

The development of probabilistic grammars in spoken English as a Foreign Language

Dissertation presented in partial fulfillment of the requirements for the degree of Docteur en langues, lettres et traductologie from the Université catholique de Louvain and Doctor in Linguistics from the KU Leuven

By Tanguy Dubois

Supervisors	Prof. dr. Magali Paquot Prof. dr. Benedikt Szmrecsanyi
UCLouvain affiliation	Faculté de philosophie, arts et lettres Institut Langage et Communication Centre for English Corpus Linguistics
KU Leuven affiliation	Faculty of Arts Department of Linguistics Quantitative Lexicology and Variational Linguistics
Examination committee	Prof. dr. Jason Grafmiller Prof. dr. Stefanie Wulff Prof. dr. Gaëtanelle Gilquin Dr. Caroline Gentens

Louvain-la-Neuve and Leuven, 2023

Table of Contents

Acknowledgements	vii
1 Introduction.....	1
2 Alternations and probabilistic grammars	7
2.1 Alternation phenomena	7
2.1.1 Probabilistic grammars.....	8
2.1.1.1 Three assumptions	8
2.1.1.2 Explaining probabilistic grammars	10
2.1.2 Previous research	14
2.1.2.1 Learners of English as a Foreign Language	16
2.1.2.2 Limitations of previous studies.....	20
2.2 Present study	24
3 The acquisition of probabilistic grammars.....	27
3.1 Probabilistic constraints as construction features.....	27
3.1.1 Acquiring construction features.....	29
3.1.1.1 Insights from cues on statistical learning	31
3.2 Constraints without learnable features	33
3.2.1 Processing-related constraints.....	36
3.3 Synopsis.....	37
3.4 Generating hypotheses.....	38
4 Methodology.....	43
4.1 Corpus-based studies.....	45
4.1.1 The corpus	45
4.1.1.1 Proficiency level in the corpus.....	46
4.1.1.2 The language production of the examiner.....	47
4.1.1.3 Register effects in the corpus	48
4.1.2 Delimiting the variable contexts	49
4.1.2.1 The future marker alternation.....	49

4.1.2.2	The genitive alternation.....	50
4.1.2.3	Native speaker interchangeability.....	50
4.1.3	Data collection & annotation.....	51
4.1.4	Statistical analysis.....	52
4.2	Supplementary rating task experiment.....	54
4.2.1	The limitations of corpus-based studies	54
4.2.2	Experiment design.....	56
4.3	The relation between corpus data and experimental data	60
5	The future marker alternation.....	61
5.1	Introduction	61
5.1.1	Previous research	62
5.1.2	Present study	64
5.2	Data and Methods	66
5.2.1	Data collection.....	66
5.2.2	Data annotation	73
5.2.2.1	Sentence type.....	73
5.2.2.2	Negation.....	75
5.2.2.3	Clause type.....	77
5.2.2.4	Persistence.....	81
5.2.2.5	Temporal proximity.....	84
5.2.2.6	Temporal adverb.....	88
5.2.2.7	Verb type	90
5.2.2.8	Person and animacy.....	94
5.2.2.9	Clause length.....	97
5.2.3	Classifying the constraints	99
5.2.3.1	Ranking the feature-based constraints	102
5.2.4	General overview	103
5.2.5	Statistical analysis.....	104
5.3	Results.....	108
5.3.1	Random effects.....	108
5.3.2	Fixed effects.....	111
5.3.2.1	Main effects	111
5.3.2.2	Interaction effects.....	113
5.3.3	Supplementary analyses	122

5.3.3.1	Model without Mexican learners	123
5.3.3.2	Mother tongue model	125
5.4	Discussion	127
5.4.1	Summary of the findings.....	127
5.4.1.1	Native speakers	127
5.4.1.2	Differences between native speakers and EFL learners.....	129
5.4.2	How does the set of constraints of learners compare to that of native speakers at different proficiency levels?.....	129
5.4.3	How does proficiency level modulate the probabilistic grammar of learners?	134
5.4.4	How do native speakers and learners at different proficiency levels compare regarding processing-related constraints?.....	137
5.4.5	Synopsis	138
5.4.6	An alternative explanation	140
6	The genitive alternation	145
6.1	Introduction	145
6.1.1	Previous research	146
6.1.2	Present study	148
6.2	Data and Methods.....	150
6.2.1	Delimiting the variable context	150
6.2.2	Data collection.....	153
6.2.2.1	Identifying potential genitives	153
6.2.2.2	Interchangeability check.....	155
6.2.2.3	Data extraction	158
6.2.3	Data annotation	167
6.2.3.1	Constituent length.....	167
6.2.3.2	Possessor discourse status.....	170
6.2.3.3	Persistence	172
6.2.3.4	Possessor thematicity.....	174
6.2.3.5	Type-token ratio	175
6.2.3.6	Possessor animacy.....	177
6.2.3.7	Semantic relation.....	181
6.2.3.8	Possessor definiteness	184
6.2.3.9	Final sibillancy	186
6.2.4	Classifying the constraints.....	188

6.2.4.1	Ranking the feature-based constraints	192
6.2.5	General overview	192
6.2.6	Statistical analysis.....	193
6.3	Results.....	197
6.3.1	Random effects.....	197
6.3.2	Fixed effects.....	201
6.3.2.1	Main effects	201
6.3.2.2	Interaction effects.....	204
6.3.3	Supplementary analyses	208
6.3.3.1	Mother tongue model.....	208
6.3.3.2	MuPDAR.....	211
6.4	Discussion	217
6.4.1	Summary of the findings.....	218
6.4.1.1	Native speakers	218
6.4.1.2	Differences between native speakers and EFL learners	220
6.4.2	How does the set of constraints of learners compare to that of native speakers at different proficiency levels?	222
6.4.3	How does proficiency level modulate the probabilistic grammar of learners?	224
6.4.4	How do native speakers and learners at different proficiency levels compare regarding processing-related constraints?	227
6.4.5	Synopsis	228
7	Rating task experiment.....	231
7.1	Introduction	231
7.2	Data and Methods.....	235
7.2.1	Materials	236
7.2.1.1	Target experimental items.....	236
7.2.1.2	Filler items.....	243
7.2.2	Participants.....	244
7.2.3	Design and procedure	248
7.2.4	Statistical analysis.....	252
7.3	Results.....	256
7.3.1	The animacy model	256
7.3.2	The corpus prediction model.....	259

7.3.3	Individual differences	263
7.4	Discussion	271
7.4.1	How does possessor animacy affect the intuitions of learners at different proficiency levels in a rating task experiment compared to correlations between animacy and genitive variant in spoken EFL?	272
7.4.2	How does the relationship between experimental ratings and corpus data vary by the proficiency level of participants and speakers?	276
7.4.3	How do the ratings of individual participants compare to the aggregated results regarding possessor animacy and the relation between intuitions and language use?	279
7.4.4	Revisiting the genitive alternation	281
8	General discussion	285
8.1	The research questions	285
8.1.1	How does the set of constraints of learners compare to that of native speakers at different proficiency levels?	285
8.1.2	How does proficiency level modulate the probabilistic grammar of learners?	289
8.1.3	How do native speakers and learners at different proficiency levels compare regarding processing-related constraints?	292
8.2	Discrepancies between the alternations	294
8.3	The acquisition account revisited	297
8.3.1	Two modifications	298
8.3.2	The acquisition of constraints in five stages	301
8.4	Pedagogical implications	303
8.5	Methods evaluation	305
9	Conclusion	309
9.1	Research contributions	309
9.2	Limitations and future research	311
	Bibliography	317
	List of Tables	337
	List of Figures	341

Acknowledgements

First of all, I would like to thank my promotors, Magali and Benedikt, for their guidance and support over the course of the past five years. Thank you, Magali, for your very thorough feedback on all my presentations, abstracts and manuscripts; I believe it has made me a better researcher. The readers should know that it is in fact in large part thanks to you that this thesis is not filled with typos (anymore). In spite of your busy schedule, you have always managed to free up a few hours of your time to discuss the project together. Benedikt, I am grateful for all the opportunities you have given me over the years, not only during my time as a doctoral student, but also before that, as a bachelor student and a master student. Given that we are trying to uncover the properties of evasive natural phenomena stemming from the minds of unaware speakers, it was easy to obsess over small methodological details. When I could no longer see the wood for the trees, you always helped me focus on the bigger picture.

I'm also grateful to the other members of the jury, Jason, Steffi, Caroline and Gaëtanelle, who had to read through these 350 pages carefully enough to pinpoint the aspects of the thesis that deserved praise and those that needed further revision. I'm particularly grateful to Jason and Angela for being so welcoming during my stay in Birmingham. Steffi, your challenging yet inspiring questions at various stages of the project have helped me improve the conceptual and theoretical aspects of the thesis tremendously.

This project was a cotutelle between two universities, which means that I have double the number of colleagues to thank for the occasional lunch and for nodding back in the hallway. The following list is probably not exhaustive, and I apologize if you cannot find your name: Rachel, Nate, Laura A., Pauline, Tove, Samantha, Yating, Laurie, Thomas P., Alice, Hung-Hsin, Robbert, Alexandra, Stefano, Alek, Isabeau, Mariana, Karlien, Laura R., Dirk, Chiara, Thomas VH., Yi, Valentijn, Maxime, Virginia, Marlieke, Jorina, Danqing, Yingying, Sumin and Xiaoyu. On the Louvain-la-Neuve side, I was given a warm welcome by Rachel, Nate, Laura A. and Pauline when I started at the CECL. As for Leuven, I'm particularly grateful to Alexandra for the laughs, for helping me with proofreading and for being a much quicker and efficient researcher than me, which allowed me to take inspiration from her

work. Sumin, Xiaoyu and Yingying, I appreciate your friendship, the various snacks you often bring me and the many dinners we shared.

Not only was I blessed with such great promotors, jury members and colleagues, many of whom became friends, I'm incredibly lucky to have support outside of work. I could always count on my close friends, Niels, Robin, Noé, Hugo, David and Virgile to listen to my endless rants about manual data annotation in Excel. I have to credit my parents, Philippe and Eva, for instilling in me the drive to work hard for my goals, and my brother, Thibault, for his useful advice. On the other side of the world, I could count on Zongxing and Yajuan to cheer me on and support me in any way possible. Finally, this thesis would not have been possible without Danqing, my first office mate, who helped me navigate life as a PhD student and who I was lucky enough to marry. Whenever I grew tired of thinking and working on linguistic alternations, I could always count on you to blow wind in my sails and help me reach the finish line.

1 Introduction

Language learning is a challenging and lengthy process during which learners continuously adapt their (implicit) understanding of the target language to gradually approximate the standard they wish to reach. A major part of this process is learning how to produce utterances that are grammatical in the target language, but beyond grammaticality, the language use of competent speakers also sounds natural. Considerations of what sounds natural and what does not implicitly come to the fore when speakers have at their availability multiple ways to express the same message in a given context. On the morphosyntactic level, multiple ways of expressing the same message are often referred to as ‘alternations’, defined as “structurally and/or lexically different ways to say functionally very similar things” (Gries 2017: 7). For example, one could either say *the dog’s tail* or *the tail of the dog*. To become competent speakers, learners must learn to choose the most natural option when faced with such choices in spoken or written language production.

Although the variants of alternations are often interchangeable, previous research discovered that “the frequency of use of one variant or another is fundamentally linked to the linguistic contexts in which they occur whether at the level of phonology, morphology, syntax, semantics, or even pragmatics” (Tagliamonte 2012: 73). In other words, the choice of variant is not random and instead depends on a myriad of constraints that interact in complex ways to determine which morphosyntactic variant is preferred in a given context. These constraints can relate to formal, semantic or contextual elements of the linguistic context on all levels of grammar. Since no constraint determines the choice of variant categorically, this interacting set of constraints forms a so-called probabilistic rather than categorical grammar (Szmrecsanyi, Grafmiller, et al. 2016). With this in mind, learners have to master the constraints that constitute probabilistic grammars in order to identify which morphosyntactic variant is more suitable in a given context.

While certain studies have recently investigated how these probabilistic grammars condition the use of alternating variants for learners of English as a Foreign Language compared to native speakers of English (e.g., Jäschke & Plag 2016; Wulff, Lester & Martinez-Garcia 2014;

Wulff 2016; Wulff, Gries & Lester 2018; Wulff & Gries 2019; Paquot, Grafmiller & Szmrecsanyi 2019; Kinne 2020), no previous research has explicitly investigated the development of probabilistic grammars over the course of language learning. Against this background, the thesis will offer a theory-based account of the acquisition and development of probabilistic grammars for learners of English as a Foreign Language (EFL).

This theory-based account adheres to usage-based approaches to Second Language Acquisition (SLA), which assume that language learning is driven by language experience (Ellis & Wulff 2019: 41), and the same applies to the acquisition of the probabilistic constraints that drive the choice between alternating variants (Szmrecsanyi, Grafmiller, et al. 2016; Paquot, Grafmiller & Szmrecsanyi 2019; Wulff & Gries 2019: 880). In particular, language learning occurs when complex language input is subjected to general cognitive mechanisms that are also responsible for other types of learning (e.g., Ellis & Wulff 2015: 76; Ellis & Ogden 2017: 607; Yi, Lu & Ma 2017: 520; Wulff & Ellis 2018: 37; Ellis & Wulff 2019: 41). Given the importance of language use in usage-based approaches, they provide a suitable framework to describe the learning process behind probabilistic grammars starting from the concrete language input learners receive until the development of abstract grammatical knowledge underlying speakers' sensitivity to the acceptability of the variants in specific contexts of use.

Predictions from this theory-based account will be tested against empirical data from two well-known alternation phenomena in English, namely the genitive alternation and the future marker alternation. The genitive alternation consists of the *s*-genitive and the *of*-genitive, exemplified in (1) and (2), whereas the future marker alternation concerns the future markers *will* and *be going to*, instantiated by (3) and (4).

- (1) I like to see the nature's beauty (TLC:7_IN_9, B2 Unknown)
- (2) We should enjoy the beauty of nature (TLC:2_7_IN_13, B2 Gujarati)
- (3) maybe they'll change one day (TLC:2_SP_1, NS)
- (4) they're not going to change everything (TLC:ME_31, C1 Spanish)

Specifically, I will carry out three separate studies to test the predictions of the theory-based account for the following three research questions:

- RQ 1: How does the set of constraints of learners compare to that of native speakers at different proficiency levels?
- RQ 2: How does proficiency level modulate the probabilistic grammar of learners?

RQ 3: How do native speakers and learners at different proficiency levels compare regarding processing-related constraints?

The first study presents a rigorous corpus-based analysis on the use of the future markers *will* and *be going to* for native speakers and EFL learners using data from the Trinity Lancaster Corpus (TLC, Gablasova, Brezina & McEnery 2019). The TLC is a learner corpus containing transcribed recordings from the GESE, an official spoken language exam between an EFL learner and an examiner, who is a native speaker of British English. The second study is a corpus-based investigation of the genitive alternation using the same data. In the third study, I complement the corpus-based analysis of the genitive alternation with a rating task experiment à la Bresnan and Ford (2010) and more recently Engel et al. (2022, 2023), where native speakers and EFL learners rate the naturalness of the genitive variants in authentic corpus excerpts. Such a mixed-methodology offers complementary perspectives on the genitive alternation (Arppe & Järvikivi 2007: 150), which increases the generalizability of the findings (e.g., Schönefeld 2011: 20–24) and offers a more comprehensive picture of the implicit grammatical knowledge of language learners. To accommodate the multifactorial nature of probabilistic grammars, the data from the three studies is analyzed with logistic and linear regression, “a technique that allows one to examine the relation between one or several predictors and a response variable” (Speelman 2014: 3).

A key aspect of the methodology is the fact that I control for the proficiency level of the learners in both corpus-based investigations and the rating task experiment. Proficiency level is operationalized as the communicative competence of EFL learners according to native speaker standards, as described for example by the Common European Framework of Reference (CEFR) (Council of Europe 2001), which categorizes learners from beginner (A1) to advanced (C2) using a wide range of descriptors for written and spoken communication. This is made possible by the metadata reported in the TLC, which assigns each learner in the corpus a proficiency level on the CEFR based on the GESE exam they pass. By entering proficiency level as a learner variable during statistical analysis, one can track the development of learners’ probabilistic grammars across different stages of language learning and test the predictions of the theory-based acquisition account.

Moreover, including proficiency level in the analysis can reveal how general learning mechanisms influence learners at specific stages of language learning (Weiss, Gerfen & Mitchel 2010; Song 2018). Previous research that used multifactorial statistical methods to research alternation

phenomena in EFL have largely neglected these effects and instead focused on L1 transfer effects by including learners' mother tongue background as the main learner variable of interest (e.g., Wulff, Lester & Martinez-Garcia 2014; Wulff & Gries 2015). In this respect, the TLC features learners from eleven different mother tongue backgrounds, which helps ensure that the findings attributed to learners' proficiency level are robust across a wide range of typologically diverse L1s.

The remainder of the thesis is structured in eight chapters as follows. In **Chapter 2**, I provide a more in-depth characterization of different types of alternation phenomena and the theoretical assumptions that underlie probabilistic grammars. Against this background, I offer an overview of previous research on alternation phenomena in English for native speakers and speakers of English as a Second Language (ESL) before turning to the few studies on English as a Foreign Language and their shared limitations.

Chapter 3 develops the theory-based account of the acquisition of probabilistic grammars. This account is based on insights from usage-based approaches such as the CREED model (Ellis 2006; see also Ellis & Wulff 2019) and the Competition Model (Bates & MacWhinney 1989; MacWhinney 2001). The account distinguishes between two types of constraints, namely processing-related and feature-based constraints. Processing-related constraints derive from the features of the brain, so they do not have to be learned. Feature-based constraints result from the fact that each variant construction incorporates a number of abstract prototypical features that determines its preferred contexts of use. These features are specific to the construction, so they need to be acquired via general learning mechanisms. With this distinction, I am able to generate hypotheses for the research questions of the thesis at the end of the chapter.

Chapter 4 presents the methodology of the three studies included in the thesis, starting with the design of the corpus-based investigations of the future marker and genitive alternation. This chapter discusses the properties of the TLC (Gablasova, Brezina & McEnery 2019) with a focus on the operationalization of proficiency level, the authenticity of the language production of the examiners and potential register effects. Then, I turn to the delimitation of the alternations and the procedure I adopted to collect and annotate corpus observations. The chapter highlights some of the limitations of corpus-based regression models, such as the fact that they mainly capture language use rather than implicit grammatical knowledge (Ellis 2017: 46; see also Schönefeld 2011: 13; Kinne 2020: 173), which can be addressed by conducting the supplementary rating task experiment in the third study. The chapter ends with a brief description of some the

methodological aspects of the experiment, including its overall design, the recruitment of the participants and the method of statistical analysis.

The first study on the corpus-based investigation of the future marker is covered in **Chapter 5**. In the introduction to the chapter, I give a brief overview of previous studies that have used multifactorial methods to model the choice of future marker. This overview shows that the choice between the future markers has been researched quite extensively for native speakers, but only Schneider (2016) and Bohmann (forthcoming) have extended this line of research to ESL speakers, and no previous study has ever modelled the choice of future marker for EFL learners. The results from the analysis are discussed in light of the research questions of the thesis and the corresponding hypotheses. Surprisingly, learners appear to be sensitive to more probabilistic constraints than native speakers, which suggests that their probabilistic grammar is more complex (Shin 2014).

The same structure is used for the corpus-based analysis of the genitive alternation in **Chapter 6**, which was also published partially as Dubois, Paquot and Szmrecsanyi (2023). The analysis revealed that EFL learners are similar to native speakers regarding most constraints, but they are less sensitive to the animacy of the possessor, which is rather surprising considering that the animacy constraint is otherwise the strongest determinant of the genitive alternation (Rosenbach 2014). The rating task experiment reported in **Chapter 7** is designed to shed more light on this finding, while at the same time providing a complementary perspective on the genitive alternation based on introspective ratings from native speakers and EFL learners at different proficiency levels. Moreover, the rating data collected from the set of carefully selected experimental items is well-suited to explore differences across individual participants. At the end of the chapter, I revisit the acquisition of the genitive probabilistic grammar with the combined insights from both types of data.

The results from the three studies are discussed once again with respect to the hypotheses and research questions of the thesis in **Chapter 8**. This revealed that the acquisition and development of the genitive alternation aligns better with the hypotheses of the theory-based account than the future marker alternation. After interpreting this discrepancy between alternations, I refine the theory-based account with the insights gathered from the three studies. Then, I discuss some of the pedagogical implications of the thesis and briefly evaluate its methodological approach. Finally, **Chapter 9** concludes the thesis with a short overview of its most important theoretical and methodological contributions before ending with some of its limitations and suggestions for future research.

2 Alternations and probabilistic grammars

In this chapter, I expand on the phenomenon under study, namely linguistic alternations on the morphosyntactic level. In the following section, I start by introducing different types of alternation phenomena, including the concept of probabilistic grammars, the theoretical assumptions that underlie them, and the way they operate. Then, I discuss previous research on probabilistic grammars in native-speaker English and World Englishes before turning to learner English. Zooming in on studies on learner English, I identify their shared limitations and explain how the present study addresses them.

2.1 Alternation phenomena

Put simply, alternation phenomena represent “two ways of saying the same thing” (Labov 1972: 271), although alternations can in principle feature more than two morphosyntactic variants (e.g., Szmrecsanyi, Biber, et al. 2016). In more technical terms, alternating constructions are often considered truth-conditionally equivalent and functionally interchangeable (see Gries 2017: 7), so whichever variant speakers choose would lead to a felicitous utterance. This is not the case for all contexts, however. For example, *s*-genitives with *own* cannot be replaced by an *of*-genitive (**the own beauty of nature*) (Hinrichs & Szmrecsanyi 2007: 446), and *will* cannot be used to express future in the past (**we would read the newspaper*) (Fehrer & Corrigan 2015: 207–208).

One can distinguish between three types of alternations as suggested by Troij, Grondelaers and Speelman (2023). First, there are substitution alternations, where one form is substituted for another. In the future marker alternation for example, the speaker can choose to substitute *will* for *be going to*. Permutation alternations involve a change of word order. In the genitive alternation for example, the possessor figures before the possessum in the *s*-genitive (*the dog's tail*), whereas the possessum is before the possessor in the *of*-genitive (*the tail of the dog*). In insertion alternations, speakers can optionally insert a linguistic element. In the optional complementizer alternation for example, speakers can choose to add *that* or to leave it out (*I think that he will play* vs. *I think he will play*).

Importantly, speakers do not pick between competing variants randomly. Instead, the most likely choice of variant in a given context is typically determined by a wide range of factors that together form a probabilistic grammar. These factors can derive from all aspects of grammar, be it formal, semantic or contextual in nature (Szmrecsanyi, Grafmiller, et al. 2016). To illustrate, the choice of future marker is influenced by the semantics of the infinitive verb following the future marker, the grammatical person of its governing subject, which is arguably a formal constraint, as well as contextual features like the temporal proximity of the future event that the future marker refers to. In addition to these three factors, the future marker alternation is influenced by at least six other factors, none of which determine which variant will be picked categorically. Rather, the constraints interact together to determine probabilistically which variant is more acceptable in a given context.

2.1.1 Probabilistic grammars

2.1.1.1 Three assumptions

The study of probabilistic grammars takes center stage in the framework of Probabilistic Grammar (e.g., Bresnan 2007; Bresnan et al. 2007; Bresnan & Ford 2010), situated “at the crossroads of corpus linguistics, variationist linguistics and theoretical linguistics” (Claes 2017: 1). The framework is guided by three main assumptions (e.g., Szmrecsanyi, Grafmiller, et al. 2016; Heller, Szmrecsanyi & Grafmiller 2017: 5; Paquot, Grafmiller & Szmrecsanyi 2019):

Assumption 1: Grammatical variation is sensitive to multiple, sometimes conflicting constraints that apply probabilistically rather than categorically. To study the effects of these constraints on linguistic structure, quantitative research is necessary.

As stated above, the choice of variant in alternation phenomena depends on a number of different factors that operate probabilistically, so it is often not very helpful to investigate the choice between the variants by simply looking at their overall frequency. Instead, one should consider the underlying factors that determine the frequency of the competing variants in different linguistic subcontexts through quantitative research (Torres-Cacoullos & Walker 2009: 329; Heller, Szmrecsanyi & Grafmiller 2017: 5). In order to do so, research on alternation phenomena within Probabilistic Grammar and allied frameworks typically take the form of large-scale corpus studies that use a methodology proper to variationist linguistic

research in the Labovian tradition (Rosenbach 2014: 216; Szmrecsanyi, Grafmiller, et al. 2016; Claes 2017: 1).

Assumption 2: The interacting constraints and the probabilistic grammar they form are the result of language experience. They are refined throughout speakers' lives as they experience language.

Probabilistic Grammar is a usage-based framework, so grammar is seen as "the cognitive organization of one's experience with language" (Bybee 2006: 711). In other words, usage-based approaches posit that language use shapes the implicit grammatical knowledge of learners. Applied to probabilistic grammars, this assumption implies that the set of constraints fluctuates throughout speakers' lives, considering that speakers experience language all the time. In fact, the Probabilistic Grammar framework emphasizes the impact of language use even more than other usage-based approaches by challenging the categorical nature of linguistic knowledge (Heller 2018: 5).

Many usage-based approaches adhere to the Principle of Isomorphism, which states that two structures cannot possibly convey the exact same semantic content (Haiman 1980). To account for interchangeable morphosyntactic variants without abandoning the Principle of Isomorphism, many construction-based approaches posit that the schematic grammatical structures underlying the variants carry semantic aspects of their own. As a result, two different constructions have slightly different semantics by default (Goldberg 1992). In this regard, it has long been claimed that the verbs *whisper* and *drag* cannot occur in the double-object dative (e.g., *drag me a box*) because these verbs are not compatible with the inherent semantics of the double-object dative construction. Yet, experimental research within the Probabilistic Grammar framework found that these verb-specific preferences could be overridden provided other constraints conspire in favor of the double-object dative (Bresnan 2007), showing that considerations of language use trump supposedly categorical mismatches between the semantics of the verb and the variant constructions.

Assumption 3: The interacting constraints form a probabilistic component that must be part of grammatical knowledge, because the likelihood of finding a variant in a given context in a corpus

is reflected by the intuition of speakers regarding the acceptability of the variants in the same context.

Although usage-based approaches assume that language use and linguistic knowledge are intrinsically linked, this does not mean that all patterns of language use necessarily derive from implicit grammatical knowledge. As stated above, alternation phenomena are often studied through large-scale corpus studies using quantitative methods. These quantitative methods reveal how probabilistic constraints are linked to the variants based on their frequency in corpus data. The question then is whether these patterns in the corpus data result from the implicit knowledge of speakers, or whether they are simply patterns of language use (e.g., Klavan & Divjak 2016; Milin et al. 2016; Divjak, Dąbrowska & Arppe 2016). Put simply, corpus-based models show what speakers say, but not necessarily what they know (Ellis 2017: 46; see also Schönefeld 2011: 13; Kinne 2020: 173). As evidence for the cognitive reality of the constraints, Bresnan et al. (2007) and Bresnan and Ford (2010) found that the predictions from corpus-based models, which reflect language use, correlated with the intuitions of speakers regarding the acceptability of the variants in a rating task experiment. The fact that these constraints determine the acceptability of the variants for speakers indicates that learners are implicitly aware of the constraints, so probabilistic grammars must be part of grammatical knowledge. This correlation is not perfect however, partially because statistical models of language use do not deal with distributional patterns in language in the same way as humans (Milin et al. 2016), not to mention models of language use can either be too precise or too vague (Klavan & Divjak 2016).

2.1.1.2 Explaining probabilistic grammars

As stated above, constraints affecting the choice of variant can occur on all levels of language (Szmrecsanyi, Grafmiller, et al. 2016), but what is it that drives such a variety of linguistic features to influence the choice between alternating morphosyntactic variants? I argue that this largely depends on the type of alternation.

Insertion alternations

One possible explanation for why speakers would choose to insert optional linguistic elements is because they try to adhere to Rohdenburg's Complexity Principle, which states that "in the case of more or less explicit grammatical option(s), the more explicit ones will tend to be favored in cognitively complex environments" (Rohdenburg 1996: 151). Provided the

variant with the additional linguistic element is more explicit, this variant should be preferred when the cognitive environment is complex. One explanation for the principle is that the cognitive resources of speakers are limited, so when some aspect of language is highly complex and cognitively demanding, speakers will choose the more explicit alternating variant, which buys the speaker more time to process during language production (see Pijpops et al. 2018). Provided language is processed incrementally, the delimitation of a cognitive environment will ultimately depend on what linguistic material is being processed at the same time as the alternating variant (see Dubois 2023).

In the case of the complementizer alternation (Wulff, Gries & Lester 2018) or the relativizer alternation (Hinrichs, Szmrecsanyi & Bohmann 2015) for example, speakers should prefer to insert *that* when the environment is cognitively complex, as this is the more explicit variant. Unsurprisingly, many of the constraints that affect the insertion of the complementizer and relativizer *that* relate to the complexity of the surrounding context, with more complex contexts preferring to retain the complementizer or relativizer. Conversely, simpler contexts tend to omit the complementizer, i.e., when the subject is structurally simple, the clause juncture is intact, the matrix verb is frequent etc. (Wulff, Lester & Martinez-Garcia 2014; Wulff 2016; Wulff, Gries & Lester 2018). In a similar vein, the relativizer *that/which* is more likely to be retained if the antecedent noun phrase and the relative clause are long, and the syntactic complexity of the variable context is high (Hinrichs, Szmrecsanyi & Bohmann 2015: 823).

Substitution alternations

For substitution alternations like the future marker alternation, it is less clear whether processing effects play a role. Nonetheless, some accounts posit that speakers might prefer *be going to* over *will* when the processing demands of language production are high because the longer *be going to* variant might buy the speaker some time to process the utterance, in accordance with Rohdenburg's Complexity Principle (Rohdenburg 1996: 151; see Szmrecsanyi 2003, 2006). Such an account could be used to explain why *be going to* is preferred in subclauses, which are more complex than main clauses. Deshors and Gries (2014) appeal to the same principle to interpret learners' use of *can* and *may*, which have a number of probabilistic constraints in common with the future markers, such as the effect of negation, the type of sentence, the type of verb, the grammatical person of the subject etc. (Deshors & Gries 2014). However, most accounts claim that the probabilistic constraints result from the incomplete grammaticalization of the future markers (Torres-Cacoullous & Walker 2009:

321). For example, the preference for *be going to* in subclauses is attributed to the fact that *be going to* was first used to refer to the future in subordinated clauses (Poplack & Tagliamonte 2000: 334; Tagliamonte, Durham & Smith 2014: 91; Gardner 2017: 143). Given that *will* was originally used to express volition, another explanation is that *will* is less suitable for subordinate clauses than *be going to* because subordinate clauses are less volitional than main clauses (Poplack & Tagliamonte 2000: 334).

Permutation alternations

Word order is related to many different processing principles, such as the principle of end-weight (Wasow 1997), the short-before-long principle (Behaghel 1909) or Easy First (MacDonald 2013). The overall pattern these principles capture is that shorter or more accessible linguistic material tends to be produced before linguistic material that is more difficult to process. In the case of permutation alternations, speakers tend to choose the variant that allows them to place more accessible linguistic material first. In the words of de Marneffe et al. (2012: 28), “syntactic alternations [...] provide a fruitful domain to investigate the multiple variables that influence production”, because they “allow us to explore the kinds of variables that lead speakers to choose between multiple possible syntactic forms that express roughly the same message” (de Marneffe et al. 2012: 28). Constituents that are animate, short, definite and discourse-given are easier to process than constituents that are inanimate, long, indefinite and discourse-new (MacDonald 2013: 3). Hence, speakers will prefer to use the *s*-genitive when the possessor is more accessible because this variant allows them to produce the possessor right away, compared to the *of*-genitive, where the possessum is produced first. The fact that the choice of syntactic variant depends on the order in which the constituents below the syntactic level are processed implies that the constituents are not processed in advance. By extension, language production is considered an incremental process that “unfolds over time” (Timmers et al. 2013: 10; Philipp et al. 2017), where constituents are produced as they are processed.

Constraints in probabilistic grammars vs. categorical grammars

Whereas recent explanations for permutation alternations relate the positioning of the constituents to principles of processing like Easy First (MacDonald 2013), previous research within Optimality Theory (Aissen 1999; Croft 2003: 84) claimed that the choice of variant tried to satisfy a so-called ‘harmonic alignment’, whereby prominent semantic-pragmatic roles according to typological scales such as for instance the animacy hierarchy

(Silverstein 1976) tend to occupy prominent structural positions in syntax. Applied to the genitive alternation, this assumption suggests that if a referent ranks highly on the typological scales of animacy and definiteness (Comrie 1989), it will be produced as a possessor in an *s*-genitive rather than an *of*-genitive, because the possessor occupies the more structurally prominent specifier position in the *s*-genitive (Rosenbach 2014: 238). However, Romano's study (2016) on the genitive alternation in native and learner language found support for the account based on processing accessibility like the Easy First principle (MacDonald 2013) rather than structural prominence as suggested by Optimality Theory.

Nonetheless, the link between probabilistic constraints and typological scales led to the implication that categorical and probabilistic aspects of grammar derive from the same source (Rosenbach 2017: 2; see also Guy & Boberg 1997; Bresnan & Nikitina 2009; Tamminga, MacKenzie & Embick 2016). Tamminga, MacKenzie and Embick (2016), however, state that variable aspects of grammar like alternations are influenced by certain constraining factors that cannot be at the origin of categorical aspects of grammar, namely 'psychophysiological constraints.' A typical example of a psychophysiological constraint is structural persistence. Structural persistence, also called priming, structural parallelism or repetition in discourse (Hinrichs & Szmrecsanyi 2007: 455), captures the fact that structures that have been used previously tend to be used again: "language users are creatures of habit" (Szmrecsanyi 2006: 1). Speakers repeat the same construction because a construction that was used recently has a higher activation level in memory, and therefore is easier to produce again (e.g., Ellis 2006: 102; Wulff & Ellis 2018: 44; Ellis & Wulff 2019: 44). It is not plausible that persistence affects language use categorically in the form of for example a rule stating that one should always reuse the genitive variant previously used in discourse. This is partially due to the fact that effects of structural persistence can occur long after a variant was produced (Bock & Griffin 2000), so persistence does not necessarily operate on the basis of linguistic elements that are available around the choice context, such as the animacy or definiteness of the possessor. Instead, these psychophysiological factors tend to influence the choice of variant because they derive from the features of working memory or language processing. Structural persistence, for example, derives from processing constructions and syntactic structures in general (Gries & Wulff 2005: 186; Pickering & Ferreira 2008: 4; Romano 2018: 276). Unsurprisingly then, structural persistence is relevant for permutation alternations like the genitive alternation (e.g., Grafmiller 2014; Heller 2018), substitution alternations

like the future marker alternation (Szmrecsanyi 2006: 123) and insertion alternations like the optional relativizer *that/which* (Hinrichs, Szmrecsanyi & Bohmann 2015: 822).

2.1.2 Previous research

At present, alternation phenomena with two or more variant constructions have been studied in various languages, including for example Chinese (Li, Szmrecsanyi & Zhang 2023), Dutch (Bloem, Versloot & Weerman 2017; Grondelaers & Speelman 2007), Russian (Divjak & Arppe 2013), Estonian (Klavan 2020) and Spanish (Mazzola, Cornillie & Rosemeyer 2022). Since the thesis focuses on English in particular, this brief literature review will focus on alternation research within this language. Typically, studies within Probabilistic Grammar and allied frameworks tend to be quantitative in nature. Many instances of the variants are collected from corpus data and then annotated for the probabilistic constraints before being analyzed by means of state-of-the-art multifactorial statistical methods, such as logistic regression (Speelman 2014) and conditional random forests and inference trees (Tagliamonte & Baayen 2012).

First, there are studies comparing the probabilistic grammars across different native varieties of English. For example, Bresnan and Ford (2010) compared how speakers of American English and Australian English choose between the variants of the dative alternation (*I gave him the book* vs. *I gave the book to him*). They found that Australian English speakers use the prepositional dative more often than American speakers, which makes them more susceptible to choose the prepositional dative when the recipient becomes complex. The future marker alternation has also been studied in various varieties of English. Often, these varieties were selected because they differ in their degree of conservativeness, which makes it possible to track the grammaticalization of the forms (e.g., Poplack & Tagliamonte 1999; Tagliamonte, Durham & Smith 2014; Fehringer & Corrigan 2015; Gardner 2017).

English alternation phenomena have also been researched in diachrony, most notably in the study by Wolk et al. (2013), who compared the genitive and dative alternation in Late Modern English. They found that although the direction of the effects does not change, the strength of the effects can fluctuate over time. In particular, the effect of referent animacy has become weaker in English in both alternations. In the genitive alternation, this is due to the fact that *s*-genitives allow for more inanimate possessors instead of animate possessors. As for the dative alternation, double-object datives tend to allow for more inanimate recipients than

before. The effect of end-weight, according to which longer constituents are placed after shorter ones (Behaghel 1909; Wasow 1997; Arnold et al. 2000), is also not stable diachronically.

A few studies have looked into the effect of mode and register. Grafmiller (2014) compared the use of the genitive variants in spoken and written language, focusing on the influence of economization and colloquialization in written language (Hundt & Mair 1999). More recently, Engel et al. (2022) compared the effect of register and mode on British English speakers' choice between the variants of the dative alternation (Engel et al. 2022) and the future marker alternation (Engel & Szmrecsanyi 2023). In general, these studies revealed that probabilistic grammars are modulated by the register and the mode, meaning that the set of constraints varies depending on the situational context of use.

Recently, research on alternation phenomena was also expanded to Outer Circle speakers for the dative alternation (Röthlisberger, Grafmiller & Szmrecsanyi 2017), the genitive alternation (Heller, Szmrecsanyi & Grafmiller 2017; Heller, Bernaisch & Gries 2017) and the particle placement alternation (Grafmiller & Szmrecsanyi 2018; see also Szmrecsanyi, Grafmiller, et al. 2016). This line of research established that different varieties of English share a core probabilistic grammar. However, there are subtle differences between native varieties and Outer Circle varieties as a result of "the complex interaction between the simplification processes due to second language acquisition and the unique cultural, political and sociolinguistic ecologies of the individual regions embedding of the syntactic pattern involved" (Szmrecsanyi, Grafmiller, et al. 2016: 111). In general, it is the effect size rather than the direction of the constraints that are not stable diachronically that also appears to fluctuate across varieties: the effect of animacy is less influential for the genitive alternation and end-weight is also not stable in the dative alternation (see Wolk et al. 2013). On a related note, Rosenbach (2017) investigated whether the constraints applying to the Afrikaans genitive alternation can influence the English genitive alternation for native speakers of Afrikaans with English as an L2. She found that speakers of Afrikaans are more likely to allow for inanimate possessors in English *s*-genitives, because inanimate possessors are more common in the genitive in Afrikaans, which is an example of "indirect transfer" (Rosenbach 2017: 17).

Finally, de Marneffe et al. (2012) investigated which probabilistic constraints affect the choice of dative variant for children learning English. They found that children have a probabilistic grammar that mirrors that of their caregivers, but it is reduced in dimensionality compared to that of

adult speakers. This indicates that there is a continuity between child and adult probabilistic grammars based on language use. Van den Bosch and Bresnan (2015) addressed the same research questions, but modelled the data using an exemplar-based machine learning method, which “offers a computational implementation of example-based or exemplar-based language processing” (Van den Bosch & Bresnan 2015).

2.1.2.1 Learners of English as a Foreign Language

Although many studies have investigated how individual factors influence the choice of variant for EFL learners, or how the variant constructions themselves are first acquired (Jäschke & Plag 2016: 486), only a limited number of studies have adopted an explicitly multifactorial approach that is necessary to study the multiple constraints that drive the choice between alternants simultaneously (Gries & Wulff 2013: 331; Gries 2015; Gries & Deshors 2015: 134; Jäschke & Plag 2016: 487). Again, the dominating approach is to extract many instances of both variants from authentic corpus data, which are then annotated for the constraints that drive the choice of variant according to research on native speakers. Using state-of-the-art multifactorial statistical methods like logistic regression (Speelman 2014), conditional random forests and inference trees (Tagliamonte & Baayen 2012), these studies compare how the constraints influence the choice of variant for native speakers and EFL learners.

A first study of this kind is Gries and Wulff (2013), who modelled the use of the genitive alternation in the written language of Chinese and German EFL learners using logistic regression. They found that Chinese EFL learners do not consider the semantic relation expressed by the possessor and the possessum, contrary to native speakers. German EFL learners, on the other hand, are more sensitive to this constraint than native speakers. Moreover, German learners prefer the *s*-genitive when the possessor refers to a specific entity, whereas possessor specificity does not really influence the choice of variant for both native speakers and the Chinese learners, who are more sensitive to the specificity of the possessum instead. In spite of these differences, they concluded that “intermediate/advanced learners’ production data already reflect target-like processing preferences at an astonishing level of complexity” (Gries & Wulff 2013: 349)

Deshors and Gries (2014) investigated the use of *can* and *may* in the written interlanguage of French EFL learners. They used the Behavioral Profile approach to compare the use *can* and *may* for native speakers and French EFL learners with French *pouvoir*. In a second step, they also analyzed how native speakers’ use of *can* and *may* compares against that of learners with mixed-effects logistic regression. They found that learners

differ from native speakers with respect to certain constraints. In negated clauses for example, learners are more likely to use *can* than native speakers. Because negated clauses are more complex than affirmative clauses, the authors claimed that learners pick the more frequent *can* in such contexts because they are more sensitive to the Complexity Principle (Rohdenburg 1996: 151).

Wulff and Gries (2015) compared the order of prenominal adjectives (*the green big jacket/the big green jacket*) in the written language production of native speakers of British English and German and Chinese EFL learners language using MuPDAR, which stands for Multifactorial Prediction and Deviation Analysis using Regressions, developed by Deshors and Gries (2014 and Gries and Adelman (2014). This method consists of first modelling the choice of variant for native speakers, and then using this model to identify in which contexts learners' choice of variant deviate from native speaker usage. They concluded that "the intermediate-advanced level Chinese and German learners [...] are quite well-aligned with native speakers' [...] preferences" (Wulff & Gries 2015: 144). Whereas German learners have fine-tuned frequency-related predictors, the Chinese learners are more native-like with respect to the segmentation of speech, i.e., the ideal ordering of the syllables. These findings were attributed to the L1s of the learners.

In a series of studies, Wulff et al. investigated the use of the optional complementizer *that* in the interlanguage of Spanish and German learners of English (Wulff, Lester & Martinez-Garcia 2014; Wulff 2016; Wulff, Gries & Lester 2018). Wulff, Lester and Martinez-Garcia (2014) modelled the alternation in the written language production of the learners using logistic regression. They found that learners omit the complementizer less often than native speakers, and that they are more sensitive to processing-related constraints relating to the Complexity Principle (Rohdenburg 1996: 151). They argued that processing factors carry more weight in learner language because learners have to juggle between multiple language systems, which is cognitively more demanding (Kroll & Dussias 2013). At the same time, learners are less sensitive to verb-specific biases for *that*-omission, meaning that they only omit *that* with the verbs that are most commonly associated with complementizer omission. Wulff (2016) expanded on Wulff et al. (2014) by including spoken learner data. Spoken language production is largely similar as for written language, but both native speakers and learners are more likely to produce *that* in writing than in speech. Wulff, Gries and Lester (2018) included an additional predictor in the analysis and they used MuPDAR. In spite of the differences between native speakers and

learners, the overall conclusion was that “German and Spanish intermediate-advanced L2 English learners display behavior that is highly similar to that of native speakers” (Wulff 2016: 132).

Gries and Deshors (2015) also used MuPDAR to explore how native speakers, EFL learners and speakers of World Englishes choose between the dative variants based on written and spoken data. Their results showed that non-native speakers make quite similar choices as native speakers. However, they deviate from native speakers in passive sentences, which the authors attributed to the fact that passive sentences are harder to process. Moreover, learners make less native-like choices in contexts that do not strongly favor either variant. The authors highlighted the fact that learners may overestimate the importance of certain constraints while downplaying that of others.

Jäschke and Plag (2016) also investigated the interlanguage of German EFL learners for the dative alternation. They annotated the experimental stimuli of Bresnan and Ford (2010) for the constraints of the dative alternation and administered the items to German EFL learners. Using linear regression, they found that the German learners are sensitive to fewer dative constraints than the native speakers. A look at the individual differences between learners revealed that the choice of dative variant is more categorical for some learners than for others.

Rautioaho, Deshors and Meriläinen (2018) looked at the progressive vs. non-progressive alternation in spoken British and American English, compared to the language production of ESL speakers (Indian, Singaporean and Nigerian English) and Finnish, French and Polish EFL learners. Using clustering techniques, they found that there is a clear division between native varieties, ESL and EFL varieties for non-progressives, whereas this distinction is less pronounced for progressives. Specifically, learners are more likely to use the progressive with inanimate subjects than native speakers. At the same time, learners use the progressive less often in passive contexts.

Finally, several studies have looked at the particle placement alternation in learner language. Paquot, Grafmiller and Szmrecsanyi (2019) compared the use of variants in spoken British English with Dutch, French, German, Greek, Italian, Spanish and Swedish EFL learners. In general, learners prefer to keep the particle next to the verb rather than placing it after the direct object, because this is supposedly the less marked order that is easier to process. Moreover, learners typically learn the phrasal verbs in decontextualized lists where the verb and the particle are presented together, which might reinforce this order in the mind of learners (see Gilquin 2015). In their data, EFL learners with Germanic L1 backgrounds

were more similar to native speakers than learners from other L1 backgrounds, who appeared to rely on a simpler probabilistic grammar.

Wulff and Gries (2019) analyzed the data from the written and spoken language production of British English native speakers and EFL learners from as many as 17 different L1 backgrounds with MuPDAR. Their results painted a “complex picture in which processing demands, input effects, and L1 typology jointly shaped the degree to which learners’ choices of constructions were nativelike or not” (Wulff & Gries 2019: 28) As in the study of Paquot, Grafmiller and Szmrecsanyi (2019), learners prefer to place the particle right after the verb rather than separating it from the verb. This order requires less processing effort because the speakers do not have to keep the particle in mind while producing the direct object. At the same time, the authors argued that this helps the hearer to determine what verb is produced quicker. Differences between native speakers and EFL learners are more pronounced in spoken language, which they attributed to the fact that processing demands are higher in spoken language. Additionally, they found that learners are less affected by an upcoming prepositional phrase, which normally triggers the split verb order. They related this result to the possibility that learners are less sensitive to linguistic material that is produced further ahead.

Finally, Kinne (2020) focused on the effect of end-weight on particle placement and the position of the prepositional phrase in the verb-prepositional construction (*she takes this fact into account* vs. *she takes into account this fact*) for native speakers of English and German EFL learners. Besides end-weight, the study also considered other probabilistic constraints such as information status and syntactic complexity. Methodologically, the study combined an analysis of corpus data containing written academic language and two experiments, namely a rating task experiment à la Bresnan and Ford (2010) and a sentence completion task where participants had to complete an excerpt with either variant. Echoing previous findings from Wulff and Gries (2019) and Paquot, Grafmiller and Szmrecsanyi (2019), Kinne concluded that learners prefer to pick a safe strategy that follows the canonical word order in English, namely the joined word order for phrasal verbs and the split order for the verb-prepositional construction. An investigation into individual differences showed that for certain learners, the effect of end-weight is almost categorical, whereas for others, end-weight is rather weak, which ties in with Jäschke and Plag's (2016) finding regarding the dative alternation. In general, end-weight is stronger in the corpus data than in the experiments, which the author

claimed is due to the fact that academic writing involves more conscious monitoring than the experiments.

To summarize, most of these studies found that the probabilistic grammar of EFL learners involve an interacting set of constraints that resembles that of native speakers. However, the probabilistic grammar of EFL learners tends to rely on fewer constraints and is in general simpler than that of learners (Jäschke & Plag 2016; Paquot, Grafmiller & Szmrecsanyi 2019). Additionally, many of the differences across alternation phenomena are due to learners being more sensitive to processing effects (Wulff, Lester & Martinez-Garcia 2014; Deshors & Gries 2014; Wulff 2016; Wulff, Gries & Lester 2018; Wulff & Gries 2019; Paquot, Grafmiller & Szmrecsanyi 2019), which in turn could be attributed to the fact that language learners have to juggle multiple language systems at all times (Kroll & Dussias 2013). In addition to processing effects, transfer from the L1 influences the choice of variant (e.g., Wulff & Gries 2015). For example, the Spanish speakers in the studies of Wulff, Lester and Martinez-Garcia (2014), Wulff (2016) and Wulff, Gries and Lester (2018) have an overall preference for retaining the complementizer compared to German learners and native speakers, which is probably due to the fact that the complementizer cannot be removed in Spanish. Moreover, the study of Rosenbach (2017) on transfer from the Afrikaans genitive to the English genitive suggested that L1 influence can also affect how the underlying probabilistic constraints affect learners. Importantly, the few analyses on individual differences between speakers regarding probabilistic grammars showed that certain constraints are more categorical for some speakers than for others (Jäschke & Plag 2016; Kinne 2020). Finally, MuPDAR analyses revealed that learners tend to make more non-native choices in contexts that less strongly favor either variant (Gries & Deshors 2015; Wulff & Gries 2019). Methodologically, most studies relied on written and/or spoken corpus data investigated through multifactorial statistical methods, with the exception of Jäschke and Plag (2016), who administered the rating task experiment of Bresnan and Ford (2010) to German learners of English, and Kinne (2020), who employed a mixed methodology combining corpus and experimental data.

2.1.2.2 Limitations of previous studies

Previous studies on alternation phenomena in EFL learners share a number of important limitations I discuss below.

Language learning and proficiency level are not controlled for

All previously mentioned studies focused on the influence of the learners' L1 without considering another important variable in language learning, namely proficiency level. In the present context, 'proficiency level' is defined as an assessment or measurement of EFL learners' communicative competence in English according to native speaker standards. A typical example of such an assessment is the Common European Framework of Reference (CEFR) (Council of Europe 2001), which classifies learners from beginner (A1) to advanced (C2) using a wide range of descriptors for written and spoken communication. One exception is Wulff and Gries (2019), who included lexical diversity as a proxy for language proficiency in their analysis of the particle placement alternation. However, the learner corpus they used, the International Corpus of Learner English (Granger et al. 2009), consists of texts written as answers to a wide variety of prompts. Given that measures of lexical complexity are often highly sensitive to topic, this measure is probably not a reliable proxy of proficiency.

Most importantly, proficiency level should be taken into account in research on alternations in EFL because it offers an indication of language learning. According to usage-based approaches to SLA, all aspects of language learning are primarily driven by exposure to language input, which is processed via general learning mechanisms that are responsible for other types of learning as well (e.g., Ellis & Wulff 2015: 76; Ellis & Ogden 2017: 607; Yi, Lu & Ma 2017: 520; Wulff & Ellis 2018: 37; Ellis & Wulff 2019: 41). Hence, if a learner reaches a higher proficiency level on proficiency tests, this means that they have developed better communicative competence in English through language learning. This, in turn, implies that more proficient learners have been exposed to more language input, which also drives the acquisition of probabilistic grammars. Therefore, including proficiency level makes it possible to track how EFL learners choose between alternating variants at different stages of language learning.

Additionally, not controlling for learners' proficiency limits the findings from previous research on alternation phenomena in EFL, which mostly attributed any differences with native speakers to influence from learners' L1. Such transfer effects have been found to be stronger for lower-proficiency learners (Morett & MacWhinney 2013: 134, 149; Wang & Xu 2015: 215), so the question arises as to whether some of these findings also apply for the more proficient learners. For example, it could be that only the less proficient Spanish learners of English have a tendency to use the complementizer more often than native speakers (Wulff, Lester & Martinez-Garcia 2014; Wulff 2016; Wulff, Gries & Lester 2018).

More importantly, learners' use of alternating variants is in all likelihood also affected by general learning mechanisms, which should apply to all learners at similar stages of language acquisition regardless of their L1 (Weiss, Gerfen & Mitchel 2010; Song 2018). For one, these mechanisms could lead to overgeneralizations or simplifications (Biewer 2011: 14; Bentz & Winter 2013). For example, learners might prefer the more analytical *of*-genitive because it involves less morphosyntactic complexity than the synthetic *s*-genitive (see e.g., Parodi, Schwartz & Clahsen 2004; Heller 2018: 178). In a similar vein, Deshors and Gries (2014) and Kinne (2020) stated that learners tend to pick the safe, more frequent variant when processing demands become high.

Such learning mechanisms could influence the effect of the individual constraints as well. As I will argue in Section 3.1.1, the mechanism that is most likely responsible for learning the association between a constraint and the choice of variant is associative statistical learning (Romberg & Saffran 2010; Misyak & Christiansen 2012; Gries & Ellis 2015: 229; Wulff & Ellis 2018: 39). Importantly, the success of associative statistical learning for a given constraint depends on a number of properties, such as the frequency, ease of processing and regularity of form of the linguistic elements involved (Romberg & Saffran 2010: 2–3; Yoshimura & MacWhinney 2010: 552; Misyak & Christiansen 2012: 304; O'Shannessy & Meakins 2012: 380; Gries & Ellis 2015). Since there is bound to be some variability among the constraints with regard to these properties, some constraints will in all likelihood be harder to master than others. As a result, the probabilistic grammar of especially lower-proficiency learners might diverge from that of native speakers with regard to the more challenging constraints. Yet, statistical models in previous research may not have detected some of these effects given that different proficiency levels were conflated together. If a model did discover such effects, then one would be inclined to wrongfully generalize the finding to the entire population of learners under study.

Limited number of L1 backgrounds

Most previous studies only featured learners from a very limited number of L1 backgrounds, with some studies only including data from a single L1 background (e.g., Jäschke & Plag 2016; Kinne 2020). The issue with including only one L1 background is that it is not possible to disentangle L1-effects from the effects of general learning mechanisms, which makes it difficult to generalize the results across learners from different L1 backgrounds.

Methodological focus

Much previous research in the field of Learner Corpus Research (LCR) compared the effect of constraints for native speakers and EFL learners on the basis of monofactorial statistical tests (Gries & Wulff 2013: 327–330; Deshors & Gries 2014: 182–184; Gries & Deshors 2015: 133; Wulff & Gries 2015: 131–133; Jäschke & Plag 2016: 487). Such an approach is not very insightful considering that many factors co-determine the preferred choice of variant in specific contexts (see e.g., Gries 2015). Therefore, many previous studies on alternation phenomena had an explicitly methodological goal, which is to promote the use of state-of-the-art multifactorial statistical methods for the investigation of alternations in learner data (Gries & Wulff 2013; Deshors & Gries 2014; Gries & Deshors 2015; Wulff & Gries 2015), including (mixed-effects) logistic regression and MuPDAR. Given this methodological focus, many of the studies did not explicitly formulate theory-based hypotheses regarding learner probabilistic grammars and instead remained largely descriptive.

Reliance on corpus data exclusively

With the exception of Jäschke and Plag (2016) and Kinne (2020), most previous studies collected many instances of the alternating variants from corpus data. These observations were annotated for the relevant probabilistic constraints and then analyzed statistically using mostly mixed-effects logistic regression. The regression model calculated how each constraint is associated with the choice of variant, how strong this association is, and whether this association is statistically significant. The constraints that are statistically significant were said to affect speakers' choice of variant. By extension, they were considered to be part of the implicit grammatical knowledge of speakers as suggested by the findings of Bresnan and Ford (2010).

The issue with this approach is that there are fundamental differences between how statistical models and humans gather knowledge about distributional patterns. For example, such mixed-effects logistic regression models have difficulty dealing with multicollinearity, yet collinearity and redundancy are an integral part of language (Milin et al. 2016: 508). Moreover, such corpus-based models are frequency-based, but the fact that a variant does not often occur in a given context does not mean that learners would find the variant unacceptable or unnatural (Kempen & Harbusch 2005: 344; Arppe & Järviö 2007: 149–150; Divjak, Dąbrowska & Arppe 2016: 26–27). In general, corpus-based models are able to “reflect the characteristics of mental processes and structures yielding usage, even though we do not know the exact form of these mental representations”

(Divjak & Arppe 2013: 230), or in simpler terms, corpus-based models show what speakers say, but not necessarily what they know (Ellis 2017: 46; see also Schönefeld 2011: 13; Kinne 2020: 173). Applied to alternation phenomena, the question arises as to whether the significant constraints in corpus-based regression models are indeed part of the implicit grammatical knowledge of speakers (Klavan & Divjak 2016), or whether they are simply frequent patterns of language use. That is, corpus-based models can capture patterns that are not cognitively relevant because they are more precise than humans at distinguishing between alternating variants. Alternatively, models can also perform less well than humans, in which case the analysis is missing some of the constraints speakers take into account when choosing between the variants (Klavan & Divjak 2016).

2.2 Present study

In view of these gaps, the first goal of the thesis is to offer a theory-based account of how EFL learners acquire the various probabilistic constraints associated with alternation phenomena, based on the insights from usage-based approaches to Second Language Acquisition like the CREED model (Ellis 2006) and the Competition Model (e.g., Bates & MacWhinney 1989; MacWhinney 2001; MacWhinney 2005). In this respect, Van den Bosch and Bresnan (2015) successfully modelled the dative alternation in child language with a memory-based machine learning algorithm, which suggests that exemplar-based theories of SLA can accurately capture alternation phenomena. The hypotheses generated by this theory-based account regarding three research questions will be tested against data from a substitution alternation, namely the future marker alternation, and a permutation alternation, viz. the genitive alternation:

- RQ 1: How does the set of constraints of learners compare to that of native speakers at different proficiency levels?
- RQ 2: How does proficiency level modulate the probabilistic grammar of learners?
- RQ 3: How do native speakers and learners at different proficiency levels compare regarding processing-related constraints?

Methodologically, I will first conduct a quantitative corpus-based analysis for both alternations. Like most previous studies, this involves collecting many corpus observations for both alternation phenomena, which I will then annotate for the relevant probabilistic constraints according to the literature. Then, the annotated datasets for both alternations will be

analyzed through state-of-the-art multifactorial statistical methods, such as mixed-effects logistic regression and MuPDAR. The corpus data I use stems from the Trinity Lancaster Corpus (Gablasova, Brezina & McEnery 2019), a 3.4-million word corpus consisting of transcribed recordings from the GESE exam. The main advantage of this corpus is that it offers a reliable measure of proficiency level, because the 12 levels of the GESE exam are designed to map onto the proficiency levels of the Common European Framework of Reference (Council of Europe 2001). Hence, if a candidate passes the 7th level of the GESE, then this learner is recognized as a B2 learner in the corpus metadata. In this way, the corpus covers language exams from more than a thousand low-intermediate (B1) to advanced (C2) learners of English. Contrary to previous studies, this makes it possible to control for proficiency level during statistical analysis and investigate how the probabilistic grammar of EFL learners develops as they become more proficient. A second advantage of the corpus is that the learners stem from as many as eleven typologically diverse L1 backgrounds, such as Spanish, Italian, Hindi, Chinese, Russian, Gujarati, Tamil, etc. Including this many different L1 backgrounds ensures that findings relating to proficiency level cannot be attributed to transfer effects from a particular L1.

