illustrations since they contain all the distinctive features listed in Table 33.

(368) *Monsieur le Président, Monsieur le Commissaire, la semaine dernière les Antilles* ***ont été ravagées par*** *l'ouragan Irma.* (PROCEP FR)
'Mr President, Commissioner, last week the Antilles **were devastated by** Hurricane' Irma.

(369) *Après quelques siècles, des efforts* ***se multiplient*** *pour recréer la race hovawart, notamment par des selectionneurs comme Bertram König.* (WIKI FR)
'After a few centuries, efforts **were multiplied** to recreate the Hovawart breed, notably by breeders such as Bertram König.'

(370) *Étymologiquement, le mot Jarnotte et tous ces dérivés est hérité d'un dialecte viking du Moyen Âge : le vieux norrois.{S} Ainsi, la Jarðhnot, comme* ***ils l'appelaient****, est une noix [hnot] de terre [jarð*] (WIKI FR)
'Etymologically, the word Jarnotte and all its derivatives is inherited from a Viking dialect of the Middle Ages: Old Norse.{S} Thus, Jarðhnot, as **they called** it, is a nut [hnot] of earth [jarð]'

(371) *Rendez-vous compte qu'en l'état actuel des choses,* ***on estime*** *que ces pratiques médicamenteuses ne sont pas loin de concurrencer, voire dépasser en volume, les prescriptions médicales destinées aux êtres humains.* (PROCEP FR)
'Do you realise that as things stand at the moment, **it is estimated** that these drug practices are not far from competing with, or even exceeding in volume, medical prescriptions for human beings.'

Turning to the characteristics of the passive-like constructions available in Spanish (the periphrastic passive, the *se*-construction, the non-anaphoric third person plural and the numeral-based construction *uno*), the analysis revealed that a high use of present tenses, which correlates with a high use of imperfective forms, is a distinctive feature of the *se*-construction compared to the periphrastic passive and the non-anaphoric third person plural. It was also interesting to note that, though the use of present tenses is relatively low in non-anaphoric third person plurals, the correlation with imperfective forms is less straightforward for this passive-like structure since it displays a relatively high number of imperfective forms. We suspect that a higher presence of *pretérito imperfecto* could partly explain this trend. The parameter *verb mode* was also found to be a distinctive feature between the three passive-like strategies. The non-anaphoric third person resorts the most to the indicative mode and the periphrastic passive the least, the *se*-construction occupying an in-between position. Regarding the type of verb process, while the periphrastic passive and the *se*-construction display an overall similar profile, our findings revealed that the non-anaphoric third person plural shows a fairly different preference as far as the type of verbs is concerned. While material verbs predominate in the periphrastic passive and the *se*-construction, though in slightly different proportions, followed by relational and mental verbs, the non-anaphoric third person plural resorts equally to material

and relational verbs. Another distinctive feature of the non-anaphoric third person plural is the large presence of verbal processes, occupying the third most frequent position whereas the other two passive-like strategies rather resort to mental verbs.

As expected, the non-anaphoric third person plural possesses a different organization as for participants. Compared to the other two structures, the non-anaphoric third person plural contains much more one-participant clauses, with an agent which is always classified as an animate human entity. In the *se*-construction, the agent is always defocused. The almost exclusive use of inanimate patients, strongly correlated with a low degree of individuation is another feature that characterizes the *se*-construction. Results also indicated that, when an overt agent is present in the periphrastic passive, animate human agents and inanimate *others* agents occur in similar proportions. The analysis thus stressed the key role of participants in portraying the Spanish passive-like structures.

The examination of the third group of parameters, namely context-related parameters, attested that a high use of these elements is closely associated with the periphrastic passive. The presence of another passive-like structure, a salient feature specific of the Spanish system, is especially dominant in the periphrastic passive. Overall, the non-anaphoric third person plural is the strategy that resorts the least to contextual elements.

In conclusion we argue that, even though the periphrastic passive and the *se*-construction share a series of features, such as the type of verb processes or the fact that the presence of an overt agent is rare, establishing two groups of passive-like structures, as is the case in French, seems less straightforward. Each of the three Spanish strategies under study appeared as possessing some very distinctive features, linked to: (i) the aspect-temporal dimension and the nature and degree of individuation of the patients for the *se*-construction; (ii) the type of verb process and number of participants for the non-anaphoric third person plural and; (iii) the nature of the agent and the high presence of contextual elements for the periphrastic passive. These features are summarized in Figure 34 and illustrated in 372 to 374 with prototypical examples including all the distinctive features.

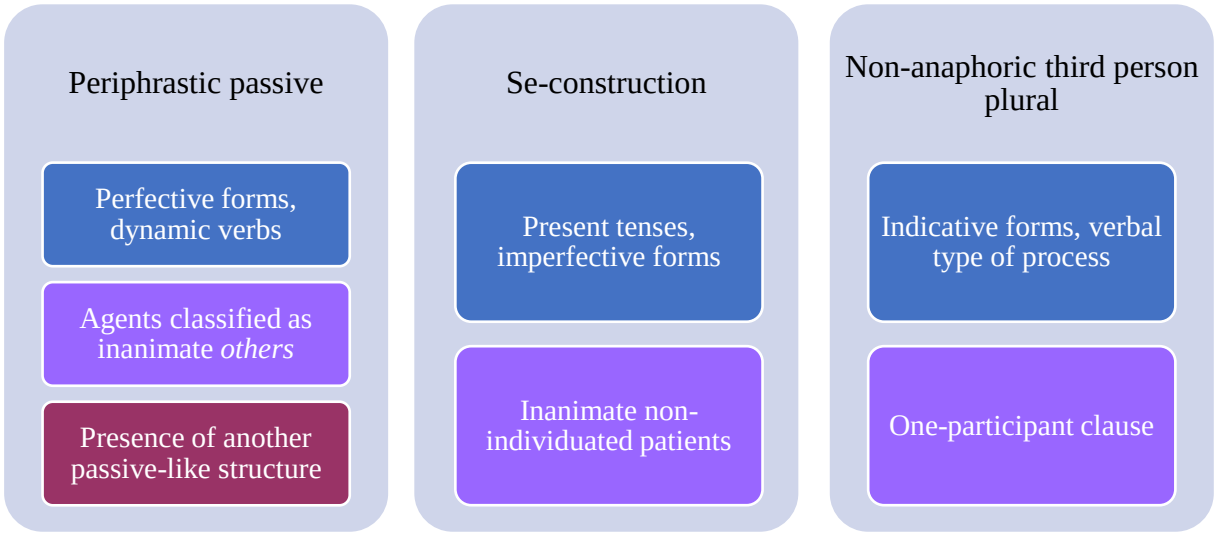



*Figure 34. Summary of the distinctive features for Spanish passive-like constructions (in blue verb-related parameters, in violet participant-related parameters, and in burgundy context-related parameters)*

(372) *Esto **fue agravado por** intereses financieros, ya que se decidió mantener la seguridad al mínimo para que esta no fuera un inconveniente para los pasajeros y no perdieran tiempo en las salidas*. (PROCEP SP)
'This **was aggravated by** financial interests, as it was decided to maintain security to a minimum so that it would not be an inconvenience to passengers and they would not lose time on departures.'

(373) *Mediante técnicas biotecnológicas **se producen** esas enzimas que pueden usarse en el reciclado de papel, disminuyendo el coste económico y la contaminación*. (WIKI SP)
'By means of biotechnological techniques, these enzymes **are produced** that can be used in paper recycling, reducing the economic cost and pollution.'

(374) *H2 una escalera que en los peldaños tiene así como unos cuadraditos y tal / **dicen** que la escalera se salvó*
*H3 [ya] HI si no= [<inint/> desaparecido]*
*H2 porque para subir carros / la habían cubierto de tierra [si no esa escalera habría] desaparecido.* (ESLORA SP)
'H2 a stairway that on the rungs has like little squares and so on / **they say** that the stairway is saved
H3 [ya] HI si no= [<inint/> disappeared]]
H2 because in order to get cars up / they had covered it with earth [otherwise the stairs would have] disappeared.'

Lastly, the close examination of the English passive-like structures, namely the periphrastic passive, and the non-anaphoric third person plural revealed the following results. Both constructions rely on a very similar proportion of present tenses. While the use of verbs in the present is correlated with the distribution of the imperfective form in the periphrastic passive, it does not appear to be the case for the non-anaphoric third person plural. A similar phenomenon was already observed in the Spanish non-anaphoric third person plural. The English non-anaphoric third person plural is also the construction that displays the highest proportion of indicative forms. The type of verb process is an additional feature that distinguishes the periphrastic passive from the non-anaphoric third person. The periphrastic passive tends to rely more on material and mental verbs than the non-anaphoric third person plural, while this latter shows a much more heterogeneous distribution with an ample use of relational, material, mental, verbal and modulation types of processes. The marked presence of material verbs in the periphrastic passive may contribute to explaining the significantly higher presence of dynamic verbs in this structure. A last distinctive verb-related feature this study uncovered is the specific affinities of the periphrastic passive with realis and affirmative clauses, which make this passive-like structure much more anchored in the real world than the non-anaphoric third person plural.
The distribution of participant-related parameters also greatly differs in the periphrastic passive and the non-anaphoric third person plural. The internal structure of the periphrastic passive

makes it a better candidate for two-participant clauses than the non-anaphoric third personal plural, which relies much more on one-participant clauses. The presence and the nature of an agent and a patient are other key factors differentiating the two passive-like structures. Whereas the non-anaphoric third person plural consists in an explicit animate human agent, the periphrastic passive defocuses its agent in the majority of the cases. When overt, the agent tends to be classified as either inanimate *others* or animate human. Concerning patients, the non-anaphoric third person plural displays a much higher proportion of inanimate patients, correlated with a lower degree of individuation.

The role of contextual elements in the English passive-like structures appeared slightly less crucial than in the other two languages since the distribution of spatio-temporal restrictions is similar across the two structures. The only substantial difference is the higher presence of prepositional phrases, mainly with agentive or locative meaning, in the periphrastic passive. The results yielded by this study prove that the English periphrastic passive and the non-anaphoric third person plural occupy rather distinctive positions on the continuum of passive-like structures. The distinctive features of the structures are summarized in Figure 35. However, our results showed that when register and mode variations are considered, the two passive-like structures share a wider range of features in some data sets. These findings will be further commented on when answering the third research question of this study.

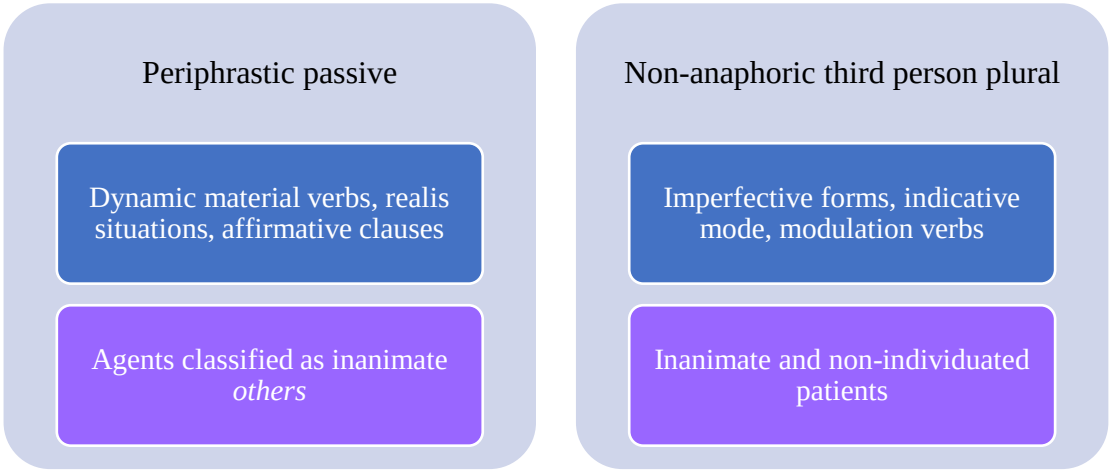



*Figure 35. Summary of the distinctive features for English passive-like constructions (in blue verb-related parameters and in violet participant-related parameters)*

Examples (375) and (376) illustrate the distinctive features of the periphrastic passive and the non-anaphoric third person plural in English.

(375) *The humane thing to do is to act in the interests of those whose lives* ***are now being torn apart by*** *climate breakdown, especially those living in low-lying territories and in the path of storms; (...)* (WIKI EN)

(376) [speaking about laptops in hospitals]

*<"S0115">any any price that would possibly be able to go up would go up and uh and yeah charging charging a patient ten pounds a day to watch TV for example is uh is is*

*uh is it is partly desperation and also of course it is very unfair because they can not um have they have no ability to to to choose another service although I do not know what they do with laptops if* ***they allow*** *them um which they probably should do</u> <"S0037" trans="overlap">yeah I know shocking opt out like yeah I think you were not <trunc>be</trunc> allowed them before but I think you sort of are now but I am not sure</u> <"* (BNC EN)

The drawing of the three language-specific portraits provided valuable information to answer the first two questions addressed in this study, namely what are the main features of the passive-like structures under examination and to which extent language is a key factor in describing passive-like constructions. This second research question was further addressed through a language-specific comparative approach to the structures. Such a method allowed to uncover the extent to which constructions with a similar surface structure share common features.

The analysis of the periphrastic passive, available in the three languages, showed that, though the three constructions share a common ground, namely the dynamic nature of the verb and the distribution of realis and affirmative clauses, each language also possesses its own characteristics. In Spanish, the periphrastic passive shows a specific association with past tenses and perfective forms that was not observed in the other two languages. Compared to its English and French counterparts, the Spanish periphrastic passive is also the one to resort the most to explicit agents and to contextual elements. Our results indicated that the French periphrastic passive can be characterized by a high use of indicative forms and a restricted presence of animate agents. A large use of inanimate patients is a third distinctive feature of the periphrastic passive in French. As far as the English periphrastic passive is concerned, the especially low presence of elements guiding the interpretation in the linguistic context seems to be its main distinguishing characteristic. Finally, the analysis pointed out a specific use of verb type processes in each language: the proportion of the most frequent category, namely material process, varies from one language to the other.

The second group of passive-like structures who received particular attention in this study was the non-anaphoric third person plural, a strategy again available in English, French and Spanish. The analysis revealed that, in addition to a different frequency of use, especially high in Spanish, the three structures share only a very limited number of features, i.e., the number of affirmative clauses, the nature of their agent and the use of spatio-temporal restrictions.

The pair formed by the *se*-construction in French and in Spanish was the last group of surface similar structures closely examined in this investigation. The present work largely confirmed the difference of frequency between the two *se*-constructions, *se* being much more frequent in Spanish than in French. Our corpus-based study partly confirmed Lamiroy's claim (1993) that the two structures differ on the transitive versus intransitive nature of the verb and the aspect-temporal dimension of the structure, the French *se* relying almost exclusively on present tenses and imperfective forms. Lamiroy also suggested differences in the degree of animacy of the nominal clauses in both languages. This distinction was not confirmed by our study since no significant difference regarding the animate versus inanimate nature of the participant could be identified. In addition to these observations, our results brought further information through the study of contextual elements, which have received little attention in the literature so far.

Compared to its Spanish counterpart, French *se* appears surrounded by much more elements in the linguistic context that help to identify possible participants, such as spatio-temporal restrictions and prepositional phrases.

Overall, the comparison of the three sets of passive-like constructions revealed that, even though they are similar regarding their surface structure, they vary not only as for their frequency but also as for the range of shared features. These results were complemented by an examination of the French man-impersonal *on*, which does not possess a structural equivalent in the other two languages included in this study. As pointed out by the literature, structures such as the non-anaphoric third person plural, the periphrastic passive or the *se*-construction are usually considered to be possible translational alternatives of *on*. Our results indicated that the closest option to the man-impersonal *on* is the other R-impersonal, namely the non-anaphoric third person plural. The French periphrastic passive and the Spanish *se*-construction, also considered to be frequent alternants, present some major differences. The results provided by this study allow to raise awareness of the fact that, even though a grammatical phenomenon may appear to be similar in two or more languages, each one possesses its own specificities and subtleties. It is important to consider these observations when carrying out translation processes. As already pointed out by several researchers (Johansson 2007, Viberg 2010, Johansson & Rawoens 2019) and confirmed in our study, divergences between seemingly comparable forms in two or more languages exist and, depending on the context, another strategy than the structural equivalent may be selected in the target language to convey the intended meaning.

The third research question formulated in this work intended to determine to which extent mode and register are crucial factors to consider in contrastive studies such as the present one. This issue was approached along two axes: (i) the first approach identifies which parameters are most influenced by mode and register and (ii) the second approach contrasts structures presented in the literature as possible equivalents. In the exploration of the second axis, the focus was placed on pairs usually described on the literature as depending on the register. More precisely, special attention was given to the alternation of the periphrastic passive and the non-anaphoric third person plural in informal data, and on the difference between the periphrastic passive and the *get*-passive, highly associated with more colloquial situations. The investigation into the frequency of the passive-like structures included in this work showed that some structures are highly register and mode sensitive and, therefore, cannot be considered to be equivalents.
The examination of the parameters in the light of mode and register variation revealed that it is crucial to consider these two factors when studying grammatical phenomena such as those surveyed in this study. It showed that some parameters are much more sensitive to variation, such as *aspect* or *tense*, while others are much more stable across the registers. Our findings also indicated that the influence of the degree of formality is stronger in oral data than in written data. Note that this key role played by the degree of formality is a phenomenon that goes beyond passive-like structures and can be applied to other linguistic phenomena. We suspect that the higher degree of standardization of written data could contribute to explaining this phenomenon. The analysis also pointed to a specific behaviour of the formal oral data set in some cases resembling informal oral data, in other cases coming closer to written data or

standing alone. In addition to the influence of the degree of colloquiality of the data, this study also highlighted the importance of the oral versus written nature of the data, each one displaying specific characteristics regarding the different passive-like structures.
The examination of the behaviour of the periphrastic passive and the non-anaphoric third person plural revealed that these two structures share a higher number of features in informal oral data than when register and mode are not considered. This finding corroborates Weiner & Labov's claim (1983). This phenomenon was especially observed for verb-parameters and context-related parameters. However, this claim appeared valid only for English, as it did not verify in our Spanish and French data. The closer qualitative study of the periphrastic passive and the *get*-passive in informal oral data revealed that, although these constructions share a slightly higher number of characteristics, they cannot be regarded as interchangeable structures. This is partly due to the fact that the *get*-passive is a much more semantically marked structure. A larger set of data, including a higher number of occurrences of the get-passive would be welcomed to confirm these tentative conclusions. Our analysis also allowed to demonstrate that the French man-impersonal *on* and the numeral-based constructions *one* and *uno* cannot be considered to be equivalents : they display a different frequency of use and are associated with a distinct degree of formality and mode of production. The inclusion of mode and register in this study contributes to highlighting the importance to consider such factors not only when carrying out contrastive analysis but also when studying linguistic phenomena in general.

## Theoretical and methodological contributions

Prior research on passive-like structures has extensively described most constructions included in this dissertation from a monolingual and theoretical perspective. As pointed out in the introductory chapter of this work, few corpus-based studies both contrast two or more languages and include a series of passive-like structures. We thus lacked a deep understanding of the actual syntactico-pragmatic behaviour of these structures from a comparative and systematic corpus-based approach. The present study pushes back the boundaries of our understanding of passive-like structures by contributing theoretically and methodologically to the field of corpus-based contrastive linguistics.

First, this research advances our understanding of defocussing mechanisms by establishing and comparing the distinctive features of each of the six passive-like structures included in the study. As an outcome, the enhancement of these features led to the establishment of a prototype for each construction across the three languages. To our knowledge, the portrayal of typical agent-defocussing strategies is a novelty in the existing body of literature on the topic. In addition, our findings provide insightful information on the different discursive functions of the six passive-like structures. We could highlight them especially through the analysis of the parameter *participant* which results to be a key parameter across our data. More especially, the notion of topicalization has been explored. Whereas the patient received the entire focus in the periphrastic passive through a promotional move, the R-impersonals maintain the agent in topic position and reduce its referentiality. As a consequence, the patient receives less attention than in a periphrastic passive. The qualitative analysis of the parameter *agent* also showed that the

adversative nature of the situation and the cohesive function of the structure explain the presence of an overt agent. The analysis of the parameters *tense* and *aspect* also largely contributes to the explanation of the pattern of use of the structures. Certain structures are more prone to be used to reflect generalizations (e.g., *se*-constructions and some instances of *on*) whereas other are preferred options to express specific situations. The choice of a structure which places the patient in topic position or which presents the situation as general rather than specific has thus different cognitive impacts on the way the receptor perceives the situation (Hansen-Schirra et al. 2012).

Second, our study contributes to the field of contrastive corpus linguistics through its solid corpus composition. By including data both from oral and written situations and from formal and informal contexts, the present analysis improves the understanding of the pattern of use of the selected passive-like structures. Since Dieck (2016) argues that information on the distribution of passive-like structures according to the degree of formality and the spoken versus written nature of the data are often lacking, this study bridges the gap and highlights the importance of including these variables. The multidimensional approach brought insightful information on the distribution of the structure across the different combinations of register/mode. The present study highlights the following tendencies: (i) the periphrastic passive shows greater affinities with formal and written productions, (ii) the Spanish *se*-construction is extremely productive in written and formal oral data while its French counterpart is not, (iii) the non-anaphoric third person plural shows a preference for informal written situations, (iv) the man-impersonal *on* is specific of oral data, especially informal ones, and (v) the *get*-passive, a semantically marked construction, is characteristic of informal oral data. This work stresses the importance of considering mode and register not only in contrastive studies but also in translation studies and teaching fields. A deep understanding of the complete language system is crucial, since, as pointed out by Steiner (2004), non-native speakers may express themselves in a comprehensible way but not sound natural as they did not select the adequate construction according to the degree of formality of the situation.

Third, this study makes significant methodological contributions in contrastive linguistic studies by adopting a multifactorial approach to the matter. The annotation scheme designed in this work combines parameters concerning syntactic (e.g., tense, aspect, and kinesis), semantic (e.g., type of verb process or animate versus inanimate nature of the participants), and contextual elements (e.g., use of spatiotemporal restrictions) but it also includes external variables (the degree of formality and the mode of production). The annotation and analysis of a large set of parameters in three languages allowed, among others, the identification of a prototype with its distinctive features for each agent-defocussing mechanism. Such undertaking would not have been possible without the inclusion of many different variables. By carrying out a multilayer analysis, this investigation positions itself among the only studies on agent-defocussing mechanisms to adopt such a fine-grained and comprehensive approach. The numerous variables included in the present work allowed us to tackle a problem raised by De Sutter & Lefer (2020). They argued that a serious weakness in corpus-based translation studies, which we extend here to contrastive corpus-based studies, is the lack of multifactorial research designs to explore the data. The authors highlight the importance of investigating a linguistic phenomenon with reference to a wide range of langue-internal and -external factors and not

only from a monofactorial approach (2020:5). In response to this plea for a multilayer method, this work uses Multiple Correspondences Analysis statistical models to explore the relationship between the parameters (syntactic, semantic, and contextual elements), the languages, the level of formality, and the mode of production of the discourse. In addition to these advanced statistical models that allow to visualize and uncover potential relationships between variables, the present work includes an interrater agreement process. The measure of the agreement, still infrequent in contrastive studies, ensures the objectivity in codifying procedures and thus the replicability and the reliability of the work.

To conclude, this thesis, through its wide scope and multi-faceted nature (three languages, six structures, two modes and two registers) has contributed to refining the understanding of the system of defocussing strategies and, by integrating such a large range of variables and factors, has ventured to do what few researchers have done so far. The special attention devoted to mode and register is one of the major strengths and contributions of this work, as the literature on passive-like structures is generally limited to the study of the characteristics of one genre. As this thesis has shown, valuable results can be obtained, which can be widely used in different fields such as language teaching or translation practices.
The present work points towards the need to include various dimensions in contrastive studies and not only to focus on the grammatical phenomenon the researchers are interested in. As we observed, elements in the linguistic context can provide further information on the object of study and thereby bring a supplementary dimension to the study. Similarly, it is crucial to consider language specificities since, as already pointed out in the literature and confirmed by this study, the usage pattern of structurally surface similar constructions differs from one language to another, even though they are closely related languages such as the two Romance languages included in this study. In his work on Swedish impersonal constructions (2010:146), Viberg already reported that the "choice is influenced by a complex set of factors having to do with the context of situation, morphological and syntactic factors, semantic properties of the verb, as well as the idiosyncratic lexical characteristics of individual verbs." The present contrastive work on English, French and Spanish passive-like constructions largely supports Viberg's observation. In addition, this study also highlights the need to consider register and mode variation when examining a linguistic phenomenon since, as demonstrated, parameters can vary considerably depending on the characteristics of the situation. This finding can be extended to the overall importance of mode and register in the selection of the data and the compilation of the corpora.

The present work has offered numerous new insights into the domain of passive-like structures in English, French and Spanish. Hereafter I suggest some possible research avenues to further deepen the current knowledge of agent-defocussing mechanisms.

### Further avenues

A first possible line of exploration could combine comparable and translational data. As pointed out by Granger (2018:185), translated language possesses its distinctive features. The

integration of bidirectional parallel corpora, using English, French and Spanish both as source and target language, would certainly be an additional asset in the study of linguistic phenomena such as those analysed in the present dissertation. The results yielded by using translational data would allow us to examine whether the differences put to the fore in this work are considered when performing the translation process. Among others, the present work has highlighted differences in aspect and tense or in the animate versus inanimate nature of agents and patients, which probably play a substantial role in the translation process. Another field worth being investigated is the extent to which the translated data take into account the linguistic environment such as the presence of spatio-temporal restrictions.

Another promising research avenue could be the exploration of the possibility of complementing the results of this study by including a second Germanic language which contains a man-impersonal construction, such as German (*man*) or Dutch (*men*). The inclusion of such a type of language would allow to examine whether the system of passive-like structures is distributed differently compared to the English one, the Germanic language included in this study, and observe whether the German or the Dutch system comes closer to the French system, which, unlike English, also contains a man-impersonal form. Given the higher presence of man-impersonals in informal data, it would also be worthwhile to examine whether the periphrastic passive and the non-anaphoric third person plural, the two most frequent strategies in English, have a distribution in German and Dutch informal data similar to the one observed in English or rather similar to the one observed in French. In a similar line of inquiry, the literature indicated that the English man-impersonal, disappeared since the fifteenth century, is re-emerging in some specific varieties of English. It would be interesting to analyse its characteristics and examine whether these are similar to those observed in the French man-impersonal use or whether the English reborn construction possesses a divergent profile.

Finally, the present work concentrated on the analysis of verb-related, participant-related and context-related parameters and described their respective nature and behaviour. However, we believe that it would be worthy to go one step further in the analysis of participants and investigate into the types of participant (hearer, speaker or a third entity) which are included in the defocussing mechanism and the possible interpretations. The analysis of these interpretations could rely on taxonomies such as the one developed and used by Siewierska and Papastathi (2011) or Aijón Oliva (2020), which allows to display the nature of the interpretation of the defocussing mechanism : universal, specific, corporate etc. Another possible avenue would be to identify whether the interpretation of the mechanisms is similar in the three languages and whether some interpretations are associated to certain modes and registers. Moreover, the abovementioned taxonomy has mainly been developed to analyse structures such as R-impersonals. A further step in the study would be to determine the extent to which such a framework can be applied to passive-like structures such as *se*-constructions or numeral-based constructions.

Overall, in the research field of passive-like structures, numerous avenues remain open to investigation. We assume that by combining such a large number of parameters and factors, the present work has highlighted the importance of considering different types of variations in contrastive corpus-based studies. Although variation can be considered as “a problem relevant

to all linguistic levels" (Neumann 2013: 1), in my opinion, variation should be viewed not as a problem but rather as a strength built on the numerous means to apprehend and decipher the world, enabling to convey our message in a most accurate and adapted way according to our situation in this world. And this creates the richness of a language.

# References


Aarts, B. & Meyer, C. 1995. *The verb in contemporary English: Theory and description.* Cambridge University Press. Cambridge.

Abdi, H. & Valentin, D. 2007. Multiple correspondence analysis. *Encyclopedia of measurement and statistics*, 2(4): 651-657.

Abouda, L. 1997. Le conditionnel : temps ou mode ? Arguments syntaxiques. *Revue Romane*, 32(2): 179-198. John Benjamins Publishing Company.

Achard, M. 2010. Fields and settings: French il and ça impersonals in copular complement constructions. De Gruyter.

Achard, M. 2015. *Impersonals and other agent defocusing constructions in French,* (50). John Benjamins Publishing Company.

Agha, A. 2004. Registers of language. *A companion to linguistic anthropology*, 23-45.

Aijón Oliva, M. A. 2020. Talking about 'others': Referential readings and pragmatic functions of non-phoric plural third persons in Spanish media discourse. *Lingua*, 243, 102906.

Akinnaso, F. N. 1982. On the differences between spoken and written language. *Language and speech*, 25(2): 2, 97-125.

Alabeda Marco, A. & Colomer, M-J. 2006. La enseñanza de los registros lingüísticos en E/LE. Una aplicación a la conversación coloquial. *MarcoELE. Revista de Didáctica Español Lengua Extranjera*, 2006(3): 1-31.

Alarcos Llorach, E. 2000. *Gramática de la lengua española*. Espasa Calpe.

Alexiadou, A. 2005. *A note on non-canonical passives: the case of the get-passive,* 13-21. *Organizing Grammar; Studies in Generative Grammar,* (86):13-21. Mouton de Gruyter.

Alonso-Ovalle, L. 2000. Is the 'arbitrary interpretation' a semantic epiphenomenon? *University of Massachussetts Occasional Papers in Linguistics*, 23(9): 155-183.

Alsaawi, A. 2019. *Spoken and Written Language as Medium of Communication: A Self-reflection.*

Altenberg, B. & Granger, S. 2002. Recent trends in cross-linguistic lexical studies. Altenberg B. and Granger S. (eds). *Lexis in Contrast. Corpus-based Approaches. Studies in Corpus Linguistics, (*7), 3-48. Amsterdam & Philadelphia. Benjamins.

Anderwald, L. 2018. Language change and cultural change: The grammaticalization of the get-passive in context. *Language & Communication,* (62): 1-14.

Arce-Arenales, M., Axelrod, M. & Fox, B. A. 1994. Active voice and middle diathesis. A Cross-Linguistic Perspective. *Voice: Form and function. Typological Studies in Language*, (27):1-21.

Ashby, W. J. & Bentivoglio, P. 1993. Preferred argument structure in spoken French and Spanish. *Language Variation and Change,* 5(01): 61.

Babcock, S. 1970. *The syntax of Spanish reflexive verbs: The parameters of the middle voice.* The Hague. Mouton.

Barcala, M., Domínguez, E., Fernández, A., Rivas, R., Santalla, M., Vázquez Rozas, V. & Villapol, R. 2018. El corpus ESLORA de español oral: diseño, desarrollo y explotación. *CHIMERA: Romance Corpora and Linguistic Studies* (5)2, 217-237.

Bach, E. 1981. *On time, tense, and aspect: An essay in English metaphysics*. University of Massachusetts. Amherst.

Bache, C. 1982. Aspect and Aktionsart: towards a semantic distinction. *Journal of Linguistics,* 18(1): 57-72.

Barrajón López, E. 2005. Un caso de impersonalidad semántica: el uso de los llamados 'singulares arbitrarios' en corpora orales. *ELUA,* (19): 47-64.

Barrenechea, A. 1979. *Operadores pragmáticos de actitud oracional: los adverbios en "-mente" y otros signos,* 39-59. Hachette. Buenos Aires.

Barrenechea, A. 1987. *El habla culta de la ciudad de Buenos Aires*. Materiales para su estudio. Instituto de Filología y Literaturas Hispánicas. Universidad Nacional de Buenos Aires. Facultad de Filosofía y Letras. Buenos Aires.

Bartra, A. 2014. Voice phenomena: how many properties behind this label?. *Catalan journal of linguistics*, (13): 3-18.

Béguelin, M. J. 1998. Le rapport écrit-oral. Tendances dissimilatrices, tendances assimilatrices. *Cahiers de linguistique française*, (20): 229-253.