One limitation of previous studies is that they tend to rely on corpus data exclusively, which primarily reveals patterns of language usage rather than speakers' implicit grammatical knowledge per se. To ascertain the cognitive plausibility of corpus-based models, it is necessary to collect behavioral data as well (e.g., Arppe & Baayen 2011; Klavan & Divjak 2016; Divjak, Dąbrowska & Arppe 2016). Therefore, the corpus-based study on the genitive alternation will be complemented by a rating task experiment à la Bresnan & Ford (2010; see also Kinne 2020; Engel et al. 2022; Engel & Szmrecsanyi 2023), where participants have to rate which variant they find most natural in different contexts. Nonetheless, experimental data has the disadvantage that it is typically less valid ecologically than corpus data (Arppe & Järvikivi 2007: 132; Gilquin & Gries 2009; Klavan & Divjak 2016: 361), so a combination of corpus data and experimental data gives the best results in terms of explanatory power and generalizability of the findings (Hoffmann 2006: 169; Arppe & Järvikivi 2007: 151–152; Gilquin & Gries 2009: 9; Schönefeld 2011: 20–24; Divjak & Arppe 2013: 265; Abbuhl, Gass & Mackey 2013: 129). More generally, data triangulation of this kind forms a desideratum both in usage-based cognitive linguistics generally (Klavan & Divjak 2016) and SLA research more specifically (Ellis 2017). To benefit from this mixed methodology to its full extent, the target items used in the experiment will stem from authentic corpus data on the genitive alternation,

allowing for a direct comparison between the results from the corpus-based analysis and the intuitions of the participants in the experiment.

3 The acquisition of probabilistic grammars

In this section, I will provide a theory-based account of the acquisition and development of probabilistic grammars for EFL learners. Using this theory-based account, I will then generate hypotheses for the research questions of the thesis.

3.1 Probabilistic constraints as construction features

A probabilistic grammar is a set of constraints, so learning a probabilistic grammar comes down to learning the relevant constraints and how they interact with each other. To offer an account of the acquisition of probabilistic grammars, one must first determine how they operate. One possibility is that probabilistic constraints are activated when there is competition between variants, i.e., when speakers have to make a choice between variants during language production. In this case, learning a probabilistic grammar is to learn how variants compete.

Inspired by the work of Huang (2022) on Chinese radicals, the second possibility is that there is no direct competition; both constructions happen to be semantically and functionally eligible in a given context. However, each of these constructions has their preferred contexts of use. Probabilistic constraints do not solve the competition between variants; they instead represent a set of prototypical features with which a given construction is associated. For example, the features prototypically associated with *s*-genitives are animate and definite possessors that do not end in a sibilant and that express a prototypical semantic relation with their possessum. By contrast, *of*-genitives occur in contexts with inanimate and indefinite possessors (e.g., Grafmiller 2014; Rosenbach 2014; Shih et al. 2015; Heller, Szmrecsanyi & Grafmiller 2017). Similarly, a prototypical instance of *be going to* is followed by a dynamic verb and figures in an interrogative, negated subclause with a second-person subject etc. *Will* prototypically occurs in affirmative, declarative main clauses with first-person subjects and stative verbs (e.g., Denis & Tagliamonte 2018; Engel & Szmrecsanyi 2023). During language production, speakers assess such features and produce the variant that provides the best fit without comparing the suitability of the variants directly (see Ellis, O'Donnell & Römer 2014;

Bidgood et al. 2020). In this case, learners primarily have to learn the features associated with each variant construction separately.

Such an account is well-suited to deal with the fact that there are often more than two variants to convey the same message, even though research on morphosyntactic alternations typically only consider two. Regarding the genitive variants for example, Szmrecsanyi, Biber, et al. (2016: 25) show that certain genitives like *director of the FBI* can be replaced not only by *FBI's director*, but also by *FBI director* or *director from the FBI*. Likewise, speakers of English have at their availability a few structures to convey futurity besides *will* and *be going to*, namely *shall*, *be to* (Declerck 2010) the present simple and the present progressive (Nesselhauf 2007), which could indirectly compete with each other depending on the context. If probabilistic grammars were activated during language production to solve competition between variants, they would have to be adapted to deal with different combinations of competing variants. If speakers do not compare the variants directly and instead simply produce the construction whose features match the context of use, it is irrelevant how many variants are eligible.

Moreover, this explanation better accounts for probabilistic constraints that are associated with only one variant construction. In the genitive alternation for example, the final sibilancy constraint captures the tendency for speakers to avoid the *s*-genitive when the possessor ends in a sibilant as in (5) (Hinrichs & Szmrecsanyi 2007: 453; Shih et al. 2015).

- (5) I would like to talk about the truth and beauty in Keats' poetry
(TLC:2_IT_3, C1 Italian)

The fact that *s*-genitives avoid such possessors does not necessarily mean that *of*-genitives attract these possessors. Rather, *of*-genitives simply do not distinguish between possessors with or without final sibilant, which implies that speakers do not actively compare both variants for this constraint. Instead, the avoidance of possessors with final sibilants is better imagined as a feature of the *s*-genitive, but not the *of*-genitive.

Importantly, the concept of construction feature does not completely overlap with the concept of probabilistic constraint. A probabilistic constraint encompasses similar linguistic elements that influence the frequency of use of one variant or another in a given context (Tagliamonte 2012: 73), without explicitly alluding to the elements' association with either alternating variant construction in the mental representations of speakers. By contrast, a construction feature refers to those linguistic elements covered by probabilistic constraints that constitute a prototypical

feature of a specific alternating construction. To illustrate, the animacy of the possessor is one of the strongest probabilistic constraints of the genitive alternation (e.g., Rosenbach 2014), meaning that it strongly influences the frequency of the genitive variants. In speakers' mental representations of the alternating constructions, the various animacy levels account for different construction features. For example, animate possessors are a prototypical construction feature of *s*-genitives. Another slightly less prototypical feature of *s*-genitive are collective possessors. Similarly, a prototypical construction feature of *of*-genitives is an inanimate possessor (e.g., Rosenbach 2014; Rosenbach 2017). From this perspective, the concept of construction feature is more fine-grained and learners must master each feature pertaining to a given probabilistic constraint in order to approximate the target probabilistic grammar with respect to said constraint. Moreover, probabilistic constraints can refer to any and all factors that influence the choice of variant, whereas construction features only overlap with the probabilistic constraints that involve linguistic elements in the context surrounding the variant constructions (see Section 3.2).

3.1.1 Acquiring construction features

According to usage-based approaches like the CREED model (Ellis 2006), language learning comes down to learning constructions, defined as form-meaning mappings used in a certain speech community (Ellis & Wulff 2015: 76). A single string of elements can be conceptualized as different constructions at multiple levels of abstraction (Ellis & Wulff 2019: 41–42). For instance, in her study of phrasal verbs, Gilquin (2015) distinguishes between three levels of abstraction. At the most abstract level, a phrasal verb structure consists of the combination of a verb and a particle. At the intermediate level, each phrasal verb structure can be categorized according to whether there is an object and whether this object comes before or after the particle. This would be the level where the variants of the particle placement alternation are distinguished (see Gries 2003). Finally, at the least abstract level, each phrasal verb structure is the combination of a specific verb with a specific particle. Likewise, one can view the *s*-genitive as a construction consisting of a possessor slot followed by the *s*-genitive marker and a possessum slot, whereas the *of*-genitive consists of a possessum slot, followed by the *of*-preposition and the possessor slot. For the future marker alternation, the construction can be conceptualized as a subject slot, *be going to* or *will*, and then the infinitive verb slot, corresponding formally to a clause.

Language learners encounter these constructions in their target language input in the form of exemplars, i.e., as concrete instances of the constructions. In these concrete exemplars, the construction features are encoded by concrete linguistic elements. In the future marker alternation for example, the type of sentence is one of the strongest constraints, with interrogative clauses almost categorically taking *be going to* (Szmrecsanyi 2003; Tagliamonte, Durham & Smith 2014; Denis & Tagliamonte 2018: 424). Hence, a feature of the *be going to* construction is that it occurs in interrogative clauses. On the surface, this construction feature is expressed by either inversion, intonation, or question words. Likewise, possessor animacy as a feature of *s*-genitives is encoded by thousands of different nouns. Importantly, these are the linguistic elements learners have access to in first instance, so they form the primary object of language learning. The question is how learners infer prototypical features of constructions from these concrete linguistic elements in their language input.

According to usage-based approaches to SLA, language learning is driven by language input, which is then exposed to a limited number of general learning mechanisms responsible for other, non-linguistic types of learning (e.g., Ellis & Wulff 2015: 76; Ellis & Ogden 2017: 607; Yi, Lu & Ma 2017: 520; Wulff & Ellis 2018: 37; Ellis & Wulff 2019: 41). The general mechanism of language learning that is responsible for learning the constructions and their features is (associative) statistical learning, which can be defined as “our ability to make use of statistical information in the environment to bootstrap language acquisition” (Rebuschat & Williams 2011a). From the very beginning, language learning is subconsciously driven by statistical regularities in the input (Romberg & Saffran 2010: 1; Misyak & Christiansen 2012: 302; Lammertink et al. 2017: 3473; Ellis 2017: 41; Ellis & Ogden 2017: 607; Wulff & Ellis 2018: 39; Ellis & Wulff 2019: 44). Statistical learning has been shown to contribute to learning relevant sounds, words, multiword units and deeper grammar (Romberg & Saffran 2010; Gullberg et al. 2010; Hay et al. 2011), in artificial languages and real language input (Romberg & Saffran 2010; Hay et al. 2011), at different stages of first and second language acquisition (Gullberg et al. 2010; Yi, Lu & Ma 2017: 526–527) for children and adults alike (see Yi, Lu & Ma 2017: 520; Misyak & Christiansen 2012: 302–303 and the references therein).

If construction features are learned through associative statistical learning, the question then arises as to whether certain features are easier to learn. In this respect, it is useful to look into the similar concept of *cues* from previous research and in particular the Competition Model (Bates & MacWhinney 1989; MacWhinney 2001; MacWhinney 2005).

3.1.1.1 Insights from cues on statistical learning

Cues “map form to functions onto each other (probabilistically) such that forms and functions are cues to functions and forms respectively” (Deshors & Gries 2014: 180). For example, in a clause like *yesterday I walked*, the function of past tense is encoded by two formal cues on the surface, namely the temporal adverb *yesterday*, which is a lexical cue, and the morphological past-tense verbal inflection *-ed* (see Ellis & Sagarra 2010). According to the Competition Model, in this case both cues conspire to indicate past tense for learners, but learners may not rely on both cues to the same extent (Ellis 2017: 56). For this example specifically, learners prefer to rely on lexical cues like the adverb rather than morphological cues in accordance with the Lexical Preference Principle (VanPatten 1996).

Cues can refer to all types of information that can form the object of statistical learning, be it social, visual (Romberg & Saffran 2010: 7; Gullberg et al. 2010: 12), or more linguistic cues, including lexical items, morphological patterns, semantics, pragmatics, word order or stress (MacWhinney 2017: 25; Duguine & Köpke 2018). In other words, cues can be considered as any feature or element in the input that co-occurs with a certain construction, so that there is a cue-outcome contingency that can form the object of (associative) statistical learning mechanisms (e.g., Romberg & Saffran 2010: 1; Misyak & Christiansen 2012: 302; Lammertink et al. 2017: 3474; Ellis 2017: 41; Ellis & Ogden 2017: 607; Wulff & Ellis 2018: 39; Ellis & Wulff 2019: 44).

Previous research revealed that statistical learning of the association between cue and construction is gradual and refined over time (Morett & MacWhinney 2013: 134, 149; see also Wang & Xu 2015: 215). Additionally, there is evidence that learners do not necessarily focus on all cues at the same time. Rather, they investigate each cue one by one, adjust the strength of a cue and then introduce a new one (Ellis & Sagarra 2010; Ellis & Sagarra 2011: 601). Applied to the linguistic elements encoding construction features, this means that learners will gradually approximate the target probabilistic grammar by adjusting how each construction feature is associated with the alternating constructions one by one.

The success of associative statistical learning is determined by a number of factors. Specifically, a cue that is more frequent, more reliable, and adjacent to the relevant construction is easier to learn than a cue that is less frequent, unreliable, and separated from the construction (MacWhinney 2001; Romberg & Saffran 2010: 2–3; Yoshimura & MacWhinney 2010: 552; Misyak & Christiansen 2012: 304; Yi, Lu & Ma 2017: 520). Most important is the reliability of the cue: “a learnable cue is one

where when the cue is there, the outcome is there, and when the cue is not there, neither is the outcome” (Wulff & Ellis 2018: 46). For example, interrogative sentences almost categorically take *be going to* (Szmrecsanyi 2003; Tagliamonte, Durham & Smith 2014; Denis & Tagliamonte 2018: 424), so if a future marker clause is interrogative, there is high likelihood the variant is *be going to*, which should make this association easy to learn. Finally, cues that are regular in form are easier to learn than cues that are irregular (O’Shannessy & Meakins 2012: 380). To illustrate, the future marker clause in (6) contains a second-person subject and a temporal adverbial *all the time*, both of which have been found to prefer *be going to* (Torres-Cacoullos & Walker 2009: 339; Fehringer & Corrigan 2015: 210; Gardner 2017: 146; Denis & Tagliamonte 2018: 423).

(6) you’re gonna be by the ocean all the time (TLC:7_ME_9, B2 Spanish)

Whereas the second-person pronoun is always represented by the linguistic element *you*, temporal adverbials come in many forms, such as *today*, *next week*, *in a few days* etc. At the same time, the second-person pronoun *you* is always adjacent to the variant construction *be going to*, whereas the temporal adverb is separated from *be going to*, which is another reason why the temporal adverb feature should be more difficult to learn.

Previous research also found that prior experience from the learners’ mother tongue can also affect the cues. For one, it is likely that learners will focus their attention on the cues that are salient in their mother tongue and show some degree of similarity to aspects of the target language to enable transfer (Morett & MacWhinney 2013: 134). This might explain the effect of “indirect transfer” found by Rosenbach (2017: 17), whereby speakers of Afrikaans are more likely to allow for inanimate possessors in English *s*-genitives, because inanimate possessors are more common in the genitive in Afrikaans. At the same time, this can lead to the phenomenon of blocking whereby cues from the target language are not noticed (Ellis 2006: 102; Ellis 2008: 236; Ellis & Sagarra 2010; Ellis & Sagarra 2011: 590; Ellis et al. 2014: 548; Ellis & Wulff 2015: 87; Ellis & Wulff 2019: 47).

Statistical learning and circularity

If this account were applied to first language acquisition, one might argue that there is an issue of circularity because copying the language use of other speakers via associative learning does not explain how the patterns of language use originate to begin with. For permutation alternations like the genitive alternation, the most likely explanation is that these patterns were originally shaped by processing principles like Easy First (MacDonald

2013). During online language production, Easy First favors placing the easier possessor as early as possible, so the more accessible animate possessors (Branigan, Pickering & Tanaka 2008) tend to figure in s-genitives. The effects of possessor final sibilancy and semantic relation are more difficult to explain in terms of such processing principles. As for the future marker alternation, a few constraints have also been explained in terms of processing. For example, most previous research investigating the effect of negation on the choice of future marker finds that *be going to* tends to be associated with negative contexts like (7) (Nesselhauf 2010: 173; Tagliamonte, Durham & Smith 2014: 98–99).

- (7) most people probably aren't gonna be able to make a living off of video games (TLC:ME_48, C2 Spanish)

Since negated clauses are more complex syntactically, the environment of the future marker variant is more complex. This causes speakers to prefer the longer *be going to* variant in language production in accordance with Rohdenburg's Complexity Principle (1996: 151; see Szmrecsanyi 2003: 317). Nonetheless, most of the variants' patterns of use are often claimed to originate from the grammaticalization history of the forms. For example, *be going to* was originally used to indicate movement (Fehrer & Corrigan 2015: 208; Gardner 2017: 142; Denis & Tagliamonte 2018: 408), so the structure was not typically followed by verbs implying motion at the early stages of grammaticalization (Danchev & Kytö 1994: 69). Previous literature has therefore posited that *be going to* might have retained its avoidance of motion verbs from the past. Hence, the construction feature of *be going to* to avoid motion verbs is originally the result of a process of language change.

3.2 Constraints without learnable features

What construction features have in common is that they are encoded by concrete linguistic elements in the immediate surrounding context of the exemplars. However, some constraints do not involve any such linguistic elements in the vicinity of the variants. To illustrate, structural persistence captures the fact that constructions that have been used previously tend to be used again (Bock 1982; Szmrecsanyi 2006) because of their higher activation level in memory (e.g., Ellis 2006: 102; Wulff & Ellis 2018: 44; Ellis & Wulff 2019: 44). Structural persistence effects can occur long after a variant is produced (Bock & Griffin 2000), meaning that the linguistic elements encoding structural persistence can be far removed from the

construction. This is one of the defining characteristics of so-called psychophysiological factors proposed by Tamminga, MacKenzie and Embick (2016). Conceptually, it is also difficult to imagine how a previous instance of a construction could be conceived of as a feature in the mental representation of a construction. This is tied to another characteristic of psychophysiological factors already discussed in Section 2.1.1.2, namely that they cannot be at the origin of categorical grammatical distinctions. For feature-based constraints, one could easily imagine a grammatical rule whereby *be going to* is used categorically in interrogative clauses for example, whereas one cannot imagine a grammar where the previous genitive variant or future marker must be repeated. Finally, psychophysiological factors “arise from cognitive [...] systems that are shared by all humans, like working memory capacity, [...] resting activation level of words [...] and so on” (Tamminga, MacKenzie & Embick 2016: 303), so they concern “effects of a range of physiological and psychological factors that govern a speaker’s language production in real time” (Tamminga, MacKenzie & Embick 2016: 305). Hence, psychophysiological factors are universal compared to feature-based constraints like sentence type or possessor animacy, which are by default language-specific. As a result, I argue that probabilistic constraints fitting the characteristics of psychophysiological factors do not have to be acquired by EFL learners through associative statistical learning, contrary to feature-based constraints. Instead, these constraints should operate by default on all speakers regardless of their proficiency level.

To some extent, whether a given constraint should be considered psychophysiological or not depends on the way it is conceptualized and operationalized. Take for example the discourse status of the possessor in the genitive alternation, which captures the fact that speakers prefer the *s*-genitive if the referent of a possessor was already mentioned in discourse (Hinrichs & Szmrecsanyi 2007: 451; Wolk et al. 2013; Rosenbach 2014: 225; Shih et al. 2015; Szmrecsanyi, Biber, et al. 2016; Heller, Szmrecsanyi & Grafmiller 2017: 9). This effect has been linked to the given-before-new principle (Gundel 1988) stating that given referents in discourse precede new referents. Hence, if the possessor is discourse-given, the *s*-genitive is preferred because this variant places the possessor first. For the corpus-based study on the genitive alternation (see Chapter 6), a possessor is considered ‘given’ if the lemma of the head of the possessor noun phrase occurs within the 100 words preceding the genitive observation. In (8), the possessor *Marvel* in *Marvel’s universe* is given, because it occurs among the 100 preceding words.

- (8) and M-Marvel has made millions of dollars from the from all these movies and they are trying to make the the universe larger but if but I guess Marvel is kind of losing their strengths because Marvel's universe is big enough [...] (TLC:2_7_CH_33, B2 Chinese)

Conceptually, the effect of discourse status can be explained in two ways. Either speakers put the given referents first because given referents are more accessible (e.g., Arnold et al. 2000) in accordance with Easy First (MacDonald 2013; Heller 2018: 74–75) or they put the given referents first because this facilitates comprehension (see discussion in Zhang & Kang 2023: 4). If given referents mostly help the speaker during production, then the constraint is primarily psychophysiological. For one, the operationalization of the constraint considers as many as 100 words preceding the genitive exemplar, so the linguistic elements involved are not necessarily part of the exemplar's immediate surrounding context. Secondly, if given possessors are more accessible because the lemma of their head noun was produced before, then the operationalization essentially captures lexical priming effects (see e.g., Cleland & Pickering 2003), which fit all the criteria of psychophysiological factors. However, if the speaker is mostly concerned with the needs of the hearer, then neither possessor accessibility nor lexical priming influence their language production, so the speaker is not affected by psychophysiological factors resulting from features of the brain.

Table 3.1 presents a general overview of the categorization of the constraints for both alternations. The rationale behind the classification of each constraint will be discussed systematically in Chapter 5 for the future marker alternation and Chapter 6 for the genitive alternation.

Table 3.1: Classification of the probabilistic constraints included in this study for the future marker alternation and the genitive alternation as either psychophysiological or not, inspired by the criteria of Tamminga, MacKenzie and Embick (2016).

	Future marker alternation	Genitive alternation
▪ Non-psychophysiological	▪ Sentence type ▪ Negation ▪ Clause type ▪ Temporal proximity of future event ▪ Presence of temporal adverb ▪ Verb type ▪ Subject grammatical person and animacy	▪ Possessor animacy ▪ Possessor definiteness ▪ Possessor final sibilancy ▪ Semantic relation between possessor and possessum
▪ Psychophysiological constraints	▪ Structural persistence ▪ Clause length	▪ Structural persistence ▪ Constituent length ▪ Possessor Discourse status ▪ Possessor thematicity ▪ Type-token ratio

For the future marker alternation, the effects of structural persistence and clause length exhibit the characteristics of psychophysiological constraints, so I argue that they are universal and do not have to be acquired by EFL learners. Regarding the genitive alternation, the same applies for structural persistence, constituent length, type-token ratio as well as the discourse status and thematicity of the possessor.

3.2.1 Processing-related constraints

As stated in Section 2.1.2.1, previous studies ascribe many of the differences between learners and native speakers across alternation phenomena to the fact that learners are more sensitive to processing effects (Deshors & Gries 2014; Wulff, Lester & Martinez-Garcia 2014; Wulff 2016; Wulff, Gries & Lester 2018; Paquot, Grafmiller & Szmrecsanyi 2019; Wulff & Gries 2019). However, it is not clear what constitutes a processing-related constraint exactly, which makes it difficult to test this finding. For the genitive alternation, the choice of variant during online language production largely depends on which constituent is processed first, so one could in principle argue that all constraints contributing to the accessibility of the constituents are processing-related. For example, the animacy constraint is often considered processing-related (Branigan, Pickering & Tanaka 2008; see also Röthlisberger, Grafmiller & Szmrecsanyi 2017), although animacy is a semantic-pragmatic feature of referents (see Rosenbach 2014: 238; Corbett 2010; Croft 2003: 128–132).

For the present study, I argue that cognitive psychophysiological constraints as defined by Tamminga, MacKenzie and Embick (2016) are suitable as processing-related constraints. These constraints are indeed strictly processing-related as they relate to the features of the brain, working memory, planning and more generally “govern a speaker’s language production in real time” (Tamminga, MacKenzie & Embick 2016: 305). By comparison, animacy is not always a factor in online language processing, for example in contexts that do not involve alternations, i.e., when animacy affects language use categorically as speakers only have one option to choose (see the discussion on probabilistic constraints vs. categorical constraints in Section 2.1.1.2). With this in mind, the psychophysiological constraints listed in Table 3.1 will henceforth be referred to as processing-related constraints.

3.3 Synopsis

In summary, I argue that constructions indirectly compete with each other in contexts where they are semantically and functionally eligible. The choice between the constructions is determined by probabilistic constraints (Szmrecsanyi, Grafmiller, et al. 2016), which I categorized into two types, namely feature-based constraints and processing-related constraints.

Feature-based constraints involve linguistic elements in the surrounding context of the variant that are statistically associated with the constructions. An example of such a feature-based constraint for the genitive alternation is the animacy constraint, encoded by the possessor noun. The grammatical person of the subject is a feature-based constraint of the future markers that can be inferred from the structure of the subject noun phrase. Through associative statistical learning, speakers can recognize the association between these linguistic elements and the constructions (e.g., Romberg & Saffran 2010: 3; Misyak & Christiansen 2012). Gradually, EFL learners infer prototypical features of the constructions across many exemplars (see Abbot-Smith & Tomasello 2006). As learners generalize across more exemplars in the input and adapt their mental representation of the construction, they ultimately approximate the construction’s mental representation in native speakers. Hence, this statistical learning account bridges the gap between the concrete linguistic elements that learners are exposed to in their language input and the mental representation of the constructions with their prototypical features (see also Divjak et al. 2021: 62).

However, research on cues suggest that the success of associative statistical learning for a given feature is dependent on various factors, so

certain construction features are in all likelihood more challenging to learn than others. Specifically, a feature that is not often available, is unreliable, highly variable in form and not adjacent to the construction it co-occurs with is more challenging to learn than a feature that is often available, reliable, regular in form and adjacent to its construction (e.g., MacWhinney 2001; Yoshimura & MacWhinney 2010: 552; Misyak & Christiansen 2012: 304). At the same time, prior experience from learners' L1 could either affect the strength of the association or block its acquisition by the learner entirely (Ellis & Sagarra 2010; Ellis & Sagarra 2011; Ellis et al. 2014).

By contrast, processing-related constraints are not encoded by any overt linguistic elements in the surrounding context of the exemplar construction, such as for example structural persistence. It is unclear how learners could acquire the association between a previous future marker exemplar and an upcoming future marker exemplar and integrate this information in their mental representation of the future marker construction through associative statistical learning. Instead, I argue that these constraints roughly correspond to psychophysiological constraints "that govern a speaker's language production in real time" (Tamminga, MacKenzie & Embick 2016: 305). Other characteristics of these factors is that they derive from the features of the brain and cannot be at the origin of categorical grammatical distinctions. As a result, EFL learners should be affected by these universal constraints from the get-go, so these constraints do not have to be learned.

3.4 Generating hypotheses

Given this theory-based account on the acquisition of probabilistic grammars, I generate hypotheses for the three research questions of the thesis.

RQ 1 How does the set of constraints of learners compare to that of native speakers at different proficiency levels?

As a first hypothesis, I argue that learners gradually approximate the target probabilistic grammar as they become proficient.

Hypothesis 1 *As proficiency increases, learners better approximate the target probabilistic grammar.*

While this hypothesis might seem circular at first, it is important to keep in mind that the proficiency level of the learners in the corpus is measured on the CEFR scale (Council of Europe 2001), which does not explicitly consider

learners' use of alternation phenomena. Hence, the relation between learners' proficiency level and the development of their probabilistic grammars is not necessarily linear. For example, one might find that learners have already approximated the target probabilistic grammar for a given alternation at the B2 level, although they are technically still intermediate learners based on the wide range of communicative skills used to assess their proficiency level. Nonetheless, controlling for proficiency level provides a way of looking into learners' language use at different stages of language learning. According to usage-based approaches to SLA, language learning of any kind is driven by input (e.g., Ellis & Wulff 2019). This means that more proficient learners have received more input, which is also responsible for learning probabilistic grammars.

Additionally, I hypothesize that learners will only be sensitive to constraints that also form part of native speakers' probabilistic grammars.

Hypothesis 2 *Learners are only sensitive to constraints that are part of the target probabilistic grammar.*

In this context, speakers are said to be sensitive to a constraint when it forms part of their implicit grammatical knowledge and influences their choice of variant as a result. Establishing sensitivity is not straightforward because one cannot directly access the implicit grammatical knowledge of speakers. Much like previous research on alternation phenomena, I study the future marker and the genitive alternation by analyzing corpus data with mixed-effects logistic regression (see Section 4), which shows associations between the constraints and the variant constructions but cannot reveal a causal relationship between the mental representation of the constraints and the variants during language production (see Section 2.1.2.2). Nonetheless, such associations provide an indication that the relevant constraints are part of implicit grammatical knowledge, and the results from the supplementary rating task experiment can provide additional evidence of this (see Section 4.2).

The hypothesis derives from the insight that language learning is rational as posited by the CREED model (Ellis 2006: 102–103), which means that learners adapt their interlanguage to accommodate the patterns they recognize in the input they have received. Applied to the acquisition of probabilistic grammars, learners should accommodate the patterns present in their target language input through associative statistical learning, while ignoring distinctions and features that are not present in the input.

Importantly, this does not apply to processing-related constraints as they are not learned.

If learners only accommodate the constraints that native speakers consider, then low-proficiency learners should be sensitive to fewer constraints than native speakers.

Hypothesis 3 *Lower-proficiency learners are sensitive to fewer constraints than more advanced learners and native speakers.*

This hypothesis is inspired by the finding that cues are learned one by one: learners adjust a given cue before turning to the acquisition of the next cue (Ellis & Sagarra 2010; Ellis & Sagarra 2011: 601). Provided cues are similar to construction features, I expect learners to acquire the construction features one by one. As a result, lower-proficiency learners might not yet have learned each of the construction features that are present in the input they receive from native speakers, so learners should be sensitive to fewer construction features. Hence, the complexity of a probabilistic grammar, operationalized as the number of constraints a speaker takes into account (Shin 2014), should be simpler for learners compared to native speakers.

While Jäschke and Plag (2016) and Paquot, Grafmiller and Szmrecsanyi (2019) do not explicitly test for this hypothesis, their findings regarding the dative alternation and the particle placement alternation respectively tie in with this hypothesis. However, this has not yet been tested on the genitive alternation, nor has it ever been tested on a substitution alternation like the future marker alternation.

Admittedly, hypotheses 2 and 3 suggest a rather simplistic acquisition process whereby EFL learners acquire each feature-based constraint one by one from scratch. As Ellis pointed out however, second language learners do not start from a ‘tabula rasa’ (Ellis 2006: 109): their language system is already tuned to their mother tongue. In principle, this means that learners could transfer some constraints from their mother tongue that may or may not be relevant in the target language (see Rosenbach 2017). Nonetheless, the TLC corpus features learners from eleven L1 backgrounds, so it is somewhat unlikely that such transfer effects would exert much influence on the overall results from the statistical analyses, unless many of the L1 backgrounds happen to converge on the same transfer effects in spite of their typological diversity.

RQ 2 How does proficiency level modulate the probabilistic grammar of learners?

For the second research question, I hypothesize that low-proficiency learners will first acquire construction features whose linguistic elements facilitate associative statistical learning compared to features that are more challenging to learn.

Hypothesis *Learners first acquire constraints that facilitate associative statistical learning.*

According to research on the associative statistical learning of cues, a cue that is frequent, reliable, regular and adjacent to its construction is easier to learn than a cue that is less frequent, unreliable, irregular and separated from the construction (MacWhinney 2001; Romberg & Saffran 2010: 2–3; Yoshimura & MacWhinney 2010: 552; Misyak & Christiansen 2012: 304; Yi, Lu & Ma 2017: 520). Provided cues are similar to construction features, I expect the same factors to apply: construction features that are more frequent, reliable, regular and adjacent should be easier to learn.

Unfortunately, assessing the learnability of the construction features in an objective way is not straightforward because it depends on several factors. However, studies generally agree that cue reliability or cue-outcome contingency is the main driver of cue strength (Yoshimura & MacWhinney 2010; MacWhinney 2014; Gries & Ellis 2015). Therefore, the learnability of the constraints will be ranked according to their overall association strength with the variant constructions as indicated by monofactorial tests (Cramér's V) for native speakers. Admittedly, this is quite a simplification given that constraints are often operationalized in different ways across studies. Moreover, some constraints are better proxies for the underlying distinctions they are meant to capture. Nonetheless, this approach provides a ranking that can be used to test the hypothesis.

The constraint that is most strongly associated with the future markers is clause type, followed by the temporal proximity of the future event (see Section 5.2.3.1). In the genitive alternation, the most learnable constraint is the animacy of the possessor and the semantic relation that holds between the possessor and the possessum (see Section 6.2.4.1). This ranking implies that if learners deviate from native speakers with regard to clause type, they should also struggle with temporal proximity, and even more so with the remaining feature-based constraints, such as verb type, sentence type etc. Likewise, if learners struggle with the animacy constraint, they should also

experience difficulties with semantic relation, possessor definiteness and final sibilancy.

RQ 3 How do native speakers and learners at different proficiency levels compare regarding processing-related constraints?

For the third research question, I hypothesize that learners should at least be as sensitive if not more sensitive to processing-related constraints than native speakers.

Hypothesis *Lower-proficiency learners are more sensitive to processing-related constraints than more advanced learners and native speakers.*

In previous research, many of the differences between learners and native speakers regarding the use of alternating variants are attributed to the fact that learners are more sensitive to processing effects (Deshors & Gries 2014; Wulff, Lester & Martinez-Garcia 2014; Wulff 2016; Wulff, Gries & Lester 2018; Wulff & Gries 2019; Paquot, Grafmiller & Szmrecsanyi 2019). However, as discussed in Section 3.2.1, almost all constraints can be related to processing since speakers choose the variant during real-time language production, implying that processing principles like the Complexity Principle (Rohdenburg 1996: 151) or Easy First (MacDonald 2013) play an important role (see Section 2.1.1.2). To tackle this issue, I argued that the characteristics of psychophysiological constraints as defined by Tamminga, MacKenzie and Embick (2016) could be used to identify constraints that are strictly processing-related. The processing-related constraints of the future marker alternation are structural persistence and clause length. For the genitive alternation, structural persistence, constituent length, type-token ratio, and the discourse status and thematicity of the possessor are considered processing-related (see Table 3.1).

One of the defining characteristics of such processing-related constraints is that they “arise from cognitive [...] systems that are shared by all humans, like working memory capacity, [...] resting activation level of words [...] and so on” (Tamminga, MacKenzie & Embick 2016: 303). Hence, these constraints should apply universally to native speakers and learners alike. This implies that learners should at least be sensitive to the same processing-related constraints as native speakers. In fact, processing-related constraints should be more influential for especially low-proficiency learners because their language production is less automatic than that of native speakers (Segalowitz 2003).

4 Methodology

One cannot offer an account of the competition between alternating variants without considering the distribution of the underlying probabilistic constraints (see Torres-Cacoullos & Walker 2009: 329; Heller, Szmrecsanyi & Grafmiller 2017: 5). Therefore, studies on alternations typically take the form of large-scale corpus studies that use a methodology proper to variationist linguistic research in the Labovian tradition (Rosenbach 2014: 216; Szmrecsanyi, Grafmiller, et al. 2016; Claes 2017: 1). In such studies, thousands of observations of the alternating constructions are extracted from spoken and/or written corpus data. These observations are then annotated for the different constraints that guide the choice of variant according to the existing literature, either manually or (semi-)automatically depending on the constraint. Afterwards, the annotated dataset is analyzed with statistical techniques that accommodate the multifactorial nature of these phenomena. Typically, the choice of variant is used as the response variable in a (mixed-effects) logistic regression model, with the annotated constraints as predictors (e.g., Gries & Wulff 2013; Wulff, Lester & Martinez-Garcia 2014; Heller, Szmrecsanyi & Grafmiller 2017; Röthlisberger, Grafmiller & Szmrecsanyi 2017; Denis & Tagliamonte 2018; Engel & Szmrecsanyi 2023). “Logistic regression is a technique that allows one to examine the relation between one or several predictors and a response variable” (Speelman 2014: 3), so that “logistic regression, then, is suited for all investigations in which the variation in the response variable is essentially probabilistic in all situations under scrutiny” (Speelman 2014: 4). Concretely, “the procedure systematically tests each factor while holding other factors in the model constant” (Hinrichs & Szmrecsanyi 2007: 460). In this way, the model determines the influence of each predictor and is able to predict the likelihood of either variant given a combination of predictors that reflect a concrete linguistic context.

Other techniques besides logistic regression include random forest and conditional inference tree analysis (Tagliamonte & Baayen 2012; Szmrecsanyi, Grafmiller, et al. 2016; Paquot, Grafmiller & Szmrecsanyi 2019; Dubois 2023), and models of memory-based machine learning (Van den Bosch & Bresnan 2015; Heller 2018). In research on learner language and World Englishes, the MuPDAR approach developed by Deshors and Gries

(2014) and Gries and Adelman (2014) is widely used. This method consists of first modelling the choice of variant for a reference group, typically native speakers, and then using this model to determine in which contexts EFL or ESL learners' choice of variant deviate from native speaker usage. Based on these contexts, one identifies the constraints learners struggle with.

The methodological design of the current study aligned with that of previous studies. That is, I conducted a corpus-based investigation of the future marker and the genitive alternation in the spoken corpus data from the TLC (Gablasova, Brezina & McEnery 2019). For both alternations, I collected many observations of the alternating variants from the learners and the native speakers in the corpus data. These observations were annotated for the relevant constraints guiding the choice of variant according to the literature. For the future marker alternation, this includes negation, structural persistence, clause length, clause type, sentence type, verb type, presence of temporal adverbs, temporal proximity of the future event, animacy and grammatical person of the subject (e.g., Denis & Tagliamonte 2018; Engel & Szmrecsanyi 2023). For the genitive alternation, relevant constraints include structural persistence, the animacy, definiteness, discourse status, final sibilancy and thematicity of the possessor, as well as the length of the possessor and the possessum, the semantic relation between the possessor and the possessum and the type-token ratio of the surrounding context (e.g., Rosenbach 2014; Heller 2018). Then, I used mixed-effects logistic regression to compare the choice of variant in the language use of the native speakers and EFL learners at different proficiency levels.

As mentioned in Section 2.1.2.2, corpus-based investigations reveal what speakers say, but not necessarily what they know (Ellis 2017: 46; see also Schönefeld 2011: 13; Kinne 2020: 173 and the discussion of sensitivity in Section 3.4). Therefore, the corpus-based study of the genitive alternation was complemented with a rating task experiment à la Bresnan and Ford (2010; see also Engel et al. 2022; Engel & Szmrecsanyi 2023), where native speakers and learners from different L1 backgrounds and proficiency levels had to rate the naturalness of the genitive variants in authentic corpus excerpts. These ratings were then compared against the predictions from the corpus-based model to compare the speakers' use of the variants with their intuitions. Combining corpus data and experimental data increases the explanatory power and generalizability of the findings (Hoffmann 2006: 169; Arppe & Järvikivi 2007: 151–152; Gilquin & Gries 2009: 9; Schönefeld 2011: 20–24; Divjak & Arppe 2013: 265; Abbuhl, Gass & Mackey 2013: 129) and therefore constitutes a desideratum both in

usage-based cognitive linguistics generally (Klavan & Divjak 2016) and SLA research more specifically (Ellis 2017).

In the following sections, I expand on each aspect of the methodology, starting with the design of the corpus-based studies, including the delimitation of the alternating variants, the data collection and annotation process as well as the method of statistical analysis. Then, I turn to the design of the experiment, which caters to some of the limitations of the corpus-based studies.

4.1 Corpus-based studies

4.1.1 The corpus

The data stems from the Trinity Lancaster Corpus (TLC, Gablasova, Brezina & McEnery 2019), a 3.4-million word corpus consisting of transcribed recordings from an official spoken language exam, namely the Graded Examinations of Spoken English (GESE). The GESE exam is designed to resemble an authentic discussion between the candidate and the examiner. Depending on the candidate's proficiency, the exam takes between 5 and 25 minutes during which participants have to complete up to 4 different tasks covering production and comprehension, namely a presentation task, a discussion task, an interactive task and a conversation task. On average, the interlocutors produce about 2,500 words during the exam, of which approximately 60% are produced by the candidate learner.

For the presentation task, the learners have to give a monologic five-minute presentation about a topic of their choice, which is (semi-)formal in nature (Gablasova & Brezina 2015: 122; Gablasova et al. 2017: 618–619). Afterwards, the candidate and the examiner discuss the presentation together in the discussion task. Given that participants can more or less decide their topic of presentation, the exams contain spoken learner language on a wide range of topics (Gablasova & Brezina 2015). The interactive task starts with a prompt from the examiner regarding an issue or an observation. The candidate then has to ask questions and make comments to find out more about the issue. An example of such a prompt is shown in (9).

- (9) scientists say that they will soon be able to prolong general life expectancy to well over the age of a hundred I'm not at all convinced of the desirability of this development (TLC:IN_12, NS)

Finally, the candidate and the examiner discuss two topics of interest in the conversation task. For more advanced learners, the examiner can choose the topic.

The corpus provides a number of advantages. First, the corpus features spoken data, which not only offers a better window into the underlying grammatical system of learners according to SLA (Myles & Mitchell 2014); psycho- and neurolinguists claim it is also more representative of processing-related effects (Kemmerer 2015). Secondly, the exams were taken in six different countries, namely Spain, Italy, Mexico, Russia, India and China, so the corpus data features learners from eleven different L1 backgrounds, namely Spanish, Italian, Russian, Hindi, Chinese, Kannada, Telugu, Gujarati, Marathi, Malayalam and Tamil. Such a typologically diverse sample of L1s ensures that any proficiency-related effects are not actually due to transfer effects from a specific L1 background. One potential issue, however, is that English learners from India would typically be considered speakers of English as a Second Language (ESL) rather than learners of English as a Foreign Language (EFL), considering that English is an official language in India used in institutions. Nonetheless, I did not differentiate between the Indian learners of English and other learners in the corpus mainly because all learners try to approximate the same standard of British English, as evidenced by the fact that all learners in the corpus pass the official GESE exam. Thirdly, given that the GESE is an official language exam, people from a wide variety of backgrounds can participate, so the participants in the corpus vary in terms of age and education. By contrast, many learner corpora contain data from a more homogeneous sample of university students enrolled in an English language degree. Finally, the main advantage of the corpus is that each candidate learner is assigned a CEFR proficiency level from low-intermediate (B1) to advanced (C2). Contrary to previous studies, this made it possible to control for proficiency level during statistical analysis and investigate how the probabilistic grammars of learners develop as they become more proficient.

4.1.1.1 Proficiency level in the corpus

The twelve levels of the GESE exam are designed to map onto the proficiency levels of the Common European Framework of Reference (Council of Europe 2001), so the proficiency level of the learners in the corpus is based on the GESE exam they pass. Hence, if a candidate passes level 7 of the GESE, then this learner is recognized as a B2 learner in the corpus metadata. In total, the corpus covers language exams from more than a thousand low-intermediate (B1) to advanced (C2) learners of English.

Operationalizing proficiency level in this way has one potential shortcoming, namely that a learner's proficiency level in the corpus ultimately depends on the GESE level the candidate wishes to pass. If a candidate requires a certificate stating that he or she speaks English at B1 level for a new job opportunity for example, they will pass the GESE exam that corresponds to this proficiency level, although this candidate might in reality have the communicative competence of a C1 learner. As a result, the corpus would underestimate the proficiency level of these learners. Moreover, the examiner can indirectly influence the performance of the candidate, and by extension, the grading of the exam (Sandlund, Sundqvist & Nyroos 2016: 20). For example, different examiners do not necessarily provide the same degree of language support to the candidates they examine by rephrasing questions, completing the candidate's turn, offering missing vocabulary, etc. (Lazaraton 1996: 158–165). As a result, one examiner may help the candidate more, resulting in a pass, while another examiner is less helpful, resulting in a fail, although both candidates might have the same communicative competence in English (Lazaraton 1996; see also Sandlund, Sundqvist & Nyroos 2016: 18). In spite of these shortcomings, I argue that an interactive language test covering both production and comprehension skills in multiple tasks like the GESE exam provides a more accurate approximation of the learners' proficiency level than other proxies often used in other learner corpora, like the number of years one has been learning English (Callies 2015), which is problematic "especially if L2 users come from different educational backgrounds and from countries with different levels of exposure to English" (Gablasova, Brezina & McEnery 2019: 139).

4.1.1.2 The language production of the examiner

To compare the probabilistic grammars of learners with native speakers, I used the data from the examiners in the corpus. This ensures that learners and native speakers discuss similar topics within the same register. In more technical terms, the data from the native speakers and the learners are nested into the corpus in the same way, which makes it easy to control for any effects relating to the joint discourse of the examiner and the learner during statistical analysis by virtue of random effects in the mixed-effects logistic regression model (see Gries 2015). Additionally, including the data from the examiners made it possible to partially control for their influence on the learners, who might for example be inclined to repeat the same variant as the examiner. Specifically, the operationalization of structural persistence will distinguish between whether it is an examiner or a candidate learner who produced the previous alternating variant.

Nonetheless, the question arises as to whether the language use of the examiners in the corpus is representative of native speakers of British English. One might expect examiners to use Teacher Talk, whereby teachers simplify their language use according to the proficiency of the learner (Patil 1994: 155; Stanley & Stevenson 2017: 2). It is characterized by the avoidance of phonetic features of connected speech such as assimilation or contraction and by repetition, the use of higher frequency vocabulary items, fewer pronouns and less complex syntactic structure (Patil 1994: 156–157; Ünel & Mirioğlu 2015: 127–128; Stanley & Stevenson 2017: 2). Closely related to Teacher Talk is Foreigner Talk, which is defined as the type of language that native speakers use to help improve communication with foreigners who have limited competence in the language (Ferguson 1975: 1; Wooldridge 2001: 622; Dashti 2013: 64). The features of Foreigner Talk are similar to Teacher Talk and include an increase of pauses and repetitions, a reduced rate of speech delivery and clearer articulation, omission of function words and inflectional endings, the use of synonyms, homonyms, parallel syntactic structures and paraphrasing, negotiation of meaning and repair, shorter sentences etc. (Ferguson 1975: 8; Lazaraton 1996: 153; Wooldridge 2001: 2001; Dashti 2013: 66). The difference between the two simplified registers is that Teacher Talk has as a goal to provide the target language in a comprehensible way, whereas Foreigner Talk is mainly used to facilitate communication between native and non-native speakers (Stanley & Stevenson 2017: 2).

Because of the interactive nature of the GESE, the language of examiners should remain grammatical with some features of both Teacher Talk and Foreigner Talk. Even so, I argue that this should not be problematic for the investigation of the alternations, because the features of these simplified registers do not concern the probabilistic constraints directly as they form part of speakers' implicit knowledge (Bresnan et al. 2007; Bresnan & Ford 2010). Admittedly however, the context of the construction may be less complex in terms of the type-token ratio for the genitive alternation (Hinrichs & Szmrecsanyi 2007: 457; Grafmiller 2014), and the clauses might be shorter, which is relevant for the future marker alternation (Szmrecsanyi 2006: 127).

4.1.1.3 Register effects in the corpus

Previous research found that register modulates the choice of genitive, dative and future marker variant in real time, both in terms of speakers' overall preference for either variant and on the level of the constraints (Grafmiller 2014; Engel et al. 2022; Engel & Szmrecsanyi 2023). In the present case, the data should a priori cover multiple spoken registers,

because “the four tasks [...] represent a variety of settings in which spoken language is used with different degrees of formality, interactiveness, topic familiarity, and different interlocutor roles” (Gablasova et al. 2017: 619). The presentation task is a (semi-)formal monologue (Gablasova et al. 2017: 618–619; Gablasova & Brezina 2015: 122), whereas the other three tasks are more interactive. Since the thesis is not concerned with how learners adapt the probabilistic constraints to fit certain registers and instead focuses on the development of the probabilistic grammars throughout language learning, I did not directly consider the effect of the exam task on the choice of variant. However, this was controlled for indirectly by virtue of random effects in the mixed-effects logistic regression models (see Gries 2015).

4.1.2 Delimiting the variable contexts

4.1.2.1 The future marker alternation

Speakers of English have at their availability a few different structures to convey futurity, namely *will*, *be going to*, *shall*, *be to* (Declerck 2010), the present simple and the present progressive (Nesselhauf 2007). However, the current study was limited to *will* and *be going to* like most previous research on alternating future markers (e.g., Torres-Cacoullos & Walker 2009; Fehringer & Corrigan 2015; Gardner 2017; Denis & Tagliamonte 2018; Engel & Szmrecsanyi 2023). Most importantly, these two variants represent most of the future marker expressions (Denis & Tagliamonte 2018: 405). Additionally, the other forms do not cover the same range of uses as *will* and *be going to*. *Shall* is very rare and only occurs 60 times in the entire corpus of 3.4 million words (17.6 times per million words); the use of *be to* is rather restricted (Bergs 2010: 226) and the futurates, viz. the simple present and the simple progressive, have a different distribution than *will* and *be going to* in native speaker English because they cannot be used to talk about unplanned future events (Copley 2009; cited in Denis & Tagliamonte 2018: 405). Admittedly, there is also some discussion about whether it is justified to consider the contracted form *'ll* as a variant of *will* rather than *shall*, since there are important differences between *will* and *'ll*. For example, *'ll* is strongly favored with first person subjects, whereas *will* is least common with pronominal subjects except for *it* (Berglund 2000: 41, see also Section 5.2.2.8). In the current study, they will nonetheless be considered variants of the same form (Szmrecsanyi 2006: 112).

Not all instances of *will* and *be going to* are interchangeable, however. For example, future-in-the-past contexts like (10) almost categorically

prefer *be going to* (Poplack & Tagliamonte 2000; Berglund 2000: 31; Aarts, Close & Wallis 2013: 36; Fehringer & Corrigan 2015)

- (10) I thought er we were going to get lost because I have never taken the subway without and adult (TLC:2_7_SP_34, B2 Spanish)

The cases that are not interchangeable according to native speaker standards were removed in accordance with Denis and Tagliamonte (2018; see also Engel & Szmrecsanyi 2023).

4.1.2.2 The genitive alternation

The genitive alternation typically refers to the competition between the *s*-genitive and the *of*-genitive. However, there are other phrasal structures one can use to express “the relation between the possessor and a possessed entity” (Rosenbach 2014: 221; see also Rosenbach 2002: 27). In the words of Szmrecsanyi et al. (2016: 25), a strictly onomasiological perspective would crack the lid of Pandora’s Box, inviting structures such as *the dog has a tail* where *have* is a potential genitive variant (see Rosenbach 2002: 27). Hence, the *s*-genitive and the *of*-genitive also share the structural characteristic that “in both constructions a possessor is linked to a possessum on the nominal level” (Rosenbach 2002: 28). This still licenses a third potential genitive variant, namely the compound noun genitive as in *community meeting*, which could also be realized as an *s*-genitive or an *of*-genitive (Szmrecsanyi, Biber, et al. 2016). Nonetheless, the present study focuses exclusively on the *s*-genitive and the *of*-genitive, which connect two separate noun phrases (Quirk et al. 1985: 1276–1278; Heller 2018: 59).

Again, not all pairs of noun phrases connected by the preposition *of* or the *s*-clitic represent interchangeable genitive variants. To determine which cases are interchangeable, the dominant approach in the literature is to exclude the cases that categorically exclude the other variant for native speakers (Kreyer 2003: 170; Rosenbach 2014: 223; Heller 2018: 16). For partitive genitives like (11), it is very unlikely that a native speaker would use *tea’s nice cup* (Szmrecsanyi & Hinrichs 2008: 296; Heller 2018: 56).

- (11) For you happiness is like sitting alone having a nice cup of tea (TLC:2_IN_16, C2 Marathi)

4.1.2.3 Native speaker interchangeability

For both alternations, I removed the cases that are not interchangeable for native speakers, but what is interchangeable for a native speaker might not be for a learner and vice versa, so learners might in fact say *tea’s nice cup*

instead of *a nice cup of tea* in (11). There are methodological and theoretical reasons for considering native speakers as the benchmark. From a methodological point of view, adhering to the native speaker standard is more efficient. That is, the categorical cases for native speakers of English are well-documented considering that the genitive alternation is arguably the best researched alternation phenomenon in English (Rosenbach 2014: 215), whereas identifying the interchangeable contexts for learners would require much work, not to mention there could be differences across L1s. Theoretically, it is justified to use native speakers as the benchmark considering that the learners in the corpus are taking an official British English spoken language exam, which suggests that this is the language variety the learners wish to approximate.

4.1.3 Data collection & annotation

Future markers and genitive observations were collected from the TLC corpus (Gablasova, Brezina & McEnery 2019) with R scripts (version 4.2.2, R Core Team 2022). Some of the categorical contexts of genitives and future markers cannot be excluded automatically via the script, so I had to manually check the interchangeability of the collected observations. Since this is rather time-consuming, I decided to reduce the workload by sampling within the more frequent variant of each alternation. That is, I only included as many *of*-genitives and instances of *will* necessary to match the number of collected observations of the less frequent variant, viz. the *s*-genitives and *be going to*. While this means that the data is not particularly insightful regarding the relative frequencies of the variants, I am mainly interested in comparing the probabilistic conditioning of the variants for native speakers and learners. Moreover, the method of analysis, mixed-effects logistic regression, is limited by the number of observations from the least frequent variant (Hosmer & Lemeshow 2000: 346–347). Therefore, having more *of*-genitives and instances of *will* does not necessarily improve the quality of the statistical analysis.

Interchangeable future marker observations were then annotated for negation, structural persistence, clause length, clause type, sentence type, verb type, presence of temporal adverbs, temporal proximity of the future event, animacy and grammatical person of the subject (e.g., Denis & Tagliamonte 2018; Engel & Szmrecsanyi 2023). Each genitive observation was also coded for structural persistence, as well as the animacy, definiteness, discourse status, final sibilancy and thematicity of the possessor, the length of the possessor and the possessum, the semantic relation between the possessor and the possessum and the type-token ratio

of the surrounding context (e.g., Rosenbach 2014; Heller 2018). To annotate the future marker observations, I largely followed the annotation guidelines of Engel (2022), whereas the annotation of most genitive constraints followed Heller (2018). Some constraints could be coded automatically with the help of R, while some had to be coded manually, as shown in Table 4.1.

Table 4.1: Overview of the coding method for the constraints of the future marker alternation and the genitive alternation.

	<i>Future marker alternation</i>	<i>Genitive alternation</i>
Manual	▪ Clause type ▪ Temporal proximity ▪ Sentence type ▪ Verb type ▪ Subject animacy ▪ Temporal adverb	▪ Possessor animacy ▪ Semantic relation
(Semi-)automatic	▪ Negation ▪ Persistence ▪ Subject grammatical person ▪ Clause length	▪ Possessor and possessum length ▪ Possessor discourse status ▪ Structural persistence ▪ Thematicity ▪ Type-token ratio ▪ Possessor definiteness ▪ Final sibilancy

For variables that are coded manually, it is important to make sure that the annotation is consistent and reliable (Révész 2012; Larsson, Paquot & Plonsky 2020). Therefore, the annotation procedure was described in a coding book in accordance with best practices in SLA research (Révész 2012) and then visualized in the form of decision trees. The coding book and decision trees were used to conduct an interrater reliability analysis. For the future marker alternation, a second coder annotated approximately 20% of the entire dataset for clause type, temporal proximity, verb type and subject animacy. For the genitive alternation, half of the dataset was double-coded for possessor animacy and semantic relation. Estimation of interrater reliability using Cohen’s kappa exceeded .8 for each of the variables, indicating almost perfect agreement (Landis & Koch 1977). Importantly, such an interrater reliability analysis does not pertain to the validity of the coding scheme, i.e., whether the operationalization properly captures the variable of interest.

4.1.4 Statistical analysis

The annotated datasets were then analyzed with mixed-effects logistic regression (Speelman 2014; Gries 2015) using the *lme4* package (Bates et

al. 2015) in R (R Core Team 2022). The models predicted the choice of variant construction based on a set of predictors, namely the probabilistic constraints for which the data was annotated. To do so, the model first calculates how each probabilistic constraint influences the choice of variant construction by systematically testing each factor in the model while holding the other predictors constant, so “regression analysis [...] is the closest a corpus linguist can come to conducting a controlled experiment” (Hinrichs & Szmrecsanyi 2007: 459).

Specifically, I used treatment coding to determine how the constraints influence the choice of variant. With treatment coding, the model chooses a reference level for each predictor and compares it against the other levels of the predictor. For example, the probabilistic constraint ‘sentence type’ in the future marker alternation refers to the fact that interrogative clauses strongly prefer *be going to*, to the extent that the preference is almost categorical (Szmrecsanyi 2003; Tagliamonte, Durham & Smith 2014; Denis & Tagliamonte 2018: 424). In the dataset, each clause was annotated as either interrogative or declarative and these were also the levels of the predictor in the model. To determine the influence of sentence type on the choice of variant, one would for example choose declarative clauses as the reference level. Then, the model compares the likelihood of the variants in interrogative clauses with their likelihood in declarative clauses and finds that *be going to* is more likely in interrogative clauses than declarative clauses. This difference between interrogative and declarative clauses is accompanied by an indication of effect size and a *p*-value for statistical significance. Once the model has estimated the influence of each probabilistic constraint, it can predict how likely *will* or *be going to* would be in unseen contexts given a certain combination of constraints.

In addition to the probabilistic constraints listed in Section 4.1.3, the model also featured the proficiency level of the speaker as a predictor, which was allowed to interact with each of the probabilistic constraints. With this design, the model was able to show whether the proficiency level of the speaker modulates the influence of the probabilistic constraints. With native speakers as the reference proficiency level, the model essentially revealed i) which probabilistic constraints influence native speakers’ choice of variant and ii) whether these probabilistic constraints have a different effect on learners at different proficiency levels.

However, the choice of variant is not only the result of probabilistic constraints. For one, each native speaker and learner in the corpus has their own idiolect, so one speaker might be more inclined to use a given variant than other speakers. Moreover, the different exam tasks vary in terms of topic familiarity, interactiveness and formality (Gablasova et al. 2017: 619),

which could also influence which variant is preferred, and the same applies to the topics the examiner and the candidate discuss during their interaction (see Section 4.1.1). As implied by previous literature, the L1 of learners is also likely to affect their choice of variant through transfer (Gries & Wulff 2013; Wulff, Lester & Martinez-Garcia 2014; Wulff & Gries 2015; Paquot, Grafmiller & Szmrecsanyi 2019). Learners from the same country might also have been taught to use the variants in a certain way. For instance, some learners might have been taught to use the *s*-genitive categorically when the possessor refers to an animate being and they should use the *of*-genitive if the possessor is inanimate, as suggested by prescriptivists (Murphy 2012; cited in Heller, Szmrecsanyi & Grafmiller 2017: 10). Previous research also found that certain verbs prefer *will* and others prefer *going to* (Engel & Szmrecsanyi 2023: 361). In the genitive alternation, the lemma of the possessor and the possessum also incorporate strong preferences for either variant. While these factors are not particularly relevant for the research goals of the thesis, it is nonetheless desirable to take these sources of variation into account because they can interfere with the results from the model. Therefore, the mixed-effects logistic regression models used for data analysis controlled for these irrelevant sources of variation and dependencies between observations via random effects (Gries 2015).

4.2 Supplementary rating task experiment

4.2.1 The limitations of corpus-based studies

Probabilistic grammars are supposedly part of speakers' implicit grammatical knowledge (Bresnan & Ford 2010, see Section 2.1.1.1). However, there are fundamental differences in the way corpus-based regression models and humans gather knowledge about distributional patterns, so the results from corpus-based models do not directly reflect the implicit knowledge of speakers. For one, mixed-effects logistic regression models have difficulty dealing with multicollinearity, yet collinearity and redundancy are an integral part of language (Milin et al. 2016: 508). For instance, most of the prototypical semantic relations that favor the *s*-genitive can only be realized by animate possessors, such as kinship relations (*the man's sister*) and body parts (*the rat's ears*) as well as legal ownership (*the businesswoman's handbag*), so the animacy of the possessor naturally correlates with the semantic relation that holds between the possessor and the possessum. If there is too much multicollinearity, one

might have to change the design of the model in a way that is less representative of actual language production.