Bennett, K. 2009. English academic style manuals: A survey. *Journal of English for Academic Purposes*, (8): 43-54.

Benzécri, J.-P. 1992. *Correspondence Analysis Handbook.* New York. Marcel Dekker.

Bermúdez, F. 2002. La estructura evidencial del castellano: elevación de sujeto y gramaticalización. *Romansk Forum*, 16(2): 19-29.

Bestgen, Y. 2017. *Getting rid of the Chi-square and Log-likelihood tests for analysing vocabulary differences between corpora*. Université Catholique de Louvain.

Bhatia, V. 1993. *Analysing genre: Language use in professional settings.* London. Longman.

Biber, D. 1988. *Variation across speech and writing.* Cambridge. Cambridge University Press.

Biber, D. 1995. *Dimensions of register variation: A cross-linguistic comparison*. Cambridge University Press.

Biber, D. 2006. Stance in spoken and written university registers. *Journal of English for Academic Purposes*, 5(2): 97-116.

Bidot Martínez, I. 2008. La desfocalización del centro deíctico personal a través de la segunda persona del singular, *Boletín de Lingüística*, 20(30): 62-87.

Blanca, D. & Lourdes, A. 2005. Oraciones impersonales en el habla de Mérida (Venezuela): varias configuraciones, un fenómeno. *Boletin de Linguistica*, 17(24): 11-17.

Blanche-Benveniste, C. 1997. La notion de variation syntaxique dans la langue parlée. *Langue française*, (115): 19-29.

Blanche-Benveniste, C. 2002. Structure et exploitation de la conjugaison des verbes en français contemporain. *Le français aujourd'hui,* (4): 11-22.

Blanco Canales, A. 2004. *Estudio sociolingüístico de Alcalá de Henares*. Alcalá de Henares. Universidad de Alcalá. Servicio de Publicaciones.

Borra, J. 2010, *La impersonalidad semántica de on y sus equivalencias en español*. Masterproef. Universiteit Gent.

Bosque, I. 2014. On resultative past participles in Spanish. *Catalan journal of linguistics*, (13): 41-77.

Boutet, J. 1986. La référence à la personne en français parlé: le cas de "on". *Langage & société*, (38): 19-49.

Briz, A. 2010. El registro como centro de la variedad situacional. Esbozo de la propuesta del grupo Val.Es.Co. sobre las variedades diafásicas. Fonte, I. & Rodríguez Alfano, L.

(compiladoras), *Perspectivas dialógicas en estudios del lenguaje,* 21-56. México: Universidad Autónoma de Nuevo León, Facultad de Filasofía y Letras.

Bruening, B. 2011. Pseudopassives, expletive passives, and locative inversion. *University of Delaware, unpublished ms*.

Cabañas Maya, R. 2006. La animicidad como factor de variación entre la pasiva perifrástica y la pasiva refleja. *Anuario de Letras, (*44): 125-146.

Cabasino, F. 2010. Des formules rituelles de l'adresse au conflit verbal personnalisé dans l'espace parlementaire. *S'adresser à autrui. Les formes nominales d'adresse en français,*169-200. Chambery. Ed. De l'Université de Savoie, Collection Langages.

Cabredo Hofherr, P. 2006. "Arbitrary" pro and the theory of pro-drop, *Agreement and Arguments,* Ackema, P., Brandt, M, Schoorlemmer, M. &Weermann, F. (eds). Oxford. Oxford University Press.

Cabredo Hofherr, P. 2008. Les pronoms impersonnels humains: Syntaxe et interprétation. *Modèles linguistiques*, (29): 57.

Calvo Montoro, M. J. 1983. *La voz pasiva*. Madrid. Coloquio.

Campos, H. 1989. Impersonal Passive"se" in Spanish. *Lingvisticae Investigationes*, 13(1): 1-21.

Carlier, A. 2002. Les propriétés aspectuelles du passif. *Cahiers Chronos*, (10): 41-63.

Carter, R. & McCarthy, M. 2006. *Cambridge grammar of English: a comprehensive guide; spoken and written English grammar and usage*. Ernst Klett Sprachen.

Casielles Suárez, E. 1996. ¿Es la interpretación arbitraria realmente arbitraria? *Revista Española de lingüística*, 26(2): 359-378.

Cela, E. & Candea, M. 2013. L'apport des indices posturo-mimico-gestuels et intonatifs à la construction de la référence de "tu" dit générique. *L'information grammaticale*, (136): 12-19.

Chafe, W. & Danielewicz, J. 1987. Properties of Spoken and Written Language, *Comprehending oral and written language*, ed. by Horowitz R. & Samuels, S. J. 83-113. San Diego. Academic Press.

Chafe, W. & Tannen, D. 1987. The relation between written and spoken language. *Annual review of anthropology*, 16(1): 383-407.

Chappell, H. 1980. Is the get-passive adversative? *Research on Language & Social Interaction*, 13(3): 411-452.

Cheshire, J. 2013. Grammaticalisation in social context: The emergence of a new English pronoun. *Journal of Sociolinguistics*, 17(5): 608-633.

Chesterman A. 1998. *Contrastive Functional Analysis*. Amsterdam. John Benjamins Publishing Company.

Clements, J. C. 2006. Transitivity and Spanish non-anaphoric se. *Functional approaches to Spanish syntax*, 236-264. Palgrave Macmillan. London.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, (20): 37-46.

Collins, P. 2012. Grammatical variation in English worldwide: the role of colloquialization. *Linguistics & the Human Sciences*, *8*(3).

Company Company, C. & Pozas Loyo, J. 2009. Los indefinidos compuestos y los pronombres genérico-impersonales *omne* y *uno*. *Sintaxis histórica de la lengua española. Segunda*

*parte: La frase nominal*. Company Company, Concepción (ed.) México. Fondo de Cultura Económica y Universidad Nacional Autónoma de México, 1073-1222.

Comrie, B. 1976. *Aspect: An introduction to the study of verbal aspect and related problems*, (2). Cambridge University Press.

Comrie, B. 1977. In defense of spontaneous demotion: the impersonal passive. *Grammatical relations*, 47-58. Brill.

Connor, U. & Moreno, A. I. 2005. Tertium Comparationis: A vital component in contrastive research methodology. Paul Bruthiaux, Dwight Atkinson, William Eggington, William Grabe & Vaidehi Ramanathan (eds.), *Directions in Applied Linguistics. Essays in Honour of Robert B. Kaplan*, 153–164. Clevedon. Multilingual Matters.

Contí Jiménez, C. 2004. Construcciones pasivas con estar. *ELUA. Estudios de Lingüística*, (18): 21-44.

Coussé, E. & Van der Auwera, J. 2012. Human impersonal pronouns in Swedish and Dutch: A contrastive study of man and men. *Languages in contrast*, 12(2): 121-138.

Coveney, A. 2000. Vestiges of "Nous" and the 1st Person Plural Verb in Informal Spoken French, *Language Sciences*, 22(4): 447-481.

Creissels, D. 2000. L'emploi résultatif de être+ participe passé en français. *Cahiers Chronos*, (6): 133-142.

Creissels, D. 2008. *Impersonal and related constructions: a typological approach.* Université de Lyon.

Crible, L. 2017. *Discourse Markers and (Dis)fluency across Registers: A Contrastive Usage-Based Study in English and French*. Doctoral dissertation. Université Catholique de Louvain.

Davison, A. 1980. Peculiar passives. *Languages*, (56).

de Araújo Júnior, B. J. 2015. Las construcciones -SE impersonales y pasivas en español y en el PB: reexaminando criterios de distinción. *SIGNOS ELE, Revista de español como lengua extranjera*, (9): 1-20

De Cock, B. 2014a. The discursive effects of Spanish uno and se. A case study of the phenomena of speaker inclusion and female-only reference. *Subjectivity and epistemicity. Corpus, discourse and literary approaches to stance*, Dylan Glynn, Mette Sjölin, 107-120. Lund. University Press.

De Cock, B. 2014b. *Profiling Discourse Participants. Forms and functions in Spanish conversation and debates.* Pragmatics & Beyond New Series, 246. Amsterdam. John Benjamins Publishing Company.

De Cock, B. 2020. Intersubjective Impersonals in Context: A Multivariate Analysis of Spanish "Uno" and "Se" in Spoken Language. *Catalan Journal of Linguistics*, 93-114.

DeLancey, S. 1987. Transitivity in grammar and cognition. *Coherence and grounding in discourse*, 53-68.

DeLancey, S. 1990. Cross-Linguistic Evidence for the Structure of the Agent Prototype. DeLancey Ed.

Delbecque, N. 2014. Passive Voice and Causal Roles in Spanish. *Nuclear and non-nuclear cases*, 217–318. Delbecque, N, Lahousse, K. & Van Langendonck, W. (eds.) Amsterdam. John Benjamins Publishing Compagny.

De los Angeles Gomez-Gonzalez, M., Lachlan Mackenzie, J. & Gonzalez Alvarez, E. M. (eds). 2008. *Languages and Cultures in Contrast and Comparison,* 3–26. Amsterdam. John Benjamins Publishing Company.

DeMello, G. 1995. Concordancia entre el verbo pronominal de tercera persona y su sustantivo: se venden flores vs. se vende flores. *Anuario de Letras. Lingüística y Filología,* (33): 59-82.

DeMello, G. 1995. Expresión impersonal + "para" + persona + infinitivo: 'es importante para Juan estudiar'. *Estudios Filológicos,* (30): 57-68.

de Miguel, E. 1999. El aspecto léxico, in Bosque. *Gramática descriptiva de la lengua española,* (2): 2977-3060. Ignacio and Violeta Demonte (eds.). Espasa Calpe. Madrid.

Desclés, J. P., & Guentcheva, Z. 1993. Le passif dans le système des voix du français. *Langages*, (109):73-102.

De Sutter, G., & Lefer, M. A. 2020. On the need for a new research agenda for corpus-based translation studies: a multi-methodological, multifactorial and interdisciplinary approach. *Perspectives*, *28*(1): 1-23.

Devís Márquez, P. 2003. La impersonalidad y las denominadas construcciones impersonales en español. *Zeitschrift für romanische Philologie (ZrP)*, 119(3): 393-442.

Devís Márquez, P. 2017. Las denominadas construcciones con se y verbos de alternancia causativa en español. *Neuphilologische Mitteilungen*, 118(2): 405-446.

Dieck, M. 2016. ¿Y qué más hace uno, pues?: la expresión de la impersonalidad en el español de Medellín. *Lingüística y Literatura*, (69).

Dixon, R. M. & Aikhenvald, A. Y. (Eds.). 2000. *Changing valency: Case studies in transitivity,* (1). Cambridge. Cambridge University Press.

Dobrovie-Sorin, C. 1998. Impersonal se constructions in Romance and the passivization of unergatives. *Linguistic Inquiry*, 29(3): 399-437.

Doró, K. 2007. The use of high-and low-frequency verbs in English native and non-native student writing. *Second language lexical processes: Applied linguistic and psycholinguistic perspectives*, 117-129.

Dowty, D. 1979. *Word Meaning and Montague Grammar*. Dordrecht. Holland. Reidel Publishing Co.

Dupont, M. 2019. *Conjunctive Markers of Contrast in English and French : From Syntax to Lexis and Discourse*. Doctoral dissertation. Université Catholique de Louvain.

Durán, J. R. 1993. Los equivalentes del "se" impersonal en inglés. *Revista Canaria de Estudios Ingleses*, (26): 105-118.

Ebeling, S. O. & Ebeling, J. 2020. Contrastive Analysis, Tertium Comparationis and Corpora. *Nordic Journal of English Studies*, 19(1): 97-117.

Espinoza, A. 1997. Contrastive analysis of the Spanish and English passive voice in scientific prose. *English for specific purposes,* 16(3): 229-43.

Evensen, L. S. 2008. The need for a tertium comparationis. *Applied Linguistics and SLA: A reply to Lantolf and Cazden*, (5): 221-227.

Fábregas, A. & Putnam, M. 2014. The emergence of middle voice structures with and without agents. *The Linguistic Review,* 31(2). De Gruyter Mouton.

Fábregas, A. & Marín, R. J. 2017. *On non-dynamic eventive verbs in Spanish,* 55(3): 451-488.

Fauconnier, S. 2012. *Constructional effects of involuntary and inanimate Agents: A cross-linguistic study*. Doctoral dissertation. Leuven.

Fenger, P. 2018. How impersonal does one get? *The Journal of Comparative Germanic Linguistics*, 21(3): 291-325.

Ferguson, C. A. 1959. Diglossia. W*ord*, 15(2): 325-340.

Fernández, S. 2007. *La voz pasiva en español: hacia un análisis discursivo*. Peter Lang.

Fernández, S. 2012. ¿Qué se esconde detrás de los usos impersonales de pronombres y desinencias personales?. *Actes du XVIIIe congrès des romanistes scandinaves/Actas del XVIII congreso de romanistas escandinavos,* 225.

Fernández, S. 2013. Impersonality in Spanish personal pronouns. *Deixis and Pronouns in Romance Languages*. Studies in language companion series 136. John Benjamins Publishing Company.

Fernández Soriano, O. & Táboas Baylín, S. 1999. Construcciones impersonales no reflejas. Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española,* (2): 1723-78. Madrid. Espasa.

Fleisher, N. 2006. The origin of passive get. *English Language and Linguistics*, 10(2): 225.

Fløttum, K., Jonasson, K., & Norén, C. 2007. *On: pronom à facettes.* De Boeck/Duculot.

Fonseca-Greber, B. & Waugh, L. R. 2003. On the radical difference between the subject personal pronouns in written and spoken European French. *Corpus Analysis,* 225-240. Brill Rodopi.

Franckel, J. J. & Paillard, D. 2007. *Grammaire des prépositions* (Vol. 1). Editions OPHRYS.

Franke, T. M., Ho, T. & Christie, C. A. 2012. The chi-square test: Often used and more often misinterpreted. *American Journal of Evaluation*, 33(3): 448-458.

Fredriksson, A.-L. 2016. *A corpus-based contrastive study of the passive and related constructions in English and Swedish*. Doctoral dissertation in English linguistics. University of Gothenburg.

Frias-Conde, X. 2006. La voix moyenne dans les langues romanes. *Revista de filologia románica*, (23): 215-228.

Gaatone, D. 1998. *Le passif en français*. Paris / Bruxelles: Editions Duculot, Champs Linguistiques.

Gaatone, D. 1999. Pour une définition restrictive du passif en français. *La langue, les signes et les êtres: actes du colloque de l'Institut d'Etudes Romanes de l'Université de Copenhague, le 3 octobre 1998, Le Passif, Etudes Romanes* (45): 15-22. Museum Tusculanum Press.

Gaatone, D. 2001. Les prépositions : une classe aux contours flous. *Travaux de linguistique,* 42-43(1-2): 23-31.

Gadet, F. et al. 2012. Un grand corpus de français parlé: le CIEL-F. *Revue française de linguistique appliquée*, 17(1): 39-54.

García, N. R. 2002. Contrastive linguistics and translation studies interconnected: The corpus-based approach. *Linguistica Antverpiensia, New Series–Themes in Translation Studies*, (1).

García Yebra, V. 1990. La voix passive en français et sa traduction en espagnol. *Meta: Journal des traducteurs /Meta:Translators' Journal*, 35(3): 510-517.

García Yebra, V. 1997. La voz pasivafrancesa y su traducción al español. *Thélème: Revista complutense de estudios franceses,* (11): 347-353.

García Yebra, V. 1997. *Teoría y práctica de la traducción.* Madrid. Gredos.

Gardner, M. 1970 *The annotated Alice.* (Revised edition.) Harmondsworth. Penguin.

Gast, V. & van der Auwera, J. 2013. Towards a distributional typology of human impersonal pronouns, based on data from European languages. *Languages across Boundaries.* De Guyter Mouton.

Gast, V. 2015. On the use of translation corpora in contrastive linguistics: A case study of impersonalization in english and german. *Languages in Contrast*, 15(1): 4-33.

Gatti, S. & Maroni, F. 2004. A profile of wine tourists in some Italian region vineyards: an application of the multiple correspondence analysis. *Colloque Econométrie (XI):* 21-22. Dijon.

Gauchola, R. 2012. Une approche typologique fonctionnelle de la voix passive en français et en espagnol. *Lidil. Revue de linguistique et de didactique des langues*, (46): 153-173.