Most importantly, the results from corpus-based models depend on how often variants occur in different contexts. Yet, the fact that a variant does not often occur in a certain context does not mean that speakers deem the variant unacceptable or unnatural (Kempen & Harbusch 2005: 344; Arppe & Järvikivi 2007: 149–150; Divjak, Dąbrowska & Arppe 2016: 26–27). If speakers gauge how well variant constructions fit a given context during language production, then corpus data is also limited in that it can only reveal which variant was ultimately produced without providing any insight regarding the processes that led speakers to choose this variant. For example, if *will* rarely occurs in interrogative clauses (see Szmrecsanyi 2003; Tagliamonte, Durham & Smith 2014; Denis & Tagliamonte 2018: 424), the assumption is that speakers strongly disfavor *will* in interrogative clauses. Although this suggests that *be going to* is indeed preferred in interrogative clauses, corpus data cannot show how strong this preference is in language production. Hence, the effect size of constraints based on frequency does not overlap perfectly with their importance in the implicit probabilistic grammar of speakers.

In general, corpus-based models are able to “reflect the characteristics of mental processes and structures yielding usage, even though we do not know the exact form of these mental representations” (Divjak & Arppe 2013: 230). Put simply, corpus-based models show what speakers say, but not necessarily what they know (Ellis 2017: 46; see also Schönefeld 2011: 13; Kinne 2020: 173). Applied to alternation phenomena, the question arises as to whether the significant constraints in corpus-based regression models are indeed part of the implicit grammatical knowledge of speakers (Klavan & Divjak 2016), or whether they are simply frequent patterns of language use (see also the discussion of ‘sensitivity’ in Section 3.4). More specifically, models might be more precise than humans at distinguishing between alternating variants. Alternatively, they might perform less well, in which case the model is missing some important constraints that speakers take into account when choosing which variant construction to produce (Klavan & Divjak 2016). Therefore, it is useful to compare the performance of any statistical method with human speakers in a more controlled setting (Arppe & Baayen 2011: 8). In this case, I complemented the corpus-based study of the genitive alternation with a rating task experiment similar to Bresnan and Ford (2010; see also Engel et al. 2022; Engel & Szmrecsanyi 2023).

4.2.2 Experiment design

The rating task experiment was designed in Qualtrics (Qualtrics 2023) and administered online. Participants were shown original corpus excerpts with both possible genitive variants. Using a slider bar, participants could indicate which variant they preferred. The position of the slider represented a rating from 0 to 100 for the variant on the right side of the slider (*the film's name* in Figure 4.1). Higher scores indicated that the variant on the right side was more natural, with a score of 100 indicating that this was the only acceptable variant in this context. Inversely, a score of 0 indicated that this variant was not acceptable at all.

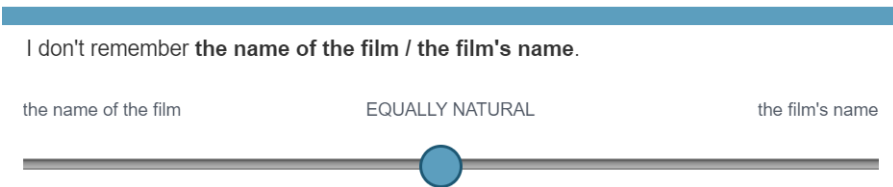


Figure 4.1: An example of a target experimental item presented to participants.

Contrary to the corpus data, which can only show which variant speakers ultimately produced (see Section 4.2.1), the position of the slider on the rating scale shows how strong the preference is for this variant relative to the competing variant. At the same time, acceptability rating tasks require introspection from the participants, so the ratings complement corpus-based models by offering a slightly different perspective on alternations (Arppe & Järvikivi 2007: 150).

Importantly, the experimental items stemmed from the corpus data, so it was possible to compare the predictions of the corpus-based logistic regression model, which capture speakers' use of the variants, with their intuitions regarding the acceptability of the variants as reflected by the ratings for the same observations (Arppe & Järvikivi 2007: 132; Gilquin & Gries 2009; Klavan & Divjak 2016: 361). The expectation is that language use correlates with intuitions, so if the corpus-based model predicts that an *s*-genitive is likely, then the *s*-genitive should also receive higher ratings in the experiment, as suggested by the findings of Bresnan and Ford (2010), Engel et al. (2022) and Engel and Szmrecsanyi (2023). This would provide additional evidence that the significant probabilistic constraints in the corpus-based model are indeed part of speakers' implicit grammatical knowledge (Bresnan & Ford 2010).

Additionally, the experiment tested the effect of possessor animacy on the ratings because the corpus-based analysis of the genitive data revealed

that B1 and B2 learners are less sensitive to this constraint than native speakers. This finding is interesting because possessor animacy should be the most learnable constraint of the genitive alternation according to the theory-based account (see Section 3.4). Moreover, possessor animacy is such a strong constraint that one would expect learners to be more sensitive to possessor animacy than native speakers. If the finding is replicated in the experiment, this would provide additional evidence that this interaction reflects the implicit grammatical knowledge of learners (Bresnan & Ford 2010).

Similar experiments have been implemented successfully to complement corpus-based research for the particle placement alternation, the position of the prepositional phrase, the dative alternation and the future marker alternation in English (Bresnan 2007; Bresnan & Ford 2010; Kinne 2020; Engel et al. 2022; Engel & Szmrecsanyi 2023; see also Verhagen et al. 2020). It has also been used to complement corpus-based models with more than two possible variants in other languages, such as Russian (Divjak & Arppe 2013) and Estonian (Klavan 2020). While these studies generally found that the results from corpus-based models and the experiments largely converge (Klavan & Divjak 2016), the predictions from the corpus-based models do not correlate in a linear fashion with the acceptability of the variants according to individual speakers. As stated before, this is partially attributable to the fact that regression models are largely concerned with how frequently the variants occur in certain contexts, which does not mean that speakers deem the variant unacceptable (Kempen & Harbusch 2005: 330; Arppe & Järvikivi 2007: 149–150; Divjak, Dąbrowska & Arppe 2016: 26–27).

Participants

The participants were recruited on Prolific (Prolific 2023), an online platform where individuals can register to complete paid online experiments. To ensure the comparability of the results, the participants in the experiment were selected to resemble the speakers featured in the TLC data, both in terms of their L1 background and in terms of their proficiency level. Specifically, participants had either Spanish, Italian, Hindi, Urdu, Chinese, Russian, Polish or Hungarian as their L1 according to the information they had entered upon registration to Prolific. To determine the proficiency level of participants on the CEFR scale (Council of Europe 2001), the Qualtrics survey included a short proficiency test as filler items, namely the Oxford Quick Placement Test (Allan 2001). Power analysis using multiple simulations (Kumle, Vö & Draschkow 2021) showed that to achieve a satisfactory statistical power of .8, one should preferably collect

data from at least 20 participants for each of the five proficiency levels featured in the TLC data, viz. B1, B2, C1, C2 and native speakers. Hence, the rating task experiment was administered to a total of 25 native speakers and 101 EFL learners.

Statistical analysis

The ratings were analyzed as the response variable in two mixed-effects linear regression models using the *lme4* package (Bates et al. 2015) in R (R Core Team 2022). The first model investigated the correlation between the ratings from the participants and the predictions from the corpus-based model. The second model focused on the effect of possessor animacy on the ratings. Both models used treatment coding to compare how the predictors influence the ratings for native speaker participants compared to the learner participants at different proficiency levels. Hence, the first model answered i) how predictions from the corpus-based models relate to the ratings of native speakers and ii) whether this correlation is the same for learners at different proficiency levels. The second model revealed i) how possessor animacy influences the ratings of native speakers, and ii) whether animacy has the same effect on learners at different proficiency levels.

Perhaps the main advantage of mixed-effects linear regression in this case is the possibility “to simultaneously consider all factors that potentially contribute to the understanding of the structure of the data” (Baayen, Davidson & Bates 2008: 409). In addition to predictors of interest, such models can control for the individual tendencies of participants. It might for example be the case that one participant does not use the entire range of the rating scale (see Schütze & Sprouse 2013: 39), or that they generally give higher ratings for either variant. Likewise, it is possible that certain items in the experiment are rated differently from others. To ensure the reliability of the results, it is recommended to include cross random effects for participant and experimental item (Baayen, Davidson & Bates 2008). This includes random intercept adjustments as well as random slopes, either by participant to account for the fact that fixed effects might influence individual participants differently, or by item to control for the fact that participants from specific proficiency levels might not rate an item in the same way (Linck & Cummings 2015).

An additional benefit of regression models is that it is easy to control for potential confounding variables (Linck & Cummings 2015). In order to replicate the corpus-based findings regarding the animacy constraint, one has to control for the other variables that could influence participants' ratings. Otherwise, it can be difficult to disentangle whether ratings are due to the animacy of the possessor or some other variable. To solve this issue,

one can add the confounding variable as a fixed effect in the regression model.

Individual differences

The results from the corpus-based investigation of the genitive alternation showed that B1 and B2 learners are less sensitive to possessor animacy than native speakers. However, the results from corpus-based models essentially present a generalization across the individual probabilistic grammars of all speakers in the data. Hence, it is possible that the probabilistic grammar of an individual speaker does not overlap perfectly with the constraints the model deems significant, as shown by Divjak et al. (2016: 28). Likewise, Kinne (2020: 128–138) found that the effect of end-weight differed across learners in the particle placement alternation and the verb-prepositional alternation, with some learners relying on the constraint almost categorically, while for others, the constraint did not influence the choice of variant at all.

With regard to possessor animacy, the question arises as to whether all B1 and B2 learners are less sensitive to possessor animacy as a group, or whether certain learners exhibit other strategies. This is all the more relevant in the context of learner language since it features even more individual variability compared to native speakers (Jäschke & Plag 2016: 511–512; Kinne 2020: 128). Investigating the effect of animacy on the ratings of individuals can provide interesting insights on inter-speaker variability for individual constraints. Similarly, the correlation between the ratings of individual speakers and predictions from the corpus-based model can show to what extent the probabilistic grammar of individuals overlaps with the set of constraints in the corpus-based model.

Inspired by Kinne (2020), I therefore measured Cronbach's alpha on the ratings of the participant at each proficiency level. Cronbach's alpha is a measurement of intra-class correlation that is often used to assess the internal consistency of instruments with multiple items. If all items in the instrument capture the same result as desired, then the internal consistency is high (Gliem & Gliem 2003). In this case, I used Cronbach's alpha to measure how consistent participants are at each proficiency level. Additionally, I compared the ratings of individual participants against the aggregate results from both mixed-effects linear regression models.

4.3 The relation between corpus data and experimental data

While the rating task experiment caters to some of the limitations of the corpus-based studies, it comes with its own shortcomings. For one, corpus data typically represents authentic language production whereas experimental data is more artificial (Arppe & Järvikivi 2007: 132; Gilquin & Gries 2009; Römer et al. 2014: 131; Klavan & Divjak 2016: 361). Moreover, the ratings of participants could be influenced by prescriptivism and their previous teaching experience. That is, learners might give higher ratings to the variant they were taught is 'right' although this does not align with their own intuitions. It is unlikely that the corpus data suffers from this shortcoming to the same extent given that learners have less time to consciously think about the choice of variant during spoken language production.

Hence, both types of data are equally important at different stages of the research process (Schönefeld 2011: 21). The quantitative corpus analysis allows for the inclusion of many interacting constraints simultaneously. The rating task experiment provides a complementary perspective on the results based on participants' introspection (Arppe & Järvikivi 2007: 152), while also shedding light on the most important finding from the corpus-based analysis of the genitive alternation, namely that B1 and B2 learners are less sensitive to possessor animacy than native speakers. Taken together, converging results from corpus data and experimental data increase the explanatory power and generalizability of the findings (Hoffmann 2006: 169; Arppe & Järvikivi 2007: 151–152; Gilquin & Gries 2009: 9; Schönefeld 2011: 20–24; Abbuhl, Gass & Mackey 2013: 129; Divjak & Arppe 2013: 265).

5 The future marker alternation

5.1 Introduction

In this first case study, I conduct a quantitative corpus-based study comparing the probabilistic grammar of native speakers with that of EFL learners at different proficiency levels regarding the future marker alternation. The alternation consists of the variants *will* and *be going to*, exemplified in (12) and (13) respectively.

(12) I'll just ask you one more question (TLC:2_7_IN_1, NS)

(13) I'm gonna ask you a real question (TLC:8_ME_12, NS)

Historically, the competition between *will* and *be going to* has been ongoing for centuries. Originally, *will* derived from Old English *willan* expressing volition (Wekker 1976: 24). Over time, the form could be used in future contexts expressing desire, purpose or intention (Engel 2022: 180). Since the Middle English period, *will* has been used to express the neutral future (Poplack & Tagliamonte 2000: 318; Tyler & Jan 2017: 423–424). *Be going to* originated from the progressive form of the verb *go*, which was then combined with a preposition (*to*) and a complement to indicate movement towards a goal (Tagliamonte, Durham & Smith 2014: 77). Gradually, this combination developed towards a general sense of prediction which led to the earliest use of future *be going to* dated in 1482 (Danchev & Kytö 1994: 61; Tagliamonte, Durham & Smith 2014: 77). Although both future markers have continued to develop, they have essentially been competing since this period onwards.

The choice between the variants is conditioned by a wide range of feature-based constraints relating to formal, semantic or contextual features of the verb, the subject and the clause (e.g., Denis & Tagliamonte 2018; Engel & Szmrecsanyi 2023), as well as processing-related constraints such as structural persistence and clause length (Szmrecsanyi 2003; Szmrecsanyi 2006). Some accounts have attributed the effect of certain feature-based constraints to Rohdenburg's Complexity Principle (Rohdenburg 1996: 151) implying that speakers should prefer the more explicit *be going to* when the cognitive environment is complex, as this could buy the speaker more processing time. Specifically, previous accounts

have suggested that *be going to* is preferred in longer, negated and subordinated clauses because they are more complex syntactically than short, affirmative main clauses (Szmrecsanyi 2003; Szmrecsanyi 2006).

However, most accounts explain the constraints with respect to the grammaticalization of the future markers. On the one hand, the markers undergo semantic bleaching, whereby the lexical meaning is eroded in favor of a grammatical meaning (Hopper & Traugott 2003; see also Bohmann forthcoming). On the other hand, the markers retain some of their features as they become more grammaticalized due to retention (Torres-Cacoullos & Walker 2009: 321). According to this view, *be going to* is preferred in subordinated clauses because *be going to* was first used as a future marker in subordinated clauses and it has retained this tendency (Poplack & Tagliamonte 2000: 334; Tagliamonte, Durham & Smith 2014: 91; Gardner 2017: 143). Additionally, subordinate clauses are less volitional than main clauses, so *be going to* is a better fit than *will*, which was originally used to express volition (Poplack & Tagliamonte 2000: 334). While *will* and *be going to* have supposedly retained some of their earlier features, they are also specializing in certain niches. Interrogative clauses, for example, favor *be going to* very strongly, and *will* is almost categorical in apodoses clauses (Torres-Cacoullos & Walker 2009: 348; Denis & Tagliamonte 2018: 405).

5.1.1 Previous research

The set of constraints and their origin has been studied extensively for native varieties of English (Poplack & Tagliamonte 2000; Tagliamonte, Durham & Smith 2014; Fehringer & Corrigan 2015; Gardner 2017; Denis & Tagliamonte 2018). Since these studies often relate the feature-based constraints to the grammaticalization of the variants, they often involve an apparent time design to compare the influence of the constraints on the choice of variants over time (e.g., Gardner 2017). Alternatively, they look into the influence of the constraints in several dialects that vary with respect to how conservative they are (Tagliamonte, Durham & Smith 2014). Since *be going to* is historically the more recent, incoming form, the underlying assumption is that differences between *will* and *be going to* are gradually being levelled through the ongoing grammaticalization of *be going to*.

Typically, these studies find that speakers of less conservative dialects and younger speakers in general are sensitive to fewer of the above-mentioned constraints related to grammaticalization, implying that the semantic bleaching of the future variants is still ongoing and that retention from earlier stages is gradually diminishing, especially for the innovative

variant *be going to*. More recently, Engel and Szmrecsanyi (2023) have compared the use of the future markers in formal and informal spoken and written British English using a mixed methodology combining a corpus-based analysis with a rating task experiment. Converging evidence from their corpus study and the experiment show that register modulates the effect of most constraints in their study, namely verb type, negation, subject grammatical person, temporal proximity and presence of temporal adverbs.

The alternation has also received some attention in research on ESL speakers (Schneider 2016; Bohmann forthcoming). Using logistic regression, Schneider compares how the future variants are used in spoken and written Ghanaian English compared to British English with regard to agentivity, negation, sentence type, subject type and the presence of definite or indefinite temporal adverbials. Schneider finds that the future markers generally follow the same patterns of use as in Ghanaian and British English, although the distribution of the forms is less systematic in Ghanaian English, which is only conditioned by sentence type (Schneider 2016). Bohmann (forthcoming) expands the analysis to 8 constraints compared across three Inner Circle varieties of English (Canada, New Zealand, and USA) and five Outer Circle varieties (Hong Kong, Singapore, Jamaica, India, Sri Lanka). While there are differences in the overall frequencies of *will* and *be going to*, the constraints affect the choice of variant in the same way. On the other hand, mode plays a role, in line with results from Engel and Szmrecsanyi (2023) regarding register effects on British English native speakers.

However, no previous study has researched how the different constraints simultaneously affect the choice between *will* and *be going to* for EFL learners, and how their probabilistic grammar compares with that of native speakers. The most similar study is the investigation of Deshors and Gries (2014) into the use of modals *can* and *may* in the written interlanguage of French learners of English, another substitution alternation also influenced by negation, the animacy of the subject and the type of sentence and verb (Deshors & Gries 2014). They first used the Behavioral Profile approach to compare the use of *may* and *can* for native speakers and French learners with French *pouvoir*, and then compared native speakers' use of *can* and *may* with that of learners using mixed-effects logistic regression. Learners were found to diverge from native speakers with respect to certain constraints. In negated clauses with plural, animate subjects, learners are more likely to use *can* than native speakers. Provided negated clauses with plural subjects are more complex than affirmative clauses with singular, the authors claim that learners pick the more frequent *can* in such contexts because they are more sensitive to the

Complexity Principle (Rohdenburg 1996: 151), reminiscent of the conclusion of Szmrecsanyi (2006) regarding the future markers.

What we do know about the acquisition of the future markers is that EFL learners acquire *will* before *be going to*, around the same time as lexical futures such as *hope to* (Bardovi-Harlig 2005). Hence, learners start using *be going to* while they already have at their availability the marker *will*, an invariable future time expression with largely the same function. The distribution of *will* and *be going to* is known to be an area of difficulty for EFL learners, often leading to unidiomatic utterances (Wekker 1976: 123; Haegeman 1989: 292). Interestingly, comparisons across studies suggest that learners' choice of future marker is not very much influenced by the learner's L1 background (Bardovi-Harlig 2000: 411).

5.1.2 Present study

The present chapter contributes to previous research by providing the first-ever multifactorial analysis of EFL learners' choice between *will* and *be going to*. Within the rather limited number of studies on alternation phenomena in EFL, only one study has investigated a similar substitution alternation, namely the study of Deshors and Gries (2014) on the use of *can* versus *may* in the written interlanguage of French EFL learners. In the current study, the future marker alternation is researched for a much wider range of L1 backgrounds in spoken language production, which offers a better window into the underlying grammatical system of learners according to SLA than written language (Myles & Mitchell 2014). Moreover, the aim of Deshors and Gries' (2014) study is to promote the use of multifactorial methods in corpus-based research on learner language and is therefore mainly methodological, whereas the current study explicitly tests hypotheses generated by the theory-based account developed in Chapter 3. Finally, the current study explicitly tests the effect of proficiency level on the choice of future marker. This answers the need expressed by previous studies on the acquisition of the future markers in English to include a broader range of proficiency levels, or to observe learners for a longer period of time in a longitudinal study design (Bardovi-Harlig 2000: 411–412).

On the methodological plane, I collected thousands of interchangeable instances of *will* and *be going to* from the TLC (Gablasova, Brezina & McEnery 2019). These were annotated for variables found to be relevant in the literature, specifically negation, structural persistence, sentence type, clause type, verb type, clause length, temporal proximity, presence of temporal adverb, as well as subject grammatical person and animacy

(Szmrecsanyi 2006; Denis & Tagliamonte 2018; Engel & Szmrecsanyi 2023). Then, the annotated dataset was analyzed with mixed-effects logistic regression to answer the research questions of the project:¹

- RQ 1: How does the set of probabilistic constraints of learners compare to that of native speakers at different proficiency levels?
- RQ 2: How does proficiency level modulate the probabilistic grammar of learners?
- RQ 3: How do native speakers and learners at different proficiency levels compare regarding processing-related constraints?

As stated in section 3.4, for the first research question I expect that i) more advanced learners will better approximate the target probabilistic grammar, ii) learners will only be sensitive to constraints that also affect native speakers, and iii) lower-proficiency learners will be sensitive to fewer constraints than native speakers. For the second research question, I hypothesize that lower-proficiency learners will struggle with feature-based constraints that are difficult to learn via associative statistical learning rather than constraints that are easy to learn. In the case of the future marker alternation, this means that learners should first acquire clause type and temporal proximity before the other feature-based constraints, namely sentence type, verb type, presence of temporal adverb, negation, subject animacy and grammatical person (see Section 3.4). For the third research question, I hypothesize that lower-proficiency learners should be more sensitive to processing effects than native speakers, which refers to structural persistence and clause length in the context of the future marker alternations.

The remainder of the chapter is structured as follows. In the following section, I present the data and methodology of the study, including the delimitation of the variants, the collection and annotation of the data, the systematic classification of the constraints as either feature-based or processing-related, and the method of statistical analysis, namely mixed-effects logistic regression. Section 5.3 then presents the results of the regression model for the random effects and the fixed effects as well as some supplementary analyses. Finally, Section 5.4 discusses the results in light of the hypotheses of the research questions.

¹All data and code are available in an online repository at <https://osf.io/6ng7x/>.

5.2 Data and Methods

5.2.1 Data collection

Delimiting the variable context

Not all instances of future markers *be going to* or *will* are interchangeable. Following the approach of Denis and Tagliamonte (2018; see also Engel & Szmrecsanyi 2023), the noun *will*, *will* in tag questions (14) and *will* with an elliptical subject or verb (15) are not considered interchangeable. The same applies for lexical uses of *go* (16) and future-in-the-past contexts (17), which almost categorically prefer *be going to* (Poplack & Tagliamonte 2000; Berglund 2000: 31; Aarts, Close & Wallis 2013: 36; Fehringer & Corrigan 2015). Additionally, the observations should be structurally complete, so they should not be missing a subject or verb as in (18). For *be going to* specifically, they should not be missing an auxiliary as in (19) and the auxiliary should not be anything other than *be*, like *have* in (20).

- (14) that will interrupt their career won't it? (TLC:2_CH_2, NS)
- (15) in conclusion abortion is and will be erm really controversial topic always (TLC:2_ME_8, C1 Spanish)
- (16) actually I'm gonna US this April (TLC:2_6_IN_105, B1 Marathi)
- (17) I thought er we were going to get lost because I have never taken the subway without and adult (TLC:2_7_SP_34, B2 Spanish)
- (18) I dunno if I'll if if I'm okay but I feel bad (TLC:2_6_ME_4, B1 Spanish)
- (19) she going to change her mind (TLC:2_7_SP_23, B2 Spanish)
- (20) I've going to talk about playing tennis my favourite hobby (TLC:2_6_CH_6, B1 Chinese)

Identifying potential future markers

First, I identified all potentially interchangeable instances of *be going to* and *will* in the TLC (Gablasova, Brezina & McEnery 2019). This was straightforward since there is a specific CLAWS6 tag for future marker *be going to*, namely 'VVGK'. For *will*, I looked for the forms *will*, *'ll* and *wo* followed by the modal verb tag 'VM'. These searches yielded a large number of potentially interchangeable instances of *will* and *be going to* as shown in Table 5.1.

Table 5.1: Distribution of the potentially interchangeable future markers across native speakers and learners. Percentages indicate the proportion of each variant. For the row total, the percentages indicate the proportion of tokens that native speakers and learners contribute to the total number of tokens.

	Will	Be going to	Row total
Native speakers	5,501 (62.1%)	3,356 (37.9%)	8,857 (54.7%)
Learners	5,918 (80.5%)	1,429 (19.5%)	7,347 (45.3%)
Column Total	11,419 (70.5%)	4,785 (29.5%)	16,204

In total, there are 16,204 potential future markers, 8,857 (54.7%) of which were produced by the examiners in the corpus. Learners contributed the remaining 7,347 (45.3%) potential future markers. For native speakers, *will* accounts for about two-thirds of all future markers, with *be going to* covering the remaining third of the data. In the learner corpus data, *be going to* is less frequent, accounting for only a fifth of the data. Table 5.2 presents the distribution of *will* and *be going to* for the learner data across proficiency levels.

Table 5.2: Distribution of the potentially interchangeable future markers across the different learner proficiency levels. Percentages indicate the proportion of each variant for a given proficiency level. For the row total, the percentages indicate the proportion of tokens that each proficiency level contributes to the total number of learner tokens.

	Will	Be going to	Row total
B1	1,908 (87.2%)	281 (12.8%)	2,189 (29.8%)
B2	2,520 (84.1%)	476 (15.9%)	2,996 (40.8%)
C1	1,152 (70.7%)	478 (29.3%)	1,630 (22.2%)
C2	338 (63.5%)	194 (36.5%)	532 (7.2%)
Column Total	5,918 (80.5%)	1,429 (19.5%)	7,347

Table 5.2 shows that the distribution of *will* and *be going to* varies considerably for learners across proficiency levels. For B1 learners, *will* accounts for almost 90% of the future markers. The proportion of *be going to* increases with the proficiency level of the learners, with C2 learners approximating native speakers, where *be going to* accounts for about one-third of the data. This suggests that *will* is used much more often by low-proficiency learners compared to more advanced learners and native speakers, possibly because *will* is the first future marker variant learners acquire (Bardovi-Harlig 2005). However, this preliminary distribution is based on a rather simplistic retrieval of the future markers, so one cannot exclude the possibility that this pattern is somehow due to certain aspects of the data at lower proficiency levels.

Learner data interchangeability check

To ensure the interchangeability of the variants in studies such as this, the collected observations need to be manually checked. This process is usually labor-intensive, especially when dealing with messy spoken learner data featuring repetitions and reformulations. To illustrate, example (21) features two interchangeable future markers *you will get* and *you'll get*. Since these future markers share the same subject, future marker *will* and infinitive verb, it is reasonable to assume that the second future marker represents a repetition or reformulation of the first. In this case, I would only include the last repetition or the reformulated form, namely *you'll get*.

- (21) school are now more like erm more bigger and you will get more
so erm you'll get more (TLC:2_8_RU_3, B2 Russian)

To keep the workload manageable, I only checked the interchangeability of a sample of the collected observations that would suffice for statistical analysis. Specifically, mixed-effects logistic regression is limited by the number of observations from the least frequent variant (Hosmer & Lemeshow 2000: 346–347), in this case *be going to*. Hence, I manually determined the interchangeability of all 1,429 instances of *be going to* in the learner data. In a second step, I identified at least as many interchangeable instances of *will* from the learners. Although this approach obscures the underlying distribution of the variants, a balanced dataset of *will* and *be going to* allows me to investigate the probabilistic conditioning of the variants in accordance with the research goals of the thesis.

In the 1,429 potentially valid instances of *be going to* in the learner data, 1,266 cases were interchangeable across 514 exams, and by extension, 514 EFL learners. To match this number, I set out to collect approximately 1,300 interchangeable cases of *will* from the learner data. To determine how many instances of *will* I had to check manually to reach this number, I first tested how many instances of *will* are interchangeable in a small sample of 200 utterances from the corpus, which contained 136 instances of *will* produced by learners. Out of these 136 instances, 104 (76%) were interchangeable. To arrive at 1,300 interchangeable cases of *will*, I therefore had to manually check approximately 1,700 potentially valid instances of *will* from the 514 exams featuring interchangeable cases of *be going to*. Collecting both future markers from the same learners ensures that they have acquired both future markers like the native speakers for a more accurate comparison of probabilistic grammars. Since these instances of *will* are embedded in the utterances of the speakers in the corpus, I randomly selected utterances containing at least one potentially

interchangeable instance of *will*. To arrive at the required 1,700 potentially interchangeable instances of *will* within these utterances, I determined that I would need to read through 75% of them. Hence, I ignored every fourth utterance I encountered. Within the remaining utterances, I found 1,570 interchangeable cases of *will*. The distribution of the valid future markers for the different proficiency levels and L1 backgrounds in the learner data is presented in Table 5.3.

Table 5.3: Distribution of the future markers across proficiency levels and L1 backgrounds. Percentages indicate the proportion of tokens that the L1 background contributes to the total tokens for a given proficiency level. For the row total, the percentages indicate the proportion of the tokens each L1 contributes to the total number of tokens.

	B1	B2	C1	C2	Row total
Chinese	59 (10.3%)	102 (9.5%)	98 (11.3%)	31 (9.8%)	290 (10.2%)
Gujarati	59 (10.3%)	31 (2.9%)	0	15 (4.7%)	105 (3.7%)
Hindi	50 (8.7%)	145 (13.5%)	0	30 (9.5%)	225 (7.9%)
Italian	49 (8.5%)	94 (8.8%)	96 (11.1%)	23 (7.3%)	262 (9.3%)
Kannada	0	0	19 (2.2%)	7 (2.2%)	26 (0.9%)
Malayalam	67 (11.7%)	0	0	0	67 (2.4%)
Marathi	38 (6.6%)	45 (4.2%)	26 (3%)	22 (6.9%)	131 (4.6%)
Russian	30 (5.2%)	78 (7.3%)	0	0	108 (3.8%)
Spanish	192 (33.4%)	429 (39.9%)	571 (65.9%)	138 (43.5%)	1,330 (47%)
Tamil	31 (5.4%)	130 (12.1%)	0	22 (6.9%)	183 (6.5%)
Telugu	0	0	6 (0.7%)	0	6 (0.2%)
Unknown	0	20 (1.9%)	50 (5.8%)	29 (9.1%)	99 (3.5%)
Column total	575 (20.3%)	1,074 (37.9%)	866 (30.6%)	317 (11.2%)	2,832

Table 5.3 shows that learners from all proficiency levels and L1 backgrounds contribute interchangeable future markers. Figure 5.1 summarizes the information from the table by visualizing how the most common L1 backgrounds are distributed across the proficiency levels. The figure shows that the largest proportion of tokens stems from Spanish learners of English for all proficiency levels and especially the C1 level, where they account for two-thirds of all observations. Chinese and Italian learners of English are represented almost equally in each of the proficiency levels. Hindi is also a prominent L1 background, although these learners do not contribute any observations at the C1 level. The same applies for Tamil,

which is mostly represented at the B2 level. The remaining tokens stem from the remaining six L1 backgrounds.

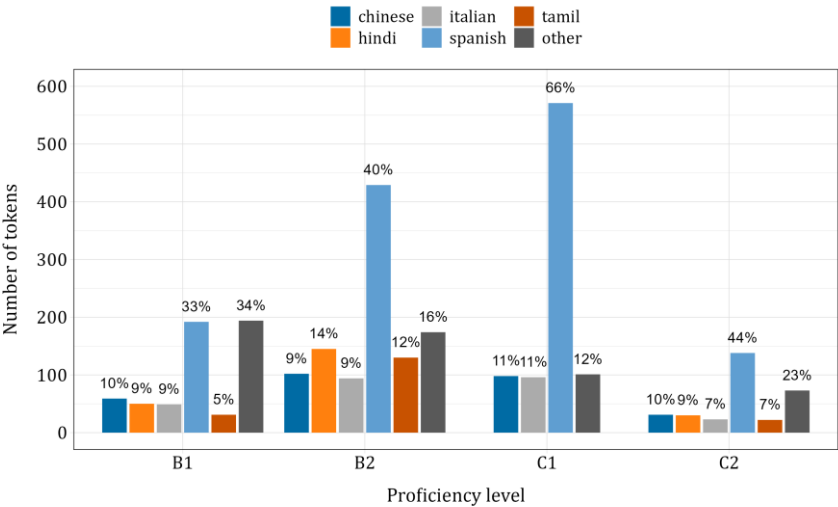


Figure 5.1: Distribution of the tokens across proficiency levels. The bars show the proportion of tokens the L1s contribute to each proficiency level.

Native speaker data interchangeability check

The native speaker data comes from the examiners in the same 514 exams where I identified the interchangeable future markers produced by learners. This facilitates the operationalization and annotation of structural persistence, which will help control for the influence the interlocutors have on each other. Additionally, this ensures that native speakers and learners speak about similar topics, which makes their data more comparable (see Section 4.1.1.2). In these exams, examiners produced 1,560 potentially interchangeable cases of *be going to* and 2,889 cases of *will*. In a random sample of 200 utterances, I first determined what proportion of future markers are interchangeable. Compared to learners, native speakers produce less structurally incomplete future markers, so a larger proportion of future markers are interchangeable, namely 94 out of 115 (82%) instances of *will* and 147 out of 163 (90%) instances of *be going to*. However, a closer look at the data revealed that as many as half of these interchangeable instances of *be going to* and about a third of the instances of *will* were scripted, i.e. they are part of instructions from the examiner, as in (22), or they appear in fixed prompts specific to a given task in the exam, as in (23). Both cases occur with very little modifications across multiple

exams and thus do not represent the spontaneous spoken language production of the examiner.

- (22) now we are going to do the interactive task okay and in this part I will tell you something and then you have to ask me questions to find out more information (TLC:2_7_CH_45, NS)
- (23) scientists say they will soon be able to prodro= prolong general life expectancy to well over the age of a hundred I'm not at all convinced of the desirability of this development (TLC:CH_30, NS)

To remove these cases, I first identified the observations that were potentially scripted. That is, I grouped together the examiner utterances that have the same subject, future marker and infinitive verb. Since examiners within a single exam might simply repeat the same utterance containing a future marker observation, future markers with the same subject and infinitive verb had to occur in at least two different exams. An example of a pair of potentially scripted future markers is illustrated in (24) and (25), which feature clauses from two different exams with the exact same subject, future marker and infinitive verb.

- (24) I'll take your point (TLC:2_IT_4, NS)
- (25) I'll take your questionnaire (TLC:2_8_IT_13, NS)

Then, I compared the similarity of the future marker clauses by counting the word tokens they have in common. Since the subject, future marker and infinitive verb are the same by default, these elements were not included in the calculation. This means that I compared *your point* for (24) and *your questionnaire* for (25), which reveals that both clauses only share the word *your*, so the similarity between the clauses is 50%. If the similarity between any two pairs of clauses was higher than 70%, they were considered scripted and subsequently removed from the data. The similarity of 70% was determined based on a test on the genitive alternation data, which showed that a threshold of 70% captures most scripted utterances, while at the same time being high enough to prevent unscripted utterances from being removed (see Section 6.2.2.3).

These scripted utterances increased the number of potentially interchangeable instances of *will* and *be going to* I had to manually check to arrive at a satisfactory number of interchangeable tokens for statistical analysis. Given that these markers are embedded in speaker utterances, this meant that I had to check more utterances containing these future markers. Ultimately, the number of non-scripted, interchangeable future markers was even lower than expected from the small sample of 200 utterances. As

a result, I had to manually check all utterances featuring *be going to* produced by the examiners in the 514 relevant exams. For *will*, I scanned through half of the relevant utterances. Out of approximately 1,250 cases of *be going to* and 1,250 cases of *will* that I checked manually, 347 interchangeable, non-scripted instances of *be going to* and 442 cases of *will* were retained. These 784 examiner observations were produced by 66 examiners in total.

Extracting interchangeable future markers

The interchangeable cases of *will* and *be going to* were extracted for data annotation. In the TLC, the data is organized in speaker turns without punctuation. This means that a clause can easily be interrupted by the interlocutor. As a result, the turn where a speaker produced a valid future marker often does not contain sufficient material to annotate for constraints requiring the entire future marker clause like clause type or clause length. At the same time, many interruptions that divide future marker clauses across multiple speakers turns do not meaningfully affect discourse (e.g., *yes, right, uh, I see*) as in (26).

- (26) Candidate: but then they restarted
 Examiner: mm
 Candidate: to have a concerts
 (TLC:2_6_RU_18, B1 Russian)

Therefore, I extracted not only the speaker turn with the relevant future marker, but also the two preceding and the two following speaker turns if present. This way, I capture the utterance of the speaker producing the future marker, as well as their previous utterance and their following utterance with any potential interruptions from the interlocutor. In total, I extracted 3,616 observations of future markers from the corpus.

Since the utterance containing the future marker with the two preceding and two following utterances often contained quite a lot of linguistic material, it was tricky to find and annotate the relevant future marker. To facilitate the annotation of the constraints, I extracted the clause containing the future marker and the clauses it is connected to with the subordinators *if, after, as, so, because, before, once, since, like*, (optional) *that, (al)though, till, until, unless, provided, except, in case, as though, whereas, while, whether* and other *wh*-words, as in (27). (Quirk et al. 1985: 998–1000).

- (27) having a university degree always will help you because you will be more prepared for some things than other people (TLC:2_SP_18, C1 Spanish)

In this way, future marker clauses were extracted with the clauses to which they are linked as main or subordinate clause, which suffices to annotate all probabilistic constraints. Moreover, this made it particularly easy to annotate the future marker for clause type, which is determined based on the presence of these subordinators (see Section 5.2.2.3). One difficulty is that subordinate clauses often precede their main clause, in which case the subordinator was not between the clauses like *when* in (28). Moreover, relative clauses (29) and complement clauses (30) often occurred without any overt relative pronoun or complementizer.

- (28) when she's sick she won't be driving right and (TLC:2_7_IN_32, B2 Hindi)
- (29) the topic erm I'm going to explain you is er about the nine eleven conspiracy theory (TLC:2_IT_17, C1 Italian)
- (30) do you think er radiation is not gonna be a big problem in the future? (TLC:ME_29, C1 Spanish)

5.2.2 Data annotation

The future markers were then annotated for the proficiency level of the speaker as reported in the corpus metadata as well as the different probabilistic constraints affecting a speaker's choice of variant, including negation, structural persistence, sentence type, verb type, clause type, clause length, temporal proximity, presence of temporal adverbs, subject grammatical person and animacy (e.g., Torres-Cacoulllos & Walker 2009; Denis & Tagliamonte 2018; Engel & Szmrecsanyi 2023). The probabilistic constraints, their effect on the choice of variant, and their operationalization are presented one-by-one in the following sections.

5.2.2.1 Sentence type

Sentence type is perhaps the strongest predictor of future marker variant (Gardner 2017: 135). Previous research has found that interrogative clauses strongly prefer *be going to*, to the extent that the preference is almost categorical (Szmrecsanyi 2003; Tagliamonte, Durham & Smith 2014; Denis & Tagliamonte 2018: 424). According to Torres-Cacoulllos and Walker (2009: 343), *be going to* is frequently used in questions about intentions

and plans decided beforehand, which stems from the fact that *be going to* was originally used to indicate that an agent is on a path towards a goal.

In the dataset, interrogative clauses are either introduced by *wh*-words or they feature subject-verb inversion (Biber et al. 1999: 212; see Engel & Szmrecsanyi 2023). Most interrogative clauses starting with *wh*-words also feature inversion, except when the interrogative clause is a main clause with a *wh*-word as its subject, as in (31). Relative clauses, complement clauses and other clausal objects also cannot feature subject-verb inversion, but their matrix clause can, as in (32). Since these subordinate clauses with future marker are tightly embedded in the matrix clause featuring inversion, I also consider them interrogative. Finally, question tags also make clauses interrogative without inversion, as in (33). All other clauses are considered declarative.

- (31) what's gonna happen (TLC:8_ME_10, B2 Spanish)
- (32) do you think the women will be very sensible? very careful with their money (TLC:2_6_IN_116, NS)
- (33) if we managed to de= to destroy this planet we'll have to live somewhere won't we? (TLC:SP_43, NS)

An overview of the distribution of the types of sentences across the future marker variants *will* and *be going to* is shown for native speakers and learners in Table 5.4 and represented visually in Figure 5.2.

Table 5.4: Contingency table for sentence type for native speakers ($X^2 = 3.61$, $df = 1$, $p = 0.057$) and learners ($X^2 = 4.16$, $df = 1$, $p = 0.042$, Cramér's $V = 0.04$). Percentages indicate the proportion of future marker variants per category of sentence type. For the row total, the percentages indicate the total proportion of declarative and interrogative clauses. If the test is significant, values that are significantly lower than expected according to standardized residuals are highlighted in orange, whereas values that are significantly higher than expected according to standardized residuals are highlighted in blue.

Native speakers	<i>Be going to</i>	<i>Will</i>	Row total
Declarative	246 (46.2%)	287 (53.8%)	533 (68%)
Interrogative	97 (38.6%)	154 (61.4%)	251 (32%)
Column total	343 (43.8%)	441 (56.2%)	784
Learners	<i>Be going to</i>	<i>Will</i>	Row total
Declarative	1,206 (44.2%)	1,524 (55.8%)	2,730 (96.4%)
Interrogative	56 (54.9%)	46 (45.1%)	102 (3.6%)
Column total	1,262 (44.6%)	1,570 (55.4%)	2,832

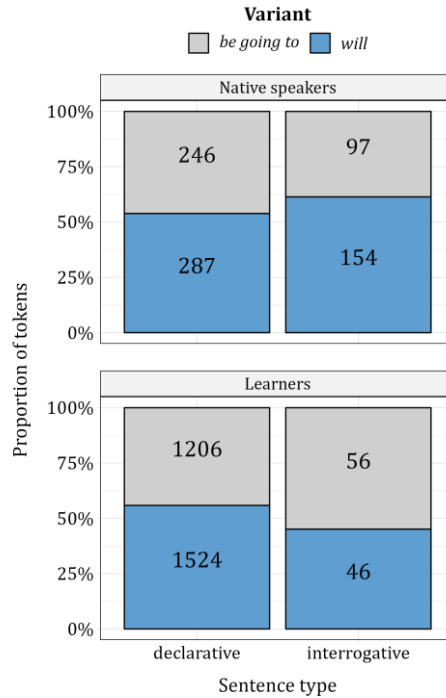


Figure 5.2: Distribution of *will* and *be going to* across levels of sentence type for native speakers (top) and learners (bottom).

Interestingly, the choice of future marker is not significantly associated with the type of sentence for native speakers. Learners, on the other hand, seem to prefer *be going to* in interrogatives as expected from previous research (Tagliamonte, Durham & Smith 2014; Denis & Tagliamonte 2018: 424). The association is very weak, however (Cramér's $V = 0.04$).

5.2.2.2 Negation

Most previous research investigating the effect of negation on the choice of future marker find that negative contexts typically favor *be going to* (Nesselhauf 2010: 173; Tagliamonte, Durham & Smith 2014: 98–99). However, this finding is not always replicated, as neither Fehring and Corrigan (2015), Gardner (2017) nor Engel and Szmrecsanyi (2023) (in spoken registers) found such an effect for negation. This is partially due to the fact that previous studies have not always considered the same negators, which do not all affect the choice of variant in the same way. While most negators prefer *be going to*, *never* strongly prefers the *will*-paradigm in structures like *I'll never...* (Torres-Cacoullous & Walker 2009: 340;

Tagliamonte, Durham & Smith 2014; Gardner 2017: 146). The overall preference for *be going to* could be explained in terms of processing: negated clauses are more complex syntactically, so speakers may prefer to use the longer *be going to* variant in accordance with Rohdenburg's Complexity Principle (1996: 151; see Szmrecsanyi 2003: 317).

For a future marker clause to be considered negative in the data, the clause featuring the future marker had to contain either *never*, *nothing*, *none*, *not*, *no*, *nothing*, *none*, *neither*, *nor*, *nowhere*, *no place* or *nobody*, as in example (34) (Huddleston & Pullum 2002: 789). In case it was not clear whether the clause is negated, I tested whether the clause could be continued with *not even when...* (see Huddleston & Pullum 2002: 789).

(34) you 'll never throw a paper or a cigarette on the floor (TLC:2_IT_6, C2 Italian)

The distribution of the future marker variants across negative and affirmative clauses is shown in Table 5.5 and visualized in Figure 5.3.

Table 5.5: Contingency table for negation for native speakers ($X^2 = 3.31$, $df = 1$, $p = 0.069$) and learners ($X^2 = 16.85$, $df = 1$, $p < 0.001$, Cramér's $V = 0.08$). Percentages indicate the proportion of future marker variants per category of negation. For the row total, the percentages indicate the total proportion of affirmative and negative clauses. If the test is significant, values that are significantly lower than expected according to standardized residuals are highlighted in **orange**, whereas values that are significantly higher than expected according to standardized residuals are highlighted in **blue**.

Native speakers	<i>Be going to</i>	<i>Will</i>	Row total
Affirmative	311 (42.8%)	416 (57.2%)	727 (92.7%)
Negative	32 (56.1%)	25 (43.9%)	57 (7.3%)
Column total	343 (43.8%)	441 (56.2%)	784
Learners	<i>Be going to</i>	<i>Will</i>	Row total
Affirmative	1,165 (45.9%)	1,374 (54.1%)	2,539 (89.7%)
Negative	97 (33.1%)	196 (66.9%)	293 (10.3%)
Column total	1,262 (44.6%)	1,570 (55.4%)	2,832

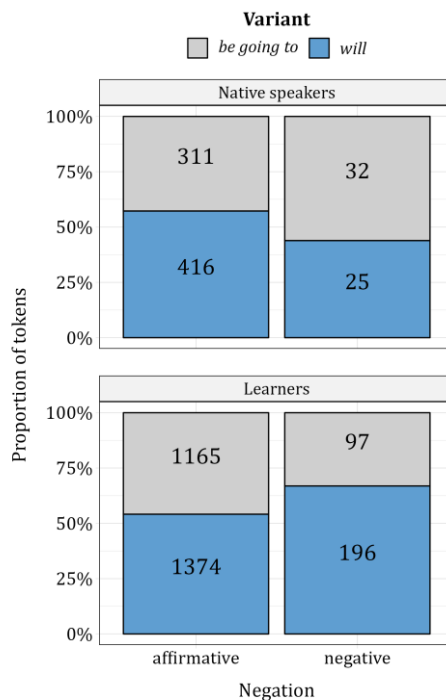


Figure 5.3: Distribution of *will* and *be going to* across levels of negation for native speakers (top) and learners (bottom).

Table 5.5 shows that there is no association between negation and the choice of future marker for native speakers. For learners, there is significant but weak association indicating that learners have a slight tendency to choose *will* when the clause is negated, which goes against most findings from previous research (Nesselhauf 2010: 173; Tagliamonte, Durham & Smith 2014: 98–99).

5.2.2.3 Clause type

Be going to is systematically preferred in subordinate clauses whereas *will* is drawn to main clauses (Tagliamonte, Durham & Smith 2014: 98; Fehrer & Corrigan 2015: 216; Gardner 2017: 153; Denis & Tagliamonte 2018). Poplack and Tagliamonte (1999: 334, see also Fehrer & Corrigan 2015: 212) suggest that *be going to* is more likely in subordinate clauses because subordinate clauses are less volitional, which is more compatible with *be going to* than *will*, which was originally used to express volition. Haegeman (1989: 308), on the other hand, observes the tendency for *be going to* to follow matrix clauses such as *I think that, I did not realize that*

(see also Denis & Tagliamonte 2018: 346 for a similar finding). She claims that this is due to the fact that these matrix clauses link the subordinate clause to the present situation, which favors *be going to*. Szmrecsanyi (2003: 306) also finds that *be going to* is preferred in subordinate clauses and that *will* is preferred in main clauses, but he explains this finding with respect to processing. According to Rohdenburg's Complexity Principle (Rohdenburg 1996: 151), the more explicit, longer variant is preferred in cognitively complex environments. Subordinate clauses are more syntactically complex than main clauses, so speakers prefer the longer *be going to* variant, which gives them more time to process the utterance while also giving more time for the hearer to process what is being said. The same explanation accounts for why the shortest variant *'ll* is preferred in main clauses. Finally, it is also possible that *be going to* is preferred in subordinate clauses because the future marker first occurred in subordinate clauses, a tendency it might have retained (Poplack & Tagliamonte 2000: 334; Tagliamonte, Durham & Smith 2014: 91; Gardner 2017: 143).

At the same time, *will* is specializing in apodoses, whereas protasis clauses introduced by *if* strongly prefer *be going to* (Gardner 2017: 153; Denis & Tagliamonte 2018; Engel & Szmrecsanyi 2023; but see Tagliamonte, Durham & Smith 2014, who only found an effect for the distinction between main clause and subordinate clause), although this effect is weaker in informal language (Szmrecsanyi 2003: 309; see also Brisard 2001: 265). Note that although the protasis strongly favors *be going to*, it is also possible for *be going to* to figure in the apodosis, especially in formal British English (Szmrecsanyi 2003: 311). One possible explanation for the preference for *will* in apodoses is that *will* conveys more uncertainty, which makes it more compatible with apodoses as they depend on whether the content of the *if*-clause is true (Fehringer & Corrigan 2015: 213; Gardner 2017: 143). In a similar vein, Brisard (2001: 267) claims that *will* is semantically compatible with apodoses because *will* is elliptical and assumes some part as understood in discourse, which is given by the protasis.

In the present data, subordinate clauses introduced by *if* like in (35) are considered protases, and their main clauses are coded as apodoses, as in (36). Other subordinate clauses are introduced by *after*, *as*, *because*, *before*, *once*, *since*, *like* (except for direct speech), (optional) *that*, (al)though, *till*, *until*, *unless*, *provided*, *except*, *in case*, *as though*, *whereas*, *while*, *whether* and other *wh*-words, as in (27). (Quirk et al. 1985: 998–1000). All other clauses are marked as main clauses.

- (35) if it's going to be a very smooth life I don't think it is gonna be interesting (TLC:CH_30, C2 Chinese)

- (36) if you've stopped taking the steroids er your body will change completely (TLC:2_SP_12, C1 Spanish)

The annotation of clause type for future marker clauses introduced by *so* depends on the way *so* is used. *So* is used as a coordinator between main clauses if one can replace *so* by *therefore* or rephrase the sentence with *because*, or when it is possible to place *and* before *so* (see Engel 2022). For instance, one can rephrase (37) as *I think ninety percent of parents will let their children to learn English because they think English is very important*. Alternatively, if *so* can be replaced with *so that*, then it is used to introduce a subordinate clause that expresses the purpose of the preceding main clause, as in (38). If *so* introduces a future marker clause that has a different sentence type than the clause to which it is connected, then I consider the *so*-clause to be a main clause as in (39).

- (37) they think English is very important so erm I think ninety per cent of parents will let erm their children to learn English (TLC:CH_11, C1 Chinese)
- (38) you say that you you plan to set up a little industry so the women of the area will earn some money for them for their families (TLC:2_6_IN_116, NS)
- (39) we're already tearing down the forest and there is no place to live so how are we going to manage? (TLC:IN_6, C2 Unknown)

Although one could argue that *she'll like* in (40) is a matrix clause, I do not consider direct speech to be a subordinate clause by default. Rather, I consider instances of direct speech in their own right, meaning that the future marker clause in direct speech can either be a main clause or a subordinate clause depending on its own syntactic structure. This means that in (40), both the matrix clause *she'll like you've gotta study study study all the time* and the clause within the reported speech *you've gotta study study study all the time* are considered main clauses.

- (40) She'll like you've gotta study study study all the time (TLC:2_7_IN_50, B2 Hindi)

It is important to note that this operationalization combines two constraints. There is a broad classification between main clauses and subordinate clauses, and then there is the more specific distinction between protasis, apodosis and other clauses (Gardner 2017: 144). This means that an apodosis, and by extension, a protasis, can in fact be subordinated, as in example (41), where *I think that* introduces the apodosis, which in turn

governs the protasis. While this technically means that these cases should be coded as subordinate clauses, the distinction between apodosis and protasis is more specific than the distinction between main clauses and subordinate clauses, so these cases are coded as 'apodosis' and 'protasis'.

- (41) I think that he will be very happy and very pleased if he's going to live there (TLC:8_ME_28, B2 Spanish)

Table 5.6: Contingency table for clause type for native speakers (Fisher's exact test $p < 0.001$, Cramér's $V = 0.16$) and learners ($X^2 = 56.13$, $df = 3$, $p < 0.001$, Cramér's $V = 0.14$). Percentages indicate the proportion of future marker variants per category of clause type. For the row total, the percentages indicate the total proportion of the different clause types. If the test is significant, values that are significantly lower than expected according to standardized residuals are highlighted in **red**, whereas values that are significantly higher than expected according to standardized residuals are highlighted in **blue**.

Native speakers	<i>Be going to</i>	<i>Will</i>	Row total
Apodosis	10 (21.7%)	36 (78.36%)	46 (5.9%)
Main clause	234 (46.5%)	269 (53.5%)	503 (64.2%)
Protasis	6 (100%)	0	6 (0.1%)
Subclause	93 (40.6%)	136 (59.4%)	229 (29.2%)
Column total	343 (43.8%)	441 (56.2%)	784
Learners	<i>Be going to</i>	<i>Will</i>	Row total
Apodosis	85 (25.8%)	244 (74.2%)	329 (11.6%)
Main clause	830 (46.8%)	942 (53.2%)	1,772 (62.6%)
Protasis	25 (61%)	16 (39%)	41 (0.1%)
Subclause	322 (46.7%)	368 (53.3%)	690 (24.4%)
Column total	1,262 (44.6%)	1,570 (55.4%)	2,832

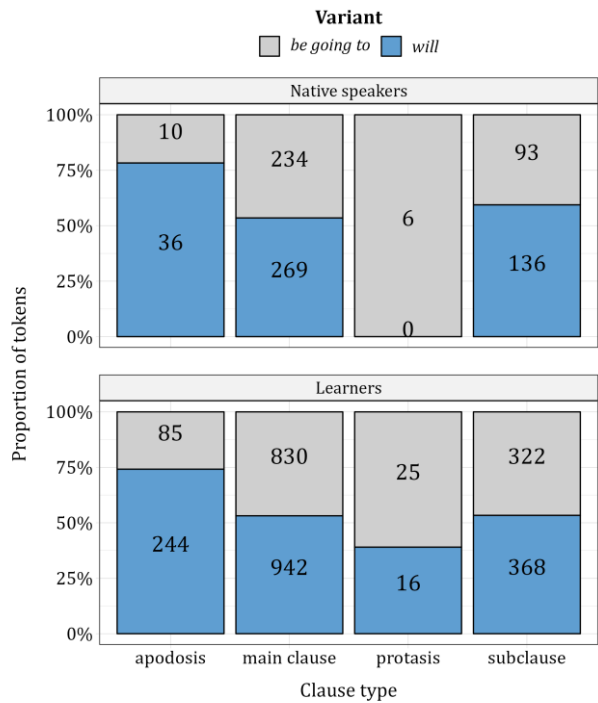


Figure 5.4: Distribution of *will* and *be going to* across levels of clause type for native speakers (top) and learners (bottom).

Table 5.6 offers an overview of the distribution of the future markers across the four types of clauses. This distribution is visualized in Figure 5.4. Table 5.6 shows that both native speakers and learners prefer *will* in apodoses and *be going to* in protases and, surprisingly, main clauses. Subordinate clauses, which have consistently been found to prefer *be going to*, do not appear to contribute to the significant association between clause type and the future markers (Tagliamonte, Durham & Smith 2014: 98; Fehringer & Corrigan 2015: 216; Gardner 2017: 153; Denis & Tagliamonte 2018). For both native speakers and learners, the effect size remains small.

5.2.2.4 Persistence

Structural persistence captures the fact that structures that have been used previously tend to be used again: “language users are creatures of habit” (Szmrecsanyi 2006: 1). In a construction grammar account of SLA, priming results from a previous use of a construction raising the activation level of said construction with the consequence that this construction is easier to produce again (e.g., Ellis 2006: 102; Ellis & Wulff 2019: 44). Structural persistence can occur through production or comprehension, so a speaker

might produce a certain structure either because they or an interlocutor produced this structure before (Bock et al. 2007).

In research on learner language within SLA specifically, priming of constructions implies that learners have an internalized representation of the construction in their grammar (e.g., Gries & Wulff 2005: 186; Romano 2018: 276). The future marker variants are obviously considered constructions given that *be going to* is perhaps the most popular instance of grammaticalization. While it seems likely that persistence affects the choice of future variant, the constraint has not received much attention in previous research. To the best of my knowledge, the only study that considers persistence is Szmrecsanyi (2006), who found that future marker choice is in fact highly sensitive to persistence. As expected, speakers tend to reuse the last future marker variant they heard or produced, and the size of this effect depends on how recently the variant was used (Szmrecsanyi 2006: 123).

For each observation, I checked which future marker was previously used by either interlocutor in the exam. Since the examiners and the learners are interlocutors in the same exam, it is possible that the examiner and the candidate influence each other. While research has found that persistence through comprehension or production exert a similar influence on structural persistence (Tooley & Bock 2014), the examiner might exert a greater influence on the candidate learner since the examiner speaks the target language and dominates the interaction in this context (Sandlund, Sundqvist & Nyroos 2016: 20; Stanley & Stevenson 2017: 2). To take this factor into account, the previous future marker was also marked for whether it is the learner or the native speaker who produced it. This means that the variable has the following five levels: examiner *will*, candidate *will*, examiner *be going to*, candidate *be going to* and 'none' in case there is no previous future marker. The distribution of the variable is presented in Table 5.7 and visualized in Figure 5.5 for native speakers and learners.

Table 5.7: Contingency table for persistence for native speakers ($X^2 = 27.27$, $df = 4$, $p < 0.001$, Cramér's $V = 0.19$) and learners ($X^2 = 465.29$, $df = 4$, $p < 0.001$, Cramér's $V = 0.41$). Percentages indicate the proportion of future marker variants per category of persistence. For the row total, the percentages indicate the total proportion of the different previous future markers. If the test is significant, values that are significantly lower than expected according to standardized residuals are highlighted in **orange**, whereas values that are significantly higher than expected according to standardized residuals are highlighted in **blue**.

Native speakers	<i>Be going to</i>	<i>Will</i>	Row total
Candidate <i>be going to</i>	72 (45.3%)	87 (54.7%)	159 (20.3%)
Candidate <i>will</i>	81 (40.3%)	120 (59.7%)	201 (25.6%)
Examiner <i>be going to</i>	75 (54%)	64 (46%)	139 (17.7%)
Examiner <i>will</i>	73 (33.3%)	146 (66.7%)	219 (27.9%)
None	42 (63.6%)	24 (36.4%)	66 (8.4%)
Column total	343 (43.8%)	441 (56.2%)	784
Learners	<i>Be going to</i>	<i>Will</i>	Row total
Candidate <i>be going to</i>	338 (59.3%)	232 (40.7%)	570 (20.1%)
Candidate <i>will</i>	166 (18.5%)	732 (81.5%)	898 (31.7%)
Examiner <i>be going to</i>	336 (69.6%)	147 (30.4%)	483 (17.1%)
Examiner <i>will</i>	350 (44.2%)	441 (55.8%)	791 (27.9%)
None	72 (80%)	18 (20%)	90 (3.2%)
Column total	1,262 (44.6%)	1,570 (55.4%)	2,832

Table 5.7 shows that if an examiner previously used *be going to*, they will be more likely to use *be going to*, and conversely, if an examiner used *will*, they are more likely to produce *will* again. However, native speakers are not more likely to repeat either future marker when it is a candidate learner who previously produced it. Candidate learners are more likely to use *be going to* when either they or an examiner previously produced the variant. Learners are more likely to use *will* if they previously used *will*, but they do not seem to reuse *will* after an examiner. In general, the constraint seems to be more strongly associated with the choice of future marker for learners (Cramér's $V = 0.41$) than for native speakers (Cramér's $V = 0.19$).