Gelabert-Desnoyer, J. J. 2008. Not so impersonal: Intentionality in the use of pronoun uno in contemporary Spanish political discourse. *Pragmatics*, 18(3): 407-424.

Gennari, S. 2003. Tense meanings and temporal interpretation. *Journal of Semantics*, (20): 35-71.

Ghenimi, A. 2005. Traduire le passif. *Linguisticae investigationes: Revue internationale de linguistique française et de linguistique générale*, 28*(*1): 37-48.

Giacalone-Ramat, A. & Sansò, A. 2007. The spread and decline of indefinite man-constructions in European languages: An areal perspective. *Europe and the Mediterranean as Linguistic Areas : Convergencies from a historical and typological perspective.* Ramat P. and Roma E. (Eds), *Studies in Language Companion,* (88): 95-131.

Gili Gaya, S. 1995. *Curso superior de sintaxis española.* Barcelona. Publicaciones y Ediciones Spes.

Giordano, R. 2015. The impersonal sentences: a proposal for a teaching unit. *Procedia-Social and Behavioral Sciences*, (178): 117-122.

Gisborne, N. J. 2001. The stative/dynamic distinction and argument linking. *Language Science,* 23(4-5): 603-628.

Givón, T. 1979. From discourse to syntax: grammar as a processing strategy. *Discourse and syntax,* 81-112. Brill.

Givón, T. 1990. *Syntax. A Functional Typological Introduction.* Vol. II, Amsterdam / Philadelphia. John Benjamins Publishing Company.

Gjesdal, A. M. 2008. *Étude sémantique du pronom on dans une perspective textuelle et contextuelle*, PhD Thesis, University of Bergen.

Goedsche, C. R. 1940. Aspect versus aktionsart. *The Journal of English and Germanic Philology*, 39(2): 189-196.

Gómez Torrego, L. 1992. *La impersonalidad gramatical: descripción y norma*. Madrid. Arco Libros.

Gómez Torrego, L. 2015. *Gramática didáctica del español*. Ediciones SM España.

Grady, M. 1969. Varieties of the English middle voice and medio-passive. *Linguistics*, 7(54): 5-12.

Granger, S. 1983. *The be+ past participle construction in spoken English with special emphasis on the passive.* PhD Thesis, Université Catholique de Louvain.

Granger, S. 2003. The corpus approach: a common way forward for Contrastive Linguistics and Translation Studies. *Corpus-based approaches to contrastive linguistics and translation studies*, (17): 29.

Granger, S. 2010. Comparable and translation corpora in cross-linguistic research. Design, analysis and applications. *Journal of Shanghai Jiaotong University*, (2): 14-21.

Granger, S. 2018. Tracking the third code: A cross-linguistic corpus-driven approach to metadiscursive markers. *The corpus linguistics discourse: In honour of Wolfgang Teubert*, 185-204.

Gregory, M. 1988. Generic situation and register: a functional view of communication. *Linguistics in a systemic perspective*, 301–329. James D. Benson, Michael J. Cummings, and William S. Greaves (eds). Amsterdam and Philadelphia. Benjamins.

Greidanus, T. 1990. *Les constructions verbales en français parlé. Étude quantitative et descriptive de la syntaxe de 250 verbes les plus fréquents*. Tübingen. Max Niemeyer Verlag.

Grevisse, G. 1994. *Nouvelle Grammaire Française, 2e édition*, 67- 71 & 108-117. Belgique. Éditions Duculot.

Gross, M. 1971. Grammaire transformationnelle et enseignement du français. *Langue française*, (11): 4-14.

Gross, M. 1975. *Méthodes en syntaxe*. Paris. Hermann.

Guirado, K. 2011. La alternancia tú ~ uno impersonal en el habla de Caracas. *Lingüística*, (26): 26-54.

Hacquard, V. 2009. On the interaction of aspect and modal auxiliaries. *Linguistics and Philosophy*, *32*(3) : 279-315.

Halliday, M. A. 1966. Some notes on 'deep' grammar. *Journal of Linguistics*, 2(1): 57-67.

Halliday, M. A. 1978. *Language as Social Semiotic. The Social Interpretation of Language and Meaning*. London. Arnold.

Halliday, M. A. 1985. *Spoken and written language*. Melbourne, Australia. Deakin University Press.

Halliday, M. A. and Ruqaiya, H. 1989. *Language, context, and text: aspects of language in a social-semiotic perspective*. Oxford. Oxford University Press.

Halliday, M. A., & Matthiessen, C. 2004. *An introduction to functional grammar*. Third edition.

Hansen-Schirra, S. et al. 2012. *Cross-linguistic Corpora for the Study of Translations*. De Gruyter Mouton.

Hatcher, A. G. 1949. To get/be invited. *Modern Language Notes*, 64(7): 433-446.

Haverkate, H. 2004. Gramática y pragmática: categorías desfocalizadores en español. *Spanish in Context*, 1(1) : 21-40.

Helland, H. P. 1995. Futur simple et futur périphrastique: du sens aux emplois. *Revue Romane 30(1)*.

Helland, H. P. 2002. *Le passif périphrastique en français moderne*, 50. Museum Tusculanum Press.

Hernández A. C. 1974. El adverbio. *Thesaurus: Boletín del Instituto Caro y Cuervo*, 29(1): 48-67.

Hernanz, M. L. 1990. En torno a los sujetos arbitrarios: La 2a persona del singular. Violeta Demonte, B. Garza Cuarón (eds.) *Estudios de lingüística de España y México,* 151-178. México. Universidad Autónoma de México.

Hollaender Jensen, M. 2002. La referencia en algunas expresiones impersonales – Diferentes lecturas de uno y la segunda persona del singular. *Romansk Forum,* 16(2): 127-138.

Honda, S. 2012. On The Origin and Development Of The Get-Passive: With Special Reference To Grammaticalization. *English Linguistics*, 29(1): 69-87.

Hopper, P. & Thompson, S. 1980. Transitivity in grammar and discourse. *Language,* 56(2): 251-299.

Huddleston, R. D. 1971. The sentence in written English; a syntactic study based on an analysis of scientific texts. England. Cambridge University Press.

Huddleston, R. D. 1984. *Introduction to the Grammar of English*. Cambridge.

Huddleston, R. & Pullum, G. 2005. *The Cambridge grammar of the English language.* Cambridge University Press.

Hugo Rojas, E. 2011. Las formas de segunda persona singular como estrategias evidenciales. *RLA. Revista de lingüística teórica y aplicada,* 49(1): 143-167.

Hupet, M. & Costermans, J. 1976. Un passif: Pour quoi faire? Quinze années de travaux psycholinguistiques. *La Linguistique,* 12(2): 3-26.

Hurtado, L. M. & Ortega-Santos, I. 2019. On the use of uno in Colombian Spanish: the role of transitivity. *Studies in Hispanic and Lusophone Linguistics*, 12(1): 35-64.

Hymes, D. 1964. *Language in Culture and Society: A reader in Linguistics and Antropology.* New York.

Irvine, J. T. 1979. Formality and informality in communicative events. *American Anthropologist*, 81(4) 773-790.

Jalenques, P. 2016. Le passif en français et le statut référentiel du sujet. *SHS Web of Conferences*, (27): 14001 . EDP Sciences.

James C. 1980. *Contrastive analysis.* Longman. Harlow.

Jaszczolt, K. M. & Katarzyna, M. 2003. On Translating 'What Is Said'. Tertium Comparationis in Contrastive Semantics and Pragmatics. Jaszczolt K.M. and Turner K. (eds), *Meaning Through Language Contrast,* (2): 441-462. Amsterdam and Philadelphia. Benjamins.

Jespersen, O. 1924. *The Philosophy of Grammar*. Henry Holt and Company. New York.

Jespersen, O. 2003. *Essentials of English Grammar: 25th impression*. Routledge.

Jiao, X. & Dou, Q. 2002. On comparison between the notional passive sentences of English and Chinese. *Journal of Northwest Sci-Tech University of Agriculture and Forestry (Social Science),* 2(2): 82-84.

Jisa, H. & Viguie, A. 2005. A developmental perspective on the role of French on in written and spoken expository texts. *Journal of Pragmatics,* (37): 125–142.

Jisa, H. & Tolchinsky, L. 2009. Developing a depersonalized stance through linguistic means in typologically different languages: Written expository discourse. *Written Language & Literacy*, 12(1): 1-25.

Johansson S. & Hasselgård, H. 1999. Corpora and cross-linguistic research in the Nordic countries. Granger et al (eds) *Contrastive Linguistics and Translation*, 145-162.

Johansson, S. 2007. *Seeing through Multilingual Corpora. On the use of corpora in contrastive studies*. Amsterdam/Philadelphia. John Benjamins.

Johansson, A., & Rawoens, G. 2019. A corpus-based contrastive study of impersonal passives in Swedish and Dutch. *Languages in Contrast*, 19(1): 2–26.

Kaiser, E. 2015. Impersonal and generic reference: A cross-linguistic look at Finnish and English narratives. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics*, 6(2): 9-42.

Karamysheva, I. D. 2019. Tertium Comparationis in Modern Contrastive Linguistics (Revising Approaches to Teaching). *Вісник Житомирського державного університету імені Івана Франка. Філологічні науки*, 1(89): 11-18.

Kaufmann, S. 2005. Conditional truth and future reference. *Journal of Semantics*, 22(3): 231-280.

Keenan, E. L. 1975. Some universals of passive in relational grammar. *Papers from the 11th Regional Meeting, Chicago Linguistic Society*, (340): 52.

Kemmer, S. 2003. Transitivity and voice. *International Encyclopedia of Linguistics. Second Edition*, 277-279. Oxford. Oxford University Press.

Kibort, A. 2005. *Passive and passive-like constructions in English and Polish.* Doctoral dissertation. University of Cambridge.

Kluge, B. 2010. El uso de formas de tratamiento en las estrategias de generalización. Hummel, M., Kluge, B. and Vázquez Laslop, M. (eds) *Formas y fórmulas de tratamiento en el mundo hispánico,* 1107–1136. México D.F.: El Colegio de México / Karl-Franzens-Universität Graz.

Kluge, B. 2012. *Referential ambiguity in discourse. The generic use of the second person singular in the Romance languages*. Habilitation thesis, Universität Bielefeld.

Kluge, B. 2016. Generic uses of the second person singular–how speakers deal with referential ambiguity and misunderstandings. *Pragmatics*, 26(3): 501-522.

Koch, P. & Oesterreicher, W. 2001. Langage parlé et langage écrit, Holtus G., Metzeltin M. & Ch. Schmitt (éd.). *Lexikon der Romanistischen Linguistik,* (I/2): 584-627. Niemeyer. Tübingen.

Koehn, P. 2005. Europarl: A parallel corpus for statistical machine translation. *MT summit* (5): 79-86.

Koenig, J.-P. 1999. On a tué le président! The nature of passives and ultra-indefinites. Fox, B., Jurafsky, D. & Michaelis, L. (eds), *Cognition and Function in Language*, 256-272. Stanford, CA. CSLI Publications.

König, E. 2012. Contrastive linguistics and language comparison. *Languages in Contrast*, 12(1): 3-26.

Krzeszowski, T. P. 1984. Tertium comparationis. Jacek Fisiak (ed.), *Contrastive Linguistics: Prospects and Problems*, 301–312. Berlin. Walter de Gruyter.

Krzeszowski, T. P. 1989. Towards a typology of contrastive studies. Oleksy, W. (ed.) *Contrastive Pragmatics*, 55-72. Amsterdam/Philadelphia. John Benjamins.

Krzeszowski, T. P. 1990. *Contrasting languages: The scope of contrastive linguistics*, 51. Berlin. Walter de Gruyter Mouton

Kurteš, S. 2009. New horizons for contrastive analysis: grammatical prototypes as tertium comparationis. *Selected papers on theoretical and applied linguistics*, (18): 233-241. Computer Science.

Laanemets, A. 2013. The passive voice in spoken and written Danish, Norwegian and Swedish: A comparative corpus-based study. *Languages in Contrast*, 13(1): 67-89.

Lagae, V. 2002. Le passif pronominal: une forme complémentaire du passif périphrastique? *Cahiers Chronos*, (10): 133-149.

Lakoff, R. 1971. Passive Resistance. *C.L.S*., 149-161.

Lamiroy, B. 1993. Pourquoi il y a deux passifs. *Langages*, 27(109): 53-72.

Lamiroy, B. 2000. Sur certains rapports entre le passif pronominal et les datifs non lexicaux. *Etudes Romanes*, (45): 135-155.

Landragin, F. & Tanguy, N. 2014. Référence et coréférence du pronom indéfini on. *Langages*, (195): 99-115.

Langacker, R. W. & Munro, P. 1975. Passives and Their Meaning. *Language*, 51(4): 789-830.

Langacker, R.W. 1987. *Foundations of Cognitive Grammar, vol. I: Theoretical prerequisites.* Stanford. Stanford University Press.

Langacker, R. W. 2011. On the subject of impersonals. Fuchs, Z., Gries, S. T. & Brdar, M. (eds.), *Cognitive linguistics: Convergence and expansion*, 179–217. Amsterdam. Benjamins.

Lapidus N. & Otheguy, R. 2005. Overt nonspecific ellos in the Spanish of New York. *Spanish in Context*, (2): 157-176.

LaPolla, R. et al. 2011. On transitivity. *Studies in Language*, 35(3): 469-492. John Benjamins Publishing Company.

Laslop, M. E. & Díaz, A. H. 2010. Impersonalidad y pasividad en las normas sobre derechos humanos: El uso de "se" y de pasivas perifrásticas en las constituciones mexicanas (1824–2001). *Revista Internacional de Lingüística Iberoamericana*, 203-231.

Lavandera, B. 1984. Tensión entre lo impersonal y lo personal en la organización del discurso. Beatriz Lavandera (éd.). *Variación y significado*, 101-24. Paris. Hachette.

Le Bellec, C. 2009. *La diathèse verbale dans trois langues romanes: vers une description dans le cadre de la Grammaire Fonctionnelle Discursive.* (Doctoral dissertation, Université Toulouse le Mirail-Toulouse II).

Le Bellec, C. L. 2015. La grammaticalisation de la construction passive en se faire. *Cahiers de praxématique*, (65).

Lee, D. Y. 2001. *Genres, registers, text types, domains and styles: Clarifying the concepts and nevigating a path through the BNC jungle*.

Leech, G. & Svartvik, J. 1975. *A Communicative Grammar of English.* Pearson Education Limited. London.

Leech, G. 2004. *Meaning and the English verb*. Routledge.

Leeman, D. 2008. Prépositions du français : état des lieux. *Langue française*, 157(1): 5-19.

Lefer, M. A. & Cartoni, B. 2011. Prefixes in contrast: Towards a meaning-based contrastive methodology for lexical morphology. *Languages in contrast*, (1): 87-105.

Legendre, G. & Theory, L. 1990. French impersonal constructions. *Natural Language & Linguistic Theory*, (8): 81-128.

Le Querler, N. 2006. Complémentation et actance : compléments en par/compléments en de. *Complétude, cognition, construction linguistique, Dermas C. (éd.).* Presses Sorbonne Nouvelle.

Le Roux, B. & Rouanet, H. 2010. *Multiple correspondence analysis; Quantitative Applications in the Social Sciences*, (163). SAGE publications, CA. Thousand Oaks.

Litosseliti, L. (ed.). 2018. *Research methods in linguistics*. Bloomsbury Publishing.

Lope Blanch, J. M. 1983. *Ensayos sobre el español de México*. Universidad Nacional Autónoma de México. México.

Lorenzo, E. 1989. Relación interpersonal y expresión impersonal. In *Comunicación y lenguaje juvenil,* 217-239. Instituto Alicantino de Cultura Juan Gil-Albert.

Love, R. et al. 2017. The Spoken BNC2014: designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics*, 22(3) : 319-344.

Luján, M. 1999. Expresión y omisión del pronombre personal. Bosque, I., Demonte, V. (Org.) *Gramática descriptiva de la lengua española.* Madrid. Espasa-Calpe.

Lyons, J. 1977. *Semantics II*. Cambridge Universsity press.

Lyons, J. 1981. Language and speech. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 295(1077): 215-222.

Malamud, S. A. 2012. Impersonal indexicals: one, you, man, and du. *The Journal of Comparative Germanic Linguistics*, 15(1): 1-48.

Malamud, S. A. 2013. (In)definiteness-driven typology of arbitrary items. *Lingua*, 126(1): 1-31.

Malchukov, A. L. 2006. Transitivity parameters and transitivity alternations. *Studies in Language Companion*; *Case, valency and transitivity*, (77): 329-357. Kulikov L., Malchukov A. & de Swart P. (eds). John Benjamins Publishing Company.

Malchukov, A. & Ogawa, A. 2011. Towards a typology of impersonal constructions. *Impersonal constructions: A cross-linguistic perspective*, 19-56. John Benjamins Publishing Company.

Malchukov, A. & Siewierska, A. 2011. *Impersonal constructions: A cross-linguistic perspective*, 124. John Benjamins Publishing Company.

Maldonado, R. 1992. *Middle voice: the case of Spanish se.* Doctoral dissertation. University of California. San Diego. Department of Linguistics.

Maldonado, R. 1999. *A media voz. Problemas conceptuales del clítico 'se'*. México. Universidad Nacional Autónoma de México.

Manole, V. 2014. De l'oral à l'écrit: quelques remarques sur les transcriptions des séances parlementaires. *Studia Universitatis Babes-Bolyai-Philologia.* 59(4): 127-141.

Marín, R. 2004. *Entre ser y estar*. Madrid. Arco libros.

Márquez, P. P. 2017. Pasivas con se e impersonales con se en español actual. *Revista Internacional de Lingüística Iberoamericana*, 207-236.

Martín Zorraquino, M. A. 1979. *Las construcciones pronominales en español*. Madrid. Gredos.

Martínez Caro, E. 2002. Impersonalisation and Reference in English and Spanish: Evidence from newspaper discourse. *Mystification of Agency and Degree of Implication in News Reports,* 71-88. J. Marín Arrese (ed.). Universidad Complutense de Madrid.

Martínez, I. A. 2001. Impersonality in the research article as revealed by analysis of the transitivity structure. *English for Specific Purposes*, 20(3): 227-247.

Matthews, R. 1993. *Papers on semantics and grammar*. Frankfurt am Main. Peter Lang.

McEnery A., Xiao, R. Z. & Tono, Y. 2006. *Corpus-based Language Studies: An advanced resource book*. Routledge. London & New York.

McEnery, T., Love, R., & Brezina, V. 2017. Introduction: Compiling and analysing the Spoken British National Corpus 2014. *International Journal of Corpus Linguistics*, *22*(3), 311-318.

McIntyre, A. 2005. The semantic and syntactic decomposition of get: An interaction between verb meaning and particle placement. *Journal of semantics*, 22(4): 401-438.

Medina, P. G. 2009. Semantic and Pragmatic constraints on the English get-passive. *Deconstructing constructions. Studies in Language Companions,* (107): 271-294. John Benjamins Publishing Company.

Melis, L. 2003. *La préposition en français*. Editions Ophrys.

Mendikoetxea, A. 1999. Construcciones con se: medias, pasivas e impersonales. I. Bosque & V. Demonte (eds), *Gramática descriptiva de la lengua española,* (2): 1631-1722. Madrid. Espasa Calpe.

Mendikoetxea, A. 2008. Clitic impersonal constructions in Romance: syntactic features and semantic interpretation. *Transactions of the Philological Society,* 106(2): 290-336.

Merk, G. 1988. Agentiel et non-agentiel en français. *Revue de linguistique romane,* 52(205-206): 143-150.

Merle, J.-M. 2001. *Etude du conditionnel français et de ses traductions en anglais.* Editions OPHRYS.

Merle, J.-M. 2001. Le conditionnel : temps ou mode ? *Linguistique contrastive et traduction ; Etude du conditionnel français et de ses traductions en anglais*, 10-23. Ophrys Editions.

Molina Redondo, J. A. 1997. *Usos de "se" : cuestiones sintácticas y léxicas*. Madrid. SGEL.

Moltmann, F. 2010. Generalizing Detached Self-Reference and the Semantics of Generic *one. Mind and Language*, 25(4): 440-473.

Monge, F. 1954. Las frases pronominales de sentido impersonal en español. *Institución Fernando el Católico de la Diputación provincial*, (9).

Moreau, M.-L. 1977. Français oral et français écrit, *Le français moderne,* 45(3).

Müller, S. & Ørsnes, B. 2013. Passive in Danish, English, and German. *Proceedings of the 20th International Conference on Head-Driven Phrase Structure Grammar,* 140-160. Freie Universität Berlin.

Næss, Å. 2007. *Prototypical transitivity,* (72). John Benjamins Publishing Company.

Nash, W. 1980. Designs in prose. *English Language Series,* (12). London. Addison-Wesley Longman Ltd.

Neumann, S., & Hansen-Schirra, S. 2012. Corpus methodology and design. In *Cross-Linguistic Corpora for the Study of Translations*. De Gruyter Mouton.

Neumann, S. 2013. *Contrastive register variation: A quantitative approach to the comparison of English and German,* Vol. 251. Walter de Gruyter.

Orts, M. Á. 2018. A Bilingual, bicultural approach to detachment and appraisal in the Law: Tracing impersonality and interaction in English and Spanish legal op-eds. *International Journal for the Semiotics of Law/Revue internationale de Sémiotique juridique*, 31(4): 805-828.

Pedersen, J. 2013. The Spanish impersonal *se*-construction. Constructional motivation for case-marking – quantitative evidence. *Deixis and Pronouns in Romance Languages*. *Studies in language companion*, (136): 109-123. John Benjamins Publishing Company.

Pence, R.W. & Emery, D. W. 1963. *A grammar of Present-Day English*, The Macmillan Campany, New York; Collier-Macmillan Ltd, London.

Pérennec, M.-H. 2002. *Le verbe en action: grammaire contrastive des temps verbaux (français, allemand, anglais, espagnol)*. UGA Éditions (ELLUG).

Posio, P. 2011. Spanish subject pronoun usage and verb semantics revisited: First and second person singular subject pronouns and focusing of attention in spoken Peninsular Spanish. *Journal of Pragmatics*, 43(3): 777-798.

Posio, P. 2016. You and we: Impersonal second person singular and other referential devices in Spanish sociolinguistic interviews. *Journal of Pragmatics*, (99): 1-16.

Posio, P. 2017. Entre lo impersonal y lo individual: Estrategias de impersonalización individualizadoras en el español y portugués europeos. *Spanish in Context*, 14(2): 209-229.

Poutsma, H. 1926. *A Grammar of Late Modern English*, Part II, Section II. P. Noordhoff. Groningen.

Pozas Loyo, J. 2010. *The Development of the Indefinite Article in Medieval and Golden-Age Spanish*. Doctoral dissertation. University of London.

Quesada, J. D. 1997. Obituary: Adios to passive in Spanish. *La linguistique*, 41-62.

Quesada, J. D. 1998. Transitivity, voice, and the middle: Spanish se revisited. *Romanische Forschungen*, 110(1): 1-36.

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. 1985. *A comprehensive grammar of the English language.*

Rabadán, R. 2006. Modality and modal verbs in contrast: Mapping out a translation (ally) relevant approach English-Spanish. *Languages in contrast*, 6(2): 261-306.

Raineri, S. 2010. *Analyse contrastive français-anglais du passif dans une perspective construc-tionnelle : Sens et fonction de BE Ven, ETRE Vé, GET Ven et SE FAIRE Ver*. Thèse. Université de la Sorbonne nouvelle - Paris III.

Ramsay, A. 1996. Aspect and Aktionsart: fighting or cooperating? *The 16th International Conference on Computational Linguistics*. *Coling* (2).

Rasson, M. 2017. *Interpretaciones y funciones del pronombre indefinido uno Análisis de tres variedades geográficas (Chile, España y México) y de tres géneros discursivos (conversaciones, foro digital y redacciones).* Université Catholique de Louvain, Belgique.

Rauch, J. L. 2018. *An analysis of the developing SE passive construction with a POR-Phrase.* Doctoral dissertation. Colorado State University. Fort Collins. Colorado.

Real Academia Española. 1973. *Esbozo de una nueva gramática de la lengua española*. Madrid. Espasa Calpe.

Reese, S. et al 2010. Wikicorpus: A Word-Sense Disambiguated Multilingual Wikipedia Corpus. *Proceedings of the 7th Language Resources and Evaluation Conference* (LREC2010). La Valleta. Malta.

Reid, T. B. 1956. Linguistics, Structuralism and Philology. *Archivum Linguisticum* 8(1): 28–37.

Ridruejo, E. 1981. "Uno" en construcciones genéricas. *Revista de Filología Española*, 61, (1/4): 65-83.

Riegel, M., Pellat, J.-C., & Rioul, R. 1994. *Grammaire méthodique du français.*

Rojo, G. 1990. Relaciones entre temporalidad y aspecto en el verbo español. *Tiempo y aspecto en español,* 17-43. Madrid. Cátedra, 1990.

Rousseau, A. 1998. *La transitivité*. Presses universitaires du Septentrion.

Rousseau, A. 2000. Formation et statut du passif. Comparaison typologique entre langues romanes et langues germaniques. L. SchØsler (ed) *Le Passif. Actes du colloque international. Institut d'Etudes Romanes. Université de Copenhague du 5 au 7 mars 1998. Etudes Romanes,* (45)*:* 117-133 Copenhagen. Museum Tusculanum Press.

Rubin, A. 1978. *A Theoretical Taxonomy of the Differences between Oral and Written Language*. Report #3731. Cambridge, MA. Bolt, Beranek and Newman.

Rühlemann, C. 2007. Lexical grammar: the GET-passive as a case in point. *ICAME Journal*, 31(1): 111-127.

Ruwet, N. 1972. *Théorie syntaxique et syntaxe du français*. Paris. Editions du Seuil.

Salkoff, M. 1999. *A French-English Grammar. A contrastive grammar on translation principles*. John Benjamins: Amsterdam & Philadelphia.

Salton, G. 1989. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Reading. Massachusetts. Addison-Wesley Publishing Company.

Safir, K. 2004. *The syntax of anaphora*. Oxford University Press on Demand.

Sansò, A. 2006. 'Agent defocusing' revisited. Passive and impersonal constructions in some European Languages. Abraham, W. & Leisiö, L. (eds), *Passivization and Typology. Form and function,* 232-273. Amsterdam-Philadelphia. John Benjamins Publishing Company.

Sasse, H. J. 1991. Aspect and Aktionsart: a reconciliation. *Belgian journal of linguistics*, 6(1): 31-45.

Schapira, C. 2006. On pronom indéfini. Corblin F. et al. (eds), *Indéfini et prédication*, 507-518. Presses de l'Université Paris-Sorbonne, Paris.

Sharp, R. 2005. A unified treatment of Spanish se. *Anaphora Processing: Linguistic, Cognitive and Computational Modelling*, 263(113).

Shibatani, M. J. 1985. Passives and related constructions: A prototype analysis. *Language,* 61(4): 821-848.

Shibatani, M. (Ed.). 1988. Passive and voice, *Typological Studies in Language*, (16). John Benjamins Publishing Company.

Siewierska, A. 1984. *The passive. A comparative linguistic analysis.* London. Croom Helm.

Siewierska, A. 2008. Introduction: Impersonalization from a subject-centred vs. agent-centred perspective. *Transactions of the Philological Society,* 106(2): 115–137.

Siewierska, A. 2008. Ways of impersonalizing: Pronominal vs. verbal strategies. *Current Trends in Contrastive Linguistics. Studies in Functional and Structural Linguistics*, 60.

Siewierska, A. 2010. From third plural to passive. Incipient, emergent and established passives. *Diachronica*, 27(1): 73-109. John Benjamins Publishing Company.

Siewierska, A. & Papastathi, M. J. 2011. *Towards a typology of third person plural impersonals*. 49(3): 575-610.

Stein, G. 1979. *Studies in the function of the passive*, 249. Gunter Narr Verlag. Tübingen. Germany.

Steiner, E. 2004. *Translated texts: Properties, Variants, Evaluations*. (SABEST 4). Frankfurt a.M.: Lang.

Steiner, E. (2017). Contrastive studies of cohesion and their impact on our knowledge of translation (English-German). In *Discourse Analysis in Translation Studies*. John Benjamins.

Stubbs, M. 1980. *Language and literacy : the sociolinguistics of reading and writing*. London/Boston. Routledge & Kegan Paul.

Suárez, E. C. 1996. ¿Es la interpretación arbitraria realmente arbitraria? *Revista Española de lingüística*, 26(2): 359-378.

Sussex, R. 1982. A note on the get-passive construction. *Australian Journal of Linguistics*, (2): 83-95.

Svartvik, J. 1966. *On Voice in the English verb.* Mouton & co., The Hague.

Svartvik, J. & Olof Sager. 1996. *Engelsk universitetsgrammatik.* Stockholm; Göteborg. Inbunden bok. Liber.

Swales, J. 1990. *Genre analysis: English in academic and research settings.* Cambridge. UK. Cambridge University Press.

Tallarida, R. J. & Murray, R. B. 1987. Chi-square test. *Manual of pharmacologic calculations,*140-142. Springer. New York.

Teich, E., Degaetano-Ortieb, S., Fankhauser, P., Kermes, H., & Lapshinova-Koltunski, E. 2016. The linguistic construal of disciplinarity: A data-mining approach using register features. *Journal of the Association for Information Science and Technology*, *67*(7): 1668-1678.

Thompson, D., Ferreira, F. & Scheepers, C. 2018. One step at a time: Representational overlap between active voice, be-passive, and get-passive forms in English. *Journal of cognition*, 1(1).

Thomson, A. J. & Martinet, A 1986. *A practical English grammar*. Oxford University Press.

Toscano Mateus, H. 1953. El español en el Ecuador. Consejo Superior de Investigaciones Científicas, Patronato Menéndez y Pelayo, Instituto Miguel de Cervantes. *Revista de Filología Española*, (61).

Toyota, J. 2008. Get-passive. *Diachronic Change in the English Passive*, 148-184. Palgrave Macmillan. London.

Tremblay, A. 2006. On the second language acquisition of Spanish reflexive passives and reflexive impersonals by French-and English-speaking adults. *Second Language Research*, 22(1): 30-63.

Tsunoda, T. 1985. Remarks on transitivity. *Journal of linguistics*, 385-396.

Ure, J. 1982. *Introduction: approaches to the study of register range.* Walter de Gruyter, Berlin/New York.

Van Hoof, H. 1989 Traduire L'Anglais. Théorie et Pratique. Duculot: Paris & Louvain-la-Neuve.

Van Olmen, D., & Breed, A. 2018. Human impersonal pronouns in West Germanic: A questionnaire-based comparative study of Afrikaans, Dutch and English. *Studies in*

*Language,* 42(4): 798-846. International Journal sponsored by the Foundation "Foundations of Language".

Van Valin, R. J. 2006. *Some universals of verb semantics*, 155-178.

Vázquez Rozas, V. 2006. Gustar-type verbs. *Functional approaches to Spanish syntax*, 80-114. Springer. NY.

Vázquez Rozas, V., Barcala, M., Domínguez Noya, E., Fernández Sanmartín, A., Rojo, G. & Santalla, P. 2020. Codificación y anotación del habla en un contexto bilingüe: el corpus ESLORA de español de Galicia". Ángel Gallego, Francesc Roca (eds.), *Dialectología digital del español*. Anexo 80 de *Verba*. Santiago de Compostela: Universidade de Santiago de Compostela. Servizo de Publicacións e Intercambio Científico, 189-224.

Vendler, Z. 1967. Linguistics in Philosophy. Ithaca. Cornell University Press. New York.

Vet, C. 1993. Conditions d'emploi et interpretation des temps futurs du français. *Verbum*, 1993/4: 71–84.

Viberg, Å. 2010. Swedish Impersonal Constructions from a Cross-linguistic Perspective. An Exploratory Corpus-based Study. *Orientalia Suecana* LIX: 122–158.

Vikner, C. 1985. L'aspect comme modificateur du mode d'action: à propos de la construction être+ participe passé. *Langue française*, (67): 95-113.

Vinay, J. P. & Darbelnet, J. 1995. *Comparative stylistics of French and English: A methodology for translation*, (11). John Benjamins Publishing Company.