Interestingly, both native speakers and learners are more likely to use *be going to* when there is no previous future marker. This is most likely due to the structure of the exam, which starts with the candidate and examiner discussing what preselected topic they will address for the presentation or discussion task of the exam, as shown in (42).

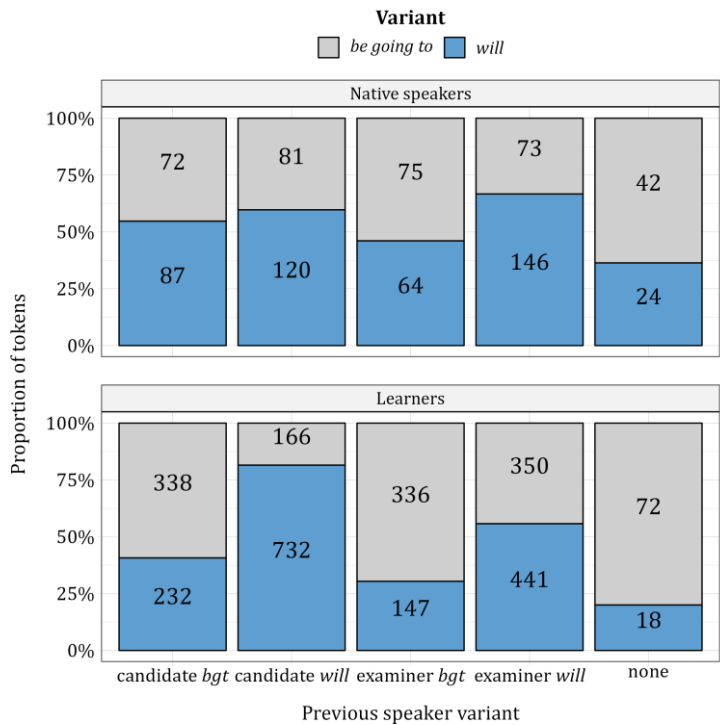


Figure 5.5: Distribution of *will* and *be going to* across levels of persistence for native speakers (top) and learners (bottom).

(42) so we're gonna begin with your topic and you're gonna talk about Indian dishes (TLC:2_7_IN_27, NS)

Out of 156 observations in the dataset without preceding future marker, 64 (42%) have exactly this structure, where *be going to* is combined with the verb *talk* to introduce the topic of the discussion or presentation. Most of these structures occur at the beginning of the exam, which means that there is no previous future marker.

5.2.2.5 Temporal proximity

Temporal proximity refers to how soon the event described by the future marker clause is expected to occur after the time of speech. The assumption is that *be going to* semantically implies more immediacy, so it is used for events occurring sooner in the future compared to *will* (Haegeman 1989: 308; Nicolle 1997; Poplack & Tagliamonte 2000: 321; Brisard 2001; Bergs 2010: 223–224). This is tied to the fact that *be going to* is typically used when there are present indications that the future event will happen

(Haegeman 1989; Brisard 2001: 266; Huddleston & Pullum 2002: 211; Bergs 2010: 223–224).

Previous studies have operationalized proximate or distal realizations of future events in various ways and with different levels of granularity. Torres-Cacoulllos & Walker (2009) distinguish between events occurring within the next minute, hour, day, week, month, year and the far future, whereas Poplack and Tagliamonte (2000) adopt a binary distinction between proximate and distal events. For example, Fehringer and Corrigan (2015: 211) found that distal contexts favor *will* using the distinction of Poplack and Tagliamonte (2000). In a similar vein, Tagliamonte, Durham and Smith (2014: 99) find that *be going to* disfavors distal contexts. However, the effect found by Fehringer and Corrigan (2015: 211) is not consistent across all of their data spanning multiple decades, with younger speakers not being sensitive to the proximity of the future event. Likewise, Gardner (2017: 155) finds that proximity is only relevant for older speakers. Engel and Szmrecsanyi (2023), who distinguish between events happening today as proximate and events happening beyond today as distal, found that *will* favors distal contexts in written registers, whereas *will* favors proximate contexts in formal spoken English.

In general, it is challenging to annotate future events for their temporal proximity because many future markers refer to habitual statements as in (43) or general statements as in (44) that do not refer to a specific time in the future.

- (43) every morning I will teach them rhymes erm rhymes
(TLC:2_6_IN_115, B1 Tamil)
- (44) different people will respond differently to situations won't they?
(TLC:ME_46, NS)

Therefore, I adopted a binary distinction between proximate events and non-proximate events. Inspired by Dahl (1984: 112–113, see also Torres-Cacoulllos & Walker 2009: 330), who states that there is a typological tendency to mark future events occurring later today differently from events occurring later than today, proximate events happen or are valid during the same day as the utterance (see also Engel & Szmrecsanyi 2023). This also includes clauses with adverbs like *forever*, *still* (see (45)) and verbs like *keep*, *stay*, *live*, *survive* and *continue* (see (46)) which often imply that an action that began before the utterance will also be valid in the future. I also coded instances as ‘proximate’ when the described future events are valid until a certain point in the future starting from the time of the utterance, as in (47). Clauses with indefinite adverbs like *one day*, *someday*,

ever, which could in principle occur the same day as the speech time but are highly uncertain, are considered non-proximate, as in (48).

- (45) we-we're gonna still continue using er regular printed books
(TLC:ME_29, C1 Spanish)
- (46) I don't think my friend will stay with me permanently
(TLC:2_7_RU_2, NS)
- (47) before I get married it's going to be difficult (TLC:2_8_ME_12, NS)
- (48) I just don't think that we'll ever get that money back you know
(TLC:SP_36, NS)

In general, temporal adverbs can often give a clear indication of whether the event is happening on the same day (*today, in a minute*) or later (*next week, tomorrow, in a few years*), with adverbs like *now* being proximate by default.

Importantly, proximate events have to occur within the same day as the time of speech, i.e., during the same day as the language exam in which the speakers of the corpus participate. Utterances that depend on a reference point other than the speech time are non-proximate by default, except in the case of direct speech as in (49). In (50), for example, the protasis *if you have a phone in the classroom* is a hypothetical scenario that serves as the temporal reference point that is not the time of speech, so the apodosis with the future marker *the teacher erm will be angry* is considered non-proximate by default. Protases, by contrast, can be proximate as in (51). Also non-proximate by default are events dependent on hypothetical scenarios without any overt marking as in (52), as well as the main clauses of hypothetical *when*-subclauses as in (53), to be distinguished from *when*-clauses that function like future adverbs and which could have a proximate main clause, as in (54).

- (49) you should invite another person with you in order to be there and
the the person that told you okay this night we are going to do this
activity for me (TLC:2_7_ME_4, B2 Spanish)
- (50) if you have a a phone er in the classroom erm the teacher erm will
be angry (TLC:2_6_RU_32, B1 Russian)
- (51) if you will ask me now erm I think it's very important
(TLC:2_8_SP_29, B2 Spanish)
- (52) I don't see the point of doing that because it will hurt anyone and
it will hurt yourself (TLC:2_8_ME_8, B2 Spanish)

- (53) when you see a pregnant girl the child the child will ask you mum why does she have the stomach so big? (TLC:2_6_ME_31, B1 Spanish)
- (54) when you look at the picture you will understand better the situation in Portman (TLC:2_6_SP_19, B1 Spanish)

An overview of the distribution of *will* and *be going to* across proximate and non-proximate future events is given in Table 5.8 and visualized in Figure 5.6 for native speakers and learners.

Table 5.8: Contingency table for temporal proximity for native speakers ($X^2 = 10.39$, $df = 1$, $p = 0.001$, Cramér's $V = 0.12$) and learners ($X^2 = 290.17$, $df = 1$, $p < 0.001$, Cramér's $V = 0.32$). Percentages indicate the proportion of future marker variants per category of temporal proximity. For the row total, the percentages indicate the total proportion of the proximate and non-proximate contexts. If the test is significant, values that are significantly lower than expected according to standardized residuals are highlighted in orange, whereas values that are significantly higher than expected according to standardized residuals are highlighted in blue.

Native speakers	<i>Be going to</i>	<i>Will</i>	Row total
Non-proximate	228 (40.1%)	340 (59.9%)	568 (72.4%)
Proximate	115 (53.2%)	101 (46.8%)	216 (27.6%)
Column total	343 (43.8%)	441 (56.2%)	784
Learners	<i>Be going to</i>	<i>Will</i>	Row total
Non-proximate	805 (36.2%)	1,418 (63.8%)	2,223 (78.5%)
Proximate	457 (75%)	152 (25%)	609 (21.5%)
Column total	1,262 (44.6%)	1,570 (55.4%)	2,832

Both native speakers and learners use *will* for non-proximate events and *be going to* for proximate events, in line with most previous research that found an effect for the constraint (Tagliamonte, Durham & Smith 2014: 99; Fehrer & Corrigan 2015: 211; Gardner 2017: 155). Figure 5.6 indicates that the distinction has more impact on learners, which is confirmed by the fact that the association is stronger for learners (Cramér's $V = 0.32$) than for native speakers (Cramér's $V = 0.12$).

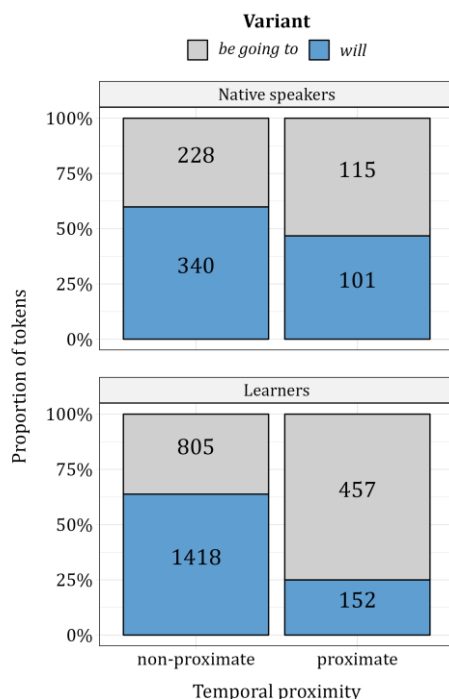


Figure 5.6: Distribution of *will* and *be going to* across levels of temporal proximity for native speakers (top) and learners (bottom).

5.2.2.6 Temporal adverb

Torres-Cacoullous and Walker (2009) found that *be going to* is less frequent if the clause contains a temporal adverb. They argue that *be going to* is more independent of temporal modification because the variant includes temporal meaning by default due to its original spatial meaning. Engel and Szmrecsanyi (2023) found the same result for written English registers, but not for spoken language, where the presence of temporal adverbs does not correlate with the choice of future marker. Similar to Engel and Szmrecsanyi (2023), I checked whether the future marker clause contained either a temporal adverb (e.g., *soon*, *sometimes*, *always*, *first*) as in (55), or a non-finite adverbial (e.g., *in the summer*, *next week*) as in (56) that conveys the time, frequency, duration or time relationship of the future event (Quirk et al. 1985: 481–482). Not included are pure connectives like *finally* in (57) (Quirk et al. 1985: 778) and adverbs that do not have scope over the future

marker like *every day* in (58), which refers to *running* and not to *be habituated*.

- (55) sometimes I will feel s-sick on the plane (TLC:7_CH_15, B2 Chinese)
 (56) in the future how are you going to achieve your dream
 (TLC:2_7_CH_41, NS)
 (57) finally I'm going to talk about the reincarnation (TLC:SP_47, C1 Spanish)
 (58) you 'll be habituated to the to the to running every day
 (TLC:2_7_IN_56, B2 Marathi)

Table 5.9: Contingency table for temporal adverb for native speakers ($X^2 = 2.79$, $df = 1$, $p = 0.095$) and learners ($X^2 = 4.08.17$, $df = 1$, $p = 0.043$, Cramér's $V = 0.04$). Percentages indicate the proportion of future marker variants per category of temporal adverb. For the row total, the percentages indicate the total proportion of the clauses with and without adverbs. If the test is significant, values that are significantly lower than expected according to standardized residuals are highlighted in **orange**, whereas values that are significantly higher than expected according to standardized residuals are highlighted in **blue**.

Native speakers	<i>Be going to</i>	<i>Will</i>	Row total
Absent	287 (45.3%)	347 (54.7%)	634 (80.9%)
Present	56 (37.3%)	94 (62.7%)	150 (19.1%)
Column total	343 (43.8%)	441 (56.2%)	784
Learners	<i>Be going to</i>	<i>Will</i>	Row total
Absent	1,022 (43.7%)	1,318 (56.3%)	2,340 (82.6%)
Present	240 (48.8%)	252 (51.2%)	492 (17.4%)
Column total	1,262 (44.6%)	1,570 (55.4%)	2,832

Table 5.9 and Figure 5.7 show that learners are more likely to use *be going to* when the future marker clause contains a temporal adverb. By extension, learners prefer *will* when there is no temporal adverb in the future marker clause. This goes against the finding from Torres-Cacoullous and Walker (2009) and Engel and Szmrecsanyi (2023) that *will* is preferred with adverbs. Although native speakers are more likely to use *be going to* when there is no adverb (see Figure 5.7), the association is not significant ($p = 0.095$).

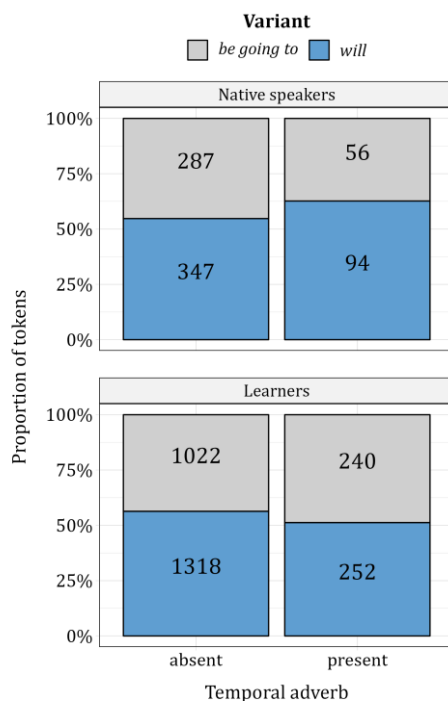


Figure 5.7: Distribution of *will* and *be going to* across levels of temporal adverb for native speakers (top) and learners (bottom).

5.2.2.7 Verb type

Originally, *be going to* was used to indicate movement (Fehrer & Corrigan 2015: 208; Gardner 2017: 142; Denis & Tagliamonte 2018: 408), so the structure was typically not followed by verbs implying motion at the early stages of grammaticalization until about the second half of the 17th century, when verbs like *come* and *go* could follow *be going to* (Danchev & Kytö 1994: 69). Previous literature has posited that *be going to* might have retained this avoidance of motion verbs from the past through retention. If this is not the case, then this means that the variant has reached an advanced stage of grammaticalization where the effect of motion verbs on the choice of variant has been levelled (Fehrer & Corrigan 2015: 199; Denis & Tagliamonte 2018: 417). Whereas Poplack and Tagliamonte (2000: 336) find that certain native varieties of English indeed avoid combining *be*

going to with motion verbs, this finding is not replicated by either Torres-Cacoullous and Walker (2009) nor Fehringer and Corrigan (2015: 212).²

To identify motion verbs, I determined whether the sense of the infinitive verb implies movement on the part of the subject, starting from the list of motion verbs by Levin (1993: 263–271) as well as some other verbs implying motion like *join* in (59). Metaphorical uses of motion verbs are also included, which covers cases where the subject cannot actually move (like *changes* in (60)), and cases where the location to which the subject is moving does not exist concretely (like *the third point* of the presentation in (61)). By contrast, cases like *go crazy* in (62) are not considered motion verbs because the sense of the verb does not imply motion, regardless of whether it is used metaphorically or not.

- (59) the two side river will join the sea (TLC:2_6_IN_114, B1 Malayalam)
- (60) the changes will come soon (TLC:SP_46, C2 Spanish)
- (61) we'll now come to the third point just my final point for today
(TLC:2_IN_10, C1 Unknown)
- (62) they'll go crazy and stuff (TLC:2_6_IN_70, B1 Marathi)

In a similar vein, some studies suggested that *be going to* should also disfavor dynamic verbs, i.e., verbs that refer to actions or events. If *be going to* does occur with dynamic verbs, this again implies that *be going to* is at an advanced stage of grammaticalization (Poplack & Tagliamonte 2000: 335). Other studies take the opposite view and argue that *be going to* should have a preference for dynamic verbs, whereas *will* supposedly favors stative verbs (Tagliamonte, Durham & Smith 2014: 78; Gardner 2017: 142). In this respect, Gries and Stefanowitsch (2004: 114) indeed found that the most distinctive collexemes of *will* are stative verbs, such as *see*, *know*, and *want*, whereas *be going to* mostly precedes dynamic verbs such as *say*, *do*, and *go*. Engel and Szmrecsanyi (2023: 366–368) also found that *will* prefers stative verbs in spoken British English. In general, the effect of verb type is not found consistently, however (see Gardner 2017: 155).

Following the approach of Engel (2022), verbs coded as ‘stative’ refer to characteristics and qualities (e.g. *be* in (63)), mental states and attitudes (e.g., *think*, *realize*, *imagine*, *dislike*, especially with a *that*-clause, as in (64)), as well as relationships and measurements (e.g., *have*, *contain*, *cost*, as in (65)) (see Quirk, Greenbaum & Leech 1985: 200–208). Importantly, it is the sense of the verb rather than the lemma itself that determines whether a

² Torres-Cacoullous & Walker (2009: 337) found that motion verbs like *come* and *go* do favor the present future.

verb is stative or dynamic. *Have*, when used for possession as in (66), is stative, but it can also be used to denote an action, as in (67), where *have* is used to refer to the dynamic act of eating or getting together to eat.

- (63) perhaps it's going to be good for her academically (TLC:2_8_ME_17, NS)
- (64) the teacher will think you are cheat you are cheating (TLC:2_6_CH_25, B1 Chinese)
- (65) this will cost so much that I mean countries will have to pull together to pay for it (TLC:2_IN_12, C1 Marathi)
- (66) we're gonna have millions of doctors and millions of engineers and nobody to clean the streets (TLC:SP_3, NS)
- (67) on Friday we're going to have a dinner (TLC:ME_20, C1 Spanish)

Sometimes, two infinitive verbs are coordinated after the future marker. In this case, I only look at the first verb, namely *go* in (68). For the future markers followed by *be*, it is sometimes difficult to assess whether *be* is used as the main verb with an adjective as in (69) or as a passive auxiliary. In this case, I checked whether it is possible to add a *by*-phrase denoting the agent of the action, as in (70), in which case *be* is an auxiliary. Instead of annotating *be* in such cases, I consider the main verb, viz. *disappointed*.

- (68) I will not go and search the ingredients here and there (TLC:2_6_IN_116, B1 Malayalam)
- (69) Chinese will be more important in the future (TLC:6_IT_27, NS)
- (70) in the future it's going to be really e-everyone are going to be disappointed by me that I don't know my own language (TLC:2_6_RU_2, B1 Russian)

The distinction between motion and non-motion verbs and between stative and dynamic verbs is combined in a single variable. Since all motion verbs are considered dynamic by default, the variable distinguishes between stative verbs, motion verbs and other dynamic verbs. The distribution of the future markers across the different verb types is shown in Table 5.10 and visualized in Figure 5.8.

Table 5.10: Contingency table for verb type for native speakers ($X^2 = 3.09$, $df = 2$, $p = 0.214$) and learners ($X^2 = 112.92$, $df = 2$, $p < 0.001$, Cramér's $V = 0.2$). Percentages indicate the proportion of future marker variants per category of verb type. For the row total, the percentages indicate the total proportion of the different types of verbs. If the test is significant, values that are significantly lower than expected according to standardized residuals are highlighted in **orange**, whereas values that are significantly higher than expected according to standardized residuals are highlighted in **green**.

Native speakers	Be going to	Will	Row total
Dynamic	249 (45.5%)	298 (54.5%)	547 (69.8%)
Motion	23 (45.1%)	28 (54.9%)	51 (6.5%)
Stative	71 (38.2%)	115 (61.8%)	186 (23.7%)
Column total	343 (43.8%)	441 (56.2%)	784
Learners	Be going to	Will	Row total
Dynamic	964 (51.1%)	921 (48.9%)	1,885 (66.6%)
Motion	90 (42.9%)	120 (57.1%)	210 (7.4%)
Stative	208 (28.2%)	529 (71.8%)	737 (26%)
Column total	1,262 (44.6%)	1,570 (55.4%)	2,832

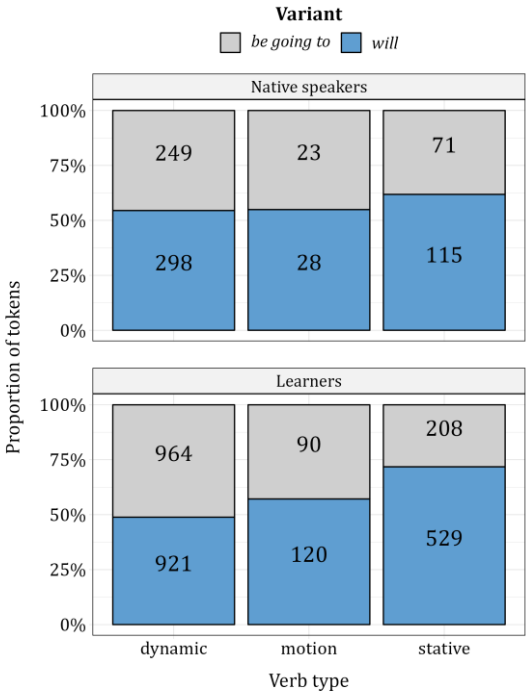


Figure 5.8: Distribution of will and be going to across levels of verb type for native speakers (top) and learners (bottom).

Table 5.10 shows that the type of verb is not associated with the choice of variant for native speakers. Learners, on the other hand, prefer *be going to* with dynamic verbs and *will* with stative verbs, in line with the findings from Tagliamonte, Durham and Smith (2014: 78), Gardner (2017: 142) and Engel and Szmrecsanyi (2023: 366–368). The association is rather weak, however (Cramér's $V = 0.2$).

5.2.2.8 Person and animacy

Since *will* was originally used to express volition, which is typically expressed in the first person, some accounts state that *will* has retained its preference for first person subjects (Fehrer & Corrigan 2015: 210; Gardner 2017: 146). It is likely that this is partially due to the reduced form *'ll*, which has a strong tendency to accompany first person subjects, whereas the full form *will* is actually the least frequent future marker with pronominal subjects (except for *it*) (Berglund 2000: 41). Moreover, Tagliamonte, Durham and Smith (2014: 78) found the opposite pattern whereby *be going to* prefers first person subjects. According to Royster and Steadman (1923: 400), *be going to* could be preferred with first-person subjects because *be going to* expresses the intention of the speakers, which is again most often realized in first person. Likewise, Torres-Cacoullós and Walker (2009: 332) argued that *be going to*, being the incoming variant, should have retained more of its original meaning, with agentive subjects most typically taking first person (Denis & Tagliamonte 2018: 418). However, most studies actually find that *be going to* is preferred with third and especially second person pronouns (Torres-Cacoullós & Walker 2009: 339; Fehrer & Corrigan 2015: 210; Gardner 2017: 146; Denis & Tagliamonte 2018: 423).

Additionally, I also distinguished between animate and inanimate subjects. Since *be going to* originally referred to a subject capable of moving itself, *be going to* was originally limited to animate subjects (Fehrer & Corrigan 2015: 199; Gardner 2017: 146). It is only when the meaning of *be going to* changed from intention to prediction that inanimate subjects started appearing (Poplack & Tagliamonte 2000: 334–335; Tagliamonte, Durham & Smith 2014: 78). In principle, *will* also originated in contexts with animate subjects because it initially expressed volition, but since *be going to* is the newer, less semantically bleached variant, the assumption is that *be going to* will favor animate subjects more. Again, the effect of subject animacy is not consistent, as neither Tagliamonte, Durham and Smith (2014: 98–99), Fehrer and Corrigan (2015) nor Engel and Szmrecsanyi (2023) find such an effect, which might again be due to the advanced stage of

grammaticalization of *be going to*. In fact, Gardner (2017: 146) found that *be going to* even prefers inanimate subjects.

Subject grammatical person and animacy were combined in a single variable that distinguishes between first person, second person, third person animate and third person inanimate. First and second person subjects are animate by default. To categorize third-person subjects as animate or inanimate, I used the annotation guidelines of Zaenen et al. (2004) and more specifically those of Wolk et al. (2013). Animate subjects refer to “things that look human and act like humans” (Zaenen et al. 2004: 4); but also non-human animates, which can broadly be understood as what one perceives to be living, such as animals (e.g., *god, carpenter, elephants*, see example (71)). Additionally, this category also includes collective nouns, like *the government, company, Italy* (see example (72)) given that these subjects also refer to groups of individuals. For third-person pronouns like *they, that* or *which*, it is often necessary to look at the referent of the pronoun in the preceding context. In example (73), the subject *they* refers to *crops*, which is inanimate.

- (71) because god will be the one who's going to judge who's going to the celestial kingdom (TLC:2_ME_26, C1 Spanish)
- (72) how far do you think the government will go in placing a tax on this junk food? (TLC:IN_6, NS)
- (73) I really believe that people would prefer to buy crops which are made in a vertical farm because transport costs will be really low and they'll cost much less than the actual thing (TLC:2_IN_12, C1 Marathi)

The distribution of the variable is summarized in Table 5.11 and visualized in Figure 5.9.

Table 5.11: Contingency table for subject grammatical person and animacy for native speakers ($X^2 = 6.4$, $df = 3$, $p = 0.094$) and learners ($X^2 = 152.39$, $df = 3$, $p < 0.001$, Cramér's $V = 0.23$). Percentages indicate the proportion of future marker variants per category of grammatical person and animacy. For the row total, the percentages indicate the total proportion of the different grammatical persons and animacy levels. If the test is significant, values that are significantly lower than expected according to standardized residuals are highlighted in **orange**, whereas values that are significantly higher than expected according to standardized residuals are highlighted in **blue**.

Native speakers	<i>Be going to</i>	<i>Will</i>	Row total
First person	116 (48.7%)	122 (51.3%)	238 (30.4%)
Second person	92 (45.3%)	111 (54.7%)	203 (25.9%)
Third person animate	57 (36.3%)	100 (63.7%)	157 (20%)
Third person inanimate	78 (41.9%)	108 (58.1%)	186 (23.7%)
Column total	343 (43.8%)	441 (56.2%)	784
Learners	<i>Be going to</i>	<i>Will</i>	Row total
First person	701 (56.1%)	549 (43.9%)	1,250 (44.1%)
Second person	147 (50.2%)	146 (49.8%)	293 (10.3%)
Third person animate	214 (31%)	476 (69%)	690 (24.4%)
Third person inanimate	200 (33.4%)	399 (66.6%)	599 (21.2%)
Column total	1,262 (44.6%)	1,570 (55.4%)	2,832

Table 5.11 shows that the association between the grammatical person and animacy of the subject and the choice of future marker is not significant for native speakers. By contrast, learners prefer *be going to* with both first and second person subjects, whereas animate and inanimate third person referents prefer *will*. The fact that first-person subjects take *be going to* over *will* implies that *be going to* has retained more of its original meaning, which is to express intention (Royster & Steadman 1923: 400; Torres-Cacoulllos & Walker 2009: 332). This association is rather weak (Cramér's $V = 0.23$).

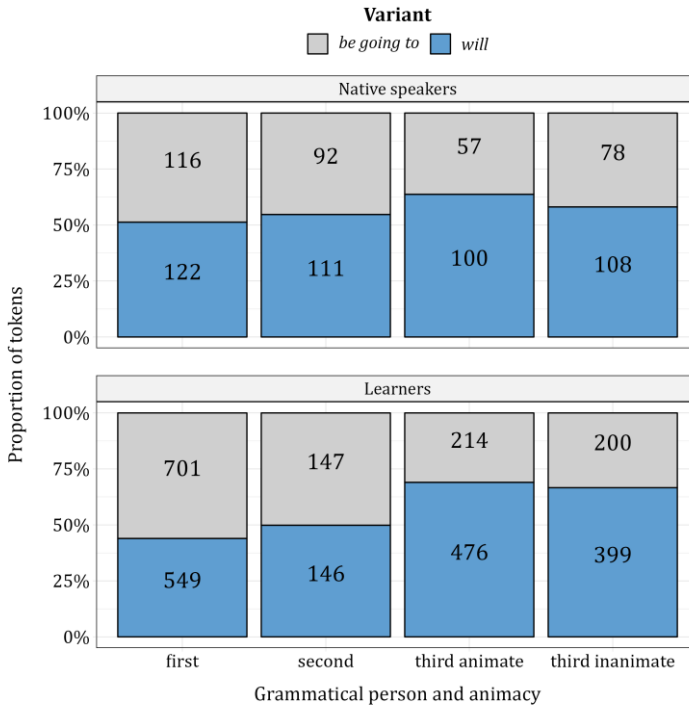


Figure 5.9: Distribution of *will* and *be going to* across levels of subject grammatical person and animacy for native speakers (top) and learners (bottom).

5.2.2.9 Clause length

Szmrecsanyi (2003: 314–315) found that *be going to* tends to be more frequent in longer sentences, a finding that was later replicated in Szmrecsanyi (2006: 116). Provided longer sentences demand more processing, speakers might prefer the longer, more explicit *be going to* variant because it gives the speaker more time to process the utterance, as suggested by Rohdenburg's Complexity Principle (Rohdenburg 1996: 151). Since the TLC (Gablasova, Brezina & McEnery 2019) consists of spoken utterances without punctuation, the observations were annotated for the character length of the clause rather than the sentence. Length can also be operationalized as the number of words or syllables (see Szmrecsanyi 2003: 311), but the effect of different length measurements have been found to be highly correlated in previous work on other alternation phenomena (see Wolk et al. 2013: 395).

For the purposes of the study, a clause consists of a subject noun phrase and a verb phrase, which in turn consists of either *will* or *be going to*, the infinitive verb, any complements and optionally some adverbs. Other elements that are not necessary to constitute a full clause according to these criteria are not included in the count. In (74) for instance, the matrix clause *I think* is not included in the count as it not necessary to complete the future marker clause *that won't be so good for your future*. If the future marker occurs in a matrix clause taking a clausal object, the entire clausal object is included in the count, because it is a complement of the verb phrase as in (75). Importantly, the clause length also counts all linguistic material that occurs within the boundaries of the clause. Hence, the matrix clause *do you think*, which is embedded in its own subordinate clause in (76), is included in the count because it occurs within the boundaries of the future marker clause.

- (74) that won't be so good for your future I think (TLC:2_7_CH_15, B2 Chinese)
- (75) all the students will then think that to be on huh a reality TV show is better than having er a career (TLC:2_IT_35, NS)
- (76) what do you think you will will be an ideal situation for future (TLC:CH_30, C2 Chinese)

Not included are certain clause-initial elements that are neither syntactically nor semantically integrated with the clause. This involves the so-called modal adverbs (e.g., *definitely, really, probably, possibly, maybe*, etc. see Quirk, Greenbaum & Leech 1985: 173) or adverbial structures (e.g., *of course, without a doubt*), which express the speakers' modality regarding the clause as a whole rather than contributing to the semantics of the clause itself, as in (77). Likewise, pure connectives like *first, for example, firstly, to begin with, to conclude, instead, although, if* etc. (see Quirk, Greenbaum & Leech 1985: 778) also do not contribute to the semantics of the clause since they place the clause in the larger discourse context. For the same reason, coordinators and certain subordinators like *because* or *as* were also removed. Vocatives like *ma'am* in (78) were also not retained.

- (77) of course they won't be studying (TLC:IT_28, C1 Italian)
- (78) if you put some money it won't be too much expensive ma'am (TLC:2_8_IN_49, B2 Tamil)

Finally, the data also features certain observations where a referent is first introduced as a full noun phrase and then replaced by a pronoun, which then figures as the subject of a clause. In this case, I only retained the later

reformulation of the referent, viz. the pronoun *it* in (79), seeing that it is the subject of the clause with the future marker. Such pronouns can even replace full noun phrases that have their own relative clause, as in (80). In this case however, the relative clause of the full noun phrase contains the future marker, so I retain the full noun phrase and the relative clause (i.e., *the woman who is gonna be the surrogate mother*) rather than the pronoun and its following clause.

- (79) The blood it will stop for one two three er one to two hours (TLC:2_7_IN_48, B2 Marathi)
- (80) for example the woman who is gonna be the surrogate mother she could have feelings for the child (TLC:2_IN_20, NS)

Figure 5.10 reveals that the proportion of *be going to* increases when the clause reaches a length of about 100 letters or more for native speakers, which ties in with previous findings for native speakers (Szmrecsanyi 2003: 314–315; Szmrecsanyi 2006: 116). Surprisingly, learners tend to prefer *will* for clauses of that length, which suggests that *will* is the easier variant for learners, possibly because it does not require the learner to inflect a verb contrary to *be going to* (see Bardovi-Harlig 2002).

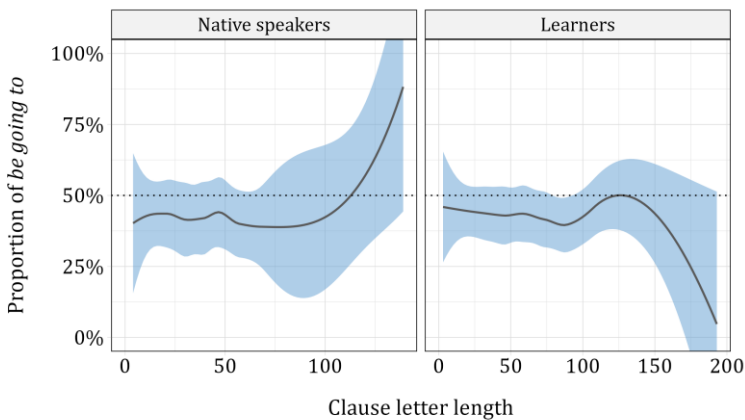


Figure 5.10: Smoothed conditional means of the proportion of *be going to* by the letter length of the clause for native speakers (left) and learners (right).

5.2.3 Classifying the constraints

The theory-based account on the acquisition of probabilistic grammars distinguishes between two types of probabilistic constraints, namely

feature-based constraints and processing-related constraints. This distinction is relevant in the present context because processing-related constraints do not have to be learned via associative statistical learning like feature-based constraints (see Section 3.3). It is therefore necessary to categorize the future marker constraints as either feature-based or processing-related to test the hypotheses of the thesis. Feature-based constraints can be identified by three characteristics: they typically involve linguistic elements in the immediate context of the choice context, they are language-specific rather than universal, and their effect can be formulated as a categorical rule of grammar (Tamminga, MacKenzie & Embick 2016: 321).

Local vs. non-local

Negation, sentence type, verb type, clause type, clause length, temporal adverb and grammatical person and animacy of the subject all refer to elements in the future marker clause. While the operationalization of temporal proximity often concerns temporal adverbs or adverbial phrases in the future marker clause, sometimes the proximity of the future event is not realized by any overt linguistic elements but is implied instead, like the proximate event in (81).

(81) I'm just going to close the window I opened it because I thought maybe it was stuffy (TLC:SP_111, NS)

Likewise, structural persistence does not necessarily derive from linguistic elements in the immediate context of the future marker, since structural persistence can occur long after a variant construction is produced (Bock & Griffin 2000).

Language-specific vs. universal

Negation, sentence type, verb type, clause type, temporal adverb and grammatical person and animacy of the subject are likely language-specific rather than universal. While English considers the type of verb, sentence or clause for example, these factors might not be relevant for other languages in the same way. Again, temporal proximity is more tricky because its operationalization is based on a widely-attested typological distinction between future events in human languages (Dahl 1984: 112–113; see Torres-Cacoullos & Walker 2009: 330). Additionally, each future event inherently has a temporal proximity. That is, if the future is perceived as a horizontal line extending from the present, each event can be placed somewhere along this line. Nonetheless, whether and how this distinction between proximate and non-proximate events is realized linguistically

depends on the individual language. In English, the preference for *be going to* in proximate events results from the historical development of *be going to*, which is language-specific (see Section 5.2.2.5). Clause length, on the other hand, is universal because it captures the cognitive complexity of the linguistic units the speaker is processing. Likewise, structural persistence results from the features of the brain (Tamminga, MacKenzie & Embick 2016).

Categorical rule

It is easy to posit a categorical rule for negation, sentence type, verb type, clause type, temporal adverb, temporal proximity, as well as the grammatical person and animacy of the subject. Hence, one could imagine a rule whereby *be going to* is exclusively used for questions, for negated contexts, for subordinate clauses, for events occurring within the same day, with certain verbs, with second-person pronouns and human referents, but not when there is a temporal adverb. By contrast, there is no plausible rule stating that *be going to* has to be used when the sentence is longer than a certain number of words, because “grammars can’t count” (Selkirk 1986; cited by Tamminga, MacKenzie & Embick 2016: 321). Likewise, structural persistence cannot have a categorical effect forcing the speaker to reuse the previous future marker.

Overview

Table 5.12 offers an overview of the constraints regarding the three classification criteria, which shows that the processing-related constraints operating on the future marker alternation are structural persistence and clause length.

Table 5.12: Overview of the classifying features for each of the future marker constraints.

	<i>Local context</i>	<i>Language-specific</i>	<i>Categorical rule</i>
Sentence type	Yes	Yes	Yes
Negation	Yes	Yes	Yes
Clause type	Yes	Yes	Yes
Persistence	No	No	No
Temporal proximity	No	Yes	Yes
Temporal adverb	Yes	Yes	Yes
Verb type	Yes	Yes	Yes
Person & animacy	Yes	Yes	Yes
Clause length	Yes	No	No

5.2.3.1 Ranking the feature-based constraints

Negation, sentence type, verb type, clause type, temporal proximity, temporal adverb, as well as the grammatical person and animacy of the subject are feature-based, which means that they are acquired via associative statistical learning. Table 5.13 presents the ranking of these constraints by their association strength with the future markers for native speakers.

Table 5.13: The feature-based future marker constraints, ranked by the strength of association according to monofactorial tests (Cramér's *V*).

Constraint	Cramér's <i>V</i>
Clause type	0.16
Temporal proximity	0.12
Sentence type	NS
Negation	NS
Temporal adverb	NS
Verb type	NS
Person and animacy	NS

According to this rather simplistic proxy for the reliability of the constraints, the most learnable constraints are clause type followed by temporal proximity. Negation, sentence type, verb type, temporal adverb, and subject person and animacy are not significantly associated with either future marker.

5.2.4 General overview

Table 5.14 presents a general overview of the future marker constraints included in the analysis, with their relevant distinctions, their effect on the choice of variant, their classification, and their learnability via associative statistical learning.

Table 5.14: Overview of the future marker constraints, as well as the relevant coding distinctions, their expected effect on the choice of variant and their learnability via associative statistical learning. ‘Not consistent’ in the effects column refers to the fact that the effect is not consistently found in the literature.

Constraints	Distinctions	(Expected) effect	Learnability
Sentence type	Declarative Interrogative	Interrogative clauses prefer <i>be going to</i>	Difficult
Negation	Affirmative Negative	Negative clauses prefer <i>be going to</i> (not consistent)	Difficult
Clause type	Main clause Subclause Apodosis Protasis	▪ Main clauses prefer <i>will</i> and subclauses prefer <i>be going to</i> ▪ Apodoses prefer <i>will</i> and protases prefer <i>be going to</i> (not consistent)	Easy
Persistence	Examiner <i>will</i> Examiner <i>be going to</i> Candidate <i>will</i> Candidate <i>be going to</i> None	▪ Previous examiner or candidate <i>will</i> prefer <i>will</i> ▪ Previous examiner or candidate <i>be going to</i> prefer <i>be going to</i>	Processing-related
Temporal proximity	Proximate Non-proximate	Proximate events prefer <i>be going to</i> (not consistent)	Easy
Temporal adverb	Present Absent	Temporal adverbs prefer <i>will</i> (not consistent)	Difficult
Verb type	Motion Dynamic Stative	▪ Motion verbs avoid <i>be going to</i> (not consistent) ▪ Dynamic verbs prefer <i>be going to</i> and stative verbs prefer <i>will</i> (not consistent)	Difficult
Person & animacy	First person Second person Third person animate Third person inanimate	▪ First person subjects prefer <i>will</i>, second and third person subjects prefer <i>be going to</i> (not consistent) ▪ Animate referents prefer <i>be going to</i> (not consistent)	Difficult
Clause length	Numerical	Long clauses prefer <i>be going to</i>	Processing-related

Clause type, verb type, temporal proximity and subject animacy involve mostly manual coding, so I conducted an interrater reliability analysis. Concretely, I developed a coding scheme for each of the variables (Révész

2012), which I then visualized as decision trees with a number of yes/no questions. Then, a second coder used the coding scheme and the decision trees to double-code a random sample of 750 future markers, corresponding to 21% of the total number of observations. The agreement with the second coder was estimated with Cohen's kappa, which stands at 0.92 for clause type, 0.91 for temporal proximity, 0.88 for verb type and 0.94 for animacy, indicating almost perfect agreement for all constraints according to the benchmark of Landis and Koch (1977).

5.2.5 Statistical analysis

The annotated dataset was analyzed with mixed-effects logistic regression using the *lme4* package (Bates et al. 2015) in R (version 4.2.2, R Core Team 2022). The model was fitted to predict the likelihood of *be going to* with the help of predictors, in this case the future marker constraints. These predictors were allowed to interact with the proficiency level of the speaker, so that it is possible to compare how the constraints affect the choice of future marker across proficiency levels.

Random effects

There are other sources of variation besides the probabilistic constraints described above. For one, speakers have their own idiolect, meaning that some speakers might be more sensitive to certain constraints or they might have an overall preference for either future marker. At the same time, the exam task could influence the choice of future marker (see Section 4.1.1). Likewise, the examiner and the participant discuss the same topics as they are interacting together, which influences the lexical items they use, the thematicity of these items and possibly the choice of variant. Previous research has found that lexical verbs also have certain preferences (Gries & Stefanowitsch 2004: 114; Engel & Szmrecsanyi 2023). Most previous research on alternations in EFL also suggests that the L1 of the learners could impact the choice of variant due to transfer. Finally, learners from the same country might have been subjected to similar teaching practices and potentially prescriptivism.

Although these sources of variation should be taken into account to ensure the reliability of the model results, including them as fixed predictors does not provide particularly interesting information, not to mention there is not enough data to do so. Admittedly, the L1 background of learners could provide interesting insights, but there is a perfect overlap between the mother tongue of the native speakers (English) and their proficiency level (native), so L1 cannot be entered as a fixed effect. Instead, the model features random effects for the individual speakers, their L1,

their country, the verb lemma, the exam, and the exam task within the exam as well as the exam task across exams. For all model parameters to be reliable with random effects, it is recommended that each level of a random effect contains at least 5 observations (Clarke 2008: 757). While this is trivial for random effects concerning the learner's country or L1, many levels from other random effects had to be conflated together. For example, 349 out of 580 (60.2%) speakers produce fewer than 5 observations each, so the data from these learners was conflated in the random effects. Although this might not seem ideal, these levels are less problematic considering that they contribute fewer dependent observations.

Exploring the data

Prior to running the model, I checked the distribution of the only numerical variable in the model, namely clause length. Because the variable was severely right-skewed, the variable was log-transformed. As recommended by Zuur, Ieno and Elphick (2010: 5), I then used Cleveland dot plots (Cleveland 1993) to visually check whether there are any outliers in the log-transformed clause length for each random effect (Baayen 2008: 243–244). No observations were removed as there were no visible outliers. Afterwards, the log-transformed clause length was first centered on the median ³. Then, the centered, log-transformed clause length was standardized by dividing the value by two standard deviations as suggested by Gelman (2008), which makes the effect of numerical variables more similar to that of binary categorical predictors in the model.

I then checked whether there are enough instances of *will* and *be going to* for each predictor level to ensure the quality of the results given that the effect of the constraints is frequency-based. For example, if the B2 data only contains instances of *will* with apodoses, then the model would conclude that apodoses categorically take *will* at the B2 level. Field, Miles and Field (2012: 340) suggest that a variable is problematic when there are fewer than 5 observations in 20% of the combinations between the response variable and the predictor. For example, the interaction between sentence type (declarative or interrogative) and proficiency level (B1, B2, C1, C2, native) combined with the response variable of the model (*will* or *be going to*) yields a crosstable with 20 possible combinations. If more than 4 of these combinations contain less than 5 observations, there is an issue of data sparsity.

³ I used the median instead of the mean because Shapiro-Wilk tests revealed that the distributions are not quite normal even after log-transformation (Heumann, Schomaker & Shalabh 2016: 42)

In the future marker data, there are no protases with *will* at the C2 level. In general, there are very few protases in the whole dataset ($N = 47$), so protases were conflated with other subclauses. At the C2 level, there are also no instances of *will* with no previous future marker. Since there are rather few such cases at the C1 level as well, one could conflate the C levels together. However, I decided to remove the 156 observations with no previous future marker in the dataset instead, because the distinction between C1 and C2 is highly relevant for the research goals of the thesis.

Treatment coding

I used treatment coding to analyze how probabilistic constraints impact the choice of future marker. This means that the model chose a reference level for each predictor and compared it against the other levels of the predictor. I chose reference levels that typically favor *will*, so the model starts from a future marker in a declarative (sentence type) and affirmative (negation) main clause (clause type) with no temporal adverbs or adverbials (temporal adverb), featuring a dynamic verb (verb type) and an inanimate third-person subject (subject grammatical person and animacy), following a previous instance of *will* produced by a candidate (persistence) and describing a proximate future event (temporal proximity). Numerical predictors were set to 0, which means that the observation had a clause length corresponding to the median. Most importantly, native speakers were the chosen reference proficiency level so the model compared the choice of variant for native speakers with each of the learner proficiency levels.

Model fitting

To fit the model, I adopted the approach laid out in Zuur et al. (2009), so I started from a beyond optimal model that includes as many predictors and random effects as possible. However, including many regressors in the model can lead to convergence issues, in which case it is recommended to select a beyond optimal model that is theoretically motivated. In this case, the beyond optimal model is generated with the following formula using the *glmer* function from the *lme4* package (Bates et al. 2015):

```
Beyond optimal model <- glmer(future marker variant ~ proficiency + sentence
type + negation + clause type + persistence + temporal proximity + temporal
adverb + verb type + person and animacy + clause length + proficiency:sentence
type + proficiency:negation + proficiency:clause type + proficiency:persistence +
proficiency:temporal proximity + proficiency:temporal adverb + proficiency:verb
type + proficiency:person animacy + proficiency:clause length + (1|country) +
(1|exam) + (1|task within exam) + (1|task across exams) + (1|speaker) +
```

```
(1|infinitive verb lemma) + (1|mother tongue), data = future marker data, family = binomial, glmerControl(optimizer = "bobyqa", maximum iterations = 1 million))
```

There were convergence issues when random slopes were included indicating data sparsity, so the beyond optimal model featured random intercept adjustments for the verb lemma, the exam, the exam task within and across exams, the speakers, their country and their L1.

To reach the optimal model, I first removed the random effects that did not contribute to the overall quality of the model according to likelihood-ratio tests with the *anova()* function, namely the exam task within or across exams and speakers' L1. To determine the fixed-effects structure of the model, I used the *drop1()* function, which checks whether each of the predictors significantly improves the quality of the model. More specifically, the *drop1()* function removes a predictor and then compares the model with and without the predictor using a likelihood ratio test. Each predictor is then assigned a *p*-value indicating whether the reduced model is significantly worse. The non-significant predictor with the highest *p*-value was removed first since this predictor is least likely to improve model quality. Sometimes, the highest *p*-value involved an interaction between proficiency level and a constraint with one or more significant regressors. Since these significant distinctions might ultimately improve the accuracy of the model over the course of model fitting, I first removed the predictors without any significant regressors. This process was reiterated until all remaining predictors significantly improved the model at a *p*-value below 0.1. The final model is generated by the following formula:

```
Optimal model <- glmer(future marker variant ~ proficiency + negation + clause type + persistence + temporal proximity + temporal adverb + verb type + person and animacy + clause length + proficiency:negation + proficiency:persistence + proficiency:temporal proximity + proficiency:temporal adverb + proficiency:verb type + proficiency:clause length + (1|country) + (1|exam) + (1|speaker) + (1|infinitive verb lemma), data = future marker data, family = binomial, glmerControl(optimizer = "bobyqa", maximum iterations = 1 million))
```

The only predictor that was removed as a whole is sentence type. Likewise, the interactions between proficiency level and clause type, subject grammatical person and animacy were not retained. The remaining six predictors, viz. negation, persistence, temporal proximity, temporal adverb, verb type and clause length are part of interactions terms with proficiency level.

This final model is significantly better than the null model ($p < 0.001$) and predicts the right future marker in 80% of the cases, which is

significantly better than the baseline of 57% ($p < 0.001$). The model's $R^2_{\text{marginal}} = 0.15$ and $R^2_{\text{conditional}} = 0.4$ and the concordance index C is .88, indicating an excellent fit of the data (Baayen 2008). The condition index κ of 26.54 indicates that there is medium collinearity in the model, but this remains below the threshold of 30 for harmful collinearity (Baayen 2008: 182). The Variable Inflation Factor (VIF) adjusted by degrees of freedom for each factor is below the threshold of 10 (Field, Miles & Field 2012: 339; Levshina 2015: 273).

5.3 Results

This section presents the results from the mixed-effects logistic regression model, starting with the random effects before turning to the main effects and the interaction effects. I will also present some supplementary analyses that helped ascertain the validity and robustness of the model results.

5.3.1 Random effects

Table 5.15: Overview of the random effects in the model, their estimated variance and their standard deviation.

Random intercept	Variance	St. dev.
Speaker	0.69	0.83
Country	0.45	0.67
Verb lemma	0.39	0.62
Exam	0.26	0.51

Table 5.15 shows that the idiolect of the speakers explains the most variance, followed by the country where the exam was taken. It is interesting that the speakers' country influences the choice of variant and not their L1. As expected, the future markers tend to co-occur with specific verbs (Gries & Stefanowitsch 2004: 114). However, the verb-specific adjustments apply for both the native speakers and the learners, so it is unclear whether certain verb biases are primarily driven by either learners or native speakers. Finally, the random effect for exam indicates that the joint discourse between learners and native speakers also explains some variability. In the following subsections, I will interpret the intercept adjustments by country and verb lemma in particular.

Country

Figure 5.11 shows that speakers are particularly prone to using *be going to* during the exams taken in Mexico, with 608 out 895 (67.9%) observations taking *be going to*.⁴

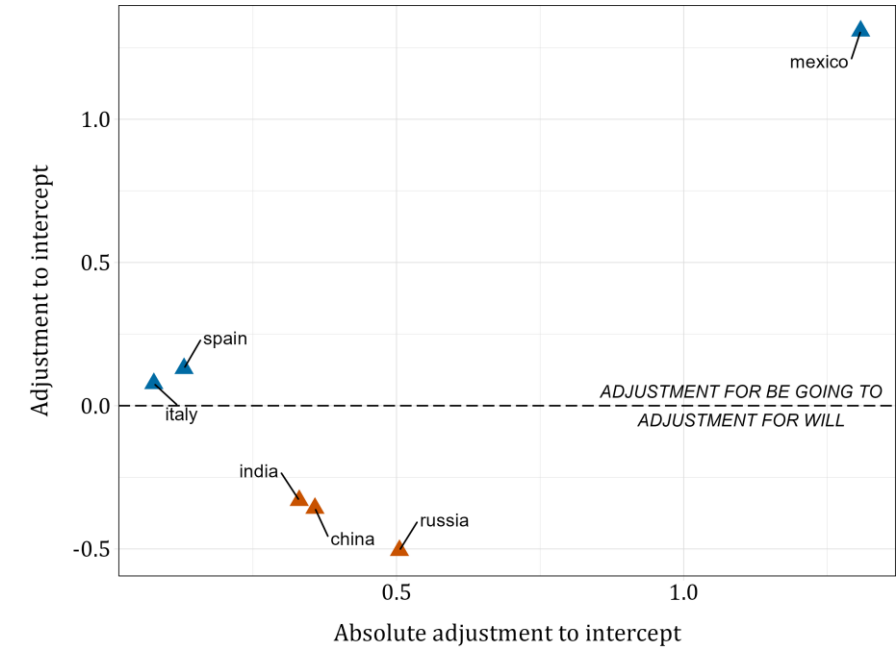


Figure 5.11: Intercept adjustments for the country where the exam was taken. Countries with positive adjustments prefer *be going to* and are depicted as blue triangles. Countries with negative adjustments prefer *will* and are depicted as orange triangles. The x-axis features the absolute value of the adjustment.

Learners are the main driving force behind this effect, with 502 out of 715 (70.2%) learner observations taking *be going to*, which is considerably more than other countries including Spain, where *be going to* accounts for only 302 out of 615 (49.1%) observations. This could point to differences in the way the future markers are taught in Mexico compared to Spain despite their common L1 background. Interestingly, the effect is partially due to native speakers, who also use *be going to* in 106 out of 180 (58.9%) cases compared to 39.2% in other countries (237 out of 604). While observations from Italy exhibit virtually no bias, those from India, China and Russia show a rather weak preference for *will*.

⁴ The visualization is inspired by the random effects plots of Röthlisberger (2018)

Verb lemma

The intercept adjustments for the lemma of the infinitive verb following the future marker are depicted in Figure 5.12.

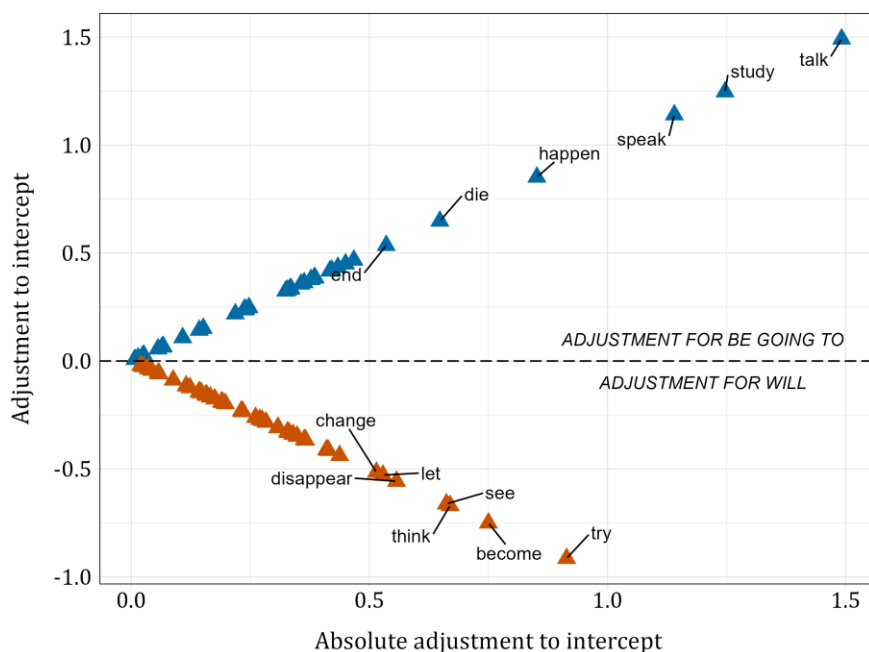


Figure 5.12: Intercept adjustments for the verb lemmas. Verb lemmas with positive adjustments prefer *be going to* and are depicted as blue triangles. Verb lemmas with negative adjustments prefer *will* and are depicted as orange triangles. Only the lemmas of the verbs with an absolute adjustment exceeding 0.5 on the x-axis are labelled.

The verb that is most biased towards *be going to* is *talk*, probably because of structures like (82), which overwhelmingly take *be going to* and are used to introduce the presentation or discussion task and therefore are very common across exams. The same explanation applies for *speak*. For *die*, the observations imply some sort of inevitability, as in (83).

(82) I 'm going to talk about sharks (TLC:2_6_SP_35, B1 Spanish)

(83) over the course of today two thousand of our brain cells will die
and never come back (TLC:2_IN_23, C2 Hindi)

The verbs favoring *will* include perception events like *see* as in (84) and (mental) states like *think* and *become* as in (85), which is reminiscent of the findings of Gries and Stefanowitsch (2004: 114). The verb *try* has the strongest bias for *will*. In 33 out of 45 (73.3%) instances of *try*, the subject is in first person, whereas these subjects account only for 41% of the

observations in the whole dataset (1484 out of 3616). Hence, it seems that the speakers in the corpus often mention they will try something in the future, as in (86).

- (84) do you think we will ever get to a point where where you will actually see through electric eyes or something (TLC:ME_15, C1 Spanish)
- (85) robbing the bank in person will become a thing of the past (TLC:2_7_RU_10, B2 Russian)
- (86) I will try to analyse the advantages and disadvantages of living abroad (TLC:2_SP_36, C1 Spanish)

5.3.2 Fixed effects

5.3.2.1 Main effects

Table 5.16 presents the main effects in the model. The coefficients and the odds ratios both represent the effect sizes and directions of the predictors. Since the model predicts the likelihood of *be going to*, positive coefficients and odds ratios above 1 indicate that the corresponding change of predictor levels makes *be going to* more likely, whereas negative coefficients and odds ratios between 0 and 1 signify that *be going to* becomes less likely.

Table 5.16: Main effects of the individual predictors on the choice of future marker. Positive coefficients indicate an increase in the likelihood of *be going to*. (Marginally) significant *p*-values are written in bold.

Predictor	Coefficient	Odds ratio	Std. error	<i>p</i> -value
(Intercept)	-0.666	0.514	0.395	0.092
Proficiency				
Native → B1	-1.152	0.316	0.372	0.002
Native → B2	-1.285	0.277	0.315	< 0.001
Native → C1	-0.712	0.491	0.331	0.031
Native → C2	-0.264	0.768	0.44	0.548
Negation				
Affirmative → negative	0.929	2.531	0.334	0.005
Clause type				
Main clause → subclause	0.328	1.389	0.108	0.002
Main clause → apodosis	-0.756	0.469	0.168	< 0.001
Persistence				
Candidate <i>will</i> → candidate <i>bgt</i>	-0.275	0.759	0.271	0.31
Candidate <i>will</i> → examiner <i>bgt</i>	-0.017	0.983	0.282	0.951
Candidate <i>will</i> → examiner <i>will</i>	-0.595	0.552	0.246	0.016
Temporal proximity				
Non-proximate → proximate	0.073	1.075	0.240	0.762
Temporal adverb				
Absent → present	-0.42	0.657	0.235	0.073
Verb type				
Dynamic → motion	0.160	1.174	0.397	0.687
Dynamic → stative	-0.424	0.654	0.272	0.119
Person & animacy				
Third inanimate → Third animate	0.007	1.007	0.148	0.962
Third inanimate → First person	0.153	1.165	0.15	0.307
Third inanimate → Second person	0.480	1.617	0.172	0.005
Clause length	-0.075	0.928	0.206	0.716

The only probabilistic constraints that do not figure in an interaction with proficiency level are the person and animacy of the subject as well as the type of clause, so these constraints affect the choice of future marker variant in the same way for both the learners and native speakers. If the future marker clause is a subclause instead of a main clause, *be going to* is 1.39 times more likely as expected from previous research (Tagliamonte, Durham & Smith 2014: 98; Fehrer & Corrigan 2015: 216; Gardner 2017: 153; Denis & Tagliamonte 2018). If the future marker figures in an apodosis rather than a main clause, the odds ratio of 0.47 indicates that *be going to* is 2.13 times less likely, which also aligns with previous findings (Gardner 2017: 153; Denis & Tagliamonte 2018; Engel & Szmrecsanyi 2023). Also in accordance with previous research, *be going to* is 1.62 times more likely with second-person subjects compared to third-person inanimates (Torres-Cacoullos & Walker 2009: 339; Fehrer & Corrigan 2015: 210; Gardner

2017: 146; Denis & Tagliamonte 2018: 423). Since there is almost no difference between inanimate and animate third person subjects, animacy does not appear to impact the choice of variant, as suggested by Tagliamonte, Durham and Smith (2014: 98–99), Fehrer and Corrigan (2015) and Engel and Szmrecsanyi (2023).

The remaining probabilistic constraints in the model, viz. negation, persistence, temporal proximity, temporal adverb, verb type and clause length, are part of an interaction with proficiency level, so the main effect for these predictors only pertain to the reference level of proficiency level, namely native speakers. These effects show that native speakers are 2.53 times more likely to use *be going to* if the future marker clause is negated, echoing the findings of Nesselhauf (2010: 173) and Tagliamonte, Durham and Smith (2014: 98–99). Examiners are 1.81 times less likely to use *be going to* if they previously used *will* rather than a candidate, meaning that examiners have a tendency to reuse *will* as expected (e.g., Bock 1982; Szmrecsanyi 2006). Finally, there is a marginally significant effect for the presence of temporal adverbs in the future marker clause indicating that *be going to* is 1.52 times less likely if future marker clause features a temporal adverb. This is reminiscent of Torres-Cacoullos and Walker (2009), who claim that *be going to* is more independent temporally than *will*.

Since proficiency level is involved in many interactions, its main effect only applies to affirmative clauses of median length describing a non-proximate future event, with a dynamic verb and no temporal adverb, preceded by a previous instance of candidate *will*. Table 5.16 shows that B1, B2 and C1 learners are less likely to use *be going to* with such observations compared to native speakers, which might point towards a general preference for *will*.

5.3.2.2 Interaction effects

Table 5.17 presents the results from the model for the interaction effects. Significant interaction effects indicate that the corresponding change in predictor levels influences the choice of future marker differently for native speakers and the relevant learner proficiency level.

Table 5.17: Interaction effects of the individual predictors on the choice of future marker. Positive coefficients indicate an increase in the likelihood of *be going to*. (Marginally) significant *p*-values are written in bold.

Predictor	Coefficient	Odds ratio	Std. error	<i>p</i> -value
Proficiency x negation				
B1 + negative	-0.322	0.725	0.591	0.586
B2 + negative	-1.219	0.296	0.452	0.007
C1 + negative	-0.931	0.394	0.44	0.034
C2 + negative	-0.757	0.469	0.585	0.196
Proficiency x persistence				
B1 + candidate <i>bgt</i>	0.923	2.516	0.435	0.034
B2 + candidate <i>bgt</i>	1.467	4.336	0.375	< 0.001
C1 + candidate <i>bgt</i>	0.881	2.413	0.387	0.023
C2 + candidate <i>bgt</i>	1.309	3.703	0.504	0.009
B1 + examiner <i>bgt</i>	1.445	4.241	0.432	< 0.001
B2 + examiner <i>bgt</i>	2.203	9.048	0.420	< 0.001
C1 + examiner <i>bgt</i>	0.712	2.037	0.421	0.091
C2 + examiner <i>bgt</i>	0.675	1.965	0.584	0.247
B1 + examiner <i>will</i>	1.273	3.57	0.419	0.002
B2 + examiner <i>will</i>	1.103	3.012	0.347	0.001
C1 + examiner <i>will</i>	1.154	3.171	0.347	< 0.001
C2 + examiner <i>will</i>	1.099	3	0.452	0.015
Proficiency x temporal proximity				
B1 + proximate	0.363	1.437	0.48	0.45
B2 + proximate	1.411	4.098	0.366	< 0.001
C1 + proximate	0.759	2.136	0.331	0.022
C2 + proximate	0.358	1.430	0.450	0.427
Proficiency x temporal adverb				
B1 + present	1.177	3.244	0.398	0.003
B2 + present	0.766	2.151	0.341	0.025
C1 + present	0.47	1.6	0.332	0.157
C2 + present	0.635	1.887	0.459	0.166
Proficiency x verb type				
B1 + motion	0.783	2.188	0.478	0.102
B2 + motion	0.049	1.05	0.51	0.924
C1 + motion	-0.509	0.601	0.562	0.365
C2 + motion	-0.54	0.583	0.702	0.442
B1 + stative	-0.392	0.676	0.372	0.292
B2 + stative	-0.811	0.444	0.312	0.009
C1 + stative	-0.381	0.683	0.313	0.223
C2 + stative	0.089	1.093	0.395	0.822
Proficiency x clause length				
B1 + clause length	-0.28	0.756	0.324	0.388
B2 + clause length	0.059	1.060	0.277	0.832
C1 + clause length	-0.553	0.575	0.275	0.045
C2 + clause length	-0.563	0.57	0.350	0.108

For example, the interaction effect with negation shows that if a future marker clause is negative rather than affirmative, B2 learners are 3.38 times less likely to use *be going to* compared to native speakers, who were 2.53

times more likely to use *be going to* in negated future marker clauses. By calculating the sum of the interaction coefficient (-1.219) and the coefficient of the corresponding main effect pertaining to native speakers (0.929), one can determine the effect of negation on B2 learners. This reveals that B2 learners are 1.33 times less likely to use *be going to* when the clause is negated rather than affirmative. In what follows, each of the interaction effects will be discussed one by one and visualized with partial effects plots.

Negation

The interaction between proficiency level and negation is visualized in the partial effects plot in Figure 5.13. Partial effects plots summarize the information from the model by changing the value of the relevant predictor while holding the other predictors in the model at their default level (Fox 2003). Hence, Figure 5.13 depicts the effect of negation by changing its value while holding the other constraints in the model at their reference level. In this way, the plot shows how a change in negation affects the likelihood of *be going to* for native speakers and learners at different proficiency levels.

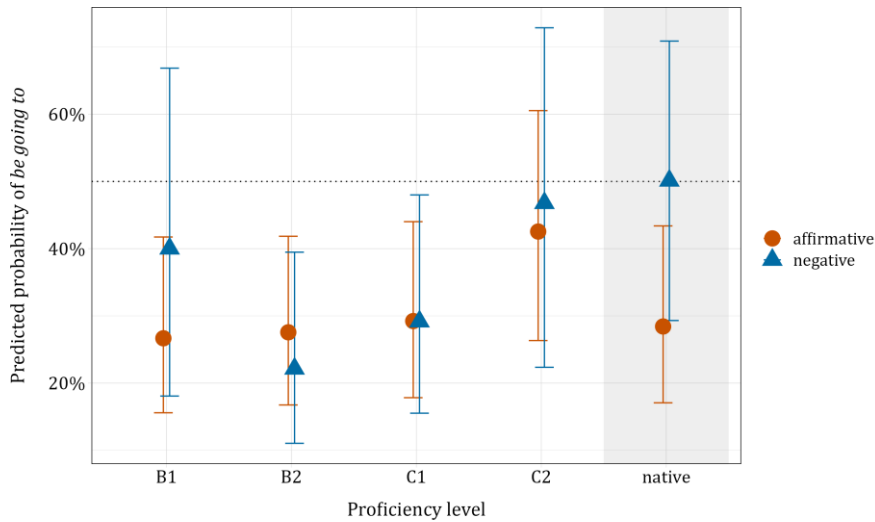


Figure 5.13: Partial effects plot of the interaction between proficiency level and negation. Probabilities on the y-axis are for *be going to*. The vertical distance between the shapes reflects the effect size of the predictor. The plotted predicted probabilities are calculated with all other categorical predictors are set at their reference level, and the numerical predictors at 0. The native proficiency level (shaded) is the reference level.

As the main effect for negation in the model shows, native speakers are 2.53 times more likely to use *be going to* when the clause is negated as expected

from most previous research (e.g., Tagliamonte, Durham & Smith 2014: 98–99). While there are not many negated clauses in the native speaker data (N = 57), this is the strongest effect in their probabilistic grammar. The interaction effect reveals that the effect of negation is much weaker for B2 and C1 learners, so they are less likely to use *be going to* in negated clauses like (87).

(87) I'm not going to throw it in the bin because it cost money
(TLC:2_SP_24, NS)

Persistence

The interaction between proficiency level and persistence is visualized in Figure 5.14.

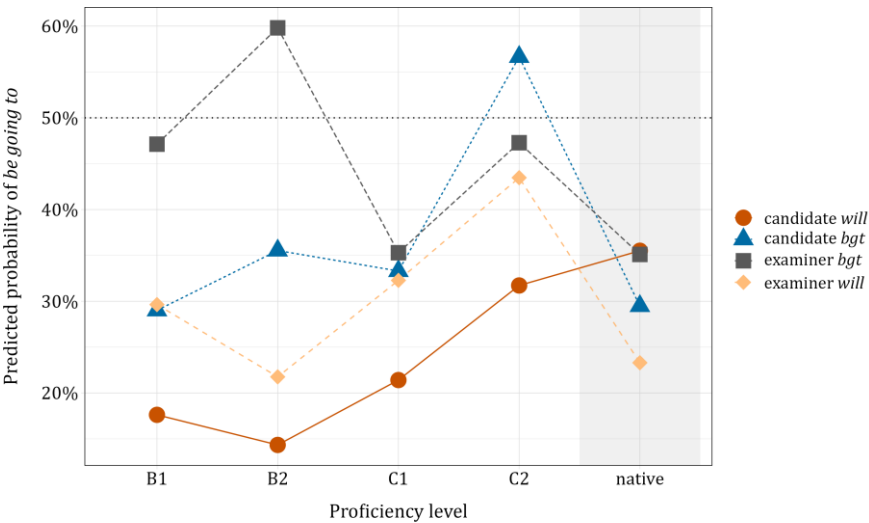


Figure 5.14: Partial effects plot of the interaction between proficiency level and persistence. Probabilities on the y-axis are for *be going to*. The vertical distance between the shapes reflects the effect size of the predictor. The plotted predicted probabilities are calculated with all other categorical predictors are set at their reference level, and the numerical predictors at 0. The native proficiency level (shaded) is the reference level.

The main effects in Table 5.16 show that examiners are more likely to use *will* only if they previously used *will* themselves. Hence, a previous instance of *be going to* does not make examiners more likely to use *be going to*, which is somewhat surprising considering that *be going to* should in principle induce stronger persistence effects than *will* as the less frequent future marker variant (Jaeger & Snider 2013). In fact, any previous future marker produced by a candidate learner does not induce persistence effects for

native speakers. By contrast, a previous instance of *be going to* by either interlocutor makes learners more likely to use *be going to*. B1 and B2 learners are particularly sensitive to previous instances of *be going to* produced by the examiner, to the extent that B2 learners are 9.1 times more likely to use *be going to* in this context. By contrast, C1 learners appear to be equally affected by *be going to* regardless of who produced the variant, whereas C2 learners are much more sensitive to their own language production. As for *will*, learners are more strongly impacted by their own previous uses of *will*, so if a learner rather than an examiner previously used *will*, learners are less likely to use *be going to*. Accordingly, Figure 5.14 shows that learners are least likely to use *be going to* when they previously produced *will*.

Temporal proximity

Figure 5.15 presents the effect of temporal proximity on the likelihood of *be going to* across proficiency levels.

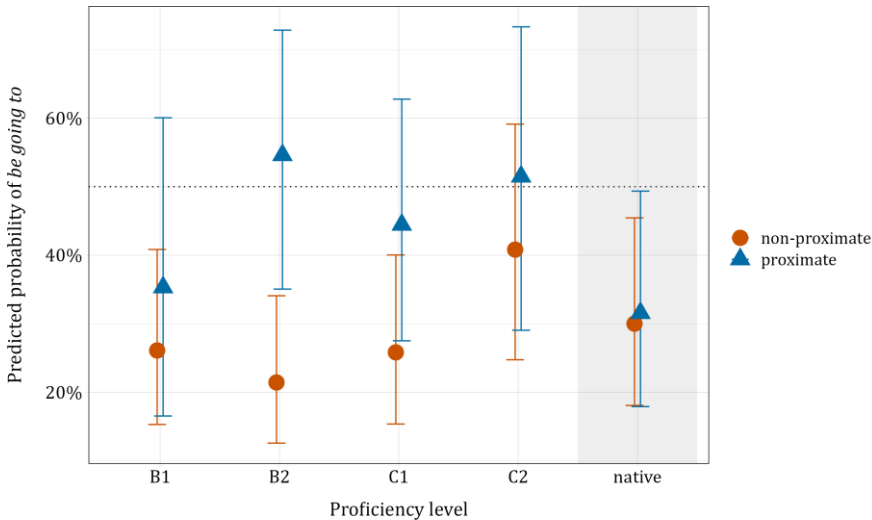


Figure 5.15: Partial effects plot of the interaction between proficiency level and temporal proximity. Probabilities on the y-axis are for *be going to*. The vertical distance between the shapes reflects the effect size of the predictor. The plotted predicted probabilities are calculated with all other categorical predictors are set at their reference level, and the numerical predictors at 0. The native proficiency level (shaded) is the reference level.

The main effects in Table 5.16 show that native speakers are not sensitive to the temporal proximity of the future event. By contrast, B2 and to a lesser extent C1 learners prefer to use *be going to* in proximate contexts as in (88), which ties in with the idea that *be going to* implies a more imminent future

than *will* (Haegeman 1989; Brisard 2001: 266; Huddleston & Pullum 2002: 211; Bergs 2010: 223–224).

- (88) actually that is the next point I'm going to say (TLC:CH_1, C1 Chinese)

A closer look at the data reveals that out of 707 proximate future events, 271 (38.3%) involve the verb *talk*, of which 185 (68.3%) occur with *be going to*. This refers to observations like (89) introducing the topic of the discussion or presentation task of the exam. These same constructions also influence the random effects in the model, where *talk* was found to be the verb with the strongest bias in favor of *be going to* (see Section 5.3.1), and they also indirectly caused a strong preference for *be going to* as the first future marker in most exams (see Section 5.2.2.5).

- (89) today for my topic phase I 'm gonna talk about parenting for which I have divided my topic into four parts (TLC:2_IN_1, C2 Gujarati)

Besides *talk*, the verbs *tell* and *speak* are also used to introduce the discussion or presentation task, so they often figure in proximate future events as well. All 22 instances of the verb *speak* and 22 out 30 (73.3%) proximate instances of *tell* take *be going to*.

Due to the frequency of these structures, it is highly likely that they are at the origin of the tendency for B2 and C1 learners to use *be going to* for proximate future events. To test this claim, I fitted a model without any instances of the verbs *talk*, *speak* or *tell*. The reduced model has a concordance index C of .86, indicating an excellent fit to the data (Baayen 2008), with no indication of harmful multicollinearity. Interestingly, the model summary reveals that while the difference between native speakers and B2 learners has become slightly smaller, the interaction remains significant for both B2 and C1 learners, confirming that B2 and C1 learners are indeed more likely to use *be going to* for proximate future events compared to native speakers.

Temporal adverb

The partial effects plot for the interaction between proficiency level and the presence of temporal adverbs is rendered in Figure 5.16.

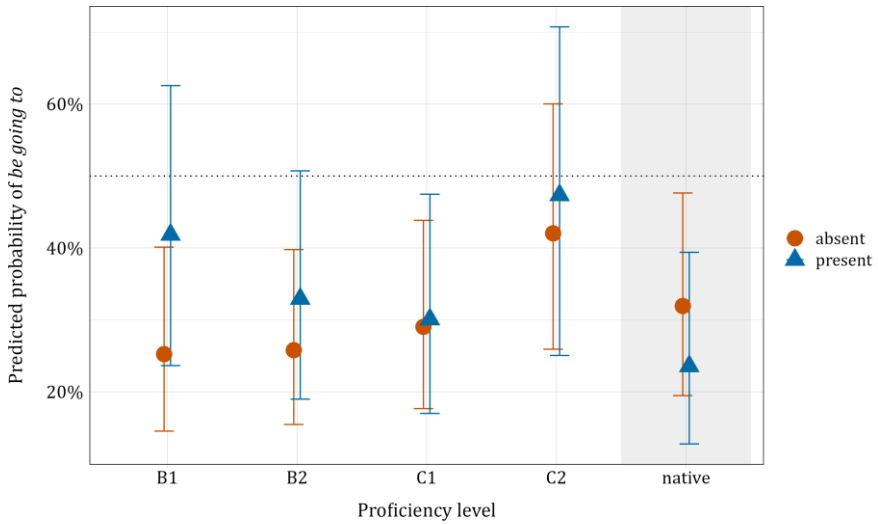


Figure 5.16: Partial effects plot of the interaction between proficiency level and temporal adverb. Probabilities on the y-axis are for *be going to*. The vertical distance between the shapes reflects the effect size of the predictor. The plotted predicted probabilities are calculated with all other categorical predictors are set at their reference level, and the numerical predictors at 0. The native proficiency level (shaded) is the reference level.

Temporal adverbs have a marginally significant effect on the choice of variant for native speakers whereby they prefer to use *will* with temporal adverbs. By contrast, B1 and to a lesser extent B2 learners are more likely than native speakers to use *be going to* in these contexts, so native speakers are less likely to use *be going to* with the temporal adverbial *this spring* in (90) compared to B1 and B2 learners. This tendency of B1 and B2 learners goes against the findings of Torres-Cacoullous and Walker (2009), who state that *be going to* is less likely to occur with temporal adverbs because the variant is more independent temporally.

- (90) this spring I'm going to take a part in a horse-riding competition
(TLC:2_6_RU_32, B1 Russian)

Verb type

Figure 5.17 portrays the effect of verb type on the choice of variant for native speakers and learners.

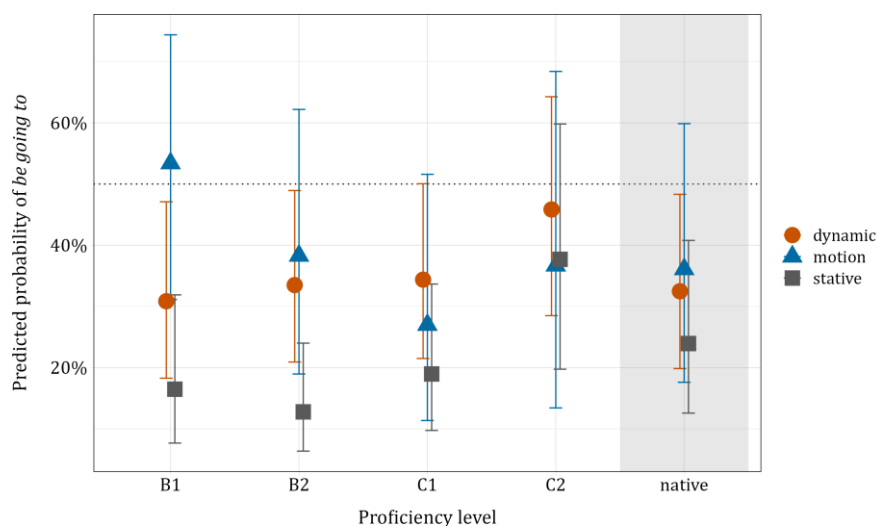


Figure 5.17: Partial effects plot of the interaction between proficiency level and verb type. Probabilities on the y-axis are for *be going to*. The vertical distance between the shapes reflects the effect size of the predictor. The plotted predicted probabilities are calculated with all other categorical predictors are set at their reference level, and the numerical predictors at 0. The native proficiency level (shaded) is the reference level.

Verb type does not play a role for native speakers, whereas B2 learners are 2.25 times less likely than native speakers to produce *be going to* if the verb is stative rather than dynamic. In other words, native speakers are more likely than B2 speakers to use *be going to* in a case like (91) with the stative verb *feel*. Figure 5.17 suggests that this is mostly due to B2 learners' stronger preference for *will* with stative verbs compared to native speakers, rather than a tendency for dynamic verbs to take *be going to*.

(91) maybe he'll feel better and try to solve his problems
(TLC:2_8_IN_18, B2 Tamil)

Clause length

Whereas previous research found that long sentences favor *be going to* (Szmrecsanyi 2003; 2006), the main effects in Table 5.16 reveal that the examiners in the corpus are not sensitive to the length of the clause. By contrast, C1 learners appear to take the length of the clause into account, but they tend to prefer *will* rather than *be going to* when the clause is long (see Figure 5.18). Hence, native speakers are less likely to use *will* in a long clause like (92) consisting of 89 letters.

- (92) because god will be the one who's going to judge who's going to the celestial kingdom who's going to the paradise (TLC:2_ME_26; C1 Spanish)

Provided the constraint is processing-related, this result suggests that *will* is the easier variant for C1 learners rather than *be going to*, possibly because it does not require *be* to be inflected (see Bardovi-Harlig 2002). If so, it is surprising that only C1 learners are sensitive to such a processing-related constraint and not the less proficient B1 and B2 learners, for whom speaking in a foreign language is arguably more cognitively demanding (Götz 2019; Götz, Wolk & Jäschke 2022: 273).

An alternative explanation is that this interaction results from the characteristics of the data. Only C1 and C2 learners gave a five-minute monologic presentation about a topic of their choice. This is the only task across the entire corpus that the learners could practice in advance, so the clauses produced by C1 learners in the presentation task appear to be significantly longer ($M = 34$, $IQR = 27$) than the clauses in the other tasks ($M = 24$, $IQR = 20$) according to a Wilcoxon test ($p < 0.001$). While preparing their presentation for the exam, the C1 learners might have consciously opted for the more formal *will* variant, resulting in an association between long clauses and the marker *will*. To explore this possibility, I removed the observations from the presentation task and ran the model again. As expected, there is no longer a significant difference between native speakers and C1 learners in this reduced model, which suggests that the presentation task does indeed contribute to the nature of the interaction effect.

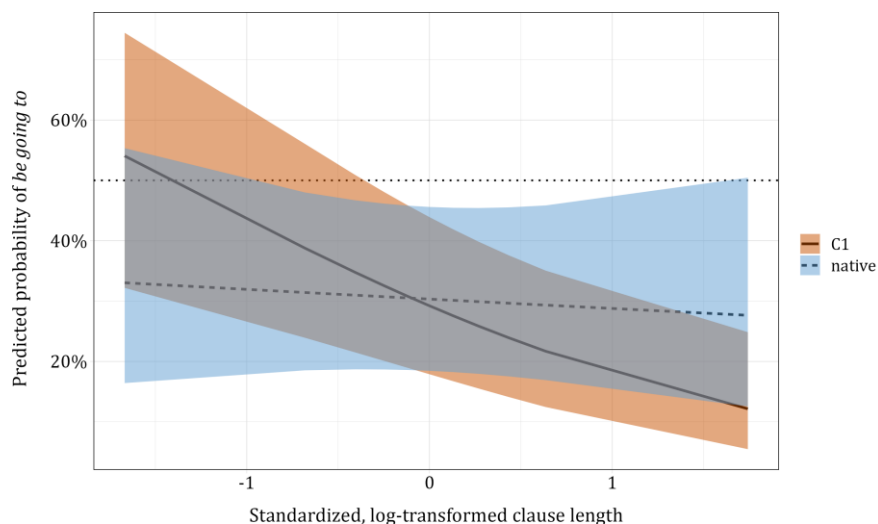


Figure 5.18: Partial effects plot of the interaction between proficiency level and clause length. Probabilities on the y-axis are for *be going to*. The values on the x-axis represent the log-transformed letter length of the clauses. The plot only shows the effect of clause length on the likelihood of *be going to* for C1 learners in orange, which differ significantly from native speakers portrayed in blue.

It is not entirely clear why C2 learners are not part of this significant interaction term, since they also produce longer clauses in the presentation task ($M = 41$, $IQR = 36$) than in other tasks ($M = 25$, $IQR = 21$) according to a Wilcoxon test ($p < 0.001$). In this regard, the relevant coefficients for the C1 and the C2 learners in Table 5.17 are very similar to each other, indicating that although the interaction term for C2 learners exceeds the threshold for marginal significance slightly ($p = 0.11$), they might share the tendency of C1 learners to use *will* in longer clauses. There are in general much fewer C2 future markers in the data than C1 observations, which might make the model less confident.

5.3.3 Supplementary analyses

As shown in Table 5.3, almost half of all the learner observations stem from Spanish L1 speakers. These Spanish EFL learners mostly have a B2 or C1 proficiency level (see Figure 5.1), where learners were found to deviate from native speakers with respect to the temporal adverb, temporal proximity, verb type, and clause length. This raises the question as to whether some of these effects are due to the Spanish learners in particular. The beyond optimal model included intercept adjustments for the mother tongue of the learners (see Section 5.2.5), but this random effect was not

retained in the model. Regardless, such an intercept adjustment could only account for any preferences Spanish learners might have for either variant; it cannot control for transfer from Spanish to the probabilistic constraints. On a related note, Mexican learners are much more likely to use *be going to* (see Section 5.3.1) than learners from other countries, which could also point to undetected differences on the level of the probabilistic constraints.

To test the stability of the model results, I fitted two additional mixed-effects logistic regression models. In the first model, I wanted to test whether the interactions remain significant without the observations from the Mexican learners. Given that the Mexican learners account for 53.8% (715 out of 1330) of the observations from Spanish L1 learners, this reduced model includes much fewer Spanish speakers. Hence, if the interaction effects remain stable, they cannot readily be attributed to Spanish speakers. The second model compares Spanish speakers with learners from other L1s more directly by allowing the probabilistic constraints to interact with L1 rather than proficiency level.

5.3.3.1 Model without Mexican learners

The reduced model was generated with the same formula as the full model:

```
Model without Mexican candidate observations ← glmer(future marker variant ~
proficiency + negation + clause type + persistence + temporal proximity +
temporal adverb + verb type + person and animacy + clause length +
proficiency:negation + proficiency:persistence + proficiency:temporal proximity
+ proficiency:temporal adverb + proficiency:verb type + proficiency:clause length
+ (1|country) + (1|exam) + (1|speaker) + (1|infinitive verb lemma), data = future
marker data without Mexican candidate observations, family = binomial,
glmerControl(optimizer = "bobyqa", maximum iterations = 1 million))
```

The resulting model has an R^2_{marginal} of 0.20 and $R^2_{\text{conditional}} = 0.39$ and the concordance index C is .88, indicating an excellent fit of the data (Baayen 2008), with no indication of harmful multicollinearity. An overview of the interaction effects in the reduced model is shown in Table 5.18.

Table 5.18: Interaction effects of the individual predictors on the choice of future marker. Positive coefficients indicate an increase in the likelihood of be going to. (Marginally) significant *p*-values are written in bold.

Predictor	Coefficient	Std. error	<i>p</i> -value
Proficiency x negation			
B1 + negative	-0.489	0.643	0.446
B2 + negative	-1.198	0.493	0.015
C1 + negative	-0.810	0.53	0.126
C2 + negative	-0.490	0.620	0.429
Proficiency x persistence			
B1 + candidate <i>bgt</i>	0.896	0.459	0.051
B2 + candidate <i>bgt</i>	1.307	0.420	0.002
C1 + candidate <i>bgt</i>	0.798	0.451	0.077
C2 + candidate <i>bgt</i>	1.452	0.544	0.008
B1 + examiner <i>bgt</i>	1.661	0.456	<0.001
B2 + examiner <i>bgt</i>	2.480	0.457	<0.001
C1 + examiner <i>bgt</i>	0.547	0.473	0.248
C2 + examiner <i>bgt</i>	0.721	0.622	0.246
B1 + examiner <i>will</i>	1.432	0.428	<0.001
B2 + examiner <i>will</i>	1.402	0.381	<0.001
C1 + examiner <i>will</i>	1.209	0.394	0.002
C2 + examiner <i>will</i>	1.376	0.480	0.004
Proficiency x temporal proximity			
B1 + proximate	0.634	0.535	0.236
B2 + proximate	1.828	0.407	<0.001
C1 + proximate	0.896	0.366	0.014
C2 + proximate	0.453	0.472	0.338
Proficiency x temporal adverb			
B1 + present	1.264	0.409	0.002
B2 + present	0.817	0.372	0.028
C1 + present	0.719	0.374	0.055
C2 + present	0.607	0.491	0.216
Proficiency x verb type			
B1 + motion	0.748	0.481	0.12
B2 + motion	-0.032	0.537	0.952
C1 + motion	-0.521	0.654	0.425
C2 + motion	-0.837	0.738	0.257
B1 + stative	-0.312	0.402	0.438
B2 + stative	-0.829	0.353	0.019
C1 + stative	-0.696	0.394	0.077
C2 + stative	-0.13	0.435	0.766
Proficiency x clause length			
B1 + clause length	-0.350	0.336	0.297
B2 + clause length	0.028	0.302	0.925
C1 + clause length	-0.755	0.322	0.019
C2 + clause length	-0.704	0.371	0.058

Table 5.18 shows that the only effect that is not retained in the reduced model is the interaction between proficiency level and negation for C1 learners. In the original model, B2 and C1 learners are found to be less

sensitive to negation than native speakers, who tend to use *be going to* when the clause is negated. In the reduced model, C1 learners no longer differ from native speakers with respect to this constraint. Additionally, certain differences between native speakers and B1 and C1 learners regarding the effect of a previous instance of *be going to* are now just above the threshold for significance, which might be due to the model having less data to work with, especially at the C1 level. All in all, the interaction effects remain stable without the Mexican data.

Interestingly, there is now a marginally significant effect for the interaction between proficiency level and temporal adverb at the C1 level, indicating that these learners are more likely to use *be going to* when there is a temporal adverb in the clause compared to native speakers, much like B1 and B2 learners. Likewise, the reduced model features a marginally significant effect showing that the C1 learners are less likely than native speakers to use *be going to* when the verb is stative, similar to B2 learners. Finally, there is a marginally significant effect for clause length in the C2 level showing that these learners prefer to use *will* when the clause is long like C1 learners. As these effects are marginally significant and based on a reduced dataset, they do not necessarily deserve more attention. Nonetheless, it is a good sign that these effects align with the patterns discovered in the full model.

5.3.3.2 Mother tongue model

For this second supplementary analysis, I fitted a mixed-effects logistic regression model where I tested whether the probabilistic constraints interact with the L1 of the learners. To do so, I used treatment coding with native speakers as the reference L1 to be compared with Spanish L1 learners and learners from all other L1 backgrounds grouped together. If the differences between native speakers and learners are truly due to proficiency level and not Spanish learners in particular, then the same effects should be present for all L1 backgrounds. That is to say, there should not be any significant differences between native speakers and Spanish learners that do not apply to the remaining ten L1 backgrounds. The model for this supplementary analysis was generated with the following formula:

Mother tongue model $\leftarrow$ glmer(future marker variant ~ proficiency + negation + clause type + persistence + temporal proximity + temporal adverb + verb type + person and animacy + clause length + mother tongue:negation + mother tongue:persistence + mother tongue:temporal proximity + mother tongue:temporal adverb + mother tongue:verb type + mother tongue:clause length + (1|country) + (1|exam) + (1|speaker) + (1|infinitive verb lemma), data =

future marker data, family = binomial, glmerControl(optimizer = "bobyqa", maximum iterations = 1 million))

The resulting model has an R^2_{marginal} of 0.14 and $R^2_{\text{conditional}} = 0.39$ and the concordance index C is .88, indicating an excellent fit of the data (Baayen 2008), with no indication of harmful multicollinearity. Table 5.19 gives an overview of the interaction terms in the model.

Table 5.19: Interaction effects of the individual predictors on the choice of future marker in the mother tongue model. Positive coefficients indicate an increase in the likelihood of *be going to*. (Marginally) significant p -values are written in bold.

Predictor	Coefficient	Std. error	p -value
L1 x negation			
Other L1 + negative	-0.858	0.416	0.039
Spanish L1 + negative	-0.895	0.412	0.03
L1 x persistence			
Other L1 + candidate <i>bgt</i>	0.948	0.355	0.007
Spanish L1 + candidate <i>bgt</i>	1.089	0.343	0.002
Other L1 + examiner <i>bgt</i>	1.401	0.367	<0.001
Spanish L1 + examiner <i>bgt</i>	1.202	0.37	0.001
Other L1 + examiner <i>will</i>	1.146	0.314	<0.001
Spanish L1 + examiner <i>will</i>	1.086	0.322	<0.001
L1 x temporal proximity			
Other L1 + proximate	1.302	0.316	<0.001
Spanish L1 + proximate	0.322	0.305	0.291
L1 x temporal adverb			
Other L1 + present	0.842	0.309	0.006
Spanish L1 + present	0.442	0.301	0.142
L1 x verb type			
Other L1 + motion	-0.096	0.444	0.83
Spanish L1 + motion	0.364	0.46	0.428
Other L1 + stative	-0.582	0.294	0.048
Spanish L1 + stative	-0.410	0.275	0.136
L1 x clause length			
Other L1 + clause length	-0.424	0.256	0.098
Spanish L1 + clause length	-0.183	0.248	0.461

The full model shows that B2 and C1 learners are less likely than native speakers to use *be going to* in negated contexts. In the reduced model, the same effect applies for both Spanish learners and learners from the other L1s. The full model also indicates that learners are more sensitive to structural persistence than native speakers. Again, the same interaction effects are found in the reduced model for both the Spanish learners and other learners. As for temporal proximity, temporal adverb, verb type and clause length, it is the learners from the other L1s who diverge from native speakers rather than the Spanish learners. Hence, the reduced model

reveals that there are in fact fewer differences between native speakers and Spanish learners than other L1s. If anything, this suggests that the data by Spanish speakers potentially obscure some of the underlying differences between native speakers and learners with other L1 backgrounds.

5.4 Discussion

This chapter presented the first-ever multifactorial corpus-based investigation of future-time reference in EFL. In total, 1,605 instances of *be going to* and 2,011 instances of *will* were collected from the examiners and learners in the TLC (Gablasova, Brezina & McEnery 2019). The examiners in the corpus are native speakers of British English, while the EFL learners have eleven different L1s with a proficiency level ranging from B1 to C2 on the CEFR scale (Council of Europe 2001). The data was annotated for a wide range of constraints that condition the choice between *will* and *be going to* according to previous literature on native speakers, including negation, structural persistence, the length of the clause, the temporal proximity of the future event, the type of sentence, the type of verb, the type of clause, the presence of temporal adverbs as well as the grammatical person and animacy of the subject (Torres-Cacoulllos & Walker 2009; Denis & Tagliamonte 2018; Engel & Szmrecsanyi 2023; Bohmann forthcoming). The data was analyzed with mixed-effects logistic regression, with the choice of future marker as response variable and the annotated constraints as predictors. Each constraint was allowed to interact with proficiency level, which enables the model to compare how constraints modulate the choice of variant for native speakers and learners across proficiency levels.

In this discussion section, I will first present a brief summary of the findings from the mixed-effects logistic regression model, starting with a discussion of the native speaker probabilistic grammar in relation to previous research before turning to the learners. Then, I discuss each of the research questions in view of the findings and assess the validity of the hypotheses. Finally, I offer a comprehensive account of the acquisition of the probabilistic grammar associated with future-time reference in EFL.

5.4.1 Summary of the findings

5.4.1.1 Native speakers

The results show that native speakers prefer *be going to* in negated, subordinated clauses with second-person subjects with no preceding *will* produced by themselves. A prototypical instance of *be going to* for native speakers would look like (93).

- (93) you know that okay you're not going to get married (TLC:2_SP_17, NS)

Conversely, prototypical instances of *will* occur in affirmative main clauses, specifically apodoses, with a third-person subject, preceded by a previous instance of *will* produced by the examiner, as in (94).

- (94) but surely if nothing's done now the problem's only going to deteriorate (TLC:2_IN_10, NS)

Although sentence type is considered one of the strongest constraints of the future marker alternation with *be going to* specializing in interrogative clauses (Torres-Cacoullos & Walker 2009: 347; Denis & Tagliamonte 2018: 424), the constraint was not retained over the course of model fitting. Additionally, there is almost no difference between inanimate and animate third person subjects, implying that animacy does not affect the choice of variant, in accordance with the results of Tagliamonte, Durham and Smith (2014: 98–99), Fehringer and Corrigan (2015) and Engel and Szmrecsanyi (2023). Likewise, native speakers ignore the type of verb, the length of the clause, the temporal proximity of the future event or the presence of adverbs in the future marker clause, all of which are not found consistently in the literature (see Table 5.14). In sum, one could say that the future markers in the spoken language production of the examiners are highly grammaticalized.

Most relevant is the study of Engel and Szmrecsanyi (2023; see also Engel 2022) because they investigate the formal and informal spoken language of British English native speakers and inspired the operationalization of most probabilistic constraints I included. In informal speech, Engel and Szmrecsanyi (2023) find that native speakers are less likely to use *be going to* with stative verbs, which is not significant in the current data. Moreover, Engel and Szmrecsanyi (2023) do find that interrogative clauses favor *be going to*. At the same time, the native speakers in Engel and Szmrecsanyi (2023) are neither sensitive to negation, grammatical person nor the presence of temporal adverbs contrary to the examiners. In formal speech, Engel and Szmrecsanyi (2023) find that first person referents are more likely to take *be going to* than third person referents. Additionally, *be going to* is less likely with proximate events in formal speech compared to informal speech. Overall, the results from the current data deviate somewhat from either informal or formal spoken British English as investigated by Engel and Szmrecsanyi (2023).

5.4.1.2 Differences between native speakers and EFL learners

The EFL learners diverge from native speakers with respect to six out of nine constraints, namely negation, structural persistence, clause length, temporal proximity, verb type and temporal adverb. Contrary to native speakers, B1 learners are more likely to use *be going to* with temporal adverbs. B1 learners are also more sensitive to virtually all effects related to structural persistence.

Most interactions involve B2 learners, so they deviate from native speakers the most. On the one hand, B2 speakers are not impacted by negation to the same extent as native speakers. On the other hand, they are influenced by the type of verb, the temporal proximity and the presence of temporal adverbs, which do not play a role for native speakers. Specifically, B2 learners prefer *be going to* in the presence of temporal adverbs when the future event is proximate, while they avoid *be going to* with stative verbs. B2 learners also exhibit the strongest structural persistence effects.

Like B2 learners, C1 learners are not sensitive to negation and they prefer *be going to* for proximate future events. However, they are not prone to using *be going to* with temporal adverbs unlike B1 and B2 learners, nor do they deviate from native speakers regarding verb type. However, C1 learners are more likely to produce *be going to* when the clause is long, although this is probably due to the data from the presentation task in particular (see 5.3.2.2). Again, C1 learners are more sensitive to structural persistence than native speakers. Structural persistence is also the only interaction involving C2 learners. However, C2 learners are not as sensitive to a previous instance of *be going to* produced by an examiner, which is otherwise the strongest persistence effect for B1 and B2 learners.

5.4.2 How does the set of constraints of learners compare to that of native speakers at different proficiency levels?

Hypothesis 1 *As proficiency increases, learners better approximate the target probabilistic grammar.*

Rationale: This hypothesis derives from the fact that language learning of any kind is driven by input according to usage-based approaches to SLA (e.g., Ellis & Wulff 2019). Learners who have reached a higher proficiency level have received more input than learners with a lower proficiency level, all other things being equal. By extension, more proficient learners have been exposed to more exemplars of *will* and *be going to* whose features, viz. the linguistic elements linked to the probabilistic

constraints, can be acquired through associative statistical learning. If the probabilistic grammar of a learner does not differ from that of native speakers with regard to any of the relevant constraints, then the learner approximates the target probabilistic grammar of native speakers. Hence, lower-proficiency learners should deviate from native speakers with regard to more constraints than more proficient learners.

B1 learners differ from native speakers with respect to two constraints, namely the effect of temporal adverbs and structural persistence. B2 learners deviate from native speakers regarding five constraints, viz. temporal adverb, structural persistence, negation, temporal proximity and verb type. Apart from clause length, C1 learners deviate from native speakers regarding three constraints, namely structural persistence, negation and temporal proximity, while C2 learners only differ from native speakers regarding structural persistence.

Given these results, I argue that the hypothesis largely holds true. Although B1 learners technically deviate less from native speakers than B2 and C1 learners, the probabilistic grammar of B2 learners is less native-like than that of C1 learners, who in turn differ more from native speakers than C2 learners. Compared to B2 learners, C1 learners no longer differ from native speakers regarding the effect of temporal adverbs, and they also better approximate native speakers regarding negation and temporal proximity as indicated by the relevant coefficients being closer to 0 (see Table 5.17). Moreover, C1 learners are less strongly influenced by previous instances of *be going to* produced by examiners, which are at the origin of some of the strongest interaction effects for B1 and especially B2 learners. The same persistence effect does not reach statistical significance for C2 learners either, who approximate the target probabilistic grammar best as they do not diverge from native speakers regarding any other constraints.

Hypothesis 2 *Learners are only sensitive to constraints that are part of the target probabilistic grammar.*

Rationale: According to usage-based approaches to SLA such as the CREED model, language learning is rational (Ellis 2006: 102–103), meaning that learners will not infer rules that go against the input they receive. In this case, learners should only learn the features associated with the probabilistic constraints that are present in the input from native speakers of British English like the examiners in the corpus. This does not apply to processing-

related constraints as they are not learned, namely structural persistence and clause length.

Like native speakers, learners prefer *be going to* with second-person subjects in subordinate clauses and avoid *be going to* with apodoses. According to the theory-based account, these features have been inferred rationally from the native speaker input via associative statistical learning. However, B1, B2 and C1 learners also consider verb type, temporal adverb and temporal proximity, which do not form part of the target probabilistic grammar and therefore cannot be acquired via associative statistical learning. Specifically, the learners from the relevant proficiency levels avoid stative verbs with *be going to*, which they use with temporal adverbs in contexts describing proximate future events.

One might wonder whether learners associate the future markers with these additional constraints because of transfer effects considering that the language system of learners is already committed to their L1 when they acquire their second language (Ellis 2006: 109). Given the many different L1s included in the data, most L1 backgrounds account for a few hundred observations. With this in mind, it is unlikely that transfer effects from a single L1 background could produce these significant effects in the model, except for Spanish L1, which accounts for about 40% of the data. Nonetheless, supplementary analyses showed that Spanish EFL learners are less sensitive to these additional constraints if anything compared to the learners from the other ten L1s. Given the typological diversity of these other L1s, it is improbable that they induce the same transfer effects so that learners from such a wide variety of L1s converged on the same non-native yet systematic distinctions between *will* and *be going to*.

A better explanation for this finding is that these constraints relate to universal features of future events. Since future events have not yet happened, they can be characterized by their temporal proximity and they always involve a certain degree of uncertainty, the latter of which can correlate with the use of adverbs: events that are more certain to happen and easier to pinpoint in the future might attract (definite) adverbs. Provided these salient features are inherent to future events, all learners regardless of their L1 could converge on these constraints and draw the same distinctions between future markers. Nonetheless, the question remains as to why learners would be driven to differentiate more between *will* and *be going to* than native speakers. To answer this question, it is important to note that learners already have *will* at their availability when they acquire *be going to*. In many ways, *will* suffices as a future marker. Not

only is it the most frequent future marker in English by far, one could also argue that it is morphologically simpler than *be going to*, which requires learners to inflect the verb *be* (see Bardovi-Harlig 2002) and deal with adverbs or negators between the auxiliary and the rest of the future marker as in (95).

- (95) it all begins with your mind and your heart so when you just have the positive attitude and you're just going to be happy (TLC:2_6_IN_125, B1 Tamil)

By comparison, the whole paradigm of *will* only consists of three fixed forms, namely *will*, *'ll* and *won't*, so *will* is possibly more advantageous in terms of processing, as implied by the fact that learners prefer this variant in long clauses (see Section 5.3.2.2). With this in mind, *be going to* mainly fulfills an expressive need as an additional future marker (see Ellis 1999: 470) rather than a functional need. Hence, B1, B2 and C1 learners might prefer to make *be going to* a more useful addition to their repertoire of future markers by restricting the form to certain and nearby future events while using *will* in all other contexts.

Hypothesis 3 *Lower-proficiency learners are sensitive to fewer constraints than more advanced learners and native speakers.*

Rationale: If language learning is rational, the maximum number of constraints learners can acquire are those constraints present in the target language input they receive. At the same time, previous research on cues showed that cues are learned one by one: learners adjust a given cue before turning to the acquisition of the next cue (Ellis & Sagarra 2010; Ellis & Sagarra 2011: 601). If construction features are similar to cues, learners should also acquire the constraints one by one. As a result, the probabilistic grammar of lower-proficiency learners should consist of fewer constraints compared to more advanced learners and native speakers, which means their probabilistic grammar is simpler (Shin 2014).

The results contradict this hypothesis, as learners are actually sensitive to more constraints than native speakers. B1 learners are influenced by the same constraints as native speakers but also take temporal adverbs into account. B2 learners are not sensitive to negation, but they consider

temporal adverbs, verb type and temporal proximity. Likewise, C1 learners take into account temporal proximity.

If B1, B2 and C1 learners pay attention to more constraints than native speakers, their use of the variants should be more predictable compared to native speakers, given that the model can rely on additional constraints to predict the choice of variant. To test this possibility, I took a random sample of 100 instances of *will* and 100 instances of *be going to* from each proficiency level in the data and compared the accuracy of the model predictions. As expected, the model is better at predicting the choice of future marker for learners than native speakers. For native speakers, the model predicts the right future marker variant in merely 73% of the cases, compared to 81% for B1 learners, 84% for B2 learner and 83% for C1 learners. With a model accuracy of 77%, the C2 level features the least predictable learners, which can be attributed to the fact that they are no longer sensitive to the presence of temporal adverbs or the temporal proximity of the future events, like the native speakers.

Given that B1, B2 and C1 learners prefer *be going to* with these rather specific features, I also expect these learners to use *be going to* less often in general compared to more advanced learners and native speakers. This claim cannot readily be confirmed with the final dataset because of the extensive sampling procedure (see Section 5.2.1), but this can be deduced from the distribution of the future markers across proficiency levels before I removed the non-interchangeable cases, shown again in Table 5.20.

Table 5.20: Distribution of the potentially interchangeable future markers for the native speakers and each of the learner proficiency levels. Percentages indicate the proportion of each variant for a given proficiency level.

	<i>Will</i>	<i>Be going to</i>
B1	1,908 (87.2%)	281 (12.8%)
B2	2,520 (84.1%)	476 (15.9%)
C1	1,152 (70.7%)	478 (29.3%)
C2	338 (63.5%)	194 (36.5%)
Native	5,501 (62.1%)	3,356 (37.9%)

As expected, the share of *be going to* increases with learners' proficiency level. While the C2 learners exhibit almost the same distribution as native speakers, with *be going to* accounting for slightly more than a third of the future markers in the data, *be going to* accounts for only 13% and 16% of the future markers in the B1 and B2 levels respectively. This suggests that

learners mostly use *will* and restrict the use of *be going to* for more specific future contexts.

If the complexity of probabilistic grammars increases with the number of conditioning variables they incorporate, then the probabilistic grammar of B1, B2 and C1 learners is supposedly more complex than that of native speakers (Shin 2014). However, this does not necessarily mean that their language use is more complex, or that their choice of variant is the result of more complex cognitive processes. Indeed, these additional constraints learners take into account rarely co-occur, so learners simply stick to *will* in the vast majority of cases as indicated by Table 5.20. Moreover, a more clear-cut distinction between variants could make the choice of variant more straightforward considering that learners tend to make more non-native choices in contexts that are less biased towards either variant according to previous studies adopting the MuPDAR approach (Gries & Deshors 2015; Wulff & Gries 2019).

5.4.3 How does proficiency level modulate the probabilistic grammar of learners?

Hypothesis *Learners first acquire constraints that facilitate associative statistical learning.*

Rationale: Previous research showed that more frequent, reliable, regular cues that are adjacent to the construction are more learnable than infrequent, unreliable, irregular, non-adjacent cues (MacWhinney 2001; Romberg & Saffran 2010: 2–3; Yoshimura & MacWhinney 2010: 552; Misyak & Christiansen 2012: 304; Yi, Lu & Ma 2017: 520). Provided construction features are similar to cues, the same factors should determine the learnability of feature-based constraints. Since these learnability factors are difficult to rank, I operationalized the learnability of a given constraint as its association strength with the variants in the language production of native speakers based on monofactorial tests. This showed that clause type is the most learnable constraint, followed by temporal proximity, which in turn should be more learnable than the remaining feature-based constraints (see Section 5.2.3.1).

According to the theory-based account, learners acquire probabilistic constraints via associative statistical learning. However, the results show that learners are sensitive to constraints that do not form part of the target

probabilistic grammar. While these constraints are not acquired through associative statistical learning from the native-speaker input, they might still be refined through associative statistical learning (Ellis & Sagarra 2010; Ellis & Sagarra 2011: 601). In other words, the learnability of the constraints could determine which of these constraints are easier to unlearn for learners to better approximate the language use of native speakers. In this way, learners are no longer sensitive to temporal adverb, verb type or temporal proximity by the time they reach the C2 level.

The feature-based constraints present in the model are negation, verb type, clause type, temporal proximity, temporal adverb, as well as the animacy and grammatical person of the subject. In line with the hypothesis, B1 and B2 learners struggle with temporal adverb and B2 learners also deviate from native speakers with respect to the type of verb. Additionally, clause type is one of only two constraints that influences native speakers and learners in the same way, which supports clause type being the most learnable constraint. However, the other constraint that even B1 learners master is the grammatical person of the subject and not temporal proximity, which is supposedly the second-most learnable constraint. Specifically, temporal proximity is problematic for B2 and C1 learners, who pay more attention to this constraint than native speakers.

While this contradicts the hypothesis, the operationalization of learnability is probably too simple to provide accurate predictions regarding the order in which the constraints are acquired. For one, association strength is meant to capture only one aspect of statistical learnability, namely the reliability of the cue. Moreover, the association strength is based on monofactorial tests and therefore fails to take into account the multifactorial nature of alternation phenomena (see Section 5.2.3.1). Perhaps unsurprisingly, the association strength between the constraints and the variants according to the monofactorial tests is not reflected by the more fine-grained model results. According to the model, negation and the grammatical person of the subject are part of native speakers' probabilistic grammar, yet these constraints are not associated with the variants according to the monofactorial tests. Conversely, monofactorial tests show that temporal proximity is significantly associated with the variants, but temporal proximity does not form part of native speakers' probabilistic grammar according to the model. A more careful assessment of learnability that takes into consideration other criteria such as cue frequency, variability and adjacency can better explain why learners deviate from native speakers regarding constraints like negation, verb type,

temporal adverb and temporal proximity rather than constraints like the type of clause and the grammatical person of the subject.

What makes the grammatical person of the subject easy to learn as a construction feature is that it is realized by a single, invariable linguistic element in the input of learners, namely the second-person pronoun *you*. Additionally, the pronoun is relatively frequent and adjacent to the future markers. By contrast, negation is expressed by several linguistic elements (e.g., *never*, *none*, *not*, *no*, *nothing*, *none*, etc.) and is rather infrequent to boot, seeing that it occurs in less than 10% of the observations (350 out of 3,616). As a result, negation does not often play a role in determining the choice of variant, which makes the constraint more difficult to (un)learn via associative statistical learning (Wang & Xu 2015: 214–215; Duguine & Köpke 2018).

Verb type, temporal adverbs and temporal proximity are hard to master mainly because they involve a wide range of linguistic elements. That is, verb type is realized by hundreds of verb lemmas with their own verb senses. For example, *have* can be stative when it is used for possession as in (66) or it can be dynamic, as in *having dinner* (see (67)). Similarly, temporal adverbs and especially adverbials are expressed with many linguistic elements that are not always adjacent to the future marker, like *one day* in (96).

- (96) are you going to get married and have children and make your own family one day (TLC:7_CH_17, NS)

The temporal proximity constraint is even more complex, since it can be encoded not only by temporal adverbs, but also through inference from the context and previous discourse.

Admittedly, subordinated clauses are introduced by several elements as well (see Section 5.2.2.3), but they are frequent considering that almost a third (29%) of all native speaker clauses in the data are subordinated (229 out of 784). By comparison, apodoses are rather infrequent (46 out of 784 (6%)), but they are strongly associated with the choice of variant and they are easily identifiable through the invariant subordinator *if*.

5.4.4 How do native speakers and learners at different proficiency levels compare regarding processing-related constraints?

Hypothesis *Lower-proficiency learners are more sensitive to processing-related constraints than more advanced learners and native speakers.*

Rationale: Since language learners constantly have to juggle between two or more language systems (Kroll & Dussias 2013), and because language production is less automatic for especially low-proficiency learners compared to native speakers (Segalowitz 2003), processing-related constraints should be stronger for learners (see e.g., Wulff, Lester & Martinez-Garcia 2014; Deshors & Gries 2014; Wulff, Gries & Lester 2018; Paquot, Grafmiller & Szmrecsanyi 2019). Given that virtually all constraints can be tied to processing, I argued that the characteristics of psychophysiological constraints as defined by Tamminga, MacKenzie and Embick (2016) can help identify strictly processing-related constraints resulting from features of the brain (see Section 3.2.1), in this case structural persistence and clause length.

At first glance, the data confirms this hypothesis given that C1 learners are more sensitive to the length of the clause than native speakers. However, if C1 learners are more sensitive to a processing-related constraint like clause length, then lower-proficiency learners, i.e., B1 and B2 learners, should a priori also be affected by this constraint since their language production is in all likelihood not more automatic or fluent than that of C1 learners. Therefore, I instead argued that the association between clause length and the future markers is due to C1 learners preferring the more formal *will* variant during the presentation task, which features longer clauses on average. A similar tendency is observed for C2 learners, who also have to give a monologic presentation although this effect does not reach statistical significance.

As expected, all learners are more sensitive to structural persistence than native speakers. Whereas native speakers only pay attention to their own previous production of *will*, learners are more likely to reuse either future marker regardless of the interlocutor. Low-proficiency learners are particularly likely to repeat *be going to* after an examiner to the extent that this is the strongest constraint for B2 learners. This can partially be

attributed to *be going to* being the more infrequent variant (Jaeger & Snider 2013), especially for B1 and B2 learners, who overwhelmingly prefer *will*. Interestingly, this effect quickly tapers down at the C1 level, which is somewhat surprising if structural persistence is strictly processing-related. Instead, the strong persistence effects at the B1 and B2 level could partially be due to these learners experimenting with the newly incorporated future marker *be going to*. Learners hear the examiner, a native speaker of British English, use *be going to* in contexts where the learner would prefer *will*, which could lead to a high degree of surprisal (Ellis & Wulff 2019: 47–49), resulting in stronger structural persistence effects than a previous instance of the more familiar *will* (Jackson 2018: 541–542). As a driver of implicit learning, these persistence effects could eventually lead to changes in the interlanguage of learners regarding their use of *be going to* (Bock et al. 2007: 455–456; Tooley & Bock 2014: 113–114).

5.4.5 Synopsis

In general, the development of the future markers in the spoken language production of EFL learners follows Ellis' account on the acquisition of alternating variants (Ellis 1999). First, learners only use one variant where native speakers would use two. This would correspond to a stage shortly before B1 when learners have acquired *will* but not yet *be going to* (Bardovi-Harlig 2002). Learners then become acquainted with the second form, viz. *be going to*, and they gradually start using both forms systematically. However, this systematicity does not necessarily correspond to the way in which variation occurs in the target language. Specifically, B1, B2 and C1 learners are influenced by constraints that are not relevant for native speakers, namely the type of verb, the presence of temporal adverbs and the temporal proximity of the future event. Additionally, native speakers and learners diverge regarding the strength of the constraints they both consider. For example, B2 and C1 learners are less sensitive to negation than native speakers, but they are more sensitive to structural persistence. Ultimately, more advanced learners at the C2 level use the forms in the same way as native speakers (Ellis 1999: 475).

According to the theory-based account, learners should only be sensitive to constraints present in the target probabilistic grammar because language learning accommodates the input learners receive in a rational way (see Section 3.4), and they should be sensitive to fewer constraints because the construction features are learned one by one (Ellis & Sagarra 2010; Ellis & Sagarra 2011). Hence, the fact that learners incorporate constraints that do not form part of native speakers' probabilistic grammar

is surprising, especially considering the wide variety of L1 backgrounds covered by the data. One possible explanation is that future events are associated with certainty and temporal proximity by default, so these characteristics can easily be incorporated as construction features by learners from any L1 background. Since *be going to* is redundant when learners already have *will* at their availability, which is more frequent and probably easier to process to boot, learners have opted to functionally distinguish *be going to* from *will* by restricting its use to future events that are certain and temporally proximate (see Bardovi-Harlig 2005), rather than fulfilling an expressive need for an additional future marker that is semantically equivalent to *will* (see Ellis 1999: 470).

Since these additional constraints do not form part of the target probabilistic grammar, associative statistical learning is not responsible for their acquisition. Instead of arising gradually through language exposure, they are implemented in a top-down fashion. However, these constraints could be refined or unlearned via associative statistical learning. As learners are exposed to more exemplars of future markers in the target language, they notice that the type of verb, the presence of temporal adverbs or the proximity of the future event are in fact not associated with the choice of variant for native speakers, so they are gradually unlearned. This in turn increases the share of *be going to*, so learners ultimately match the distribution of the variants in the target language by the time they reach the C2 level. The success of statistical learning is determined by the reliability, frequency, variability and adjacency of the relevant construction features and the constructions (e.g., MacWhinney 2001; Yoshimura & MacWhinney 2010: 552; Misyak & Christiansen 2012: 304). For instance, temporal proximity is hard to refine because it must be inferred either from temporal adverbs and adverbials or from the preceding context, whereas the grammatical person of the subject is realized by invariant pronominal forms whose association with the future markers is easier to notice.

In this account, learners diverge from native speakers because they have to figure out how *be going to* relates to *will*, which is by far the dominant form in learner language. The fact that the function of *be going to* in the language system is not well-understood is reflected by learners' sensitivity to structural persistence, which shows that they are particularly attentive to instances of *be going to* produced by the examiner, probably because learners would have opted for *will* instead (Jackson 2018: 541–542). As a driver of implicit learning (Bock et al. 2007: 455–456; Tooley & Bock 2014: 113–114), these structural persistence effects could contribute

to the development of learners' language use towards the native-speaker standard, who alternate between the markers more freely.