Warburton, I. 1975. The passive in English and Greek. *Foundations of Language*, 13(4): 563-578.

Weber, E. 2014. La traducción de la voz pasiva francesa al español: ¿cuestión de lengua o cuestión de traducción?. *Mutatis Mutandis: Revista Latinoamericana de Traducción*, 7(2): 368-385.

Weiner, E. J. & Labov, W. 1983. Constraints on the Agentless Passive. J*ournal of Linguistics*, 19(1): 29-58. Cambridge University Press.

Werner, E. 1980. *Die Verbalperiphrase im Mittelfranzösischen. Eine semantisch-syntaktische Analyse.* Frankfurt: Lang.Wilmet, M. 1991. L'aspect en français: essai de synthèse. *Journal of French language studies*, 1(2): 209-222.

Wilmet, M. 1998. *Grammaire critique du français.* Louvain-la-Neuve. Hachette-Duculot. Paris.

Wonder, J. P. 1990. Las formas reflexivas y la voz pasiva. *Hispania*, (73): 401-411.

Xiao, R. J., McEnery, T., & Qian, Y. 2006. Passive constructions in English and Chinese: A corpus-based contrastive study. *Languages in contrast*, 6(1): 109-149.

Xiao, R. J. 2009. *Multidimensional analysis and the study of world Englishes*, 28(4): 421-450.

Yao, K. 2016. Referencias y equivalencias del pronombre on francés en el español. *Revista de Lenguas Modernas*, (25): 115-130.

Yus, F. 2001. *Ciberpragmática. El uso del Lenguaje en Internet*. Ariel Lingüística. Barcelona.

Zanettin, F. 2011. Translation and corpus design. *SYNAPS – A Journal of Professional Communication*, (26). Università di Perugia.

Zhang, Y. 2017. Transitivity analysis of Hillary Clinton's and Donald Trump's first television debate. *International Journal of Applied Linguistics and English Literature*, *6*(7): 65-72.

Zhu, L. 2017. On the origin of the term tertium comparationis. *Language & History, 60*(1): 35-52).

Zribi-Hertz, A. 1982. La construction" Se-Moyen" du français et son statut dans le triangle moyen-passif-réfléchi. *Lingvisticae Investigationes*, 6(2): 345-401.

**Corpus references**

British National Corpus online. Available online at http://www.natcorp.ox.ac.uk/

Contrastive Corpus of Questions and Answers. 2009. Compiled by Hendrik De Smet. Department of Linguistics, University of Leuven.

Corpus International Ecologique de la Langue Française. Available online at http://www.ciel-f.org/corpus.html (Last accessed 16/05/21)

Corpus de LAngue Parlée en Interaction. Available online at http://clapi.ish-lyon.cnrs.fr/ (Last accessed 16/05/21)

Corpus para el estudio del español oral <http://eslora.usc.es>, versión 2.0 de septiembre de 2020, ISSN: 2444-1430.

Davies, Mark. 2015. The Wikipedia Corpus. Available online at https://www.english-corpora.org/wiki/. (Last accessed 16/05/21)

Wikipédia-Corpus (FR). 2009 : Corpus developed by Université de Toulouse. Available via http://redac.univtlse2.fr/corpus/wikipedia.html (Last accessed 16/05/21)

# Appendices

## Appendix 1: FSGraph created in Unitex for data extraction

### FSGraphs created to extract French data

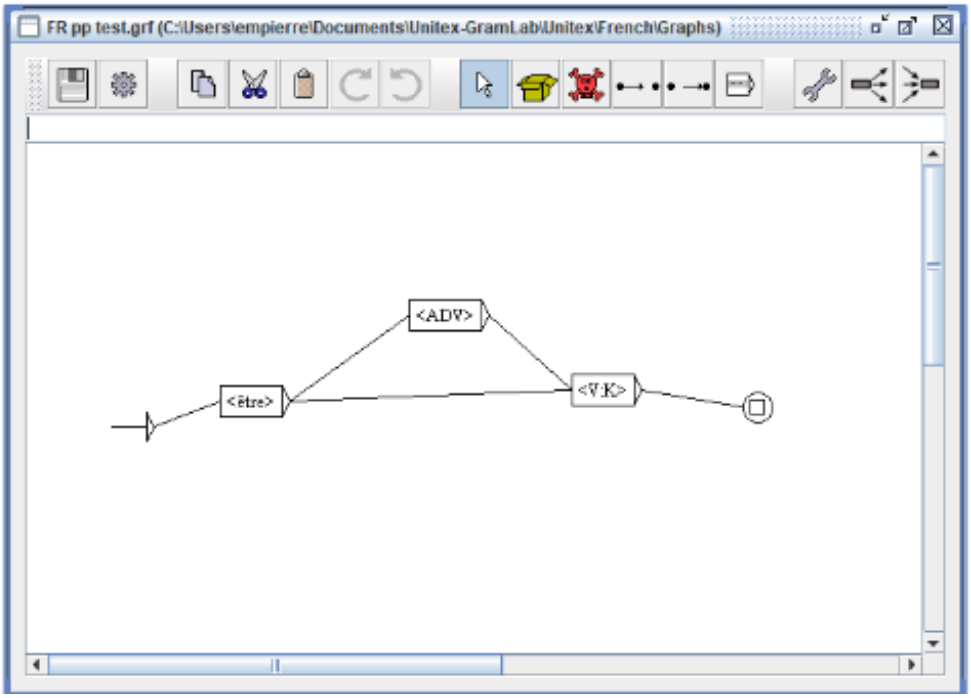



*Figure 37. Graph used to extract the periphrastic passive in French*

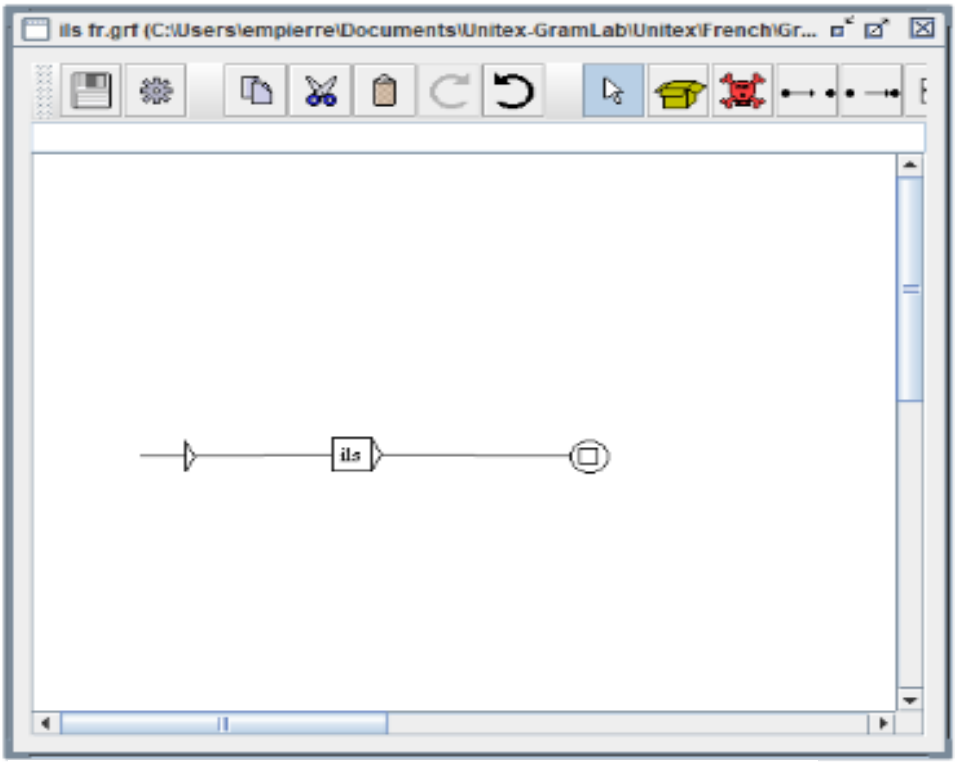



*Figure 36. Graph used to extract the non-anaphoric third person plural in French*