5.4.6 An alternative explanation

Probably the most surprising finding is that learners are sensitive to certain constraints that do not influence native speakers, namely the type of verb, the presence of temporal adverbs and the temporal proximity of the future event. Given that this contradicts the assumption that learners rationally infer the constraints from their language input through associative statistical learning, I argued that learners purposely restricted *be going to* to certain events taking place in the nearby future to distinguish it from the dominant form *will*. However, an alternative explanation is that there is a mismatch between the language production of the native speakers in the corpus and the input the learners received.

First, examiners might exhibit features of Teacher Talk, whereby the examiners simplify their language use according to the proficiency of the learner (Patil 1994: 155; Stanley & Stevenson 2017: 2), possibly in combination with features from Foreigner Talk (Ferguson 1975: 1; Wooldridge 2001: 622; Dashti 2013: 64), which prioritizes communication efficiency. Since the probabilistic constraints are largely implicit, I argued in Section 4.1.1.2 that it is unlikely these adaptations influence the constraints to such an extent.

Hence, a second possibility is that learners are not exposed to the type of input the examiner produce. For example, learners might have received more input from another variety of English, which would have to be more conservative than the British English variety of the examiners given that constraints like verb type and temporal proximity have not yet been levelled completely via grammaticalization. However, one would expect these differences to persist across all proficiency levels if this were the case. Yet, the results clearly show that the learners ultimately approximate the probabilistic grammar of the examiners as they become more proficient seeing that C2 learners do not deviate from native speakers regarding any of the feature-based constraints.

The more likely possibility is that learners picked up these constraints from EFL textbooks, which can be an important source of target language input (see Le Foll 2023: 259–260). Provided language input drives language learning, textbooks could exert considerable influence on the language use of the learners, even though these books do not always adhere to recent findings from SLA (Mishan 2022: 497) nor does their language use reflect authentic corpus data (Burton 2023: 25). For example, Winter and Foll

(2022) found that the way *if*-clauses are taught and exemplified in EFL textbooks poorly captures conditionals in authentic (British) English. Nonetheless, the conditionals produced by French learners exposed to these textbooks are clearly influenced by these rules. Le Foll's (2023) study on causatives (e.g., *make someone do something*) largely corroborates these findings as well: EFL learners overuse causatives that are overrepresented in EFL textbooks and they underuse causatives that receive less attention.

EFL textbooks are particularly relevant in the present context because future markers are very often part of textbook material, to the extent that Burton (2023) considers future-time reference to be representative of the 'canon' of EFL textbooks. Against this background, the tendency for *will* to take stative verbs, or for *be going to* to describe temporally proximate events with temporal adverbs might derive from prescriptivist rules. Moreover, learners from a wide variety of L1 backgrounds as in the TLC data could be influenced by the same rules given that many popular textbooks are meant for a global audience, which is more cost-efficient than publishing localized versions of the textbooks (see Mishan 2022).

To test this hypothesis, I looked through a number of highly popular global EFL textbooks, including certain installments of the *Headway* series (Soars & Soars 2005, 2009a, 2009b, 2009c, 2019), as well as *face2face* (Redstone & Cunningham 2005a, 2005b, 2005c, 2012) and *New English File* (Oxenden & Latham-Koenig 2006), which are also included in the case study by Burton (2023) on future-time reference in EFL textbooks. Additionally, I consulted the English reference grammars *Practical English Usage* (Swan 2005) and *English Grammar in Use* (Murphy 2019). Finally, I looked into two series specifically aimed to help EFL learners pass the GESE exams, namely *Pass Trinity Now* (Cochrane 2015; Clyde & Parker 2015; Clyde & West 2015; Hansen 2015) and *Ready for Trinity* (Humphries 2017a, 2017b). Table 5.21 presents an overview of the uses associated with *will* and *be going to* in these textbooks.

Table 5.21: The uses of *will* and *be going to* according to a sample of popular EFL textbooks and reference grammars (see also Burton 2023: 126).

<i>Will</i>	<i>Be going to</i>
▪ Future decisions or intentions made at the moment of speech ▪ Neutral future (facts in the future) ▪ Predictions based on personal opinion without present indications ▪ Predictions with words like <i>probably</i> and after verbs indicating a decision or judgment (e.g., <i>think</i>) ▪ Offers, requests, agreements, promises, orders, threats, announcements of decisions ▪ Asking and giving information about the future	▪ Future decisions, intentions or plans made before the moment of speech ▪ Predictions based on present indications ▪ Not with motion verbs ▪ Asking questions about planned future events ▪ Commands and refusals ▪ Not used in short answers

One could argue that the recommendation to use *be going to* for ‘predictions based on present indications’ listed in Table 5.21 is at the origin of the temporal proximity effect at the B2 and the C1 level (see also Haegeman 1989; Brisard 2001: 266). Likewise, the suggestion to use *will* in offers, requests, agreements etc. and with ‘verbs indicating a decision or judgment (e.g., *think*)’ could also explain why B2 learners are more likely to use *will* with stative verbs. As learners become more proficient and are exposed to more authentic English input, they gradually unlearn these textbook rules and the constraints they entail through associative statistical learning.

While this explanation seems plausible, not all rules listed in Table 5.21 appear to influence learners. For example, the textbooks claim that *be going to* is preferred for questions about plans, whereas *will* should be used when asking about information about the future. While previous research does not draw such a distinction between types of interrogatives, sentence type is consistently an influential constraint (e.g., Szmrecsanyi 2003; Tagliamonte, Durham & Smith 2014; Denis & Tagliamonte 2018: 424). Yet, learners do not differ from native speakers regarding sentence type, which implies that learners do not rely on this prescriptivist rule. Moreover, learners also do not seem to avoid combining *be going to* with motion verbs as opposed to the recommendation of many textbooks. More generally, all EFL textbooks first introduce *be going to* before *will*, which calls into question the impact of EFL textbooks considering that *will* is always learned first regardless (Bardovi-Harlig 2005). In sum, the relation between the learners’ language use and the recommendation of EFL textbooks remains

unclear and requires a more in-depth analysis, which is unfortunately beyond the scope of the thesis.

6 The genitive alternation

6.1 Introduction

In this second case study, I conduct a quantitative corpus-based study to compare the probabilistic grammar of native speakers with that of EFL learners at different proficiency levels regarding the genitive alternation, which consists of the *s*-genitive (97) and the *of*-genitive (98).

(97) The city's name (TLC:2_6_CH_32, B1 Chinese)

(98) The name of our city (TLC:2_6_CH_53, NS)

In the *s*-genitive, the possessor (*city*) is followed by the genitive marker (*'s*) and then the possessum (*name*). In the *of*-genitive, the order of the constituents is reversed: the possessum *name* is followed by the possessor, which is part of a prepositional phrase introduced by *of*. Hence, the genitive alternation is a permutation alternation according to the classification of Troij, Grondelaers and Speelman (2023).

Historically, both constructions have coexisted since the 9th century. The *s*-genitive stems from the genitive case in Old English, where the *s*-ending formed one part of the inflectional paradigm of the genitive case. At the time, the possessor of an inflectional genitive could not only precede but also follow its possessum similar to the contemporary German inflectional genitive. Due to the deflexion of the English language, the postposed genitive was increasingly replaced by the preposed inflectional genitive and by the *of*-genitive, which had been on the rise with support from the French *de*-genitive. Originally, the *of*-genitive was a marginal variant in the early Old English period that had developed from the meaning *out of* (Allen 2003; Hinrichs & Szmrecsanyi 2007: 437; Heller 2018: 11). By the 14th century, the *of*-genitive became the most frequent variant. Around this period, the *s*-genitive was reanalyzed as a clitic, possibly because the *s*-ending had been generalized to all noun classes. Since the 15th century, the *s*-clitic has been on the rise, eventually ousting the inflectional genitive completely during the Early Modern English period (Allen 2003; Rosenbach 2014: 234; Heller 2018: 12). The frequency of the *s*-genitive diminished in the late Early Modern English period due to a decrease in the possessors and semantic relations that take the *s*-genitive such as legal ownership relations (e.g., *the*

woman's handbag) (Szmrecsanyi 2013). Ever since this dip, the frequency of the *s*-genitive has increased at varying rates depending on genre and variety, which is mostly due to the fact that the animacy constraint, the most important constraint of the *s*-genitive (Hinrichs & Szmrecsanyi 2007: 450; Rosenbach 2017: 14; Heller 2018: 106), is gradually becoming weaker. *S*-genitives are increasingly acceptive of possessors from less animate noun classes (Wolk et al. 2013: 408; Rosenbach 2014: 236; Heller 2018: 12).

The choice of genitive variant is jointly determined by a myriad of constraints, some of which are reminiscent of well-known semantic-pragmatic typological scales, such as the animacy and definiteness scales (e.g., Comrie 1989). Recent accounts of the genitive alternation tend to explain the constraints in terms of language processing. An oft-mentioned processing principle in this regard is Easy First (MacDonald 2013; see e.g., Heller 2018), which states that speakers prefer to produce constituents that are easy to process first. In the context of the genitive alternation, speakers prefer the *s*-genitive when the possessor is easy to process, i.e., when it is short, animate, definite and discourse-given, because this variant allows speakers to produce the possessor right away. Other genitive constraints are more difficult to explain in terms of processing, such as the effect of final sibilancy whereby *s*-genitives tend to be avoided when the possessor ends in a sibilant (Hinrichs & Szmrecsanyi 2007: 453; Shih et al. 2015).

6.1.1 Previous research

At present, the genitive alternation has been studied very extensively in previous research on native speakers, to the extent that it is arguably the best researched of all syntactic alternations (Rosenbach 2014: 215). For an extensive overview of previous research on the genitive alternation, see Appendix A in Rosenbach (2014). Most relevant are studies that take into account the multifactorial nature of the genitive alternation in their analysis by including multiple probabilistic constraints and investigating their effect on the choice of genitive variant.

Hinrichs and Szmrecsanyi (2007) and Szmrecsanyi and Hinrichs (2008) looked into the factors that underlie the ongoing increase of the *s*-genitive by comparing spoken and written data in American English and British English. They found that the lexical density of texts has increased considerably in press language, leading to stronger economization effects, which refers to the pressure to be concise in writing. Specifically, thematic possessors prefer the *s*-genitive, and so do lexically dense texts. The effect of thematicity is stronger in American English whereas possessor animacy is more pronounced in British English.

Szmrecsanyi (2013) researched the competition between the *s*-genitive and the *of*-genitive in British English press language over time and the same data was used by Wolk et al. (2013) to compare the development of the genitive alternation and the dative alternation. These studies showed that the strongest constraint of the genitive alternation, viz. possessor animacy, is gradually weakening. That is, inanimate possessors are increasingly hosted by *s*-genitives, and a similar trend is visible in the dative alternation. Another common observation is that the topics discussed in the texts change over time, which also influences the probabilistic grammar.

Shih et al. (2015) investigated how speakers choose between the genitive variants in spoken American English and focused in particular on rhythmicity as an additional constraint. They found that while rhythmicity plays a role in the choice of variant in that speakers prefer the genitive variant that best approximates an ideal weak-strong alternation of stressed syllables, this effect remains rather weak. Grafmiller (2014) also included rhythmicity in his investigation of the genitive variants across different registers of spoken and written language. The results largely aligned with previous research, although there is no direct evidence of economization through constraints such as possessor thematicity and lexical density.

Besides native speakers, the genitive alternation has also been studied in ESL varieties. Hundt and Szmrecsanyi (2012) compared the effect of possessor animacy on the genitive alternation in New-Zealand English and British English and found that the constraint is generally stronger in New-Zealand English. Similarly, Rosenbach (2017) focused on the role of animacy in the English and Afrikaans genitive alternation. A first step revealed that Afrikaans speakers are less sensitive to possessor animacy than speakers of English. In a follow-up experiment, she found that native speakers of Afrikaans with English L2 transfer this weaker animacy constraint to English (Rosenbach 2017: 17). Heller (2018; see also Heller, Szmrecsanyi & Grafmiller 2017) researched the genitive alternation in the written and spoken language of speakers from nine varieties of English, which revealed that the ESL varieties are less sensitive to possessor animacy than native speakers. Finally, Heller, Bernaisch and Gries (2017) tried to identify linguistic epicenters in Asia by analyzing the genitive alternation. A common finding in these studies is that possessor animacy, although generally the strongest constraint of the genitive alternation, is also the most variable.

The only study on the genitive alternation in EFL is Gries and Wulff (2013), who compared the written language use of British English native speakers with that of Chinese and German EFL learners. They found that

native speakers pay attention to the semantic relation expressed by the possessor and the possessum when choosing between the genitive variants, whereas Chinese learners do not take this constraint into account. By comparison, German learners are more sensitive to this constraint than both native speakers and the Chinese learners. Moreover, German learners are also more responsive to the specificity of the possessor, while native speakers and Chinese learners are more sensitive to the specificity of the possessum. These differences were mostly attributed to the L1 of the learners. Overall, they concluded that “intermediate/advanced learners’ production data already reflect target-like processing preferences at an astonishing level of complexity” (Gries & Wulff 2013: 349).

6.1.2 Present study

The present study on the genitive alternation addresses some of the limitations of Gries and Wulff (2013) and other studies involving (permutation) alternations. First and most importantly, neither Gries and Wulff (2013) nor any other studies on alternation phenomena in EFL controlled for the effect of learners’ proficiency level on speakers’ choice of variant, so it is unclear whether some of the findings are limited to one or more proficiency levels. For example, it might be the case that only the lower-proficiency Chinese EFL learners do not take the semantic relation between possessor and possessum into account, whereas more proficient Chinese EFL learners approximate the target probabilistic grammar perfectly. At the same time, previous research may also have overlooked certain effects related to proficiency level given that it is difficult to detect effects proper to a specific proficiency level when data from all proficiency levels are conflated together during statistical analysis. Secondly, the current study covers a much wider range of typologically diverse L1 backgrounds, which increases the generalizability of the findings, and helps ensure that L1 does not overlap with proficiency level during statistical analysis. Thirdly, the present study features spoken language production rather than written language, which offers a better window into the underlying grammatical system of learners than written language according to SLA (Myles & Mitchell 2014). Finally, the current study explicitly tests hypotheses based on the theory-based account whereas most previous studies had a primarily methodological focus.

Concretely, I collected thousands of interchangeable instances of *s*-genitives and *of*-genitives from the TLC (Gablasova, Brezina & McEnery 2019). The collected observations were annotated for structural persistence, the lexical density of the surrounding context, the animacy,

definiteness, final sibilancy, discourse-status, thematicity of the possessor, the length of the possessor and the possessum as well as the semantic relation that holds between them (Rosenbach 2014; Heller 2018). The annotated dataset was then analyzed with mixed-effects logistic regression to answer the research questions of the thesis:⁵

- RQ 1: How does the set of constraints of learners compare to that of native speakers at different proficiency levels?
- RQ 2: How does proficiency level modulate the probabilistic grammar of learners?
- RQ 3: How do native speakers and learners at different proficiency levels compare regarding processing-related constraints?

For the first research question, there are three hypotheses (see Section 3.4): i) more advanced learners will better approximate the target probabilistic grammar, ii) learners will only be sensitive to constraints that also affect native speakers, and iii) lower-proficiency learners will be sensitive to fewer constraints than native speakers. As for the second research question, lower-proficiency learners are expected to struggle more with constraints that are difficult to learn via associative statistical learning than with constraints that are easy to learn. Applied to the genitive alternation, learners should first master the animacy constraint and the semantic relation of the possessor and possessum. The remaining feature-based constraints should prove more challenging, namely the definiteness of the possessor and its final sibilancy. For the third research question, I hypothesize that lower-proficiency learners should be more sensitive to processing effects than native speakers, namely the thematicity and discourse status of the possessor, the length of the possessor and the possessum and the lexical density of the context surrounding the genitive variant.

The remainder of this chapter is structured as follows. In the next section, I present the data and methodology of the study, which covers the delimitation of the variants, the collection and annotation of the data, the classification of the constraints as either feature-based or processing-related as well as the method of statistical analysis, viz. mixed-effects logistic regression. The results from the regression model are presented in Section 6.3, starting with the random effects before turning to the fixed effects and several supplementary analyses including MuPDAR (see Gries &

⁵All data and code are available in an online repository at <https://osf.io/6ng7x/>.

Adelman 2014; Gries & Deshors 2014). The chapter concludes with a discussion of the results and how they relate to the hypotheses.

6.2 Data and Methods

6.2.1 Delimiting the variable context

As discussed in Section 4.1.2, the present study focuses exclusively on the competition between the *s*-genitive and the *of*-genitive (Quirk et al. 1985: 1276–1278; Heller 2018: 59). However, not all pairs of noun phrases connected by *of* or the *s*-clitic are interchangeable genitive variants, so one has to remove the categorical cases that exclude the other variant (Kreyer 2003: 170; Rosenbach 2014: 223; Heller 2018: 16). In the following subsections, I discuss each of these categorical cases.

Double genitives

Double genitives consist of both an *of*-preposition and a following *s*-genitive without possessum, as in (99). It is very similar to independent *s*-genitives as in (100), which also lack a possessum noun phrase. Double genitives were omitted because they are not structurally equivalent to either the *s*-genitives or the *of*-genitives while also having idiosyncratic properties of their own (Rosenbach 2014: 222; see e.g., Szmrecsanyi & Hinrichs 2008: 295–296; Grafmiller 2014: 475; Heller, Szmrecsanyi & Grafmiller 2017: 8).

(99) The best friend of John's (Rosenbach 2014: 222)

(100) His car is faster than John's (Kreyer 2003: 171)

Indefinite genitives

The possessum of *s*-genitives is definite by default by virtue of the possessor, which takes the role of determiner (Huddleston & Pullum 2002: 477; Hinrichs & Szmrecsanyi 2007: 446). Therefore, an *of*-genitive with an indefinite possessum like (101) cannot alternate. In fact, any determiner other than the definite article *the* in the possessum noun phrase of *of*-genitives makes the structure non-interchangeable (Rosenbach 2002: 30; Grafmiller 2014: 475).

(101) This is an example of your English test (TLC:2_6_CH_25, NS)

Fixed genitives

Fixed genitives seem to alternate at first glance, but in reality only appear in one of the genitive variants (e.g., Grafmiller 2014: 475; Heller, Szmrecsanyi & Grafmiller 2017: 8). A further distinction exists between idiomatic expressions (102) and titles or fixed names (103). If there were

no reliable instances of the other variant on Google or the GloWbE corpus (Davies & Fuchs 2015), the observation was considered non-interchangeable (Heller 2018: 56).

(102) Within arms' reach (Shih et al. 2015: 211)

(103) this is the final scene of The Name Of The Rose a novel written by Umberto Eco (TLC:2_IT_29, C1 Italian)

Classifying genitives

Classifying genitives differ from other *s*-genitives in that the possessor does not refer to a specific referent and instead denotes the class to which the possessum belongs (Quirk et al. 1985: 327; Szmrecsanyi & Hinrichs 2008: 295–296; Grafmiller 2014: 475; Heller 2018: 55). Moreover, classifying genitives are often idiomatic (Quirk et al. 1985: 328) so they tend to become compound nouns (Rosenbach 2002: 31). To identify classifying genitives, one can paraphrase the *s*-genitive as a prepositional structure with *for* (Heller 2018: 55), so one could paraphrase *school's uniform* in example (104) as *the uniform for school*.

(104) well I think it's good er er to have er some r-rules er about er er school er school's uniform (TLC:2_6_RU_8, B1 Russian)

Pronominal and non-nominal genitives

Pronominal genitives refer to *s*-genitive and *of*-genitives where either noun phrase is pronominal. This includes structures like *her style* (Heller 2018: 56), which do not alternate because the *of*-genitive equivalent *the style of her* is very rare (Rosenbach 2002: 30). Pronominal genitives also refer to categorical cases like (105) where the possessum *that* is a pronominal demonstrative referring to *face* (Wolk et al. 2013: 387). The only pronouns included are indefinite pronouns like *one*, *everybody*, *anyone* etc. in order to retain cases like (106). Also excluded from the analysis are genitive structures where either constituent does not contain a nominal head to take pseudo-nominal verb phrases such as (107) into account.

(105) Her face is as ugly as that of a dog (Wolk et al. 2013: 387)

(106) You cannot guess the future of someone by looking at the star (TLC:2_8_ME_14, B2 Spanish)

(107) I think that erm they want to discover again the beauty of making things by hand because it's not funny for machine do everything (TLC:2_IT_27, C1 Italian)

Partitive genitives

Partitive genitives are used as a cover term for a number of different structures that typically involve some sort of measurement (Szmrecsanyi & Hinrichs 2008: 296; Heller 2018: 56). In example (108), the fused-head quantifying determiner *some* fills the position of possessum followed by *of* and a partitive oblique noun phrase (Huddleston & Pullum 2002: 412). In examples (109) and (110) the partitive genitives feature full-fledged noun phrases consisting of a determiner and a nominal head. Finally, some *s*-genitives like *a few years' time* in (111) can also convey measurement.

- (108) What do you think are some of the advantages and disadvantages of living in a village (TLC:2_7_SP_51, NS)
- (109) Currently the majority of water on the Earth's surface is not suitable for human consumption (TLC:SP_67, C1 Spanish)
- (110) For you happiness is like sitting alone having a nice cup of tea (TLC:2_IN_16, C2 Marathi)
- (111) I think I was going to be volunteer a few years' time and also I the volunteer to be er a translator (TLC:2_6_CH_43, B1 Chinese)

Appositive genitives

Appositive genitives refer to *of*-genitives where the head of the *of*-complement is coreferential with the head of the noun phrase in the possessum slot and usually describes it further (Heller 2018: 55). To identify appositive genitives, one can either paraphrase example (112) as *the princely state that is Travancore* (Heller 2018), or turn the *of*-genitive into a sentence with *be*, as in *Travancore is a princely state* (Quirk et al. 1985: 1284).

- (112) He's a painter from the princely state of Travancore (TLC:IN_11, C1 Unknown)

Modifying *of*-genitive

In modifying *of*-genitives, the *of*-complement modifies the preceding noun phrase. One can optionally distinguish between material genitives (113) (Shih et al. 2015: 211) and descriptive genitives (114) (Rosenbach 2002: 29). Their prenominal equivalent would be an adjectival structure rather than an *s*-genitive (e.g., *a golden crown*) (Rosenbach 2002: 31).

- (113) A crown of gold (Shih et al. 2015: 211; Kreyer 2003: 171)
- (114) All persons of quality here present (Wolk et al. 2013: 388)

Own

Finally, I excluded *s*-genitives with possessums that are premodified by *own*, as in (115). *Own* is unique in that it solely occurs after *s*-genitives and possessive pronouns (Huddleston & Pullum 2002: 482), so these structures do not alternate with prepositional genitives (Hinrichs & Szmrecsanyi 2007: 446).

- (115) The parents should respect the the kid's own ideas yeah actually that is the kid's own life not the parents' life so I think the kids can choose their er his own way his own life (TLC:CH_1, C1 Chinese)

6.2.2 Data collection

The genitive observations were collected from the TLC (Gablasova, Brezina & McEnery 2019) in three steps. First, potential *s*-genitives and *of*-genitives were identified in the corpus data with the help of an R script (R Core Team 2022), which also excluded some of the categorical genitives automatically. The remaining non-interchangeable genitives were removed in a second step. Then, I extracted the interchangeable genitives from the corpus data based on their noun phrase boundaries.

6.2.2.1 Identifying potential genitives

The main goal of the R script was to mark the potentially interchangeable genitives with a high recall. Ideally, regular expressions would identify a noun phrase, followed by the genitive marker, viz. either the clitic 's, an apostrophe ', or *of*, and then another noun phrase. However, it is difficult to identify the noun phrase preceding the genitive marker with regular expressions, given that postnominal modifiers like relative clauses include too many different linguistic elements for the expression to consider. Instead, the script focused on finding the genitive markers and the following noun phrase, which can be detected more easily by virtue of the more restricted prenominal structure of noun phrases.

To find *of*-genitives, the script looked for predeterminers (*all*, *both*), determiners including definite articles (*the*), indefinite articles (*a*, *an*), demonstratives (e.g., *these*) and possessive pronouns (e.g., *my*, *your*), and after-determiners (e.g., *such*, *same*). These elements were optionally followed by cardinal numbers (e.g., *the two men*), or an adjective phrase in simple, comparative or superlative form preceded optionally by an adverb. This prenominal structure had to be followed by either a nominal (e.g., *chair*, *bacteria*), a pseudo-nominal temporal adverb (e.g., *today*, *tomorrow*) or an indefinite pronoun (e.g., *anybody*, *somebody*). For the *s*-genitive, the script looked for the same elements, except for the predeterminer and determiner

slot, which are occupied by the possessor by default, and for the fact that *own* could not follow the clitic. Hence, the script took virtually all potential noun phrases into account, except for adverbial phrases as in (116) and sentential modifiers as in (117) (Quirk et al. 1985: 1336–1337).

(116) A come-and-fight-me attitude (Quirk et al. 1985: 1336)

(117) He asked I don't know how many guys (Quirk et al. 1985: 1337)

Since the script looked for genitive markers immediately followed by a noun phrase, several categorical cases were automatically excluded, including double genitives, independent genitives, as well as certain pronominal and non-nominal genitives.

In order to further narrow down non-interchangeable *of*-structures, the R script incorporated part-of-speech constraints before *of* (see Heller 2018: 57) to ensure that cardinal numbers (118), determiners (119) and verbs (120) could also not precede the *of*-preposition.

(118) The scariest well o-one of my favourites it's The The Shining (TLC:2_6_ME_59, B1 Spanish)

(119) My family says I must throw away some of my old things I'm finding this difficult (TLC:2_8_ME_10, B2 Spanish)

(120) We know that er human slumber is made of cycles and each cycle has a period of rapid eye movement which is always connected with dreaming (TLC:IT_53, C1 Italian)

Similarly, certain lexical items were not allowed to precede *of* so as to exclude frequent collocations that are unlikely to be part of interchangeable *of*-genitives. Inspiration for this list of lexical items came from the GloWbE corpus (Davies & Fuchs 2015), where I looked through lexical items that have an MI score higher than 3 with a following *of*-preposition. The list consisted of adjectives (121), prepositions (122) and specific nouns (123). These constraints also greatly reduced the number of partitive genitives.

(121) Are they aware of those rights really? (TLC:2_ME_20, NS)

(122) Uhu yeah yeah you see erm his father wants the child out of the house (TLC:2_8_IT_23, NS)

(123) The areas of these kind of transplants (TLC:2_SP_25, C2 Spanish)

In total, the script returned 10,630 potential *of*-genitives (6,176 (58%) from learners, 4,454 (42%) from examiners) and 1,826 potential *s*-genitives (907 (50%) from the learners, 919 (50%) from the examiners).

Script recall

To check the recall of the script, I selected 10 exams that featured a high number of potential genitives. In total, these exams covered approximately 37,000 words and contained 353 *of*-genitives and 61 *s*-genitives according to the script. Manual retrieval of the *of*-genitives showed that these 10 exams featured 186 interchangeable *of*-genitives, 173 of which were among the 353 potential *of*-genitives the script retrieved, which means that the script had a recall of 93% for *of*-genitives in the corpus. Out of 54 interchangeable *s*-genitives identified manually, 49 were among the 61 *s*-genitives retrieved by the script for a recall of 91%.

The interchangeable genitives that the script could not detect belonged to three types. Firstly, there were a few cases where a pronoun followed *of* (124) which the script removed automatically because they are very rare compared to the equivalent with a possessive pronoun (*its timing*). Secondly, some elements were assigned the wrong part-of-speech tag in the corpus, resulting in the regular expression missing the genitive. Finally, the genitive structure was sometimes interrupted by intervening linguistic material the regular expression did not account for, like *okay* in (125).

(124) you can use the timer if you need help with the timing of it
(TLC:2_IN_6, NS)

(125) let's look at some of the neg-negative aspects okay of technology
(TLC:IN_11, NS)

6.2.2.2 Interchangeability check

Once the potentially interchangeable genitives in the corpus data had been identified, I checked their interchangeability. Given the precision of the script (53%), about half of the 4,454 examiner *of*-genitives were interchangeable. Since this provided more than enough data and manually checking the interchangeability of the tokens is time-consuming, I decided to inspect the interchangeability of the examiner genitives in approximately a third of the exams. The concrete interchangeability criteria I checked are shown for *of*-genitives and *s*-genitives in Table 6.1.

Table 6.1: Criteria used to manually determine on a case-by-case basis whether the potential *of*-genitives and *s*-genitives identified by the script were interchangeable.

<i>of</i> -genitives	<i>s</i> -genitives
▪ There is a noun phrase on either side of the <i>of</i>-preposition ▪ The determiner of the possessum noun phrase is the definite article <i>the</i> ▪ The <i>of</i>-genitive is semantically valid, so it is not partitive, appositive or modifying ▪ The <i>of</i>-genitive is not fixed	▪ The <i>s</i>-genitive is semantically valid, so it is not classifying or partitive ▪ The <i>s</i>-genitive is not fixed

While checking the interchangeability of the potential genitives, I came across a number of borderline cases. By far the most common problematic case are missing determiners in either noun phrase in *of*-genitives, as in (126), or in the possessor noun phrase of *s*-genitives, as in (127). These cases suggest that learners are struggling with the challenging English determiner system (Ionin, Zubizarreta & Philippov 2009: 337; Gries & Wulff 2013: 348). These cases were discarded because it is unclear whether the learner intended to produce *the*, which is the only option for interchangeable *of*-genitives (Rosenbach 2002: 30).

(126) I would like to spend some time on some unique phenomenons of current Chinese education (TLC:CH_26, C1 Chinese)

(127) I think a year ago I went to erm colleague's hometown (TLC:2_7_CH_30, B2 Chinese)

For the *s*-genitive, there was one rather unusual pattern characterized by the presence of a central determiner, either a possessive pronoun (128) or an article (129), in addition to the genitive determiner. While the literature states that the possessors in *s*-genitives cannot be followed by noun phrases with central determiners (Heller 2018: 59), the additional determiner instead precedes the genitive possessor in this case. In native speech, such structures are considered erroneous and do not occur according to Rosenbach (2002: 14).

(128) Maybe our China's first er president (TLC:8_CH_7, B2 Chinese)
(129) I like the Roald Dahl's books very much it is very interesting and amazing (TLC:2_6_IN_92, B1 Gujarati)

The question arises as to whether the extra central determiner is a feature proper to the genitive constructions for these learners, for instance because

the genitive possessor does not make the possessum definite by default in their interlanguage and therefore definiteness is encoded by an extra determiner (see Koptjevskaja-Tamm 2001: 65 on so-called *adjectival-genitives*), or whether some learners are not aware that determiners cannot precede these proper nouns. The second scenario seems more plausible, so I retained these observations. For one, the determiner system is a notoriously difficult aspect of English grammar for learners (Ionin, Zubizarreta & Philippov 2009: 337; Gries & Wulff 2013: 348). Moreover, half of the relevant speakers used proper nouns with determiners or possessive pronouns in non-genitive constructions, suggesting that learners primarily struggle with determiners rather than the genitive variants specifically.

There were three borderline cases involving the *of*-genitive. First, some *of*-genitives seem to better alternate with compound nouns than *s*-genitives as illustrated by example (130), where the more natural equivalent of the *of*-genitive *topics of discussion* is not an *s*-genitive (*discussion's topic*), but a compound noun (*discussion topic*). These cases were retained because all three variants can alternate with each other, as shown by Szmrecsanyi et al. (2016). Moreover, one can easily find instances of *discussion's topic* on Google.

(130) the media influence our daily life that is er er by creating the topics of discussion (TLC:CH_28, C1 Chinese)

A substantial number of *of*-genitives involved possessum noun phrases such as *base*, *top*, *bottom* or *middle* referring to a certain part of the possessor. In (131) for example, the possessum *middle* refers to a specific part of the possessor *cake*. The fact that the possessum does not introduce a wholly new referent is reminiscent of partitive genitives, although the possessum points to a certain part of the possessor in the present case rather than quantifying it (Heller 2018: 56). These cases were retained because Google searches returned reliable instances of their *s*-genitive counterparts.

(131) the tradition by ca= by pla= by placing cand= a candle in the middle of the cake er came from a a G= a Ger= came from Germany erm er er i= er it's representing the lights the lights of life (TLC:2_7_RU_5, B2 Russian)

Finally, some *of*-genitives featured (*with*) *the help of* (132) or (*with*) *the use of* (133). These possessums resemble fixed, complex prepositions (see Huddleston & Pullum 2002: 616–626), which raised the question as to

whether they should be considered interchangeable *of*-genitives. In (132) for instance, one could replace *with the help of games* by *by means of games* (Huddleston & Pullum 2002: 618), whereas *with games' help* arguably sounds rather unnatural. However, the *s*-genitive counterpart clearly seems acceptable when it is followed by the animate possessor *my parents* in (132) (*with my parents' help*). Hence, I argue that these observations are interchangeable, although the animacy of the possessor plays a significant role in their perceived acceptability, which is not unexpected given the importance of the constraint (Rosenbach 2014: 231; Rosenbach 2017: 14; Heller, Szmrecsanyi & Grafmiller 2017: 10).

(132) with the help of the games with the help of my parents I recognise that I would like to be an architect (TLC:2_8_RU_11, B2 Russian)

(133) But he was all alone to save er to save the girl and the whole country and with her the the er the use of his brains (TLC:2_8_IN_10, B2 Hindi)

6.2.2.3 Data extraction

The noun phrase boundaries

Once the interchangeable genitives had been identified, they had to be extracted from the corpus data for variable annotation. To automatically extract the genitives, I used the boundaries of the genitive structures, viz. the beginning of the first noun phrase and the end of the last noun phrase. However, identifying the boundaries of the genitive structures is not always straightforward, especially in spoken language. In (134) for instance, the possessum noun phrase *the the the reasons* features the definite article *the* three times.

(134) the the the reasons of the firing are er are several I think (TLC:2_IT_9, C1 Italian)

If all repetitions of *the* were included when counting the length of the constituents, this constituent would be considered rather long and complex (e.g., Heller, Szmrecsanyi & Grafmiller 2017: 12) although the same word is simply being repeated. Therefore, if repetitions occurred at the beginning of the first noun phrase as in (134), I only extracted the genitive observation from the last repetition, viz. the third instance of *the*. Speakers in the corpus also tend to reformulate certain parts of the relevant noun phrases, including the head noun, as in (135), where *language* was reformulated as *language*. In general, I extracted the reformulated linguistic material for variable annotation. If the entire genitive structure was reformulated as

something other than a genitive variant, the observation was discarded. For example, the *of*-complement in (136) was replaced by a complement phrase headed by *from*, so the reformulated structure is no longer an *of*-genitive.

(135) today English is the most important language language of er the world (TLC:2_6_IT_15, B1 Italian)

(136) but they er just ignore the real er real value of university from university life i-is to it is to er equip them with experience (TLC:CH_34, C2 Chinese)

Finally, linguistic material is occasionally marked as 'unclear' in the corpus data, indicating that the relevant words in the exam recordings were unintelligible. If a genitive observation contained any unclear linguistic material, like *udder* in (137), it would not be possible to annotate the data for certain variables reliably, so these genitive observations were discarded.

(137) Dolly was created by replacing the DNA of the one sheep's egg with the DNA of another sheep's (unclear text=udder) (TLC:2_7_RU_4, B2 Russian)

Additionally, the relevant noun phrases were often morphosyntactically complex. Firstly, noun phrases that are coordinated in the possessor or possessum with *and* or *or* as in (138) were extracted as a whole, whereas subsequent noun phrases without overt coordinator were treated like reformulations. Genitives can also be embedded in each other as in (139), where the *of*-genitive *the attitude of the today's young generation* contains a valid *s*-genitive *the today's young generation*, both of which were extracted with their own boundaries.

(138) it is India's biggest and scariest roller coaster I love to sit on it and the best thing about is that it has three loops (TLC:2_7_IN_30, B2 Hindi)

(139) live-in relationship is more in urban areas compared to rural areas because of the lifestyle of the urban areas er because of the the attitude of the today's young generation (TLC:2_IN_4, C2 Kannada)

The possessors and possessums also exhibited several kinds of postnominal modifications, including prepositional complements as in (140) and relative clauses without or with relativizer, as in (141). If part of the postnominal structure was unclear or unfinished as in (142), the postnominal modification was not included for extraction. Nonetheless, the observation was not discarded as long as an otherwise structurally complete noun

phrase remained and the genitive observation as a whole was interchangeable.

- (140) she's Aphrodite she's the goddess of love and beauty with fertility (TLC:2_6_IN_86, B1 Marathi)
- (141) this is er this film er ends with the er the death of the hero that decide to sacrifice himself (TLC:2_IT_30, C1 Italian)
- (142) if our neighbour's place in a for= if our relative is in foreign we should learn from him (TLC:2_6_IN_9, B1 Hindi)

A pervasive issue with all postnominal modifications is that it is often unclear how they attach to the *of*-genitive as several noun phrases are available, as shown in Figure 6.1.

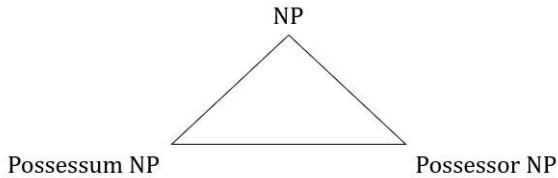


Figure 6.1: Schematic representation of the noun-phrase structure of the *of*-genitive

To illustrate, the prepositional phrase *in Beijing* in (143) can either attach to the immediately preceding noun phrase *the main cause*, it could be a postponed realization of a postnominal modification to *heavy traffic*, or it can modify the whole *of*-genitive represented by the top NP node in Figure 6.1, viz. *the main cause of heavy traffic*.

- (143) I think the main cause of heavy traffic in Beijing might be the increasing number of private cars (TLC:2_7_CH_11, B2 Chinese)

Transforming the *of*-genitive into an *s*-genitive can help solve the ambiguity of such postnominal modifiers. As shown in Figure 6.2, postnominal modifiers in the *s*-genitive can only attach to two noun phrases, namely the noun phrase within the determiner phrase itself, or the entire noun phrase consisting of the genitive determiner and its head noun, viz. the possessum.

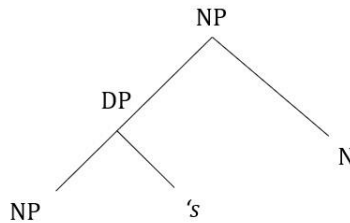


Figure 6.2: Schematic representation of the noun-phrase structure of the *s*-genitive

Since the head of the determiner phrase is the *s*-marker, which does not take any complements (Sag, Wasow & Bender 2003), it is impossible for a noun phrase that follows the *s*-genitive to modify the possessor noun phrase within the determiner phrase. Instead, the possessor head in the *s*-genitive has to be immediately followed by its postnominal modifiers. Importantly, this representation of the *s*-genitive implies that the possessum is never modified because nouns by themselves cannot take prepositional phrases. As a result, I assume that *in a village* in (144) can only modify the whole *s*-genitive and not the possessum *house*.

(144) I've been to my aunt's house in a village it's in Madhya Pradesh in north (TLC:2_7_IN_21, B2 Hindi)

With this in mind, I determined the structure of (143) in two steps. First, I tested whether *in Beijing* can follow the *s*-genitive with the paraphrase *heavy traffic's main cause in Beijing*. If so, then *in Beijing* can modify the entire *of*-genitive, meaning that it attaches to the top NP node in Figure 6.1. Then, I checked whether *in Beijing* could also modify the possessor noun *heavy traffic* with a second paraphrase, viz, *heavy traffic in Beijing's main cause*. If neither paraphrase is felicitous, I concluded that *in Beijing* modifies the remaining noun phrase, i.e., the possessum *main cause*. In the latter case, the postnominal modifier *in Beijing* is separated from its head noun, which would be problematic for the annotation of certain constraints, so these observations were discarded. If both paraphrases were correct and it was unclear whether the postmodifier attaches to the entire structure or either noun phrase, I assumed that it belongs to the whole structure. If so, the modification does not attach to either the possessor or the possessum noun phrase, so it falls outside the noun phrase boundaries and was therefore not extracted. If only the second paraphrase is correct, indicating that it is the possessor that is modified, then I extracted

the genitive observation with its postmodified possessor as *heavy traffic in Beijing*.

After identifying the noun phrase boundaries of the interchangeable genitives, I extracted a total of 3,831 learner genitive constructions (3,000 (78%) *of*-genitives and 831 (22%) *s*-genitives) and 1,402 native speaker genitive constructions (1,035 (74%) *of*-genitives and 367 (26%) *s*-genitives).

Scripted utterances

Some of the interchangeable genitive observations occurred within scripted utterances produced by the examiner as in (145), where the examiner asks the learner a preplanned question after reading about earworms during the listening task.

(145) what is the speaker's attitude towards earworms? (TLC:2_IT_13, NS)

Such utterances do not represent the spontaneous spoken language production of the examiners in the corpus, so the genitive observations they contain were not retained for further analysis. First, I grouped the genitive observations that shared the same possessor, possessum and genitive marker and occurred in at least two exams, considering that examiners within a single exam might simply repeat the same scripted utterance containing a genitive structure. To illustrate, (146) and (147) are examples from the conversation tasks of two different exams featuring an *s*-genitive with the same possessor *your grandparents* and the same possessum *day*, so they are potentially scripted.

(146) what difference do you feel there are between society today and society in your grandparents' day? (TLC:2_8_IN_26, NS)

(147) what differences are there between society today and society in your grandparents' day (TLC:2_8_IN_29, NS)

Then, I compared the similarity of the clauses in which these genitive variants are embedded by counting how many words the clauses have in common without including the genitive structure itself since it was the same by default. To illustrate, the clause in (146) contains 13 words, 10 of which are also present in (147), so the similarity between clauses stands at 76.9% (10/13).

To determine how similar clauses had to be in order to be considered scripted, I tested different similarity thresholds for a sample of utterances that I considered scripted by default and another sample that was not

scripted. The scripted sample stemmed from the genitive observations in the listening task, where the examiners ask preplanned questions after a reading. For the non-scripted sample, I randomly chose utterances that do share the same genitive observation. The threshold that gave the best results was 70%, capturing 91% of the scripted utterances while ignoring 97.3% of the non-scripted utterances. Hence, clauses featuring the same genitive construction and whose similarity exceeded this threshold were removed from further analysis, including examples (146) and (147). Out of 1,402 interchangeable examiner observations, I removed 363 scripted observations, plus the 70 observations from the listening task that I used for the scripted sample, which left me with 969 examiner observations (714 (74%) *of*-genitives, 255 *s*-genitives (26%)).

Overview of the interchangeable genitives

Table 6.2 shows the distribution of the *of*-genitives and the *s*-genitives across proficiency levels.

Table 6.2: Distribution of the interchangeable genitive variants across the different proficiency levels. Percentages indicate the proportion of each variant for a given proficiency level. For the row total, the percentages indicate the proportion of the tokens that each proficiency level contributes to the total number of tokens.

	<i>of</i> -genitive	<i>s</i> -genitive	Row total
B1	590 (75%)	193 (25%)	783 (16%)
B2	941 (77%)	284 (23%)	1,225 (26%)
C1	1,117 (80%)	272 (20%)	1,389 (29%)
C2	352 (81%)	82 (19%)	434 (9%)
Native	714 (74%)	255 (26%)	969 (20%)
Column Total	3,714 (77%)	1,086 (23%)	4,800

The distribution of the genitive variants seems rather homogeneous across proficiency levels except for the fact that the share of *of*-genitives slightly increases with proficiency level, which raises the question as to whether this could be attributed to certain L1s. Hence, Table 6.3 provides an overview of the L1 backgrounds and their contribution across proficiency levels.

Table 6.3: Distribution of the genitive observations across proficiency levels and L1 backgrounds. Percentages indicate the proportion of tokens that the L1 background contributes to the total tokens for a given proficiency level. For the row total, the percentages indicate the proportion of the tokens each L1 contributes to the total number of tokens.

	B1	B2	C1	C2	Row total
Chinese	86 (11%)	135 (11%)	147 (10.6%)	32 (7.4%)	400 (10.4%)
Gujarati	64 (8.2%)	64 (5.2%)	0	25 (5.8%)	153 (4%)
Hindi	86 (11%)	157 (12.8%)	10 (0.7%)	40 (9.2%)	293 (7.6%)
Italian	196 (25%)	337 (27.5%)	445 (32%)	64 (14.7%)	1,042 (27.2%)
Kannada	0	1 (0.1%)	14 (1%)	14 (3.2%)	29 (0.8%)
Malayalam	3 (0.4%)	4 (0.3%)	0	0	7 (0.2%)
Marathi	26 (3.3%)	45 (3.7%)	12 (0.9%)	46 (10.6%)	129 (3.4%)
Russian	42 (5.4%)	79 (6.4%)	0	0	121 (3.2%)
Spanish	269 (34.4%)	349 (28.5%)	605 (43.6%)	131 (30.2%)	1,354 (35.3%)
Tamil	10 (1.3%)	27 (2.2%)	11 (0.8%)	18 (4.1%)	66 (1.7%)
Telugu	0	0	5 (0.4%)	0	5 (0.1%)
Unknown	1 (0.1%)	27 (2.2%)	140 (10.1%)	64 (14.7%)	232 (6.1%)
Column total	783 (20.4%)	1,225 (32%)	1,389 (36.3%)	434 (11.3%)	3,831

Table 6.3 indicates that the genitive variants are produced by learners from eleven L1 backgrounds. Figure 6.3 visualizes the proportion of data that the five most commonly attested L1s in the data contribute to each of the proficiency levels, namely Spanish, Italian, Chinese, Hindi and Gujarati, which together account for 3,242 of the 3,831 (85%) learner observations. The remaining L1s are conflated in Figure 6.3 as ‘other’.

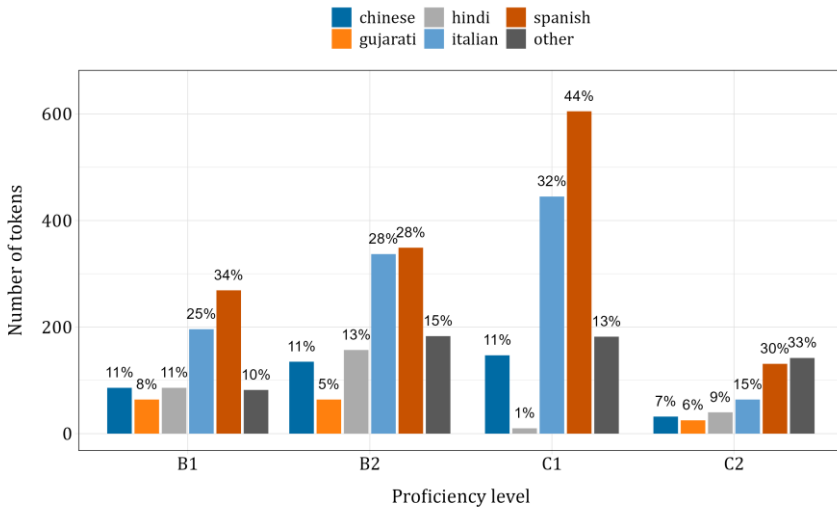


Figure 6.3: Distribution of the tokens across proficiency levels. The bars show the proportion of tokens the L1s contribute to each proficiency level.

Figure 6.3 reveals that Italian and Spanish EFL learners account for a considerable portion of the genitive data, which increases with proficiency level. At the C1 level, Spanish EFL learners account for almost half of the genitives (44%). However, only 113 out of 1,042 (11%) genitive observations in the Italian subsample are *s*-genitives. Likewise, only 192 out of 1,354 genitive observations from Spanish EFL learners take the *s*-genitive. While this might explain the increase in *of*-genitives from B1 to B2 to C1, it does not really explain the equally high proportion of *of*-genitives in the C2 data. However, the C2 data only contributes 434 observations, 64 (15%) of which stem from learners with unknown L1 backgrounds, so it remains somewhat unclear whether L1 plays a role here as well.

Data sampling

In both the learner data and the examiner data, *of*-genitives account for approximately three-quarters of the genitive observations. Since variable annotation is time-consuming, I decided to only annotate approximately as many *of*-genitives as there are *s*-genitives in the data to reach a balanced sample. This is justified theoretically because I am primarily interested in the probabilistic conditioning of the genitive variants rather than their relative distribution. From a methodological point of view, the method of statistical analysis, viz., mixed-effects logistic regression, is limited by the number of observations of the least frequent variant (Hosmer & Lemeshow

2000: 346–347), in this case the *s*-genitive, so including more *of*-genitives would not necessarily improve the quality of the statistical analysis.

To reach a balanced sample, I counted how many *s*-genitives occurred in a given L1 background and selected an equal number of *of*-genitives. All *of*-genitives preferably stemmed from the same learners as the *s*-genitives, which helps ensure that the learners under study have acquired both genitive constructions like native speakers. If the learners who produced *s*-genitives somehow produced fewer *of*-genitives, I randomly selected the remaining *of*-genitives from other learners with the same L1 background. Through this process, a total of 514 *of*-genitives coming from mostly the Spanish and the Italian L1 backgrounds were sampled out. All *of*-genitives from native speakers and L1 speakers of Chinese and Hindi were retained since the distribution of the variants in these sub-corpora was rather balanced to begin with. An updated overview of the distribution of the genitive observations by L1 background and proficiency level is shown for the learner data in Table 6.4.

Table 6.4: Updated distribution of the genitive observations across proficiency levels and L1 backgrounds. Percentages indicate the proportion of tokens that the L1 background contributes to the total tokens for a given proficiency level. For the row total, the percentages indicate the proportion of the tokens each L1 contributes to the total number of tokens.

	B1	B2	C1	C2	Row total
Chinese	86 (25.7%)	135 (22.4%)	147 (24.2%)	32 (14%)	400 (22.6%)
Gujarati	45 (13.5%)	52 (8.6%)	0	16 (7%)	113 (6.3%)
Hindi	60 (18%)	113 (18.7%)	10 (1.6%)	40 (17.5%)	223 (12.6%)
Italian	47 (14.1%)	83 (13.7%)	99 (16.3%)	8 (3.5%)	237 (13.4%)
Kannada	0	0	8 (1.3%)	7 (3.1%)	15 (0.8%)
Marathi	16 (4.8%)	34 (5.6%)	4 (0.7%)	25 (11%)	79 (4.5%)
Russian	17 (5.1%)	56 (9.3%)	0	0	73 (4.1%)
Spanish	57 (17.1%)	88 (14.6%)	241 (39.7%)	55 (24.1%)	441 (24.9%)
Tamil	6 (1.8%)	22 (3.6%)	8 (1.3%)	9 (3.9%)	45 (2.5%)
Telugu	0	0	5 (0.8%)	0	5 (0.3%)
Unknown	0	21 (3.5%)	85 (14%)	36 (15.8%)	142 (8%)
Column total	334 (18.8%)	604 (34.1%)	607 (34.2%)	228 (12.9%)	1,773

The final dataset contains a total of 2,302 genitive constructions (1,216 (53%) *of*-genitives and 1,086 (47%) *s*-genitives). Learners produce

1,773 genitive observations (942 (53%) *of*-genitives and 831 (47%) *s*-genitives) and native speakers account for the remaining 529 observations (274 (52%) *of*-genitives and 255 (48%) *s*-genitives). The data stems from 469 learners and 45 examiners.

6.2.3 Data annotation

The 2,302 genitive variants were then annotated for the proficiency level of the speaker as indicated by the corpus metadata, as well as the constraints affecting the choice of genitive variant according to the literature, namely structural persistence, the animacy, definiteness, thematicity, discourse status and final sibilancy of the possessor, the semantic relation between the possessor and the possessum and the length of the possessor and the possessum (e.g., Gries & Wulff 2013; Heller, Szmrecsanyi & Grafmiller 2017). Not included were branching direction, segment alternation and rhythmic alternation, although they influenced the choice of genitive variant in the study by Gries and Wulff (2013). These constraints are difficult to operationalize and time-consuming to annotate, so I instead opted to include the most well-established and strong predictors following Rosenbach (2014), most of which could be annotated automatically. The probabilistic constraints, their effect on the choice of variant, and their operationalization are presented one-by-one in the following sections.

6.2.3.1 Constituent length

The length of the constituents, viz. the possessor and the possessum, influences the choice of genitive variant (e.g., Hinrichs & Szmrecsanyi 2007; Shih et al. 2015; Heller, Szmrecsanyi & Grafmiller 2017). Specifically, if the possessor is short and the possessum is long, speakers prefer the *s*-genitive, whereas they prefer the *of*-genitive when the possessum is short and the possessor is long (Kreyer 2003: 178). Several principles could be at the origin of this tendency (Szmrecsanyi & Hinrichs 2008: 292–293), such as the principle of end-weight or heaviness, which states that heavier constituents tend to be produced after lighter constituents (Behaghel 1909; Arnold et al. 2000). More generally, the principle of Easy First states that speakers produce easier linguistic elements before harder ones (MacDonald 2013). From a hearer-oriented perspective, Hawkins (2004) formulates the principle of immediate constituents (see also Arnold et al. 2000: 31), which states that placing shorter constituents before longer ones facilitates comprehension because the hearer can recognize the different constituents quicker. Regardless, the consensus in the literature is that the length of constituents is related to language processing (Rosenbach 2014:

240; Grafmiller 2014: 474). Speakers try to accommodate processing by producing the genitive variant that places the shorter constituent before the longer constituent. Nonetheless, the strength of the factor is not stable both synchronically (Kinne 2020) and diachronically (Wolk et al. 2013: 408). Moreover, Szmrecsanyi's (2006: 106) study on spoken British English found that both long possessors and long possessums are associated with *of*-genitives, whereas long possessums should normally favor the *s*-genitive, which he attributed to the fact that the *of*-genitive could be the easier variant to produce when the genitive is heavy.

The length of the constituents was operationalized as the number of letters they contain, excluding hesitations and false starts. In example (148), the possessum *fortune* consists of seven letters and the possessor *the family* consists of nine letters because the hesitation *er* is excluded. There are other ways to operationalize the length of constituents like counting the number of words or syllables, but previous research has shown that these different operationalizations are very highly correlated (Wolk et al. 2013: 395).

(148) on New Year's Eve all the vacuums er brooms and dusters er
should be set aside for fear the er family's fortune would be swept
swept away (TLC:2_8_CH_9, B2 Chinese)

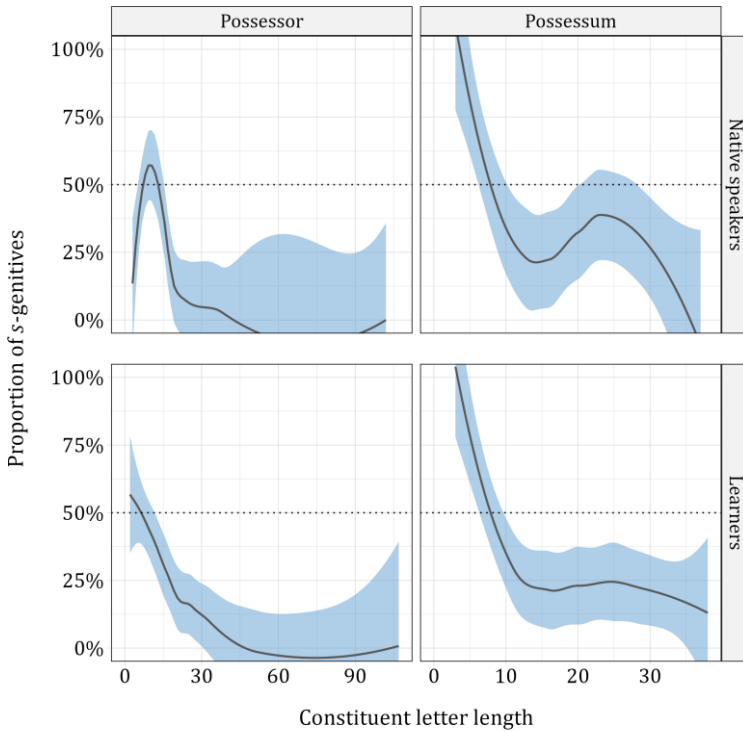


Figure 6.4: Smoothed conditional means of the proportion of *s*-genitives by the letter length of the possessor (left) and the possessum (right) for native speakers (top) and learners (bottom).

As expected, Figure 6.4 shows that both native speakers and learners avoid the *s*-genitive with longer possessors. Interestingly, they also avoid the *s*-genitive when the possessum is approximately 10 letters or more. When the possessum is short, the *s*-genitive is particularly dominant, which suggests that speakers are generally more sensitive to the length of the possessum. In general, Figure 6.4 shows that the *s*-genitive is favored when both constituents are small, whereas the *of*-genitive is preferred when the constituents are long, in conflict with the above-mentioned principles of end-weight or Easy First (Behaghel 1909; MacDonald 2013), which assume that the genitive variant is determined by relative length of the constituents. Instead, the pattern is reminiscent of Szmrecsanyi (2006: 106) and aligns better with Rohdenburg's Complexity Principle (1996: 151) stating that speakers prefer the more explicit variant, in this case the analytic *of*-genitive, when the linguistic environment is complex, as is the case when both constituents are long.

6.2.3.2 Possessor discourse status

Previous research has found that discourse-given possessor referents prefer the *s*-genitive (see Hinrichs & Szmrecsanyi 2007: 451; Wolk et al. 2013; Rosenbach 2014: 225; Shih et al. 2015; Szmrecsanyi, Biber, et al. 2016; Heller, Szmrecsanyi & Grafmiller 2017: 9). However, this tendency is not very strong and sometimes the constraint is not significant at all (Kreyer 2003: 172; Hinrichs & Szmrecsanyi 2007: 461). When the effect is present, it is often explained as being the result of the given-before-new principle (Gundel 1988) or more generally Easy First (MacDonald 2013) since a referent that has been produced before in discourse is more accessible and therefore easier to produce. Placing the more accessible discourse-given constituent first could also facilitate comprehension (Arnold et al. 2000: 30–32; Bresnan & Ford 2010: 168).

The possessor was coded as either ‘given’ or ‘new’ depending on whether the head of the possessor noun phrase was mentioned among the 100 words preceding the genitive construction. In example (149), *Marvel* is given, because the lemma *Marvel* is present among the preceding 100 words.

(149) and M-Marvel has made millions of dollars from the from all these movies and they are trying to make the the universe larger but if but I guess Marvel is kind of losing their strengths because Marvel’s universe is big enough [...] (TLC:2_7_CH_33, B2 Chinese)

Table 6.5 and Figure 6.5 show that there are more or less as many given and new possessors in either variant for both native speakers and learners. Accordingly, the chi-square tests revealed no significant association between possessor discourse-status and the variants for either native speakers ($p = 0.77$) or learners ($p = 0.57$).

Table 6.5: Contingency table for possessor discourse status for native speakers ($X^2 = 0.09$, $df = 1$, $p = 0.77$) and learners ($X^2 = 0.42$, $df = 1$, $p = 0.571$). Percentages indicate the proportion of genitive variants per category of possessor discourse status. For the row total, the percentages indicate the total proportion of the different levels of discourse status. If the test is significant, values that are significantly lower than expected according to standardized residuals are highlighted in **orange**, whereas values that are significantly higher than expected according to standardized residuals are highlighted in **blue**.

Native speakers	s-genitive	of-genitive	Row total
New	141 (47.5%)	156 (52.5%)	297 (56.1%)
Given	114 (49.1%)	118 (50.9%)	232 (43.9%)
Column total	255 (48.2%)	274 (51.8%)	529
Learners	s-genitive	of-genitive	Row total
New	430 (46.1%)	503 (53.9%)	933 (52.6%)
Given	401 (47.7%)	439 (52.3%)	840 (47.4%)
Column total	831 (46.9%)	942 (53.1%)	1773

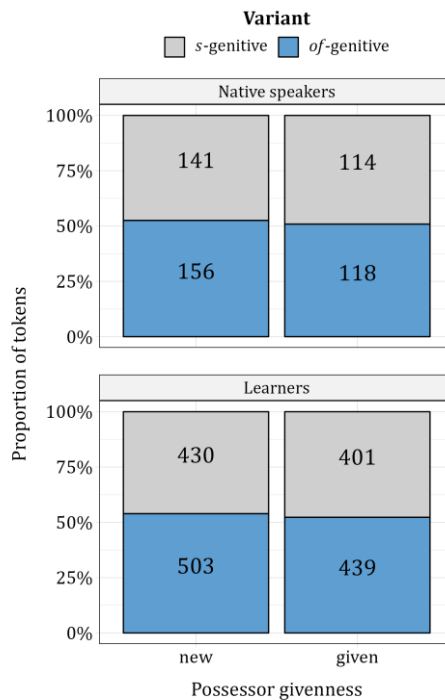


Figure 6.5: Distribution of the s-genitive and the of-genitive across levels of possessor discourse status for native speakers (top) and learners (bottom).

6.2.3.3 Persistence

In the context of the genitive alternation, persistence refers to the fact that speakers are likely to reuse the genitive variant they produced before (Szmrecsanyi 2006: 87–105; Heller 2018: 70–71). A previous genitive exemplar raises the activation level of the construction, so the construction is easier to produce again (Ellis 2006: 102; Wulff & Ellis 2018: 44; Ellis & Wulff 2019: 44). If the native speakers and the learners are sensitive to structural persistence, this implies that they have an internalized representation of the genitive constructions (Gries & Wulff 2005: 186; Romano 2018: 276). Structural persistence can be induced through production or comprehension, so a speaker might produce a certain genitive variant either because they or the interlocutor produced the variant before (Bock et al. 2007).

The operationalization of the constraint was the same as for the future marker alternation. For each observation, I checked which genitive variant was previously used by either interlocutor in the exam if any. Although research has found that persistence is equally strong regardless of whether it is caused by comprehension or production (Tooley & Bock 2014), it is possible that the examiner exerts more influence on the learner given that they dominate the interaction as examiners in this context (Sandlund, Sundqvist & Nyroos 2016: 20; Stanley & Stevenson 2017: 2). Therefore, the annotation also specifies who produced the previous genitive. The variable then consists of the following five levels: candidate *s*-genitive, candidate *of*-genitive, examiner *s*-genitive, examiner *of*-genitive and ‘none’ in case there were no previous genitives. In example (150), the genitive *a fool’s fake* would be annotated as ‘candidate *s*-genitive’, because it is preceded by the *s*-genitive *a friend’s frown* produced by the same candidate learner.

(150) a true friend may at times frown at you because of some reason
mm but I’ll better have a friend’s frown than a fool’s fake praise
(TLC:2_7_IN_55, B2 Gujarati)

The distribution of the variable levels is presented for both genitive variants in Table 6.6 and visualized in Figure 6.6 for native speakers and learners.

Table 6.6: Contingency table for persistence for native speakers ($X^2 = 19.62$, $df = 4$, $p < 0.001$, Cramér's $V = 0.19$) and learners ($X^2 = 55.27$, $df = 4$, $p < 0.001$, Cramér's $V = 0.18$). Percentages indicate the proportion of genitive variants per category of persistence. For the row total, the percentages indicate the total proportion of the different previous genitive variants. If the test is significant, values that are significantly lower than expected according to standardized residuals are highlighted in **orange**, whereas values that are significantly higher than expected according to standardized residuals are highlighted in **blue**.

Native speakers	s-genitive	of-genitive	Row total
Candidate s-genitive	34 (57.6%)	25 (42.4%)	59 (11.2%)
Candidate of-genitive	67 (46.5%)	77 (53.5%)	144 (27.2%)
Examiner s-genitive	72 (62.6%)	43 (37.4%)	115 (21.7%)
Examiner of-genitive	40 (36.7%)	69 (63.3%)	109 (20.6%)
None	42 (41.2%)	60 (58.8%)	102 (19.3%)
Column total	255 (48.2%)	274 (51.8%)	529
Learners	s-genitive	of-genitive	Row total
Candidate s-genitive	221 (54.2%)	187 (45.8%)	408 (23%)
Candidate of-genitive	279 (40.8%)	404 (59.2%)	683 (38.5%)
Examiner s-genitive	94 (70.1%)	40 (29.9%)	134 (7.6%)
Examiner of-genitive	91 (38.1%)	148 (61.9%)	239 (13.5%)
None	146 (47.2%)	163 (52.8%)	309 (17.4%)
Column total	831 (46.9%)	942 (53.1%)	1773

As expected, Table 6.6 shows that the native speakers are significantly more likely to repeat the s-genitive if they used an s-genitive before, and they are more likely to repeat an of-genitive if they used an of-genitive before. However, previous genitive variants produced by the learner candidates do not trigger structural persistence effects for the native speakers. By contrast, learners are more likely to repeat both previous s-genitives and of-genitives produced either by themselves or by the examiner. In fact, a quick look at the percentages in the learner data of Table 6.6 suggests that structural persistence effects are more pronounced for learners if the examiner previously produced a genitive variant, especially the s-genitive. Overall, the association with previous genitive variants is equally strong for native speakers (Cramér's $V = 0.19$) and learners (Cramér's $V = 0.18$).

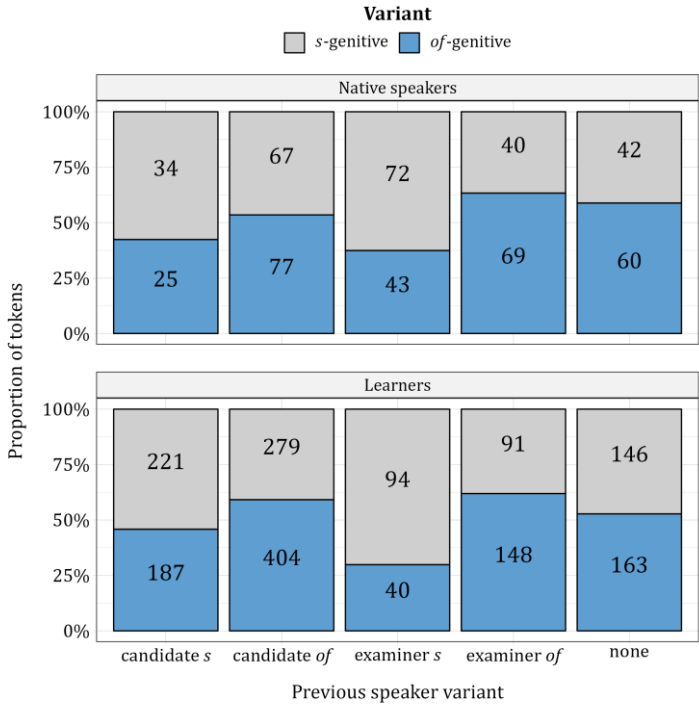


Figure 6.6: Distribution of the *s*-genitive and the *of*-genitive across levels of persistence for native speakers (top) and learners (bottom).

6.2.3.4 Possessor thematicity

Previous research on the genitive alternation has found that when the possessor is thematic, it is more likely to be produced as an *s*-genitive (Osselton 1988; see also Szmrecsanyi & Hinrichs 2008: 307). In a book about dogs for example, *the dog's tail* is more likely than *the tail of the dog*. To reflect the degree to which a possessor represents the theme of an exam, I counted the number of times the head of the possessor noun phrase occurred in the preceding context, normalized by the average word length of the exams in the corpus (Heller, Szmrecsanyi & Grafmiller 2017: 11).

The effect of thematicity mostly operates on written language and in particular press language, where the constraint reflects the phenomenon of economization. Economization refers to the need for writing to be concise due to the large amount of information there is to convey in today's world (Biber 2003: 170). That is, writers prefer to use the more concise *s*-genitive

to save space when the possessor is thematic in a text. Over the past decades, the influence of thematicity has been increasing rapidly (Hinrichs & Szmrecsanyi 2007: 469), leading to an increase in the number of *s*-genitives in written language (Hinrichs & Szmrecsanyi 2007; Grafmiller 2014).

In spoken language, economization is not expected to play an equally important role. Instead, highly thematic referents are also more frequent in the current operationalization, so that thematic possessors are more accessible and easier to produce again. Hence, thematic possessors should favor the *s*-genitive in spoken language because the variant allows speakers to directly produce the easier constituent in accord with Easy First (MacDonald 2013). Figure 6.7 visualizes the proportion of *s*-genitives for increasingly thematic possessors.

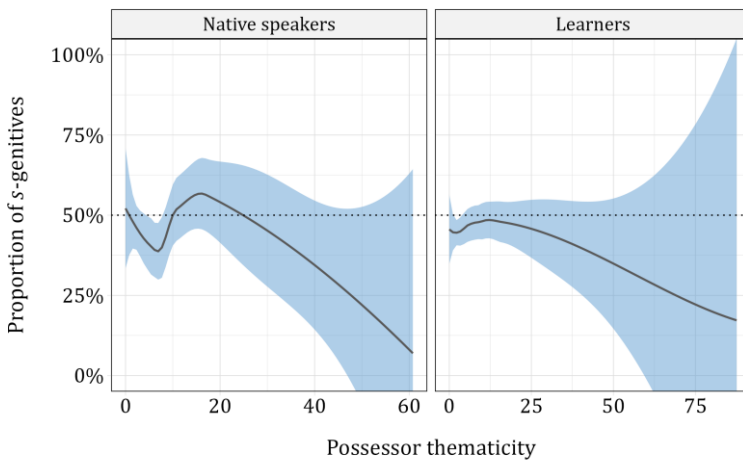


Figure 6.7: Smoothed conditional means of the proportion of *s*-genitives by the thematicity of the possessor for native speakers (left) and learners (right).

Contrary to the expectations, the figure shows that the proportion of *s*-genitives decreases as the possessor becomes more thematic. However, the confidence intervals for possessors occurring more than 20 times are already rather large, so this interpretation of thematicity remains tentative. Possessor thematicity does not seem to impact the proportion of *s*-genitives for possessors occurring up to 20 times in the exam.

6.2.3.5 Type-token ratio

Previous research has found that the lexical density of the material surrounding the genitive choice context, operationalized as the ratio of word tokens to word types, influences the choice of variant. Like

thematicity, the effect of type-token ratio has been linked to the increase of the *s*-genitive in certain written genres due to economization. The higher the ratio of types to tokens, the more lexically dense the text is, and the more likely the author will also pick the more dense genitive variant to save space (Hinrichs & Szmrecsanyi 2007: 457; Szmrecsanyi & Hinrichs 2008: 300–301; Grafmiller 2014: 479).

In spoken language, it is more sensible to view the effect of lexical density on the choice of variant in terms of language processing. According to Rohdenburg's Complexity Principle (1996: 151), speakers prefer to use the more explicit variant in cognitively complex environments. Provided lexically dense environments are cognitively complex as they require the speaker to retrieve more lexical items from memory, speakers should prefer the *of*-genitive. Importantly, it is not entirely clear how much linguistic material such cognitive environments encompass (see Dubois 2023).

The operationalization of type-token ratio follows Heller (2018: 69). That is, I calculated the ratio of the types to tokens by counting the number of unique tokens and dividing this number by the total number of tokens in the surrounding context consisting of 50 words preceding the genitive alternation and 50 words following it. If there were fewer than 50 words on one side of the construction, I expanded the surrounding context on the other side of the construction to arrive at a total of 100 words. To illustrate, the 100 words surrounding the *of*-genitive *the diabetes capital of the world* in (151) feature 77 unique tokens, so the type-token ratio is 0.77 (77/100).

(151) [...] special offers but eventually eating junk food we'll be paying with not only our wallets but our waistlines as well many Indians today feel it's more fashionable to order a pizza rather than eating chapattis or rice all this naturally takes a toll on our health India is the diabetes capital of the world more and more young adults are collapsing with heart failures obesity is on the rise yes even the children and now we know exactly why and if we don't amend our present lifestyles we could soon look like this ending up here in the hospital or even here the [...] (TLC:IN_6, C2 Unknown)

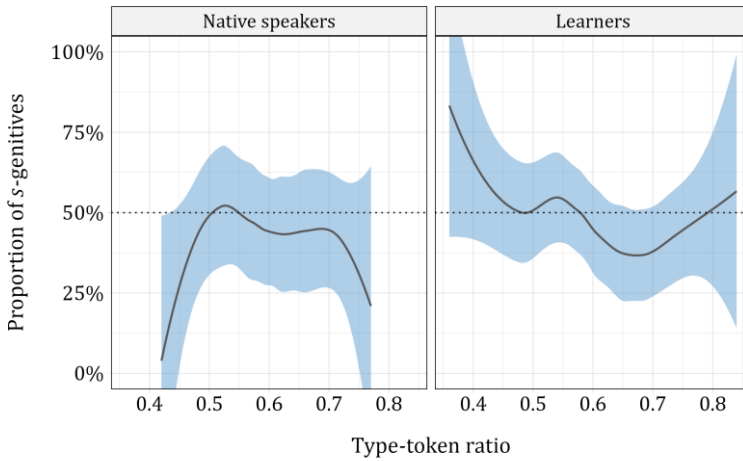


Figure 6.8: Smoothed conditional means of the proportion of *s*-genitives by the type-token ratio of the surrounding context for native speakers (left) and learners (right).

Figure 6.8 suggests that type-token ratio does not exert much influence on the choice of variant. The confidence intervals are rather large for observations with low and high type-token ratios. In the middle range, the proportion of *s*-genitives is stable and covers about half of the observations.

6.2.3.6 Possessor animacy

Animacy is arguably the strongest constraint of the genitive alternation (Shih et al. 2015; Rosenbach 2017: 14), to the extent that prescriptivists advise to use the *s*-genitive categorically when the possessor is a living referent and the *of*-genitive if the possessor is inanimate (Murphy 2012; cited in Heller, Szmrecsanyi & Grafmiller 2017: 10). According to Branigan, Pickering and Tanaka (2008), animate entities are cognitively more accessible and therefore retrieved more easily (see also Röthlisberger, Grafmiller & Szmrecsanyi 2017: 10). As a result, Easy First predicts that speakers prefer to use the *s*-genitive with animate possessors, because this variant allows speakers to produce the more accessible possessor right away (MacDonald 2013).

Interestingly, the strength of this constraint has been found to be fluctuate across different varieties (Hinrichs & Szmrecsanyi 2007: 450), modality (Grafmiller 2014: 484) and time (Wolk et al. 2013), even though the direction of the effect remains constant (Rosenbach 2017: 14). For native speakers, the animacy constraint has been growing weaker since at

least the 17th century as a result of the *s*-genitives becoming increasingly acceptive of inanimate noun classes (Wolk et al. 2013: 408). Heller, Szmrecsanyi and Grafmiller (2017) also found that ESL speakers are generally less sensitive to the animacy of the possessor than native speakers. In general, these findings point to the fact that the animacy constraint is not understood very well (Zaenen et al. 2004; Rosenbach 2017: 14).

Based on the work of Zaenen et al. (2004), the animacy of the possessor was classified as either animate, collective, locative, temporal or inanimate (see also Szmrecsanyi & Hinrichs 2008: 298; Wolk et al. 2013: 395). The category ‘animate’ refers to all animates, which include human animates or entities represented as humans (“things that look human and act like humans”) (Zaenen et al. 2004: 4); but also non-human animates, which can broadly be understood as what one perceives to be living, such as animals, as in (152). Other examples include *god, carpenter, elephants* etc.

(152) he woke up and he dreamt that outside his front door of his house
there was er a little white dog eating the cat’s food (TLC:2_8_SP_8,
NS)

Collective nouns refer to groups, collectives or organizations of humans that display some degree of group identity. Typical examples include nouns such as *corporation, community, organization, church*, as in (153). However, the annotation of possessor animacy is not determined solely by the noun itself, but also by the way it is used in the context. Hence, even plural nouns such as *parents* can show group identity if they share the same purpose or have a collective voice. In (154) for instance, the parents in *your parents’ decision* represent a collective entity because they made a decision with a collective mind or voice. Alternatively, typical collective nouns such as *committee* may not always represent a collective possessor. For example, in *each of the committee’s members went home*, the committee is not sharing one same goal; each member happens to do the same thing separately. Put simply, collective possessors differ from animate possessors in that they share a common goal or voice in the context.

(153) d-do you think it’s the government’s responsibility to s= do
something about global warming (TLC:2_6_ME_69, NS)

(154) is that your parents’ decision did people influence you did people
tell you or is it your own decision (TLC:8_ME_6, NS)

Locative possessors should truly imply a location in the context. In an example like *she left a letter at my house*, *house* is understood as a place or a

spatial reference point, but not in *my house was built in 1964*. Moreover, locative possessors should refer to places accessible for humans. For example, in *the other side of the coin*, *coin* is used as a spatial reference point with different sides, but coins are not places for humans, so it is not considered locative, contrary to *China* in (155). Overall, the tag for locative possessors was used rather restrictively, meaning that possessors are typically not coded as locatives if the possessor could fit another category.

- (155) Qingdao is in the south south of the China (TLC:2_7_CH_10, B2 Chinese)

Temporal possessors refer to a time period or a specific point in time. Mostly, this concerned possessors whose head noun inherently refers to time such as *today*, *last month*, *this period*, *the past* etc., as in (156). Finally, all other noun phrases were considered inanimate, like *physics* in (157).

- (156) do you have sympathy for today's generation because we we are always negative about them (TLC:2_IN_24, NS)
 (157) Albert Einstein was er the father of the modern physics (TLC:2_8_IN_34, B2 Marathi)

The distribution of the *s*-genitives and the *of*-genitives across the five animacy categories is presented in Table 6.7 and visualized in Figure 6.9 for native speakers and learners separately.

Table 6.7: Contingency table for possessor animacy for native speakers ($X^2 = 247.62$, $df = 4$, $p < 0.001$, Cramér's $V = 0.68$) and learners ($X^2 = 652.82$, $df = 4$, $p < 0.001$, Cramér's $V = 0.61$). Percentages indicate the proportion of genitive variants per category of possessor animacy. For the row total, the percentages indicate the total proportion of the different types of possessor animacy. If the test is significant, values that are significantly lower than expected according to standardized residuals are highlighted in **orange**, whereas values that are significantly higher than expected according to standardized residuals are highlighted in **blue**.

Native speakers	s-genitive	of-genitive	Row total
Animate	173 (82.8%)	36 (17.2%)	209 (39.5%)
Collective	61 (59.2%)	42 (40.8%)	103 (19.5%)
Inanimate	7 (4.2%)	160 (95.8%)	167 (31.6%)
Locative	3 (12.5%)	21 (87.5%)	24 (4.5%)
Temporal	11 (42.3%)	15 (57.7%)	26 (4.9%)
Column total	255 (48.2%)	274 (51.8%)	529
Learners	s-genitive	of-genitive	Row total
Animate	576 (79.8%)	146 (20.2%)	722 (40.7%)
Collective	136 (43.2%)	179 (56.8%)	315 (17.8%)
Inanimate	48 (8.9%)	490 (91.1%)	538 (30.3%)
Locative	28 (23.9%)	89 (76.1%)	117 (6.6%)
Temporal	43 (53.1%)	38 (46.9%)	81 (4.6%)
Column total	831 (46.9%)	942 (53.1%)	1773

In the native speaker data, the choice of variant is significantly associated with each of the animacy categories except for temporal possessors. As expected, animate and to a lesser extent collective possessors prefer the s-genitive. Conversely, inanimate and locative possessors strongly favor the of-genitive (see Heller 2018: 64). The animacy categories pattern with the choice of genitive variant in the same way in the learner data, except for the fact that collective possessors are not associated with either variant like temporal possessors. As expected, the overall association between possessor animacy and the genitive variants is strong for both native speakers (Cramér's $V = 0.68$) and learners (Cramér's $V = 0.61$).

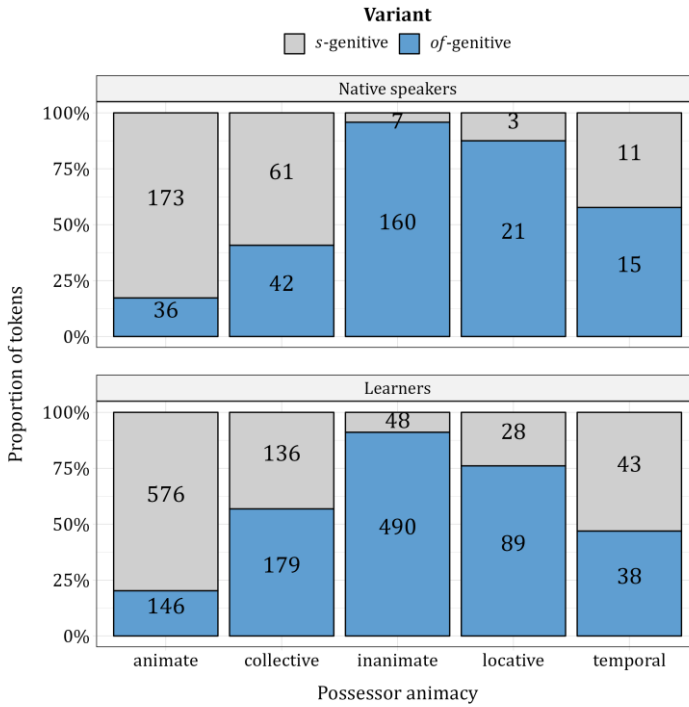


Figure 6.9: Distribution of the *s*-genitive and the *of*-genitive across levels of possessor animacy for native speakers (top) and learners (bottom).

6.2.3.7 Semantic relation

A probabilistic constraint proper to the genitive alternation concerns the semantic relation that holds between the possessor and the possessum. Specifically, most studies distinguish between prototypical relations and non-prototypical relations following Rosenbach (2002: 121; Rosenbach 2014: 229). This distinction is informed by research on adnominal possessives in typology, notably Koptjevskaja-Tamm (2001), who posits that prototypical relations at their core involve “an exclusive asymmetric long-term relation between two entities” (Taylor 1989; cited by Koptjevskaja-Tamm 2001: 964). In such prototypical relations, the adnominal possessor is suitable as a reference point to identify the head, viz. the possessum, so prototypical relations should favor the *s*-genitive (Koptjevskaja-Tamm 2001: 964). The prototypical relations that should favor *s*-genitives include kinship relations (158), body-part relation (159), legal ownership relations (160) and part-whole relations (161) (Rosenbach

2002: 121; Wolk et al. 2013; Shih et al. 2015). Any other relation was considered non-prototypical.

- (158) Alfred's father started the fortune by inventing landmines and sea mines which are bombs and which explode if people walk over them (TLC:2_7_RU_2, B2 Russian)
- (159) well I mean there are scientists who really believe that erm modern children's brains are being radically altered (TLC:2_8_IN_8, NS)
- (160) so what should I bring if I go to a friend's house for meal (TLC:2_7_IT_27, NS)
- (161) currently the majority of water on the Earth's surface is not suitable for human consumption we have the technology to obtain fresh water but not all countries can afford that (TLC:SP_67, C1 Spanish)

First, I manually identified the prototypical relations in the data. These relations correlate with the animacy of the possessor (Heller 2018: 77): kinship terms, body parts and legal ownership all involve animate possessors, while part-whole relations concern inanimate possessors. The different types of prototypical relations were then conflated in single 'prototypical' category, resulting in a binary variable that distinguishes between prototypical and non-prototypical relations.

The distribution of prototypical and non-prototypical semantic relations across genitive variants is shown for the native speakers and the learners in Table 6.8 and Figure 6.10. As expected, prototypical semantic relations appear to be associated with the *s*-genitive, whereas non-prototypical relations pattern with the *of*-genitive for both native speakers and learners. Although the association is rather weak in general, this tendency appears to be slightly more pronounced for the learners (Cramér's $V = 0.16$) compared to native speakers (Cramér's $V = 0.10$).

Table 6.8: Contingency table for the semantic relation between possessor and possessum for native speakers ($X^2 = 5.42$, $df = 1$, $p = 0.02$, Cramér's $V = 0.10$) and learners ($X^2 = 46.54$, $df = 1$, $p < 0.001$, Cramér's $V = 0.16$). Percentages indicate the proportion of genitive variants per category of semantic relation. For the row total, the percentages indicate the total proportion of the different levels of semantic relation. If the test is significant, values that are significantly lower than expected according to standardized residuals are highlighted in **orange**, whereas values that are significantly higher than expected according to standardized residuals are highlighted in **blue**.

Native speakers	s-genitive	of-genitive	Row total
Non-prototypical	187 (45.4%)	225 (54.6%)	412 (77.9%)
Prototypical	68 (58.1%)	49 (41.9%)	117 (22.1%)
Column total	255 (48.2%)	274 (51.8%)	529
Learners	s-genitive	of-genitive	Row total
Non-prototypical	592 (42.6%)	798 (57.4%)	1390 (78.4%)
Prototypical	239 (62.4%)	144 (37.6%)	383 (21.6%)
Column total	831 (46.9%)	942 (53.1%)	1773

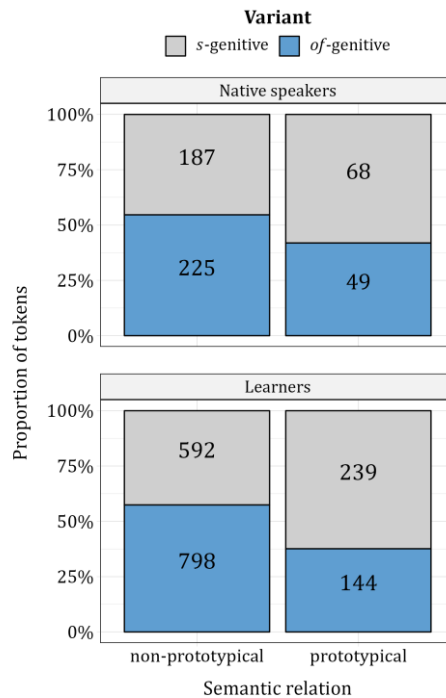


Figure 6.10: Distribution of the s-genitive and the of-genitive across levels of semantic relation for native speakers (top) and learners (bottom).

6.2.3.8 Possessor definiteness

Previous research found that definite possessors tend to take the *s*-genitive (Wolk et al. 2013; Heller 2018: 19). Provided definite possessors are more easily accessible than indefinite possessors, this preference for the *s*-genitive could once again be attributed to Easy First (MacDonald 2013) given that this variant allows the speaker to produce the definite possessor right away.

To annotate the observations for possessor definiteness, I followed the approach of Heller (2018: 72–74). Firstly, possessors that were coded as ‘temporal’ for the animacy constraint like *today* in (162) were considered definite by default. Also definite are proper nouns like *Youtube* in (163). Finally, possessors with definite determiners like *my neighbour* in (164) are definite as well. In practice, this operationalization of definiteness is very similar to the operationalization of specificity in Gries and Wulff (2013)

(162) in today's er fast-moving world people have less and less time spent on eating (TLC:2_6_RU_34, B1 Russian)

(163) she doesn't like the ideology of Youtube at all (TLC:2_7_IN_54, B2 Marathi)

(164) my first memory er er is er is maybe about er erm about erm erm a-an accident I had with a with a dog er a dog er my neighbour's dog er hurt me he he he bite (TLC:2_7_IT_29, B2 Italian)

The distribution of definite and indefinite possessors in the genitive variants is shown for native speakers and learners separately in Table 6.9 and Figure 6.11. The definiteness of the possessor does not appear to be associated with the genitive variants in the native speaker data ($p = .094$). Learners, on the other hand, follow the expected pattern whereby definite possessors occur more frequently in *s*-genitive and indefinite possessors in *of*-genitives. However, definiteness does not appear to be a very influential factor (Cramér's $V = 0.12$) as reported by other studies (see Heller 2018: 104).

Table 6.9: Contingency table for possessor definiteness for native speakers ($X^2 = 0$, $df = 1$, $p = 0.944$) and learners ($X^2 = 25.60$, $df = 1$, $p < 0.001$, Cramér's $V = 0.12$). Percentages indicate the proportion of genitive variants per category of possessor definiteness. For the row total, the percentages indicate the total proportion of the definite and indefinite possessors. If the test is significant, values that are significantly lower than expected according to standardized residuals are highlighted in **orange**, whereas values that are significantly higher than expected according to standardized residuals are highlighted in **blue**.

Native speakers	s-genitive	of-genitive	Row total
Indefinite	96 (47.8%)	105 (52.2%)	201 (38%)
Definite	159 (48.5%)	169 (51.5%)	328 (62%)
Column total	255 (48.2%)	274 (51.8%)	529
Learners	s-genitive	of-genitive	Row total
Indefinite	217 (38.1%)	353 (61.9%)	570 (32.1%)
Definite	614 (51%)	589 (49%)	1203 (67.9%°)
Column total	831 (46.9%)	942 (53.1%)	1773

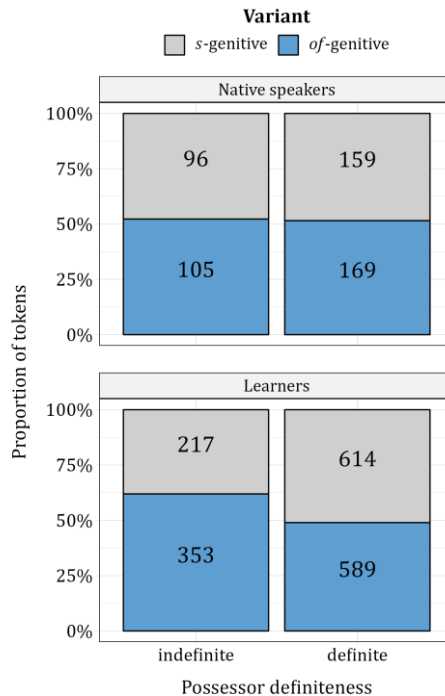


Figure 6.11: Distribution of the s-genitive and the of-genitive across levels of possessor definiteness for native speakers (top) and learners (bottom).

6.2.3.9 Final sibilancy

Previous research found that plural possessors tend to avoid the *s*-genitive, but it was later determined that this is probably due to a more general avoidance of *s*-genitives when the possessor ends in a sibilant (Gries & Wulff 2013: 334–335). This avoidance has been explained as a phonological horror aequi effect (Hinrichs & Szmrecsanyi 2007: 453; Shih et al. 2015), reminiscent of the obligatory contour principle (OCP) which states that adjacent sounds are disfavored (Guy & Boberg 1997).

The operationalization of final sibilancy is based on Grafmiller (2014: 478) and Heller (2018: 71–72). First, I looked up the last sound of the final word of the possessor in the Carnegie Mellon University Pronouncing Dictionary (Lenzo 2014). The possessor has a final sibilant if the dictionary shows that the last word of the possessor ends in either [s], [z], [ʃ], [tʃ], [ʒ], or [dʒ]. In case the final word has no corresponding entry in the dictionary, I checked its spelling. If the word ends in either *s*, *z*, *ce*, *sh*, *tch*, *dge* or the digit 6, the possessor was considered to have a final sibilant (see Grafmiller 2014: 478; Heller 2018: 71). In example (165), the possessor *his er ancestors* has a final sibilant because it ends in [z] and is spelled with *s*.

(165) there's this machine called the Animus er which lets him er relive
his er ancestors' memories so with that he ge= he knows where the
 er artefact is and he retrieves it (TLC:2_8_IN_9, B2 Hindi)

The distribution of final sibilancy across *s*-genitives and *of*-genitives is given in Table 6.10 and Figure 6.12.

Table 6.10: Contingency table for the final sibilancy of the possessor for native speakers ($X^2 = 0.78$, $df = 1$, $p = 0.377$) and learners ($X^2 = 10.05$, $df = 1$, $p = 0.002$, Cramér's $V = 0.08$). Percentages indicate the proportion of genitive variants per category of possessor final sibilancy. For the row total, the percentages indicate the total proportion of the possessors with and without final sibilant. If the test is significant, values that are significantly lower than expected according to standardized residuals are highlighted in **orange**, whereas values that are significantly higher than expected according to standardized residuals are highlighted in **blue**.

Native speakers	s-genitive	of-genitive	Row total
Absent	215 (49.2%)	222 (50.8%)	437 (82.6%)
Present	40 (43.5%)	52 (56.5%)	92 (17.4%)
Column total	255 (48.2%)	274 (51.8%)	529
Learners	s-genitive	of-genitive	Row total
Absent	690 (48.8%)	724 (51.2%)	1414 (79.8%)
Present	141 (39.3%)	218 (60.7%)	359 (20.2%)
Column total	831 (46.9%)	942 (53.1%)	1773

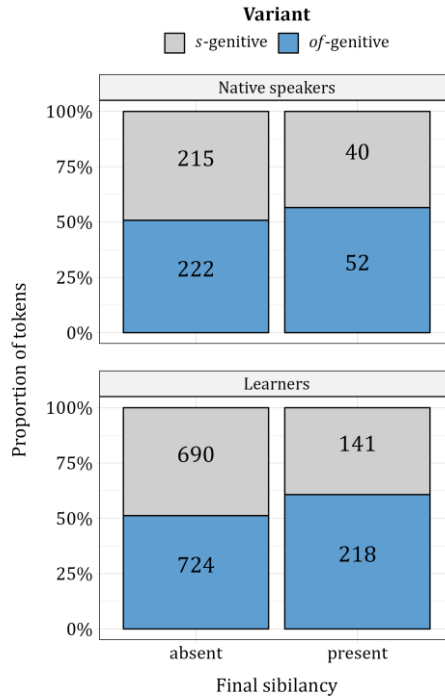


Figure 6.12: Distribution of the *s-genitive* and the *of-genitive* across levels of possessor final sibilancy for native speakers (top) and learners (bottom).

The final sibilancy of the possessors does not seem to be significantly associated with the choice of genitive variant in the native speaker data ($p = 0.38$). By comparison, learners tend to avoid the *s-genitive* when the possessor ends in a sibilant, as expected. This association is rather weak however (Cramér's $V = 0.08$).

6.2.4 Classifying the constraints

The theory-based account distinguishes between feature-based constraints and processing-related constraints. Whereas feature-based constraints are claimed to be acquired via associative statistical learning, processing-related constraints do not have to be learned as they derive from the features of the brain (see Section 3.3). To test the relevant hypotheses, it is necessary to classify the genitive constraints in either category. With this in mind, feature-based constraints can be identified by three characteristics: they typically involve linguistic elements in the context surrounding the choice context, they are language-specific rather than universal, and their

effect can be formulated as a categorical rule of grammar (Tamminga, MacKenzie & Embick 2016: 321).

Local vs. non-local

Most genitive constraints involve linguistic elements in the immediate context of the genitive variants given that they relate to the features of the possessor and possessum, including the definiteness, final sibilancy and animacy of the possessor, the semantic relation between possessor and possessum, the length of the constituents and the type-token ratio. By comparison, the operationalization of structural persistence, possessor discourse status and thematicity involves much linguistic material preceding the genitive variants, so they are not local.

Language-specific vs. universal

While animacy and definiteness are well-known in the typological literature as semantic-pragmatic scales (see Croft 2003: 128–132; Corbett 2010; Rosenbach 2014: 238), the concrete way in which these cross-linguistically attested patterns are realized in English is language-specific. For example, possessor animacy does not operate in exactly the same way in closely related languages such as Dutch and German, which also feature competition between a remnant of the inflectional genitive case and a prepositional variant (see Allen 2003; Gries & Wulff 2013: 350–351).

The operationalization of semantic relation is also based on typological research (Koptjevskaja-Tamm 2001). The general pattern is that adnominal possessive structures like the *s*-genitive favor prototypical relations between the possessor and possessum. In this respect, body-part and kinship relations are considered more prototypical than part-whole relations, which in turn are more prototypical than legal ownership relations (Koptjevskaja-Tamm 2001: 965). While all of these prototypical relations can take the *s*-genitive in English, other languages might only license adnominal possessive structures with the most prototypical relations like body part or kinship (Koptjevskaja-Tamm 2001: 964), so this constraint is also language-specific.

One could argue that final sibilancy is universal because it is physiologically difficult to produce subsequent sibilants. In reality, the sibilant of the possessor is not adjacent to the sibilant of the genitive marker because the genitive clitic becomes [iz], so this is a language-specific tendency instead.

Although constituent length operates in the same way as animacy or definiteness at first glance, I argue that the length of the constituents is strictly processing-related. These constraints are explained with respect to

Easy First, which states that easy constituents are produced first (MacDonald 2013). However animate or definite possessors indirectly influence processing because they are more accessible (e.g., Branigan, Pickering & Tanaka 2008). Additionally, 87% of the *s*-genitives in the data (946 out of 1,086) have a collective or animate possessor, so frequency and semantic prototypicality effects could also explain why it is efficient to produce *s*-genitives with animate possessors (see Ellis, O'Donnell & Römer 2014; Bidgood et al. 2020). By contrast, I argue that the tendency for shorter constituents to be produced first simply results from the basic fact that shorter constituents require speakers to retrieve less material from memory combined with the fact that human language is processed incrementally (Timmers et al. 2013: 10; Philipp et al. 2017).

Similarly, I argue that the effect of type-token ratio is processing-related. Contexts with a high type-token ratio are lexically dense, meaning that speakers have to retrieve more lexical items from memory. As a result, lexically dense contexts form cognitively complex environments, which should favor the *of*-genitive following Rohdenburg's Complexity Principle (1996: 151).

Discourse status captures whether the referent of the possessor has been mentioned in the preceding context, while possessor thematicity captures how often the referent was mentioned in the entire preceding discourse to reflect the overall saliency of the referent as a topic. Therefore, one can argue that discourse-given, thematic possessors are easier to produce again because of universal lexical priming effects (e.g., Branigan, Pickering & Tanaka 2008). The same explanation applies for structural persistence, which can occur as long as the speakers have an internalized representation of the relevant constructions (Gries & Wulff 2005: 186; Romano 2018: 276).

Categorical rule

It is possible to formulate a categorical rule for the animacy, definiteness and final sibilancy of the possessor as well as the semantic relation that holds between the possessor and the possessum. Regarding animacy, prescriptivist grammars have already proposed a categorical rule stating that animate possessors should always take the *s*-genitive (Murphy 2012; Heller, Szmrecsanyi & Grafmiller 2017: 10). One could also easily imagine such a rule for semantic relation given that languages vary regarding the types of semantic relations that license adnominal possessive structures (Koptjevskaja-Tamm 2001: 964). The same applies for definiteness, which also reflects a semantic-pragmatic scale. For final sibilancy, one could formulate a rule whereby 'the *s*-clitic cannot be preceded by the sounds [s],

[z], [ʃ], [tʃ], [ʒ], or [dʒ]'. For this rule to apply, speakers should be aware of the pronunciation of the last word in the possessor while planning syntax since the choice of genitive variant involves a change of word order. While this might seem unlikely at first glance, the same assumption underlies final sibilancy when it applies probabilistically rather than categorically.

By contrast, I argue that it is not possible to formulate a categorical rule for the effect of constituent length, type-token ratio, structural persistence, discourse status or possessor thematicity. For constituent length, such categorical rules are not realistic because "grammars can't count" (Selkirk 1986; cited by Tamminga, MacKenzie & Embick 2016: 321). For the same reason, it seems unlikely that languages can keep track of the lexical density surrounding the genitive construction, or the degree to which a possessor is thematic based on its frequency. Likewise, it would not make sense to always repeat the last genitive variant in discourse. In principle, one can imagine a rule whereby a discourse-given referent should be produced in an *s*-genitive, but this rule becomes less realistic as the previously mentioned referent is further removed from the choice context.

Overview

Table 6.11 offers an overview of the constraints regarding the three classification criteria, which shows that the processing-related constraints operating on the genitive alternation are constituent length, discourse possessor status, structural persistence, possessor thematicity and type-token ratio.

Table 6.11: Overview of the classifying features for each of the genitive constraints.

	<i>Local context</i>	<i>Language-specific</i>	<i>Categorical rule</i>
Constituent length	Yes	No	No
Discourse status	No	No	No
Persistence	No	No	No
Possessor thematicity	No	No	No
Type-token ratio	Yes	No	No
Possessor animacy	Yes	Yes	Yes
Semantic relation	Yes	Yes	Yes
Possessor definiteness	Yes	Yes	Yes
Final sibilancy	Yes	Yes	Yes

6.2.4.1 Ranking the feature-based constraints

The animacy, definiteness and final sibilancy of the possessor, as well as the semantic relation holding between the possessor and the possessum are the feature-based constraints of the genitive alternation that are acquired via associative statistical learning. Table 6.12 offers an overview of these constraints ranked by their association strength with the genitive variants for native speakers.

Table 6.12: The feature-based genitive constraints, ranked by the strength of association according to monofactorial tests (Cramér’s *V*).

Constraint	Cramér’s <i>V</i>
Possessor animacy	0.68
Semantic relation	0.10
Possessor definiteness	NS
Final sibilancy	NS

According to this rather simplistic proxy for the reliability of the constraints, the most learnable constraint is possessor animacy, followed by semantic relation. Possessor definiteness and final sibilancy should be less learnable as they are not significantly associated with either construction according to the monofactorial tests.

6.2.5 General overview

Table 6.13 offers a general overview of the genitive constraints included in the analysis, with their relevant distinctions, their effect on the choice of variant, their classification, and their learnability via associative statistical learning. Only possessor animacy and semantic relation required extensive manual coding, so I conducted an interrater reliability analysis for these constraints. To do so, I developed a coding scheme detailing the annotation process for both variables (Révész 2012), which I then visualized as decision trees with several yes/no questions. A second coder used the coding scheme and decision trees to double-code a random selection of 1,200 genitive observations for both constraints, corresponding to 52% of the entire genitive dataset. The interrater reliability was estimated with Cohen’s kappa, which showed an almost perfect agreement of 0.83 for possessor animacy and 0.85 for semantic relation between coders (Landis & Koch 1977).

Table 6.13: Overview of the genitive constraints, as well as the relevant coding distinctions, their expected effect on the choice of variant and their learnability via associative statistical learning. 'Not consistent' in the effects column refers to the fact that the effect is not consistently found in previous literature.

Constraints	Distinctions	(Expected) effect	Learnability
Constituent length	Numerical	- Long possessors favor the <i>of</i>-genitive - Long possessums favor the <i>s</i>-genitive (not consistent)	Processing-related
Discourse status	Given New	Given possessors favor the <i>s</i> -genitive (not consistent)	Processing-related
Persistence	Examiner <i>s</i> -genitive Examiner <i>of</i> -genitive Candidate <i>s</i> -genitive Candidate <i>of</i> -genitive None	- Previous examiner or candidate <i>s</i>-genitive prefer the <i>s</i>-genitive - Previous examiner or candidate <i>of</i>-genitive prefer the <i>of</i>-genitive	Processing-related
Possessor thematicity	Numerical	Thematic possessors prefer the <i>s</i> -genitive (not consistent)	Processing-related
Type-token ratio	Numerical	Environments with high type-token ratio prefer the <i>s</i> -genitive	Processing-related
Possessor animacy	Animate Collective Inanimate Locative Temporal	- Animate and collective possessors favor the <i>s</i>-genitive - Inanimate, locative and temporal possessors prefer the <i>of</i>-genitive	Easy
Semantic relation	Prototypical Non-prototypical	Prototypical relations favor the <i>s</i> -genitive	Easy
Definiteness	Definite Indefinite	Definite possessors favor the <i>s</i> -genitive	Difficult
Final sibilancy	Present Absent	Possessors with final sibilant avoid the <i>s</i> -genitive	Difficult

6.2.6 Statistical analysis

The annotated dataset was analyzed with mixed-effects logistic regression using the *lme4* package (Bates et al. 2015) in R (version 4.2.2, R Core Team 2022). The model was fitted to predict the likelihood of the *s*-genitive with the probabilistic constraints as predictors. These predictors were allowed to interact with the proficiency level of the speaker, so that it is possible to compare how the constraints affect the choice of genitive variant across proficiency levels.

Random effects

The choice of genitive variant is influenced by other sources of variation besides the probabilistic constraints. For the most part, these are the same as for the future marker (see Section 5.2.5), so they include the idiolect of the speakers, the country and L1 of the learner, the type of exam task and the joint discourse established by the learner and the examiner during the exam. However, the alternations diverge with respect to lexical biases, which are tied to the verb for the future marker alternation. In the genitive alternation, it is the head noun of the possessor and possessum that are biased towards either variant (e.g., Wolk et al. 2013; Heller, Szmrecsanyi & Grafmiller 2017).

While these sources of variation should ideally be controlled for, there is not enough data to include them as predictors in the model. Moreover, most of these sources of variation are not particularly interesting except for L1, which can also not be entered as a predictor in the model because both proficiency level and L1 overlap perfectly for native speakers. Therefore, the model instead featured random effects for the idiolect of the speakers, their L1, their country, the exam, the exam task within the exam and the exam task across exams, as well as the head of the possessor and the possessum. Again, it is recommended to include at least 5 observations for each level of a random effect for model parameters to be reliable (Clarke 2008: 757), which can be problematic for certain sources of variation like the idiolect of the speakers, given that only 157 out of the 514 (31%) speakers in the corpus contributed more than 5 genitives to the data.

Exploring the data

With the exception of type-token ratio, the numerical variables were severely right-skewed, indicating that while most constituents are rather short and not particularly thematic, there are outliers. To reduce the skew, these variables were log-transformed. Following the recommendation of Zuur, Ieno and Elphick (2010: 5), I then used Cleveland dot plots (Cleveland 1993) to visually identify whether any outliers remained in the distribution of the log-transformed numerical variables for each random effect (Baayen 2008: 243–244). This revealed no outliers, so none of the observations were removed. All numerical variables were then centered on the median before being standardized by dividing the value by two standard deviations to make the effect of numerical variables more similar to that of binary categorical predictors in the model (see Gelman 2008).

Then, I explored the categorical predictors to determine whether each level contained enough *s*-genitives and *of*-genitives to ensure reliable results. Like for the future marker data, I adopted a slightly stricter version

of Field, Miles and Field's (2012: 340) suggestion and considered variables problematic when there are fewer than 10 to 15 observations in 20% of the combinations between the response variable and the predictor. To illustrate, the animacy constraint consisting of five predictor levels is allowed to interact with proficiency level, which also features five levels, to predict the binary choice of variant, yielding a total of 50 combinations. 12 of these 50 (24%) combinations contained fewer than 10 genitive observations, so I decided to conflate possessor animacy from a five-level predictor to a binary predictor. Specifically, animate and collective possessors were conflated as 'animate' while inanimate, locative and temporal possessors were combined in an 'inanimate' category (see Röthlisberger 2018: 61). Through the same process, I found that there are few *s*-genitives with final sibilant and no previous genitive variant in the C2 level, so I decided to conflate the data from the C2 speakers with the C1 data.

Treatment coding

I again used treatment coding to analyze how probabilistic constraints influence the choice of genitive variant, meaning that the model chose a reference level for each predictor to be compared against the other levels of the predictor. Since the model predicts the likelihood of the *s*-genitive, I chose the reference level that typically favors the *of*-genitive. Alternatively, I picked the predictor level that incorporates the most genitive observations. The model proceeds from an observation with an inanimate (animacy), indefinite (definiteness), discourse-new (discourse status) possessor with no final sibilant (final sibilancy), a non-prototypical semantic relation between the possessor and the possessum (semantic relation) preceded by an *of*-genitive produced by a learner (persistence). Numerical predictors are set to 0 by default, which means that the observation has a type-token ratio, possessum length, possessor length and thematicity corresponding to the median in the dataset.

Model fitting

The model was again fitted following the approach of Zuur et al. (2009). That is, I started from a beyond optimal model that includes all possible predictors as well as random slopes and intercepts. If including this many regressors leads to convergence issues, the recommendation is to select a beyond optimal model that is theoretically motivated. The beyond optimal model was generated with the following formula using the *glmer* function from the *lme4* package (Bates et al. 2015):

```
Beyond optimal model ← glmer(genitive variant ~ proficiency + possessor animacy + final sibilancy + semantic relation + possessor length + possessum length + possessor definiteness + type-token ratio + possessor discourse status + possessor thematicity + persistence + proficiency:possessor animacy + proficiency:final sibilancy + proficiency:semantic relation + proficiency:possessor length + proficiency:possessum length + proficiency:possessor definiteness + proficiency:type-token ratio + proficiency:possessor discourse status + proficiency:possessor thematicity + proficiency:persistence+ (1|country) + (1|exam) + (1|task within exam) + (1|task across exams) + (1|speaker) + (1|mother tongue) + (1|possessor head lemma) + (1|possessum head lemma), data = genitive data, family = binomial, glmerControl(optimizer = "bobyqa", maximum iterations = 1 million))
```

To arrive at the optimal model, I first removed the random effects that do not significantly contribute to the quality of the model according to likelihood-ratio tests as implemented by the *anova()* function in R, namely the intercept adjustments for the exam, the task across exams, the speaker as well as the learners' country and L1. The fixed-effects structure of the model was optimized with the *drop1()* function (see Section 5.2.5). In short, this function allowed me to identify and remove the fixed effect that is least likely to contribute to the quality of the model iteratively until all predictors remaining in the model significantly improve the model at a *p*-value below 0.1. The final model is generated with the following formula:

```
Optimal model ← glmer(genitive variant ~ proficiency + possessor animacy + final sibilancy + possessor length + possessum length + possessor definiteness + type-token ratio + persistence + proficiency:possessor animacy + proficiency:possessor definiteness + (1|task within exam) + (1|possessor head lemma) + (1|possessum head lemma), data = genitive data, family = binomial, glmerControl(optimizer = "bobyqa", maximum iterations = 1 million))
```

The thematicity and discourse status of the possessor as well as the semantic relation between the possessor and the possessum were removed from the model. Likewise, the interaction between proficiency level and final sibilancy, structural persistence, type-token ratio as well as possessor and possessum length were not retained as they did not improve the model. Hence, the only constraints that interact with proficiency level are possessor animacy and definiteness.

This final model is significantly better than the null model ($p < 0.001$) and predicts the right future marker in 88% of the cases, which is significantly better than the baseline of 53% ($p < 0.001$). The model's $R^2_{\text{marginal}} = 0.40$ and $R^2_{\text{conditional}} = 0.74$ and the concordance index *C* is .96, indicating an excellent fit of the data (Baayen 2008). The condition index κ

of 20.30 indicates that there is medium collinearity in the model, but this remains below the threshold of 30 for harmful collinearity (Baayen 2008: 182). The VIF adjusted by degrees of freedom for each factor is below the threshold of 10 (Field, Miles & Field 2012: 339; Levshina 2015: 273).

6.3 Results

This section presents the results from the mixed-effects logistic regression model, starting with the random effects. Then, I turn to the main effects and the interaction effects involving the probabilistic constraints. At the end of the section, I also report the results from supplementary analyses including MuPDAR (Gries & Adelman 2014; Gries & Deshors 2014).

6.3.1 Random effects

Table 6.14: Overview of the random effects in the model, their estimated variance and their standard deviation.

Random intercept	Variance	St. dev.
Possessum head lemma	3.02	1.74
Possessor head lemma	1.47	1.21
Exam task within exams	0.88	0.94

Table 6.14 shows that the head of the possessum explains the most variance, followed by the head of the possessor and the exam tasks within specific exams. The influence of the exam task implies that there are exams in which the interlocutors showed a preference for a variant during a specific task. This effect is rather puzzling considering that many possible explanations are controlled for. For one, the effect cannot be attributed to the nature of the exam task, considering that the intercept adjustment for exam tasks across exams (e.g., all presentation tasks across exams grouped together) did not contribute to the quality of the model. Similarly, it cannot be due to the country or L1 of the learners for which the random effects were removed as well. A brief look at some of these exam tasks suggests that the interlocutors do not simply repeat the exact same genitive structure during the task either. Perhaps the effect is tied to the specific topic the learner and examiner were discussing during a given exam task, but the head of the possessor and possessum should in principle partially control for such effects as well.

In the following subsections, I will interpret the intercept adjustments for the head of the possessor and possessum. Importantly, it is again not clear whether certain lexical biases are mostly driven by the tendencies of either the learners, the examiners, or both.

Possessor head

Figure 6.13 visualizes the lexical adjustments for the head nouns of the possessors.

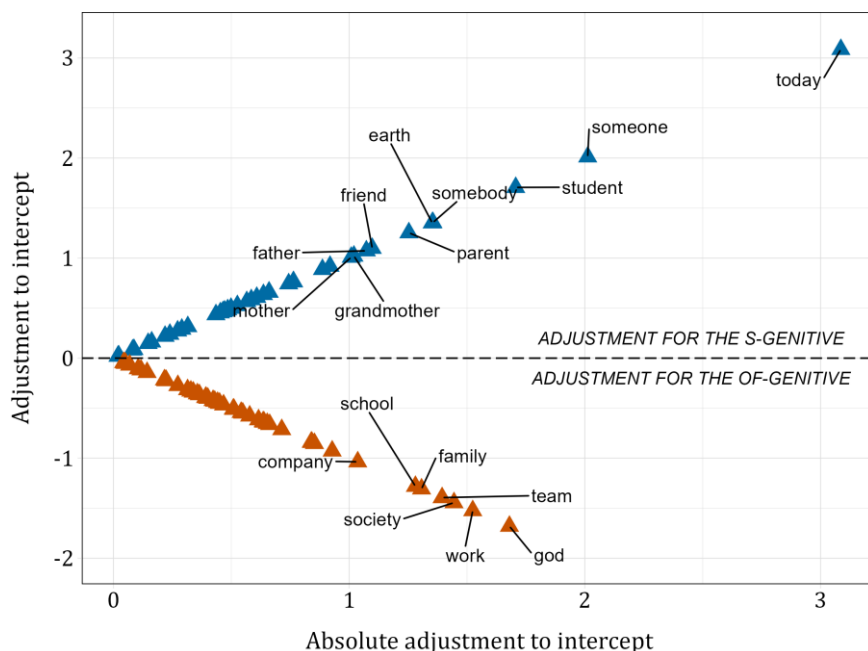


Figure 6.13: Intercept adjustments for the heads of the possessors. Possessors with positive adjustments prefer the *s*-genitive and are depicted as blue triangles. Possessors with negative adjustments prefer the *of*-genitive and are depicted as orange triangles. Only the lemmas of the possessors with an absolute adjustment on the *x*-axis exceeding 1 are labelled.

The possessor with the strongest bias for the *s*-genitive is the temporal possessor *today* as in (166). Similarly, the indefinite pronouns *someone* and *somebody* as in (167) evince a strong tendency towards the *s*-genitive. In general, most of these possessors refer to rather prototypical animates like *student* and many refer to kinship terms in particular like *parent*, *friend*, *father*, *grandmother* and *mother* like as in (168).

(166) er m= in today's society er there are s-so many antisocial
behaviour (TLC:2_8_SP_52, B2 Spanish)

(167) for example he has a really cool bowtie that can change his voice into someone else's voice (TLC:7_CH_5, B2 Chinese)

(168) mm erm my mother's parents erm live not far from we but my father's grandparents erm not live in Hangzhou in an-another city yes (TLC:2_6_CH_53, B1 Chinese)

The possessors that favor the *of*-genitive seem to often imply a collective group of people, like *team*, *society*, *family*, *company* and *school* as in (169). Interestingly, the possessor *god* (170) is the possessor with the strongest bias for the *of*-genitive although it refers to a singular animate entity.

(169) I mean if you're if you're in the bottom part of society wondering about where your next meal comes from is happiness really relevant (TLC:2_ME_19, NS)

(170) I believe in this and I think it's it's giving the message of god it is very simple (TLC:2_8_IN_26, B2 Hindi)

The lexical biases of the possessors in the current data share similarities with the possessors in different World Englishes. *Today* is also the possessor with the strongest bias for the *s*-genitive in Heller (2018: 113), who also finds that *god* and collective nouns like *society*, *church*, *school*, *team*, *department* etc. prefer the *of*-genitive.

Possessum head

The intercept adjustments for the heads of the possessums are pictured in Figure 6.14.

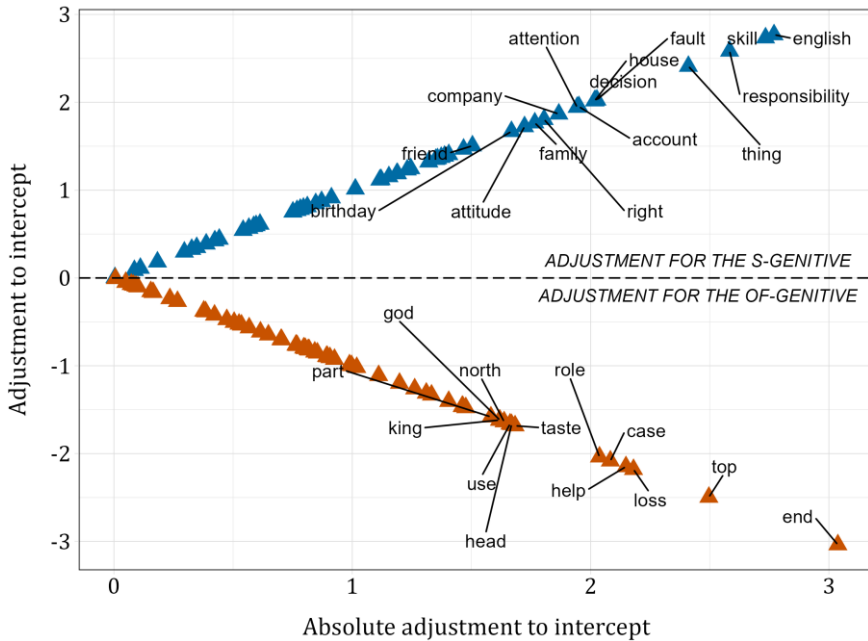


Figure 6.14: Intercept adjustments for the heads of the possessums. Possessums with positive adjustments prefer the *s*-genitive and are depicted as blue triangles. Possessums with negative adjustments prefer the *of*-genitive and are depicted as orange triangles. Only the lemmas of the possessums with an adjustment exceeding 1.5 on the x-axis are labelled.

Perhaps unsurprisingly, the possessums that favor the *s*-genitive tend to take animate possessors, like *birthday*, *friend* and *family* as in (171). Hence, 122 out of 151 (80.6%) observations containing the labelled possessums that favor the *s*-genitive in Figure 6.14 are animate, whereas animate possessors only figure in 40.4% of the genitive observations across the entire dataset (931 out of 2,302). As a result, these possessums tend to occur in *s*-genitives.