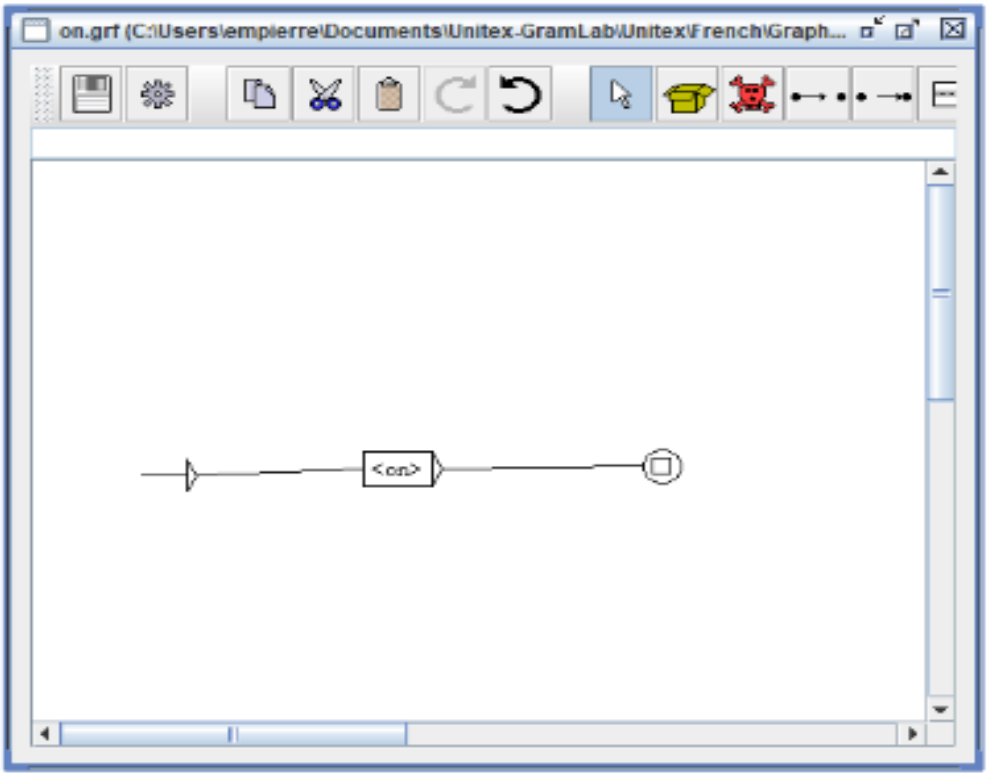



*Figure 38. Graph used to extract the man-impersonal on*

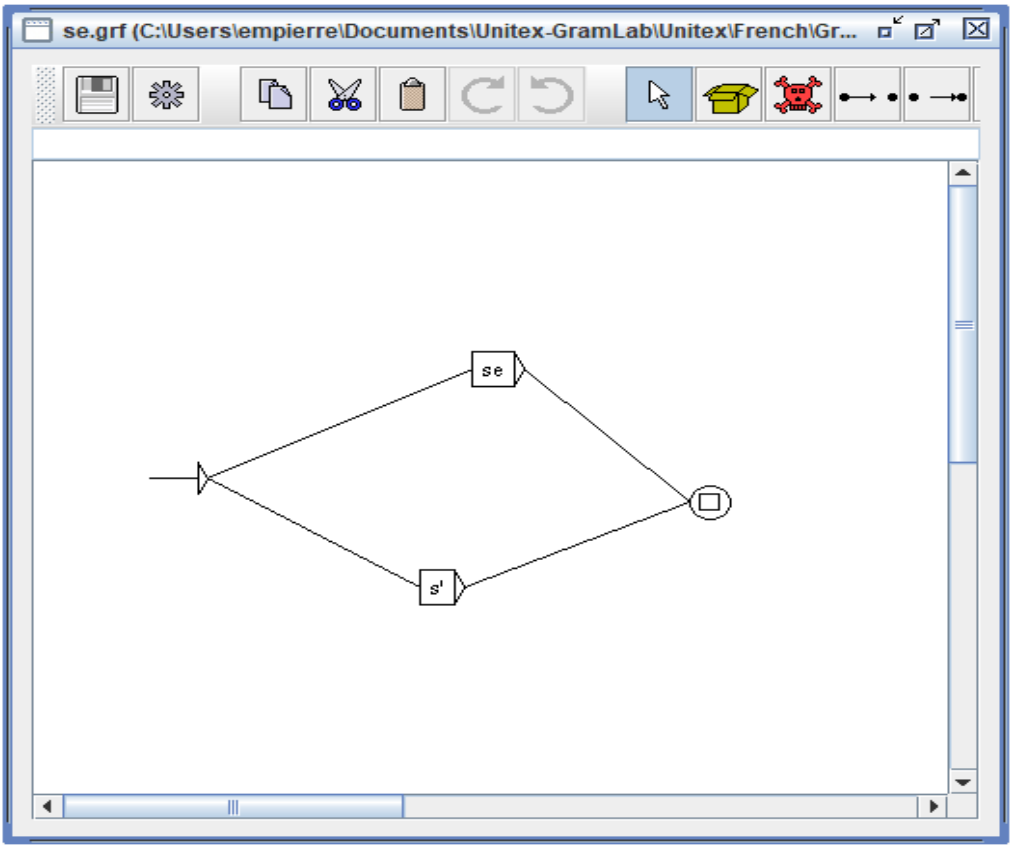



*Figure 39. Graph used to extract the se-construction in French*

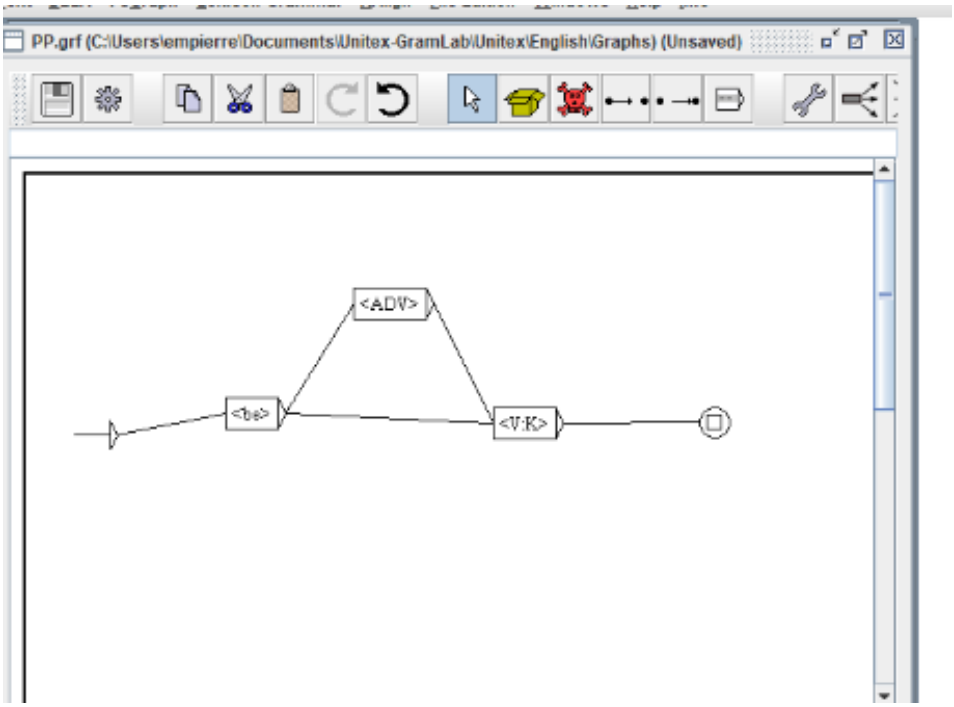



*Figure 41. Graph used to extract the periphrastic passive in English*

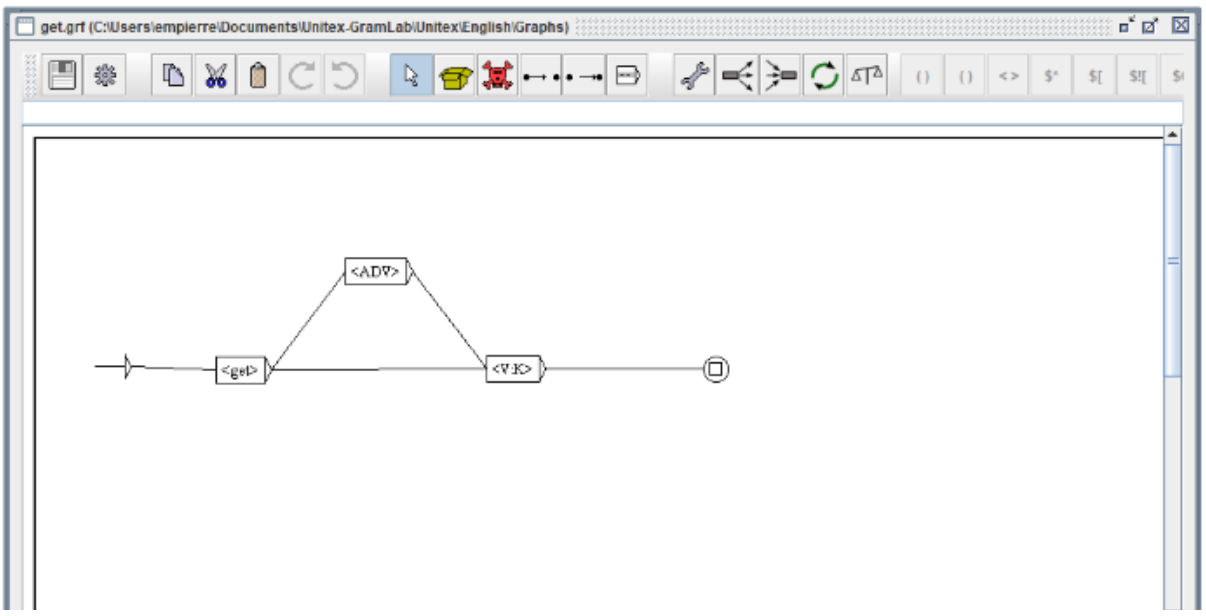



*Figure 40. Graph used to extract the get-passive*

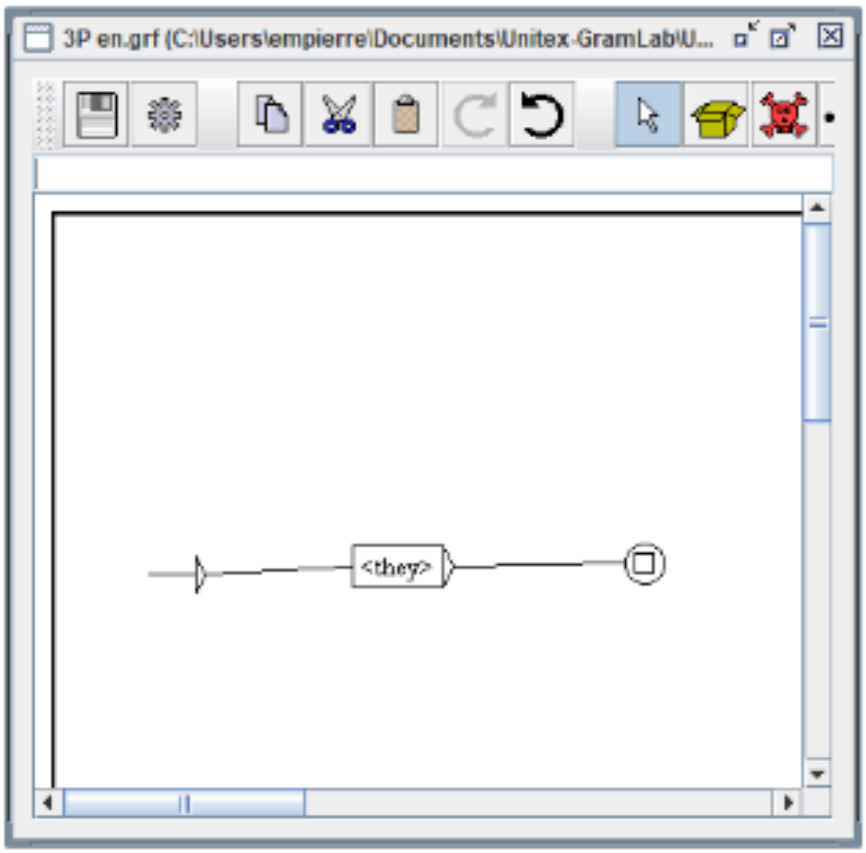



*Figure 42. Graph used to extract the non-anaphoric third person plural in English*

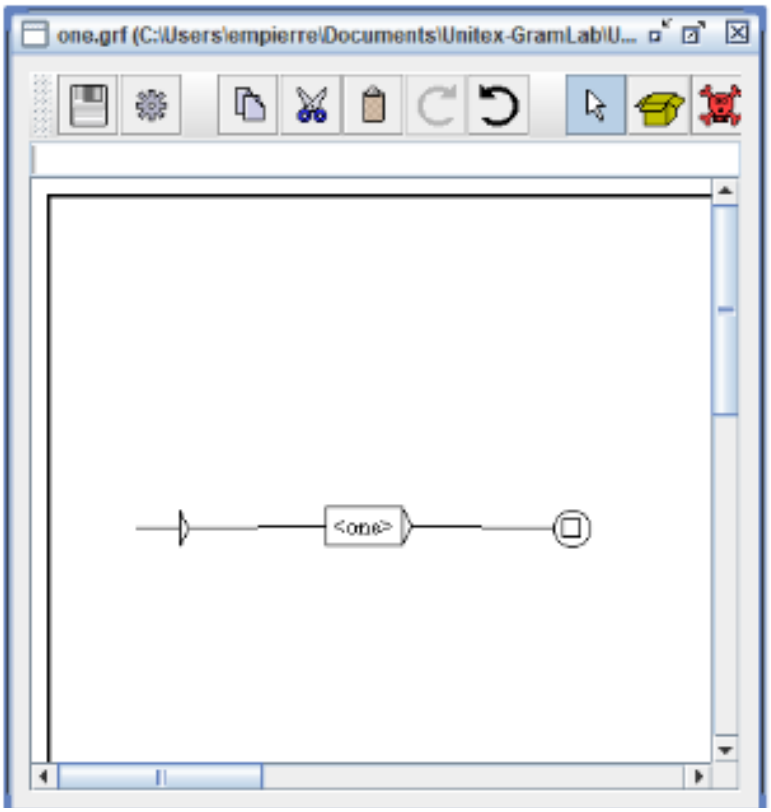



*Figure 43. Graph used to extract the numeral-based one*

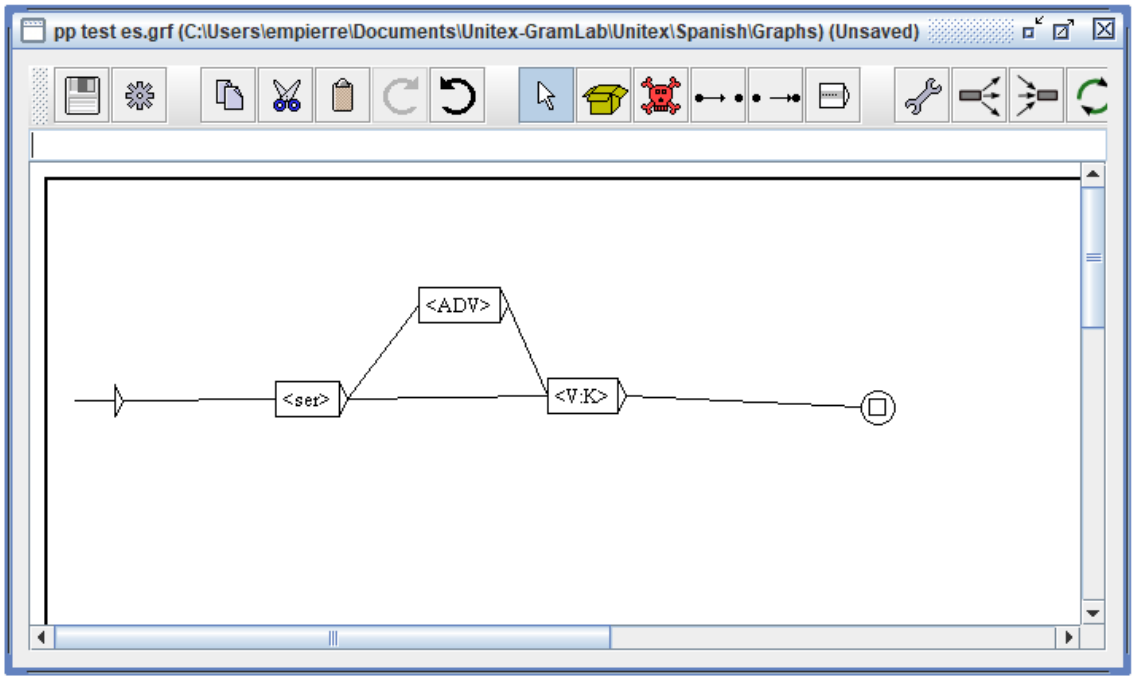



*Figure 44. Graph used to extract the periphrastic passive in Spanish*

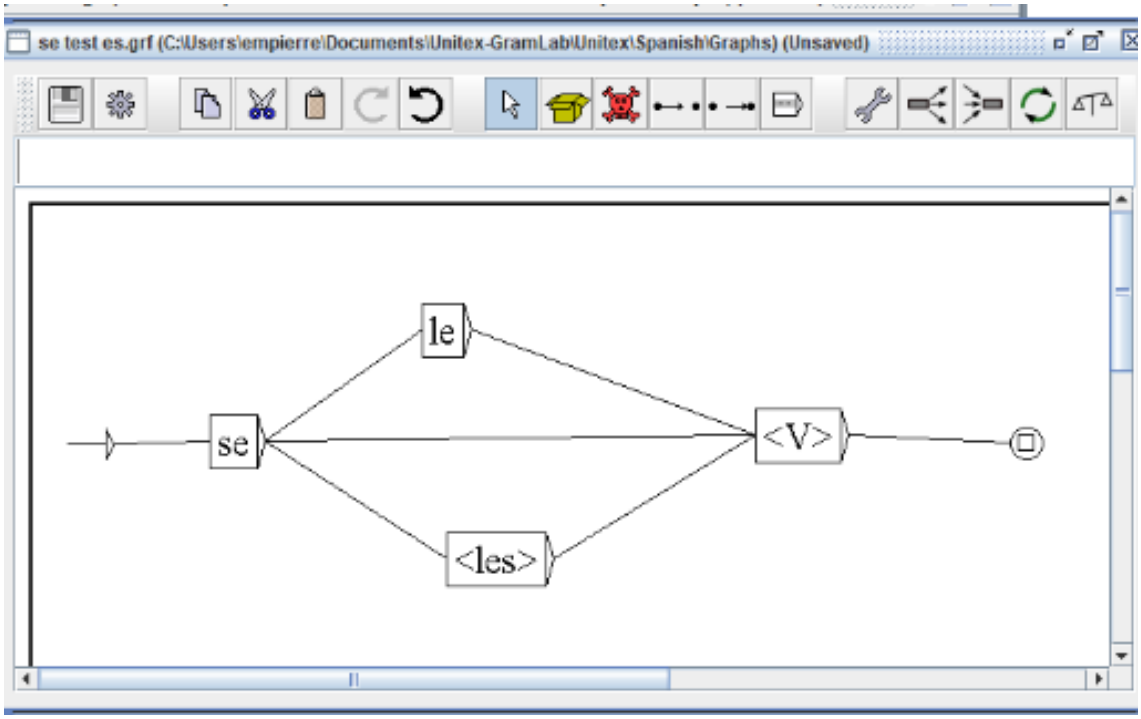



*Figure 45. Graph used to extract the se-construction in Spanish*

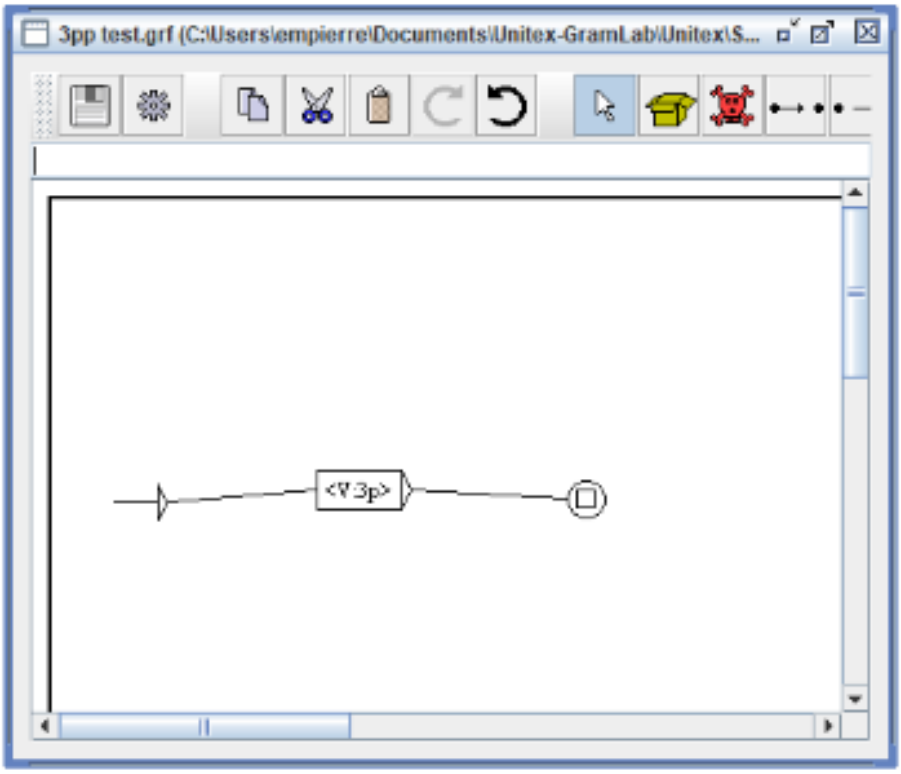



*Figure 46. Graph used to extract the non-anaphoric third person plural in Spanish*

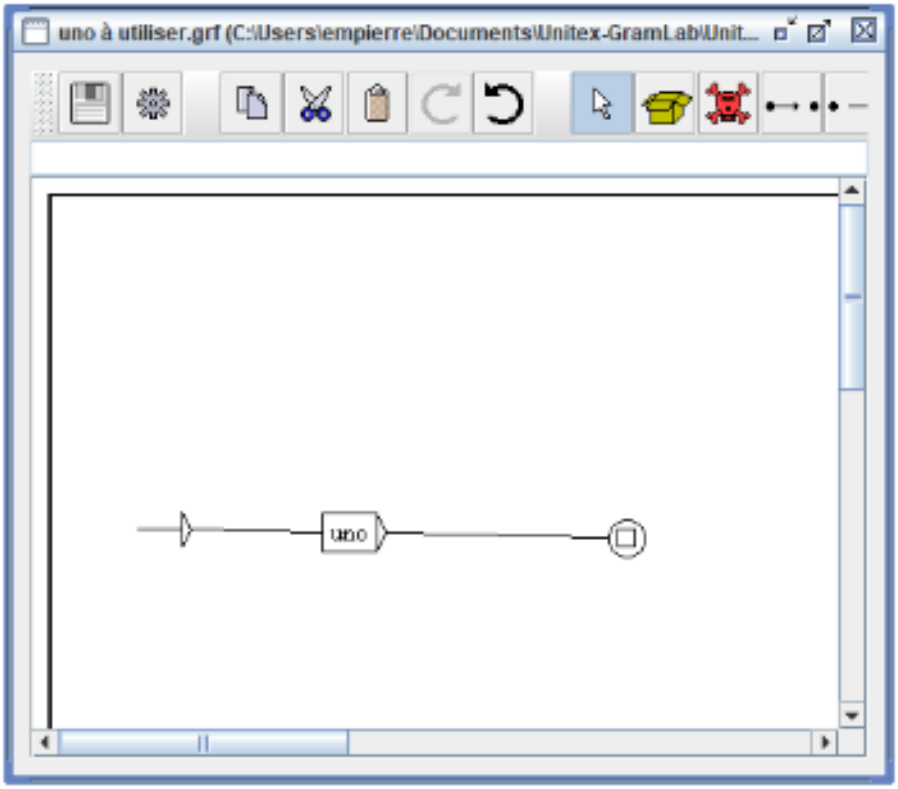



*Figure 47. Graph used to extract the numeral-based uno*

Appendix 2: Code  used for the language, the passive-like structure and the mode/register

| Language | Code |
|---|---|
| Spanish | 1 |
| French | 2 |
| English | 3 |

*Table 38. Code used for each language*

| Object of study | Code |
|---|---|
| Periphrastic passive | 1 |
| Non-anaphoric third person plural | 2 |
| Construction based on the reflexive pronoun *se* | 3 |
| Man-impersonal *on* | 4 |
| *Get*-passive | 5 |
| Numeral-based indefinite pronoun *uno* | 6 |
| Numeral-based indefinite pronoun *one* | 7 |

*Table 39. Code used for each passive-like structure*

| Mode and Register | Corresponding corpus | Code |
|---|---|---|
| **ORAL INFORMAL** | BNC (English) | 1 |
| | Ciel (French) | |
| | ESLORA (Spanish) | |
| **ORAL FORMAL** | Europarl | 2 |
| **WRITTEN FORMAL** | Wikipedia | 3 |
| **WRITTEN INFORMAL** | Yahoo Q&A | 4 |

*Table 40. Code used for each register/mode*

Appendix 3: Example of script used to obtain kappa score in Rstudio

```
interannot_emeline <-function()
{
 # S'assurer d'installer la librarie avec install.packages("irr")
 library("irr")

 # Creation of a log file
 # CHANGER
 logFile = "~/Doctorat/Interrater annotation/res_log3PP.txt"
 sink(logFile)

 # Chemin du fichier à  changer
 file = "~/Doctorat/Interrater annotation/interrater3PP.csv"
 data = read.csv(file, header = TRUE, sep = ";")
 print (names(data))


 cat("################## IRR: Tense  ###############\n")
 tense_data = data[, c(1,2)]

 res_tense = kappa2(tense_data)

 cat("InterIRR for multiclass comparison : ", res_tense$value, "\n")
 print (res_tense)

 cat("################## IRR: Mode  ###############\n")
 mode_data = data[, c(3,4)]

 res_mode = kappa2(mode_data)

 cat("InterIRR for multiclass comparison : ", res_mode$value, "\n")
 print (res_mode)

 cat("################## IRR: Aspect  ###############\n")
 aspect_data = data[, c(5,6)]

 res_aspect = kappa2(aspect_data)

 cat("InterIRR for multiclass comparison : ", res_aspect$value, "\n")
 print (res_aspect)


 cat("################## IRR:  Type of process  ###############\n")
 typeofprocess_data = data[, c(7,8)]

 res_typeofprocess= kappa2(typeofprocess_data)

 cat("InterIRR for multiclass comparison : ", res_typeofprocess$value, "\n")
 print (res_typeofprocess)

 cat("################## IRR:  Participant  ###############\n")
 participant_data = data[, c(9,10)]
```

```
res_participant = kappa2(participant_data)

cat("InterIRR for multiclass comparison : ", res_participant$value, "\n")
print (res_participant)


cat("################## IRR: Kinesis  ###############\n")
kinesis_data = data[, c(11,12)]

res_kinesis = kappa2(kinesis_data)

cat("InterIRR for multiclass comparison : ", res_kinesis$value, "\n")
print (res_kinesis)

cat("################## IRR: Punctuality  ###############\n")
punctuality_data = data[, c(13,14)]

res_punctuality = kappa2(punctuality_data)

cat("InterIRR for multiclass comparison : ", res_punctuality$value, "\n")
print (res_punctuality)


cat("################## IRR: Volitionality  ###############\n")
volitionality_data = data[, c(15,16)]

res_volitionality = kappa2(volitionality_data)

cat("InterIRR for multiclass comparison : ", res_volitionality$value, "\n")
print (res_volitionality)


cat("################## IRR: Affirmation  ###############\n")
affirmation_data = data[, c(17,18)]

res_affirmation = kappa2(affirmation_data)

cat("InterIRR for multiclass comparison : ", res_affirmation$value, "\n")
print (res_affirmation)


cat("################## IRR: Mode TR  ###############\n")
modetr_data = data[, c(19,20)]

res_modetr = kappa2(modetr_data)

cat("InterIRR for multiclass comparison : ", res_modetr$value, "\n")
print (res_modetr)


cat("################## IRR: Agency  ###############\n")
agency_data = data[, c(21,22)]

res_agency = kappa2(agency_data)
```

```
  cat("InterIRR for multiclass comparison : ", res_agency$value, "\n")
  print (res_agency)


  cat("################## IRR: Affectedness  ###############\n")
  affectedness_data = data[, c(23,24)]

  res_affectedness = kappa2(affectedness_data)

  cat("InterIRR for multiclass comparison : ", res_affectedness$value, "\n")
  print (res_affectedness)


  cat("################## IRR: Individuation  ###############\n")
  individuation_data = data[, c(25,26)]

  res_individuation = kappa2(individuation_data)

  cat("InterIRR for multiclass comparison : ", res_individuation$value, "\n")
  print (res_individuation)


  cat("################## IRR: Presence agent complement  ###############\n")
  agentcmplt_data = data[, c(27,28)]

  res_agentcmplt = kappa2(agentcmplt_data)

  cat("InterIRR for multiclass comparison : ", res_agentcmplt$value, "\n")
  print (res_agentcmplt)


  cat("################## IRR:  Patient   ###############\n")
  patient_data = data[, c(29,30)]

  res_patient = kappa2(patient_data)

  cat("InterIRR for multiclass comparison : ", res_patient$value, "\n")
  print (res_patient)


  cat("################## IRR: Prepositional phrase  ###############\n")
  prepphrase_data = data[, c(31,32)]

  res_prepphrase = kappa2(prepphrase_data)

  cat("InterIRR for multiclass comparison : ", res_prepphrase$value, "\n")
  print (res_prepphrase)


  sink()

}
```

Appendix 4: Interrater agreement results

| | SE1 | % agree | PP | % agree | ON | % agree | 3PP | % agree |
|---|---|---|---|---|---|---|---|---|
| Aspect | 1 | 100 | 0.572 | 91.1 | 1 | 100 | 0.87 | 94.8 |
| type process | 0.819 | 84.2 | 0.822 | 84 | 0.864 | 87,9 | 0.84 | 86.6 |
| participant | 0.368 | 76.8 | 0.069 | 89.3 | (-)0.0784 | 39,4 | 0.77 | 91.5 |
| kinesis | 0.562 | 80 | 0.452 | 80.4 | 0.06 | 48,5 | 0.59 | 82.9 |
| punctuality | 0.052 | 72.6 | 0.371 | 72 | 0 | 84,8 | 0.15 | 79.3 |
| volitionality | 0.48 | 82.1 | 0.56 | 89.8 | 0.355 | 75.8 | 0.36 | 92.6 |
| affirmation | 0.883 | 98.9 | 0.916 | 98.7 | 0.872 | 97 | 1 | 100 |
| (ir)realis | 0.642 | 94.7 | 0.828 | 92 | 0.471 | 72.7 | 0.81 | 95.1 |
| affectedness | 0.536 | 88.4 | 0.08 | 92 | 0.68 | 81.8 | 0.56 | 76.8 |
| individuation | 0.63 | 92.6 | 0.51 | 76.9 | 0.75 | 84.8 | 0.66 | 76.8 |
| agent | NAN | 98.9 | 0.91 | 98.2 | NAN | 100 | NaN | 100 |
| patient | 0.594 | 88.4 | 0.91 | 96 | 0.58 | 78.8 | 0.79 | 85.4 |
| agent | 0.495 | 97.9 | 0.58 | 87.6 | NAN | 100 | NaN | 100 |
| prep phrase | 0.70 | 82.1 | 0.51 | 67.6 | 0.459 | 87.9 | 0.68 | 91.5 |
| personref | 0.627 | 87.4 | n.a | n.a | 0.59 | 87.9 | n.a | n.a |
| adverb | 0.585 | 86.3 | 0.66 | 90.7 | 0.59 | 81.8 | n.a | n.a |
| spatiotemp | 0.248 | 68.4 | 0.33 | 70.2 | 0.52 | 81.8 | n.a | n.a |
| another pls | 0.289 | 62.1 | 0.53 | 76 | 0.47 | 72.7 | 0.09 | 48.8 |
| type of cnt | 0.419 | 83.2 | 0.51 | 90.2 | 0.72 | 90.6 | n.a | n.a |

*Table 41. Results interrater agreement (kappa score and % of agreement)*

Appendix 5: MCA graphs excluding the man-impersonal *on*

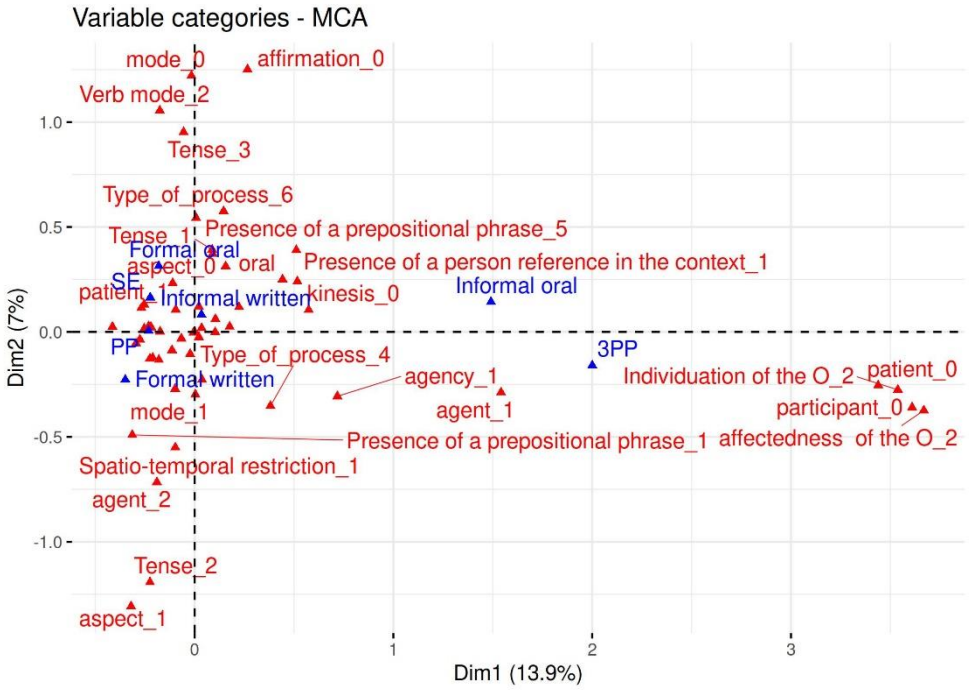



*Figure 48. Distribution of parameters in French without on*

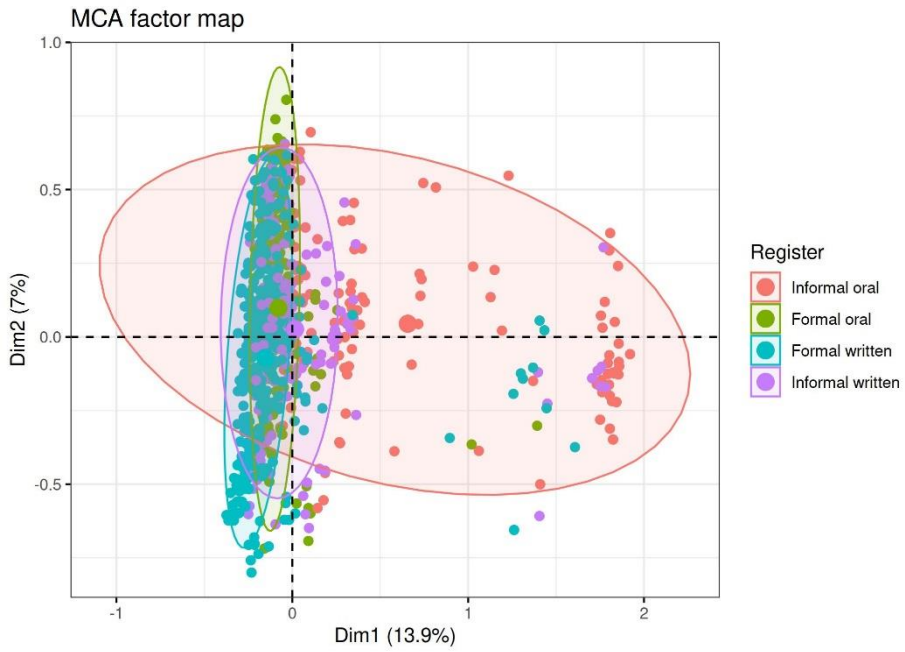



*Figure 49. Factor map for mode and register in French without on*

# Appendix 6: Chi-square tests used in Chapter 10

### Periphrastic passives

Tense : $\chi^2 = 6.554$. $p < .05$ (Spanish vs French vs English)
Aspect: $\chi^2 = 17.5569$. $p < .05$ (Spanish and French), $\chi^2 = 2.3504$. $p > .05$ (Spanish vs English)
Verb mode: $\chi^2 = 6.2053$. $p < .05$ French vs English
Defocused agent: $\chi^2 = 9.2763$. $p < .05$ (Spanish vs English)
Inanimate patient: $\chi^2 = 19.8$. $p < .05$ (French vs English)
Spatiotemporal restriction: $\chi^2 = 2.9463$. $p > .05$ (French vs English), $\chi^2 = 6.9014$. $p < .05$ (Spanish vs English)

### Non-anaphoric third person plurals

Tense : $\chi^2 = 4.1926$ $p < .05$ (French versus English), $\chi^2 = 3.1489$ $p > .05$ (Spanish vs English)
Aspect : $\chi^2 = 7.7597$ $p < .05$
Verb mode: $\chi^2 = 4.8548$ $p < .05$ (English vs French vs Spanish)
Kinesis: $\chi^2 = 7.4721$ $p < .05$ (French vs English), $\chi^2 = 9.5697$ $p < .05$ (French vs Spanish)
Patient : $\chi^2 = 4.5835$ $p < .05$ (English vs French)
Non individuation : $\chi^2 = 4.6985$ $p < .05$ (English vs French)
Prepositional phrases: $\chi^2 = 6.0549$ $p < .05$ (Spanish vs English); $\chi^2 = 1.6661$ $p > .05$ (Spanish vs French)

### Se-constructions in Spanish and in French

Present tense : $\chi^2 = 13.6222$. $p < .05$
Imperfective : $\chi^2 = 11.5458$. $p < .05$
Kinesis : not significant $\chi^2 = 2.8936$. $p > .05$
Realis: $\chi^2 = 6.3043$. $p < .05$
Numb participant: $\chi^2 = 5.2244$. $p < .05$
Individuation: not sign: $\chi^2 = 2.4246$ $p > .05$
Prepositional phrase: $\chi^2 = 28.9942$. $p < .05$
Spatiotemporal restriction: $\chi^2 = 4.2074$. $p < .05$

### Spanish periphrastic passive and Spanish *se*-construction

Present tense : $\chi^2 = 30.6833$ $p < .05$
Imperfective: $\chi^2 = 42.6327$ $p < .05$
Verb mode : $\chi^2 = 20.2923$ $p < .05$
Kinesis: $\chi^2 = 8.8908$. $p < .05$
Two-participant clause: ($\chi^2 = 38.2965$. $p < .05$
Prepositional phrase : ($\chi^2 = 55.3048$. $p < .05$
Other passive-like structure : ($\chi^2 = 154.3812$. $p < .05$

### Man-impersonal *on* vs French periphrastic passive

Tense: $\chi^2 = 9.8087$, $p < .05$
Imperfective : $\chi^2 = 40.8553$, $p < .05$
Indicative: $\chi^2 = 19.3652$, $p < .05$
Affirmation: $s\chi^2 = 13.7764$, $p < .05$
Spatiotemporal restriction: $\chi^2 = 9.2093$, $p < .05$

### Man-impersonal *on* vs French non-anaphoric third person plural

Tense: $\chi^2 = 1.2031$, $p > .05$
Realis: $\chi^2 = 3.8592$, $p < .05$
Participant: $\chi^2 = 2.2634$, $p > .05$
Patient: $\chi^2 = 10.0739$, $p < .05$