(171) yeah but socially people don't talk about work do they they talk about their family and friends and that's boring I don't want to hear about other people's friends and family (TLC:2_IN_18, NS)

Many of the possessums that favor the *of*-genitive fit the borderline cases discussed in Section 6.2.2.2. That is, possessums like *end*, *top*, *part* etc. typically appear in *of*-genitives to point to a certain part of the possessor (172). Likewise, the possessums *help* and *use* exhibit a strong preference for the *of*-genitive, especially when combined with inanimate possessors and

preceded by *with* or *without* as in (173), in which case the possessum resembles complex prepositions like *by means of* (see Section 6.2.2.2). As a possessum, *god* also appears to favor the *of*-genitive to describe what kind of god the referent is as in (174).

(172) er training er is the most important part of sports (TLC:2_6_SP_65, B1 Spanish)

(173) we couldn't probably er imagine nowadays er our routine without s= without the use of its erm technologies such as smartphones er computers (TLC:2_ME_2, C1 Spanish)

(174) he's Ares he's the god of war (TLC:2_6_IN_86, B1 Marathi)

6.3.2 Fixed effects

6.3.2.1 Main effects

The main effects in the model are presented in Table 6.15. The coefficients and the odds ratios both represent the effect sizes and directions of the predictors. Given that the model predicts the likelihood of the *s*-genitive, positive coefficients and odds ratios above 1 indicate that the corresponding change of predictor levels makes the *s*-genitive more likely, whereas negative coefficients and odds ratios between 0 and 1 mean that the *s*-genitive becomes less likely.

Table 6.15: Main effects of the individual predictors on the choice of genitive variant. Positive coefficients indicate an increase in the likelihood of the *s*-genitive. (Marginally) significant *p*-values are written in bold.

Predictor	Coefficient	Odds ratio	Std. error	<i>p</i> -value
(Intercept)	-2.406	0.09	0.52	<0.001
Proficiency				
Native → B1	0.927	2.528	0.595	0.119
Native → B2	-0.359	0.698	0.516	0.486
Native → C	-0.278	0.758	0.476	0.56
Possessor animacy				
Inanimate → animate	3.284	26.678	0.411	<0.001
Possessum length (logged)	-3.258	0.038	0.219	< 0.001
Possessor length (logged)	-1.118	0.327	0.187	<0.001
Persistence				
Candidate <i>of</i> → candidate <i>s</i>	-0.027	0.973	0.205	0.894
Candidate <i>of</i> → examiner <i>s</i>	0.314	1.369	0.278	0.258
Candidate <i>of</i> → examiner <i>of</i>	-0.623	0.536	0.223	0.005
Candidate <i>of</i> → None	0.124	1.132	0.218	0.571
Final sibilancy				
Absent → present	-0.543	0.581	0.22	0.013
Type-token ratio	0.391	1.479	0.158	0.013
Possessor definiteness				
Indefinite → definite	-0.235	0.79	0.375	0.531

Except for the animacy and the definiteness of the possessor, the probabilistic constraints that were retained in the model do not interact with the proficiency level of the speaker. In other words, the length of the possessor and the possessum, structural persistence, final sibilancy and type-token ratio all influence the choice of variant in the same way for native speakers and learners regardless of their proficiency level.

For the length of the constituents, the model shows that the *s*-genitive is 3.06 times less likely when the possessor is comparatively long. Put simply, longer possessors take the *of*-genitive as predicted by the short-before-long principle (Behaghel 1909) or Easy First (MacDonald 2013). When the possessum is comparatively long, the *s*-genitive is 26 times less likely than with the average possessum. Hence, *of*-genitives are particularly accommodating to long possessors and especially long possessums as in (175).

(175) do you think then that the internet is the most important scientific development of the last fifty years (TLC:ME_36, NS)

This is surprising as it goes against the short-before-long principle (Behaghel 1909) and Easy First (MacDonald 2013). Szmrecsanyi (2006)

found the same result in his investigation of the genitive alternation in spoken British English. As stated in Section 6.2.3.1, this pattern aligns with Rohdenburg's Complexity Principle (1996: 151), whereby the more explicit *of*-genitive is favored when the cognitive environment is highly complex, as is the case when both constituents are long.

Regarding structural persistence, the model shows that the *s*-genitive is less likely if an examiner previously used an *of*-genitive, as expected. This is the only persistence effect reported by the model, which has a number of implications. First, it seems that previous *s*-genitives do not necessarily cause speakers to produce an *s*-genitive again, which is surprising considering that the less frequent *s*-genitive should in principle induce stronger persistence effects than *of*-genitives (Jaeger & Snider 2013). Second, the finding suggests that learners are more reactive to previous genitive variants produced by the examiner rather than themselves. Conversely, the genitive variants produced by learners do not seem to cause persistence effects for native speakers. Overall, the results are reminiscent of structural persistence in the future marker alternation (see Section 5.3.2.2).

The effect for final sibilancy indicates that the *s*-genitive is less likely to be used when the possessor ends in a sibilant, as expected (e.g., Heller, Szmrecsanyi & Grafmiller 2017; Grafmiller 2014). Finally, speakers are more likely to use the *s*-genitive when the type-token ratio is high. In written language, this effect is attributed to the pressures of economization: the *s*-genitive helps save space in lexically dense environments (Hinrichs & Szmrecsanyi 2007: 457; Szmrecsanyi & Hinrichs 2008: 300–301). However, it is somewhat unclear how this explanation pertains to spoken language where one would expect processing factors such as Rohdenburg's Complexity Principle (1996: 151, see Section 6.2.3.5) to induce a preference for the *of*-genitive instead.

Possessor animacy and definiteness interact with proficiency level. Hence, the main effects for possessor animacy and definiteness only apply for the reference proficiency level, viz. native speakers. The effect for possessor animacy shows that when the possessor is animate rather than inanimate, native speakers are almost 27 times more likely to use the *s*-genitive. Hence, animacy is arguably the strongest constraint guiding the choice of genitive variant according to the model in accordance with previous research (Shih et al. 2015; Rosenbach 2017: 14), followed closely by the length of the possessum. The definiteness of the possessor does not appear to affect native speakers, but since it is part of an interaction term, there is in all likelihood a significant difference with one or more learner

proficiency levels that helps explain variability in the data. Conversely, the main effect for proficiency level only applies for the reference levels of possessor animacy and definiteness. Hence, these effects reveal that the *s*-genitive is equally likely across proficiency levels with inanimate and indefinite possessors.

6.3.2.2 Interaction effects

Table 6.16 gives an overview of the model results for the interactions between the animacy and definiteness of the possessor and the proficiency level of the speaker. Significant regressors indicate that the corresponding group of learners is not affected by the constraint in the same way as native speakers.

Table 6.16: Interaction effects of the individual predictors on the choice of genitive variant. Positive coefficients indicate an increase in the likelihood of the *s*-genitive. (Marginally) significant *p*-values are written in bold.

Predictor	Coefficient	Odds ratio	Std. error	<i>p</i> -value
Proficiency x possessor animacy				
B1 + animate	-1.213	0.297	0.553	0.028
B2 + animate	-1.204	0.3	0.486	0.013
C + animate	0.738	0.478	0.469	0.116
Proficiency x possessor definiteness				
B1 + definite	0.406	1.501	0.566	0.473
B2 + definite	1.226	3.409	0.477	0.010
C + definite	0.847	2.333	0.444	0.056

To illustrate, possessor definiteness influences B2 learners differently from native speakers, who are not sensitive to this constraint. Specifically, the coefficient and odds ratio indicate that B2 learners are 3.41 times more likely than native speakers to use the *s*-genitive when the possessor is definite rather than indefinite. This coefficient can be summed up with the coefficient of the corresponding main effect applying to native speakers in order to calculate the effect of possessor definiteness on B2 learners. This shows that B2 learners are 2.69 times more likely to use the *s*-genitive when the possessor is definite rather than indefinite, as expected from the literature (Wolk et al. 2013; Heller 2018: 19). In the following subsections, I focus on both interactions separately and visualize their effect with partial effects plots.

Possessor definiteness

The partial effects plot in Figure 6.15 visualizes the interaction between proficiency level and possessor definiteness. Partial effects plots

summarize the information from the model by changing the value of the relevant constraint while holding the other constraints in the model at their default level (Fox 2003). In the case of Figure 6.15, the plot shows how a change of definiteness affects the likelihood of the *s*-genitive across proficiency levels while holding the other constraints in the model at their reference level.

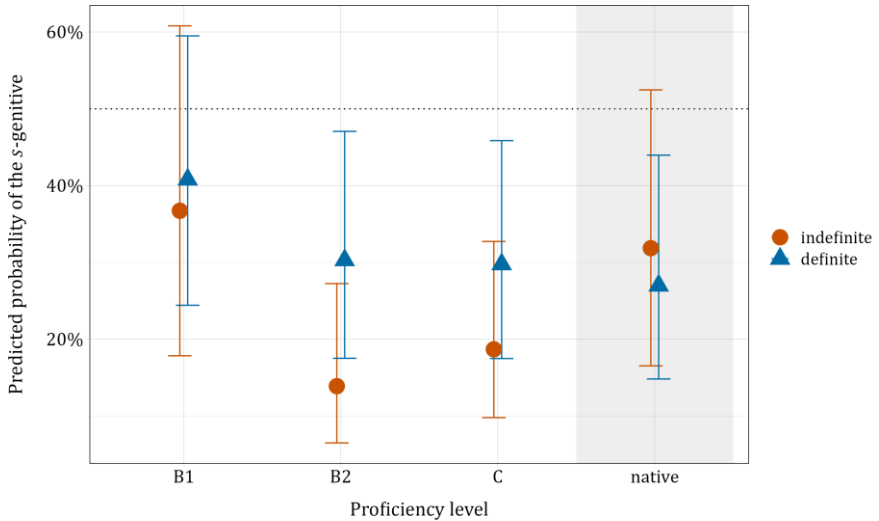


Figure 6.15: Partial effects plot of the interaction between proficiency level and possessor definiteness. Probabilities on the y-axis are for the *s*-genitive. The vertical distance between the shapes reflects the effect size of the predictor. The plotted predicted probabilities are calculated with all other categorical predictors are set at their reference level, and the numerical predictors at 0. The native proficiency level (shaded) is the reference level.

As discussed in the previous section, B2 learners are 3.41 times more likely than native speakers to use the *s*-genitive when the possessor is definite rather than indefinite. Table 6.16 shows that there is another marginally significant regressor indicating that C-level learners also have a slight tendency to use the *s*-genitive when the possessor is definite compared to native speakers. B1 learners do not differ from native speakers regarding this constraint, so they are in all likelihood not sensitive to possessor definiteness as well.

A closer look at the dataset reveals that this lack of an effect for possessor definiteness in the native speaker data may be due to the examiners' frequent use of two constructions that go against the expected tendency for definite possessors to figure in *s*-genitives. First, the examiners often place definite possessors like *the exam* or *the test* in *of*-genitives rather than *s*-genitives to refer to parts of the exam, as in (176).

(176) Now we've arrived at the final part of the exam the discussion phase erm did you prepare from list A or B? (TLC:2_IN_24, NS)

Second, the examiners almost always place the indefinite possessor *people* in *s*-genitives rather than *of*-genitives to introduce new topics of discussion with the candidate, as shown in (177).

(177) Do you think that erm values change from generation to generation? So for example when your grandmother was very young do you think people's values were different from values today? (TLC:2_8_IN_7, NS)

Both of these constructions alone account for 72 out of the 529 (13.6%) native speaker observations, so they may have counteracted the tendency for definite possessors to figure in *s*-genitives in the remaining observations. As a result, possessor definiteness does not significantly affect native speakers' choice of variant according to the model.

To test this possibility, I fitted the model again without both of these constructions. This model reached an accuracy of 89% with a concordance index *C* of 0.96, indicating an excellent fit of the data (Baayen 2008). This reduced model no longer features a significant interaction between definiteness and proficiency level. Instead, the native speakers now also prefer the *s*-genitive when the possessor is definite, which lends support to the claim that both constructions have caused the full model to miss the effect of possessor definiteness in native speakers. If it is the case that native speakers are in fact sensitive to possessor definiteness if not for these two constructions, then only the B1 learners are not sensitive to the definiteness constraint.

Possessor animacy

The main effects in Table 6.15 show that when the possessor is animate rather than inanimate, native speakers are almost 27 times more likely to use the *s*-genitive. The coefficients of the interactions in Table 6.16 reveal that this tendency is weaker for B1 learners and B2 learners. Specifically, B1 and B2 learners are approximately three times less likely than native speakers to choose the *s*-genitive when the possessor is animate rather than inanimate. Nonetheless, both B1 and B2 learners still prefer to use the *s*-genitive when the possessor is animate rather than inanimate, and this effect remains rather strong: both B1 and B2 learners are about 8 times more likely to use the *s*-genitive when the possessor is animate. Overall, it is surprising that learners would be less sensitive to such a strong constraint, since it would be easier for learners to simply rely on the

animacy of the possessor categorically if anything, as suggested by prescriptivists (Murphy 2012; Heller, Szmrecsanyi & Grafmiller 2017: 10).

Figure 6.16 reveals that when the possessor is animate, B1 learners are more or less as likely as native speakers to use the *s*-genitive. However, B1 learners are more likely to use the *s*-genitive than native speakers when the possessor is inanimate. As a result, possessor animacy is weaker for B1 speakers, as shown by the smaller vertical distance between the shapes at the B1 level than at the native speaker level in Figure 6.16.

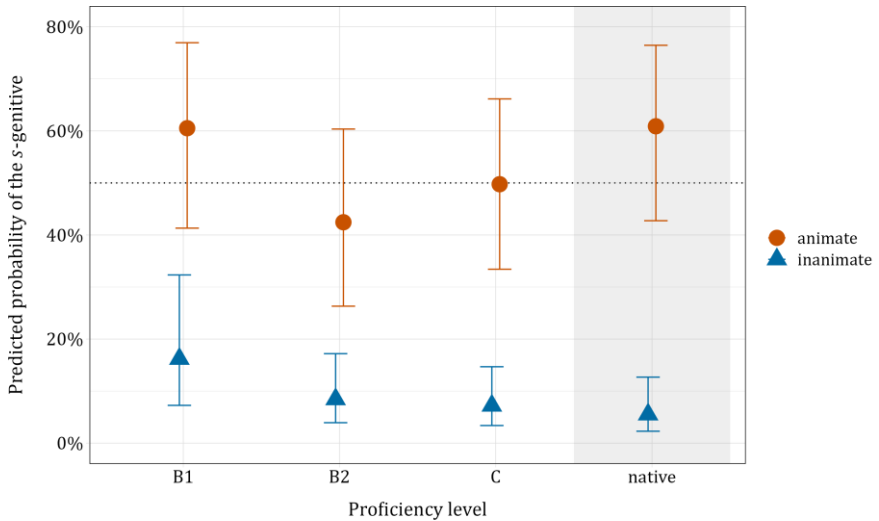


Figure 6.16: Partial effects plot of the interaction between proficiency level and possessor animacy. Probabilities on the y-axis are for the *s*-genitive. The vertical distance between the shapes reflects the effect size of the predictor. The plotted predicted probabilities are calculated with all other categorical predictors are set at their reference level, and the numerical predictors at 0. The native proficiency level (shaded) is the reference level.

More concretely, the finding implies that learners sometimes use the *s*-genitive with inanimate possessors where native speakers would produce the *of*-genitive instead. For instance, the possessor *the London* in (178) is inanimate yet figures in an *s*-genitives whereas a native speaker would have rather used *pictures of (the) London*.

(178) Examiner: I'm from England I'm from London in England have you been to London?

Candidate: No but I I see the London's pictures

(TLC:2_6_CH_23, B1 Chinese)

As for B2 learners, it seems that they deviate from native speakers because they use the *s*-genitive less often than native speakers with animate possessors. Hence, *of*-genitives featuring animate possessors as in (179) are more common for B2 learners.

(179) without the cooperation of the people erm government will won't be able to do anything (TLC:2_8_IN_40, B2 Hindi)

6.3.3 Supplementary analyses

In this section, I present some supplementary analyses of the genitive data. First, I modelled the data with a focus on the influence of the learners' L1 rather than their proficiency level. Second, I also studied the effects of proficiency level and L1 using the MuPDAR approach (Gries & Deshors 2014; Gries & Adelman 2014; Gries & Deshors 2020).

6.3.3.1 Mother tongue model

To investigate how L1 modulates the choice of variant, I fitted another mixed-effects logistic regression on the same 2,302 genitive observations, but I allowed the probabilistic constraints to interact with speakers' L1 rather than their proficiency. In order to compare native speakers with learners, the reference L1 in this model is English. Hence, the model shows which constraints influence native speakers and whether any constraint influences the choice of genitive variant differently for learners from one or more different L1 backgrounds. Much like the supplementary analyses for the future marker alternation (see Section 5.3.3), this model could help ascertain that the results from the proficiency-based model are not due to tendencies of the underlying L1s.

Model design and fitting

The roles of proficiency level and L1 are essentially reversed in the design of the L1 model. In the main analysis, proficiency level is a fixed effect that is allowed to interact with each probabilistic constraint, whereas intercept adjustments are included in the beyond optimal model for learners' L1. These intercepts were ultimately removed over the course of model fitting as they did not improve the quality of the model. In the L1 model, the beyond optimal model includes L1 as a fixed effect and in interaction terms, while proficiency level is partially controlled for in the form of intercept adjustments. For this design, I had to make sure each L1 background features enough *s*-genitives and the *of*-genitives. An exploration of the data revealed that many L1s do not contain many observations in general, so they were conflated together, including Gujarati, Kannada, Malayalam,

Marathi, Russian, Tamil, Telugu and any unknown L1s. The numerical variables were log-transformed and standardized in the same way as for the proficiency-based model. The beyond optimal model was generated with the following formula:

```
Beyond optimal L1 model ← glmer(genitive variant ~ L1 + possessor animacy +
final sibilancy + semantic relation binary + possessor length + possessum length
+ possessor definiteness + type-token ratio + possessor discourse status +
possessor thematicity + persistence + L1:possessor animacy + L1:final sibilancy +
L1:semantic relation binary + L1:possessor length + L1:possessum length +
L1:possessor definiteness + L1:type-token ratio + L1:possessor givenness +
L1:possessor thematicity + L1:persistence+ (1|country) + (1|exam) + (1|task
within exam) + (1|task across exams) + (1|speaker) + (1|proficiency level) +
(1|possessor head lemma) + (1|possessum head lemma), data = genitive data,
family = binomial, glmerControl(optimizer = "bobyqa", maximum iterations = 1
million))
```

As recommended by Zuur et al. (2009), I first determined the random effects structure before turning to the fixed effects. The final model is generated as follows:

```
Optimal L1 model ← glmer(genitive variant ~ L1 + possessor animacy + final
sibilancy + possessor length + possessum length + possessor definiteness + type-
token ratio + possessor thematicity + persistence + L1:type-token ratio +
L1:possessor thematicity + (1|possessor head lemma) + (1|possessum head
lemma), data = genitive data, family = binomial, glmerControl(optimizer =
"bobyqa", maximum iterations = 1 million))
```

The optimal L1 model has a concordance index C of .95, indicating an excellent fit of the data (Baayen 2008), with no indication of harmful multicollinearity (condition index $\kappa = 9.07$) (Baayen 2008: 182).

Results and discussion

Table 6.17 presents an overview of the main and interaction effects in the L1 model.

Table 6.17: Main and interaction effects of the individual predictors on the choice of genitive variant in the mother tongue model. Positive coefficients indicate an increase in the likelihood of the *s*-genitive. (Marginally) significant *p*-values are written in bold.

Predictor	Coefficient	Std. error	<i>p</i> -value
(Intercept)	-2.011	0.38	<0.001
Mother tongue			
English → Chinese	0.134	0.244	0.583
English → Hindi	-0.547	0.280	0.051
English → Italian	0.447	0.286	0.119
English → Spanish	-0.286	0.227	0.209
English → Other L1s	-0.341	0.237	0.150
Possessor animacy			
Inanimate → animate	2.418	0.198	<0.001
Possessum length (logged)	-3.15	0.208	< 0.001
Possessor length (logged)	-1.095	0.181	<0.001
Persistence			
Candidate <i>of</i> → candidate <i>s</i>	0.011	0.195	0.953
Candidate <i>of</i> → examiner <i>s</i>	0.301	0.267	0.260
Candidate <i>of</i> → examiner <i>of</i>	-0.519	0.218	0.017
Candidate <i>of</i> → None	0.246	0.211	0.242
Final sibilancy			
Absent → present	-0.464	0.212	0.028
Type-token ratio	0.213	0.328	0.515
Possessor definiteness			
Indefinite → definite	0.512	0.179	0.004
Possessor thematicity	0.011	0.322	0.974
Mother tongue x type-token ratio			
Chinese + type-token ratio	-0.517	0.464	0.266
Hindi + type-token ratio	-0.710	0.557	0.202
Italian + type-token ratio	1.133	0.573	0.048
Spanish + type-token ratio	0.434	0.436	0.32
Other L1s + type-token ratio	0.085	0.444	0.849
Mother tongue x possessor thematicity			
Chinese + possessor thematicity	0.708	0.472	0.134
Hindi + possessor thematicity	-0.087	0.545	0.873
Italian + possessor thematicity	-0.241	0.547	0.659
Spanish + possessor thematicity	-0.128	0.43	0.766
Other L1s + possessor thematicity	-0.86	0.436	0.049

Table 6.17 shows that the main effects largely match with the proficiency-based model: possessors with final sibilants avoid the *s*-genitive, both long possessors and long possessums favor the *of*-genitive, and previous *of*-genitives produced by examiners discourage the use of *s*-genitives. The main effects for possessor animacy and definiteness indicate that all speakers prefer to use the *s*-genitive when the possessor is animate and definite. The proficiency-based model painted a more complex picture of these constraints as they interact with proficiency level. Specifically, B1 and

B2 learners are less sensitive to possessor animacy than native speakers. As for definiteness, native speakers are at first glance not impacted by this constraint, which I attributed to the frequent occurrence of two specific constructions in the language production of examiners. The fact that both constraints do not interact with L1 helps ascertain that the effects found in the proficiency-based model are not due to underlying L1s.

L1 appears to interact with possessor thematicity and type-token ratio, so the main effects for these two predictors only apply for the reference L1, namely the native speakers of English, and show that neither constraint influences native speakers' choice of variant. This partially contradicts the results from the proficiency-based model, which suggested that all speakers prefer the *s*-genitive in lexically dense environments with high type-token ratios. Instead, the L1 model reveals that this effect is mostly driven by Italian EFL learners. As for possessor thematicity, learners from the 'other' L1 backgrounds that had been conflated together are less likely to use the *s*-genitive than native speakers when the thematicity of the possessor is high. In other words, these learners prefer the *of*-genitive with thematic possessors. This is surprising because thematic possessors are highly frequent and therefore in principle more accessible, so they should induce *s*-genitives according to Easy First (MacDonald 2013) (see Section 6.2.3.4).

In general, it is interesting that learners' L1 does not play a more important role in the choice of genitive variant considering the importance L1 has received in previous studies comparing learners against native speakers (Gries & Wulff 2013; Kinne 2020). Provided transfer from the L1 is stronger at the early stages of acquisition (Morett & MacWhinney 2013: 134,149; Wang & Xu 2015: 215), one possible explanation for this lack of L1 effects is that a large portion of the data comes from the more advanced C levels, which could counteract any transfer effects occurring at the lower proficiency levels. To identify such transfer effects, one would ideally fit a model with three-way interactions between each probabilistic constraint, L1 and proficiency level, but this was not possible with the current data because proficiency level and L1 overlap perfectly for native speakers (see Section 6.2.6).

6.3.3.2 MuPDAR

A second series of supplementary analyses investigates the genitive data using the MuPDAR approach (Gries & Deshors 2014; Gries & Adelman 2014; Gries & Deshors 2020). The approach has been used to uncover differences between native speakers and learners regarding alternation phenomena

like the particle placement alternation (Wulff & Gries 2019), *can* vs. *may* (Gries & Deshors 2014) and optional *that*-complementation (Wulff, Gries & Lester 2018). Likewise, I use MuPDAR to compare learners with native speakers regarding the genitive alternation. With this in mind, the approach consists of three steps.

Step 1: Modelling the native speaker data

First, I fitted a model on the genitive observations from the native speakers only that tries to predict the likelihood of the *s*-genitive based on the genitive constraints. As recommended by Zuur et al. (2009), I started from a beyond optimal model including all probabilistic constraints and relevant random effects. Then, I removed the random effects that did not improve the model before turning to the fixed effects structure. The native speaker model was generated with the following formula:

```
Optimal native speaker model ← glmer(genitive variant ~ possessor animacy +
possessor length + possessum length + (1|possessor head lemma) +
(1|possessum head lemma), data = native speaker data, family = binomial,
glmerControl(optimizer = "bobyqa", maximum iterations = 1 million))
```

The concordance index *C* of the model is .95, indicating an excellent fit of the data (Baayen 2008), with no indication of multicollinearity. A summary of the model results is given in Table 6.18.

Table 6.18: Effects of the individual predictors on the choice of genitive variant in the native speaker data. Positive coefficients indicate an increase in the likelihood of the s-genitive. (Marginally) significant p-values are written in bold.

Predictor	Coefficient	Std. error	p-value
(Intercept)	8.721	1.611	<0.001
Possessor animacy Inanimate → animate	3.451	0.430	<0.001
Possessum length (logged)	-3.177	0.443	< 0.001
Possessor length (logged)	-1.702	0.475	<0.001

Results show that native speakers prefer to use the *s*-genitive when the possessor is animate and both constituents are short.

Step 2: Predicting learner observations

The second step is to use the native speaker model to predict the choice of variant in the learner data. Since this model is trained on native speaker data, this essentially shows which variant the native speaker would have produced in the learner observations. Take for instance the learner

observation in (180) featuring an *s*-genitive. Contrary to the Italian B1 learner who produced this utterance, the native speaker model only predicts a 0.2% chance for the *s*-genitive in this context, probably because the possessor *location* is inanimate and the possessum is very long. In other words, learners used an *s*-genitive where a native speaker would in all likelihood have produced an *of*-genitive.

(180) all tourist are responsible for protecting and respecting a location's economic environmental and social culture erm er er social culture (TLC:6_IT_26, B1 Italian)

Concretely, I introduced a binary variable indicating whether native speakers would have chosen the same variant as learners for each learner observation. Out of 1,773 learner observations, learners only choose the non-native like variant in 380 observations (21.4%).

Step 3: Predicting non-native choices

In the third step, this binary variable is used as the response variable in a second mixed-effects logistic regression model, which also uses the probabilistic constraints as predictors. This second model reveals which constraints cause learners to make non-native choices of variant, indicating that learners have not mastered these constraints yet. Since I am primarily interested in the acquisition and development of probabilistic grammars in EFL, I allowed these probabilistic constraints to interact with proficiency level. These interactions essentially show whether a given constraint is more likely to cause non-native choices for learners with a certain proficiency level compared to others, so these interactions would essentially reveal which proficiency levels have not yet mastered the relevant constraints.

To fit the second model on the influence of proficiency level, I log-transformed the length of the constituents and the thematicity of the possessor in the learner data, and then standardized all numerical variables following Gelman (2008). Due to data sparsity issues, the C1 and C2 data were conflated, animacy was entered as binary distinction between animate and inanimate possessors, and structural persistence was also simplified to a three-way distinction between a previous instance of an *s*-genitive, *of*-genitive, or no previous genitive, without identifying which interlocutor produced the variant. Since there are no native speakers in the data of this model, I chose the B1 learners as the new reference proficiency level. The model predictions are for non-native-like choices. The approach laid out in Zuur et al. (2009) was used to fit the model, so I started from a beyond

optimal model and determined the random effects structure before the fixed effects structure. The final model was generated with the following formula:

```
Optimal proficiency MuPDAR model ← glmer(agreement ~ proficiency +
  possessor animacy + semantic relation binary + possessor length + possessum
  length + possessor definiteness + type-token ratio + proficiency:possessor
  animacy + (1|possessor head lemma) + (1|possessum head lemma), data =
  learner genitive data, family = binomial, glmerControl(optimizer = "bobyqa",
  maximum iterations = 1 million))
```

The final model is significantly better than the null model ($p < 0.001$) and predicts whether learners and native speakers agree on the same variant in 81% of the cases, which is significantly better than the baseline of 78% ($p < 0.01$). The concordance index C is .82, indicating an excellent fit of the data (Baayen 2008). The condition index κ of 18.6 shows some degree of multicollinearity but remains below the threshold of 30 for harmful collinearity (Baayen 2008: 182). Finally, the VIF adjusted by degrees of freedom for each factor is also below the threshold of 10 (Field, Miles & Field 2012: 339; Levshina 2015: 273). The results from the model are summarized in Table 6.19.

Table 6.19: Main and interaction effects of the individual predictors for non-native choices in the learner data. Positive coefficients indicate an increase in the likelihood of a non-native choice. (Marginally) significant *p*-values are written in bold.

Predictor	Coefficient	Std. error	<i>p</i> -value
(Intercept)	-2.721	0.366	<0.001
Proficiency			
B1 → B2	-0.401	0.334	0.229
B1 → C	-0.656	0.335	0.050
Possessor animacy			
Inanimate → animate	1.057	0.339	0.002
Semantic relation			
Non-prototypical → prototypical	-0.772	0.204	<0.001
Possessum length (logged)			
0 → 1	-3.177	0.443	< 0.001
Possessor length (logged)			
0 → 1	-1.702	0.475	<0.001
Possessor definiteness			
Indefinite → definite	0.346	0.162	0.033
Type-token ratio			
0→1	-0.23	0.138	0.095
Proficiency x animacy			
B2 + animate	0.68	0.405	0.093
C + animate	0.925	0.394	0.019

The main effects of the model pertain to all learners. If the possessor and the possessum are long and the semantic relation between the possessor and the possessum is prototypical rather than non-prototypical, learners are more likely to make the same choice as native speakers. On the other hand, if the possessor is definite, then learners are less likely to pick the same variant as native speakers.

Animacy is part of a significant interaction term with proficiency level and animacy, so its main effect only applies to the reference proficiency level, viz. B1. Conversely, the main effect for proficiency level only pertains to inanimate possessors. The main effect for animacy shows that if the possessor is animate rather than inanimate, B1 learners are more likely to make a non-native choice of variant. However, animate possessors are even more likely to cause non-native choice for B2 learners (marginally significant) and C learners compared to B1 learners. Conversely, the main effect for proficiency indicates that when the possessor is inanimate, C level learners are less likely than B1 learners to choose the non-native genitive variant. All in all, this interaction effect ties in with the fact that B1 learners mostly struggle with inanimate possessors, but use the *s*-genitive as often as native speakers when the possessor is animate (see Section 6.3.2.2 and Figure 6.16 in particular).

In addition to this MuPDAR analysis focusing on how proficiency modulates the likelihood of non-native choices, I fitted a second MuPDAR model where L1 was allowed to interact with the constraints in order to reveal whether learners from certain L1 backgrounds are struggling with specific constraints. The only L1 interaction this analysis revealed is that Chinese EFL learners are more likely to make non-native choices when the possessor is long compared to other learners, but they are more similar to native speakers with short possessors. The main effects in both MuPDARs largely overlap, except for the fact that long possessors decrease non-native choices in the proficiency-based MuPDAR while they increase non-native choices in the L1-based MuPDAR. This should in principle not be possible since the main effects in both MuPDAR models essentially refer to the same learner data.

Some comments about MuPDAR

In principle, the MuPDAR approach captures the same thing as the mixed-effects logistic regression that compares native speakers and learners with proficiency interactions. That is, the MuPDAR approach identifies the constraints that result in learners making non-native-like choices, which implies that these constraints affect the choice of variant differently for native speakers and learners. By contrast, the proficiency interactions in the original model reveal whether and how constraints affect the choice of variant differently for learners across proficiency levels compared to native speakers, which could potentially cause non-native choices. Hence, one advantage of MuPDAR is that it highlights the constraints that result in meaningful differences between speaker populations, rather than focusing on how constraints operate differently across speaker populations regardless of whether this actually causes learners to deviate from native speakers (Gries & Deshors 2020: 70).

With this in mind, it is surprising that most significant constraints in the MuPDAR models operate in the same way across native speakers and learners according to the original model, including the type-token ratio, semantic relation and length of the constituents. In his study of the genitive alternation in ESL varieties, Heller (2018: 188) also found that MuPDAR detects more significant effects than mixed-effects logistic regression models with interactions, which led him to conclude that MuPDAR is sensitive to very small effects. This is tied to the fact that MuPDAR is “maybe too ‘divisive’ or ‘harsh’ around the cut-off point” (Gries & Deshors 2020: 91–92): if the native speaker model gives a prediction of 0.51 for the *s*-genitive, an *of*-genitive produced by a learner would be considered a non-native choice although 0.51 represents a very weak preference for the *s*-genitive,

to the extent that a native speaker would probably find an *of*-genitive equally natural. If both variants are more or less equally acceptable in certain contexts, speakers would alternate between the variants more freely, resulting in learners making more non-native choices.

A simple way to solve this issue would be to posit a middle range where both variants are equally acceptable (Gries & Deshors 2020), so that non-native choices only count when there is a considerable degree of deviation between native speakers and learners. As suggested by Gries et al. (2020: 93), I therefore considered that both genitive variants are acceptable in contexts for which the prediction from the native speaker model is between 0.4 and 0.6, in which case native speakers agree with learners by default.

After implementing this change, the number of non-native choices in the data decreased from 380 (21.4%) to 304 (17.1%), which caused two issues when I tried fitting both MuPDAR models again. For one, this further limited the number of regressors one can enter into the model (Hosmer & Lemeshow 2000: 346–347) and increased the risk of overfitting. Secondly, this improved the baseline of the model up to 83%, which neither MuPDAR model was able to surpass. This essentially means that the MuPDAR model including the genitive constraints does not do a better job at explaining why learners make non-native choices than a model that always assumes learners make the same choice as native speakers. To summarize, the analysis of the genitive alternation with the MuPDAR approach was rather unsuccessful because learners actually choose the same genitive variant as native speakers in the vast majority of cases.

6.4 Discussion

In this chapter, I compared how EFL learners at different proficiency levels choose between the alternating variants of the genitive alternation compared to native speakers of British English. To do so, I collected a balanced sample of 2,302 genitive observations from the TLC (Gablasova, Brezina & McEnery 2019), which consists of transcribed recordings of an official language exam. The genitive observations from the examiners in the corpus represent native speakers of British English. The candidates who take the exam are EFL learners from eleven different mother tongue backgrounds with a proficiency level ranging from B1 to C2 on the CEFR scale (Council of Europe 2001).

The collected data was annotated for the most relevant probabilistic constraints guiding native speakers' choice of genitive variant according to the literature, namely structural persistence, the lexical density of the

surrounding context, the animacy, definiteness, final sibilancy, discourse status, and thematicity of the possessor, the length of the possessor and the possessum as well as their semantic relation (Rosenbach 2014; Heller 2018). To analyze the data, I fitted a mixed-effects logistic regression with the choice of variant as response variable and the annotated constraints as predictors. Proficiency was allowed to interact with each of the probabilistic constraints so that the model could detect whether and how the constraints influence the choice of variant differently for native speakers and learners across proficiency levels.

In this discussion section, I first summarize the findings from the regression model in relation to the existing literature. Then, I address each of the research questions separately and assess the validity of the hypotheses in view of the findings. Finally, I combine these insights to provide an account of the acquisition of the genitive probabilistic grammar in EFL.

6.4.1 Summary of the findings

6.4.1.1 Native speakers

The results show that native speakers are sensitive to structural persistence, the animacy and final sibilancy of the possessor, the length of the possessor and the possessum and the lexical density of the surrounding context. In short, native speakers prefer to use the *s*-genitive when both constituents are short, the possessor is animate and does not end in a final sibilant, an examiner did not previously use an *of*-genitive and the context surrounding the genitive observation is lexically dense. A prototypical instance of an *s*-genitive for native speakers is shown in (181).

(181) what other resources can we use to to improve women's status in society (TLC:2_IN_19, NS)

By extension, a prototypical *of*-genitive for native speakers has an inanimate possessor ending in a final sibilant,⁶ a long possessor and possessum, is preceded by an *of*-genitive produced by an examiner and occurs in a context that is not lexically dense, as exemplified by (182).

⁶ If constraints are interpreted as construction features, it is not necessarily the case that they pertain to both variant constructions conceptually. Hence, I argued in Section 3.1 that final sibilants are not necessarily a prototypical feature of *of*-genitives simply because they avoid *s*-genitive.

(182) I want you to summarise the main what you think are the main benefits of modern communications (TLC:2_CH_2, NS)

Most of these effects align with results from previous studies: animate possessors favor the *s*-genitive, final sibilants avoid the *s*-genitive and structural persistence causes variants to be reused. As for the effect of type-token ratio, it makes sense to use the *s*-genitive when the context surrounding the genitive variant is lexically dense in written language to save space, especially in press language (Hinrichs & Szmrecsanyi 2007: 457; Szmrecsanyi & Hinrichs 2008: 300–301), but it is somewhat unclear how this explanation pertains to spoken language where one would not expect editorial pressure to play a role but rather processing factors such as Rohdenburg's Complexity Principle (1996: 151): lexically dense environments are more demanding cognitively, which should drive speakers to use the more explicit *of*-genitive (see Section 6.2.3.5, Dubois 2023). While this is surprising, Szmrecsanyi's (2006: 97) analysis of the genitive alternation in spoken British English uncovered the same pattern.

The effect direction of possessum length is unexpected as it runs counter to the short-before-long principle: long possessums favor the *of*-genitive in the same way as long possessors do. Again, the same effect was found by Szmrecsanyi (2006). This effect can be explained by Rohdenburg's Complexity Principle (1996: 151): when the possessor and the possessum are long, the cognitive environment is highly complex, so speakers prefer to use the more explicit *of*-genitive.

A last somewhat interesting effect for the native speakers concerns structural persistence. The results showed that native speakers are only influenced by persistence effects when they previously produced an *of*-genitive themselves, in which case they tend to reuse the *of*-genitive as expected. This implies that previous *s*-genitives do not lead to persistence effects, although the *s*-genitive is less frequent and should therefore induce stronger persistence effects than the *of*-genitive (Jaeger & Snider 2013). Moreover, it seems that examiners are not affected by persistence from either variant when it is produced by a learner.

The thematicity and discourse status of the possessor as well as the semantic relation of the possessor and the possessum were not retained in the model, so these constraints do not condition the choice of genitive variant for neither native speakers nor learners in the present data. Regarding the discourse status of the possessor, the constraint is generally rather weak in the genitive alternation so its effect is not consistently significant (Kreyer 2003: 172; Hinrichs & Szmrecsanyi 2007: 461). As for the thematicity of the possessor, its effect is mostly restricted to written

language as a result of economization (see Grafmiller 2014). However, as discussed in Section 6.2.4, possessor thematicity is operationalized in such a way that it captures lexical priming effects in spoken language. One possible reason why thematicity does not influence the outcome is that the constraint takes into account linguistic material too far removed from the choice context to exert an influence, considering that lexical priming effects are short-lived compared to syntactic priming (Hartsuiker et al. 2008). For example, it is unlikely that the lemma of a possessor mentioned many times at the beginning of the exam would induce priming effects in a genitive choice context 20 minutes later, even though its thematicity is high. In principle, a similar explanation can apply to the discourse status of the possessor as well, which is also operationalized on the occurrence of lexical items in a smaller chunk of the previous context. Possessor definiteness is part of an interaction term and is therefore discussed in the next section.

6.4.1.2 Differences between native speakers and EFL learners

Learners deviate from native speakers with respect to only two constraints, namely the animacy of the possessor and its definiteness. More specifically, it seems that B1 and B2 speakers are less sensitive to the animacy constraint than native speakers. A visualization of the interaction effect suggests that B1 learners overuse the *s*-genitive when the possessor is inanimate, whereas B2 speakers tend to use the *of*-genitive with animate possessors (see Section 6.3.2.2). While animacy is weaker, the effect direction is the same: animate possessors prefer the *s*-genitive. This finding represents yet another instance of possessor animacy being a strong yet unstable constraint in the literature (e.g., Hundt & Szmrecsanyi 2012; Wolk et al. 2013; Heller, Szmrecsanyi & Grafmiller 2017; Rosenbach 2017). Within the present context however, this result is surprising because possessor animacy should be highly learnable.

B2 learners and to a lesser extent C learners also diverge from native speakers with respect to possessor definiteness. Native speakers are not sensitive to this constraint, whereas B2 learners follow the expected pattern whereby definite possessors prefer the *s*-genitive. The same tendency is marginally significant for C learners. Hence, the prototypical *s*-genitive for B2 or C learners and native speakers are highly similar, except that the possessor is preferably definite, as in (183).

(183) I'm worried yeah about India's future (TLC:2_8_IN_31, B2 Gujarati)

However, a closer look at the dataset revealed that native speakers are probably not sensitive to the definiteness of the possessor according to the

model because of two constructions they frequently use as part of their role as examiners that go against the expected tendency for definite possessors to figure in *s*-genitives. If native speakers are indeed sensitive to possessor definiteness if not for these constructions, as suggested by a supplementary analysis (see Section 6.3.2.2), then only B1 learners are not sensitive to the definiteness of the possessor.

Although structural persistence is not part of an interaction term, its main effect implies differences between native speakers and learners because of the way the constraint is operationalized. As stated above, native speakers are more likely to use the *of*-genitive if they previously used the variant. According to the regression model, the exact same effect applies to learners, so they are also more likely to use the *of*-genitive when an examiner produced an *of*-genitive. The difference lies in the fact that examiners consider their own production of *of*-genitives, whereas learners are influenced by the production of their interlocutor, the examiner. This is reminiscent of the fact that B1 and B2 learners are affected more strongly by previous instances of *be going to* produced by examiners than their own.

There are no differences between native speakers and learners regarding the remaining constraints, viz. the final sibilancy of the possessor, the lexical density of the surrounding context and the length of the constituents. Even for the probabilistic constraints where native speakers and learners exhibit dissimilarities, the direction of the effect remains the same. Hence, the study generally echoes the conclusion of Gries and Wulff (2013) that “intermediate/advanced learners’ production data already reflect target-like processing preferences at an astonishing level of complexity” (Gries & Wulff 2013: 349).

The study by Gries and Wulff (2013) revealed no difference between the native speakers and the Spanish or German EFL learners with respect to either constraint.⁷ Since even the native speakers in their data were not sensitive to possessor definiteness, it is no surprise that they did not find a difference between native speakers and learners, who are not sensitive to the constraint at the B1 level. The effect of definiteness may not have been found for native speakers in their analysis due to the fact that the constraint is less influential in written language compared to spoken language, or because it is rather weak in general (Heller 2018: 104). As far as possessor animacy is concerned, which significantly affects both native speakers’ and learners’ choice of genitive variant in their study, it is much more likely that

⁷ Strictly speaking, Gries and Wulff (2013) did not annotate their observations for possessor definiteness, but their operationalization of possessor specificity is similar.

their model was not able to detect a similar interaction because they conflated data from low-proficiency learners with that of more advanced learners who might already master the animacy constraint (see Section 2.1.2.2).

6.4.2 How does the set of constraints of learners compare to that of native speakers at different proficiency levels?

Hypothesis 1 *As proficiency increases, learners better approximate the target probabilistic grammar.*

Rationale: This hypothesis derives from the fact that language learning of any kind is driven by input according to usage-based approaches to SLA (e.g., Ellis & Wulff 2019). Learners who have reached a higher proficiency level have received more input than learners with a lower proficiency level, all other things being equal. By extension, more proficient learners have been exposed to more *s*-genitive and *of*-genitive exemplars whose features, viz. the linguistic elements linked to the probabilistic constraints, can be acquired through associative statistical learning. If the probabilistic grammar of a learner does not differ from that of native speakers with regard to any of the relevant constraints, then the learner approximates the target probabilistic grammar of native speakers. Hence, lower-proficiency learners should deviate from native speakers with regard to more constraints than more proficient learners.

The model revealed that B1 and B2 are less sensitive to possessor animacy than native speakers. At the same time, native speakers do not pay attention to the definiteness of the possessor, contrary to B2 and C learners, who prefer to use the *s*-genitive when the possessor is definite, in accordance with the literature (Wolk et al. 2013; Heller 2018: 19). Hence, B1 learners struggle with one constraint, B2 learners with two constraints and C learners with one constraint. The fact that B1 learners better approximate the target probabilistic grammar than the more proficient B2 learners does not align very well with the hypothesis.

However, a closer look at the data suggests that the model was probably unable to detect the effect of possessor definiteness for native speakers because of two specific structures they frequently use within their role as examiners to organize discourse. In these structures, specific possessors figure in *of*-genitives and the indefinite possessor *people* almost

exclusively takes the *s*-genitive, so they counteracted the tendency for definite possessors to favor *s*-genitives. A supplementary analysis without these two constructions showed that like B2 and C learners, native speakers prefer the *s*-genitive when the possessor is definite, as expected (see Section 6.3.2.2).

If native speakers are indeed influenced by possessor definiteness, then only the B1 learners do not take this constraint into consideration. Hence, they are less sensitive to both possessor definiteness and possessor animacy compared to native speakers. Like B1 learners, B2 learners deviate from native speakers regarding the animacy of the possessor, but they have implemented possessor definiteness. At the C level, learners no longer deviate from the native speakers regarding possessor animacy either. In summary, B1 learners diverge from native speakers regarding two constraints, B2 learners differ only with respect to possessor animacy, and C learners do not deviate from native speakers at all. Hence, the learners better approximate the target probabilistic grammar as they become more proficient, in accordance with the hypothesis.

Hypothesis 2 *Learners are only sensitive to constraints that are part of the target probabilistic grammar.*

Rationale: According to usage-based approaches to SLA such as the CREED model, language learning is rational (Ellis 2006: 102–103), meaning that learners will not infer rules that go against the input they receive. In this case, learners should only learn the features associated with the probabilistic constraints that are present in the input from native speakers of British English like the examiners in the corpus. This does not apply to processing-related constraints as they are not learned, namely structural persistence, constituent length, possessor thematicity and discourse status.

At first glance, this hypothesis is false given that native speakers do not take into account the definiteness of the possessor, contrary to B2 and C learners. However, I argued that native speakers may very well prefer the *s*-genitive when the possessor is definite, but the model could not detect this effect because of specific structures in the examiner data counteracting this tendency. If native speakers are indeed sensitive to possessor definiteness, then there are no constraints that only affect learners, so this hypothesis is not contradicted by the data.

Hypothesis 3 *Lower-proficiency learners are sensitive to fewer constraints than more advanced learners and native speakers.*

Rationale: If language learning is rational, the maximum number of constraints learners can acquire are those constraints present in the target language input they receive. At the same time, previous research on cues showed that cues are learned one by one: learners adjust a given cue before turning to the acquisition of the next cue (Ellis & Sagarra 2010; Ellis & Sagarra 2011: 601). If construction features are similar to cues, learners should also acquire the constraints one by one. As a result, the probabilistic grammar of lower-proficiency learners should consist of fewer constraints compared to more advanced learners and native speakers, which means their probabilistic grammar is simpler (Shin 2014).

Since B2 and C learners pay attention to the definiteness of the possessor, they technically rely on more constraints than native speakers, for whom definiteness is a priori irrelevant. As mentioned before however, the examiners are in all likelihood not impacted by possessor definiteness according to the model because of two specific structures they frequently use to organize discourse during the exam. As indicated by a supplementary analysis, native speakers could very well prefer the s-genitive for definite possessors if not for these two structures. In this case, only B1 learners ignore possessor definiteness, which supports the hypothesis that less proficient learners might not yet have learned all relevant probabilistic constraints.

6.4.3 How does proficiency level modulate the probabilistic grammar of learners?

Hypothesis *Learners first acquire constraints that facilitate associative statistical learning.*

Rationale: Previous research showed that more frequent, reliable, regular cues that are adjacent to the construction are more learnable than infrequent, unreliable, irregular, non-adjacent cues (MacWhinney 2001; Romberg & Saffran 2010: 2–3; Yoshimura & MacWhinney 2010: 552; Misyak & Christiansen 2012: 304; Yi, Lu & Ma 2017: 520). Provided construction features are similar to cues, the same factors should determine the learnability of

feature-based constraints. Since these learnability factors are difficult to rank, I operationalized the learnability of a given constraint as its association strength with the variants in the language production of native speakers based on monofactorial tests. This showed that possessor animacy is the most learnable constraint, followed by semantic relation, which in turn should be more learnable than the remaining feature-based constraints (see Section 6.2.4.1).

The data contradicts this hypothesis in multiple ways. Most importantly, possessor animacy should be the easiest feature-based constraint to learn since it is most strongly associated with the genitive variants. In reality, animacy appears to be the most challenging constraint for learners considering that they deviate from native speakers until they reach the C level. By comparison, the second constraint learners struggle with, possessor definiteness, is already mastered at the B2 level. Like possessor definiteness, final sibilancy should cause more difficulties for learners than possessor animacy, yet even B1 learners have acquired the constraint, so final sibilancy is the feature-based constraint learners grasp first.

Like for the future marker alternation, I argue that association strength as a proxy for cue reliability probably cannot adequately capture the learnability of the constraints. This is evident from the fact semantic relation is supposedly the second-most learnable constraint for native speakers, yet it does not even significantly contribute to the choice of variant according to the model. On the one hand, cue reliability is only one aspect of statistical learnability. On the other hand, monofactorial tests neglect the inherently multifactorial nature of alternations, so the association strength might not even provide an accurate representation of cue reliability.

Hence, a more detailed assessment of statistical learnability based on more criteria besides reliability, such as frequency, variability and adjacency (e.g., MacWhinney 2001; Yoshimura & MacWhinney 2010: 552; Misyak & Christiansen 2012: 304) might better account for the results. For possessor definiteness, the main issue is indeed probably that the constraint is rather weak and therefore unreliable: whether the possessor is definite or not does not give a strong indication as to which variant is preferred. This can make it more difficult for low-proficiency learners to notice the association between definite possessors and either variant (O'Shannessy & Meakins 2012: 380; Gries & Ellis 2015; Ellis & Wulff 2019: 44).

The reason why possessor animacy is more challenging for learners than other feature-based constraints is that it involves much more variability. The linguistic elements associated with final sibilancy are easy to recognize because they represent a few sounds. Likewise, the definiteness of the possessor is encoded by a definite determiner in almost half of the cases (179 out of the 328) in the native speaker data. By contrast, the animacy of the possessor cannot be derived from any recurrent structural patterns. Instead, learners have to figure out the animacy of each noun phrase based on the concept they convey. This in turn, implies that the acquisition of the animacy constraint correlates with the lexical knowledge of learners. Moreover, the animacy of the noun phrase is context-dependent, so learners have to encounter the nouns in various contexts sufficiently often to assign the right animacy category. This raises the question as to whether nouns from certain animacy categories are more problematic to classify than others.

Although the animacy constraint was entered in the model as a binary distinction between animate and inanimate possessors, the genitive variants pattern distinctly with at least five categories, namely animate, collective, inanimate, locative and temporal possessors for native speakers (e.g., Wolk et al. 2013; Heller 2018) (see Section 6.2.3.6). The categories that should be most challenging for learners at first glance are the so-called “borderline cases” (Yamamoto 1999: 139–146), i.e., locatives and collectives, because they can sometimes be hard to distinguish. In the context of example (184) for instance, it is unclear whether *South Africa* is used as a locative or as a collective. Other examples include possessors like *province, country, region, village, India* etc.

(184) In May nineteen ninety four for the first time in South Africa's history all the races voted in democratic election (TLC:2_6_IT_100, B1 Italian)

Moreover, locative and collective possessors are less distinctly associated with the genitive variants compared to animate and inanimate possessors according to Table 6.7. As a result, they represent less reliable cues that are more difficult to learn via associative statistical learning (see Yoshimura & MacWhinney 2010; MacWhinney 2014; Gries & Ellis 2015 on cue-outcome contingency).

Confusing collective possessors with locative possessors is problematic because they do not associate with the same genitive variant: collective possessors take the *s*-genitive whereas locatives prefer the *of*-genitive. Hence, learners would produce *s*-genitives with locative

possessors and *of*-genitives with collective possessors. Since collective possessors were conflated with animate possessors in the ‘animate’ category and locative possessors were combined with inanimate and temporal possessors in the ‘inanimate’ category, the model would conclude that learners are less sensitive to possessor animacy. An instance of such an *s*-genitive with a locative possessor is given in (178), copied below as (185).

(185) Examiner: I’m from England I’m from London in England have you been to London?

Candidate: No but I I see the London’s pictures

(TLC:2_6_CH_23, B1 Chinese)

However, the current analysis cannot confirm this theory because I cannot investigate the five animacy levels underlying the binary constraint entered in the model. Therefore, one of the main goals of the rating task experiment in the next chapter is to provide more insight regarding this puzzling finding. Interestingly, Heller (2018) found that ESL speakers are also less sensitive to the animacy constraint, which raises the question as to whether these speakers face similar difficulties when learning the animacy constraint (see Biewer 2011; Deshors 2014).

6.4.4 How do native speakers and learners at different proficiency levels compare regarding processing-related constraints?

Hypothesis *Lower-proficiency learners are more sensitive to processing-related constraints than more advanced learners and native speakers.*

Rationale: Since language learners constantly have to juggle between two or more language systems (Kroll & Dussias 2013), and because language production is less automatic for especially low-proficiency learners compared to native speakers (Segalowitz 2003), processing-related constraints should be stronger for learners (see e.g., Wulff, Lester & Martinez-Garcia 2014; Deshors & Gries 2014; Wulff, Gries & Lester 2018; Paquot, Grafmiller & Szmrecsanyi 2019). Given that virtually all constraints can be tied to processing, I argued that the characteristics of psychophysiological constraints as defined by Tamminga, MacKenzie and Embick (2016) can help identify strictly processing-related constraints resulting from features of the

brain (see Section 3.2.1), in this case structural persistence, constituent length, possessor thematicity and discourse status.

The thematicity and discourse status of the possessor are not relevant for neither native speakers nor EFL learners. As mentioned in Section 6.4.1.1, this could be due to the fact that the type of lexical priming effects these constraints capture indirectly in spoken language are rather short-lived compared to syntactic priming (Hartsuiker et al. 2008), so the operationalization for both constraints possibly looks too far into the preceding context for lexical priming to occur.

Native speakers and learners exhibit no dissimilarities regarding the length of the constituents. That is, both prefer the *of*-genitive when both constituents are long, in line with Rohdenburg's Complexity Principle (Rohdenburg 1996: 151). As for structural persistence, both native speakers and EFL learners avoid the *s*-genitive when an examiner previously used an *of*-genitive. While this means that native speakers are only sensitive to their own previous *of*-genitives and that EFL learners only repeat the *of*-genitives of their interlocutor, this does not suggest that persistence has a more pronounced impact on learners.

In short, learners are affected by the same processing-related constraints as native speakers, but there is no evidence that these constraints exert more influence on the choice of genitive variant for learners. With that being said, the probabilistic grammar of learners is highly similar to that of native speakers even at the B1 level, which raises the question as to whether processing-related effects were more prominent at earlier stages of language learning.

6.4.5 Synopsis

Overall, the results showed that learners are already strikingly similar to native speakers even at the B1 level, which echoes the conclusion of Gries and Wulff regarding the genitive alternation in the written language of EFL learners (Gries & Wulff 2013: 350). This highlights a major shortcoming of the study, namely that I do not have access to lower proficiency learners at the A levels (Council of Europe 2001) who might still be fine-tuning constraints like final sibilancy. Nonetheless, I will offer an overview of the development of the genitive probabilistic grammar, which largely corresponds to the expectations of the theory-based account as evidenced by the fact that the results align with most hypotheses.

EFL learners first encounter both genitive variants through concrete exemplars in their language input. Supposedly, they should originally be

influenced mostly by processing-related constraints like structural persistence and the length of the constituents. At the same time, processing principles like Easy First (MacDonald 2013) encourage learners to produce more accessible referents first. Hence, if the possessor is animate and definite and therefore accessible, learners will prefer the *s*-genitive because the possessor is produced right away. However, these processing effects alone do not suffice to approximate the language use of native speakers, given that the concrete realization of possessor animacy and definiteness is language-specific and must therefore be fine-tuned for the English genitive alternation (see Section 3.1). To refine possessor animacy and definiteness with the goal to match native speakers, learners have to infer how the types of possessors are associated with the variants as features of the constructions through associative statistical learning (see e.g., Romberg & Saffran 2010; Rebuschat & Williams 2011b; Misyak & Christiansen 2012). This mechanism is also responsible for the acquisition of the final sibilancy constraint.

Provided the construction features associated to probabilistic constraints are similar to cues, they are acquired and adjusted one by one (Ellis & Sagarra 2010; Ellis & Sagarra 2011). At the B1 level, the learners have noticed that final sibilants avoid the *s*-genitive. They have also learned that animate possessors prefer the *s*-genitive, although they have difficulty mastering this constraint because it relates to lexical knowledge and requires large amounts of input to properly classify the nouns. However, they have not yet picked up on the fact that definite possessors prefer the *s*-genitive, probably because the constraint is rather weak. As learners become more proficient and reach the C level, they have adjusted all constraints and their choice of genitive variant is native-like.

Throughout the learning process, language learning is gradual and rational (Ellis 2006: 102–103), meaning that learners infer the patterns of language use from the input they receive in the target language. This is evidenced by the fact that all constraints learners take into account also form part of native speakers' set of constraints.

Contrary to the future markers, the use of the *s*-genitive is barely discussed in the EFL textbooks I consulted (see Section 5.4.6), so it is unlikely that prescriptivism is at the origin of the patterns in the learner data. In fact, reference grammars recommend using the *s*-genitives with animate possessors categorically (Murphy 2012; Heller, Szmrecsanyi & Grafmiller 2017: 10), which suggests that learners should be more sensitive to possessor animacy than native speakers, rather than less sensitive as in the present analysis.

7 Rating task experiment

7.1 Introduction

In this chapter, I present the results of a rating task experiment meant to complement the corpus-based investigation of the genitive alternation reported in the previous chapter. The design of the rating task experiment is based on Bresnan and Ford (2010) and more recently Engel et al. (2022) and Engel and Szmrecsanyi (2023). Concretely, 25 native speakers and 101 EFL learners with various proficiency levels and L1 backgrounds were asked to rate the acceptability of the genitive variants in authentic corpus excerpts extracted from the TLC (Gablasova, Brezina & McEnery 2019) for the corpus-based investigation of the genitive alternation. Using a slider bar, participants could indicate which of the two genitive variants they preferred in the excerpt. The position of the slider bar corresponds to a numerical value, so the participants were essentially distributing points to both variants by moving the slider.

Conducting this follow-up experiment has three main aims tied to specific research questions. First, experiments can shed more light on patterns discovered in the corpus-based analysis (Schönefeld 2011: 24). In the present case, the rating task experiment was designed to replicate the main finding from the corpus-based analysis of the genitive alternation, namely that B1 and B2 learners are less sensitive to possessor animacy than native speakers. This finding was surprising because possessor animacy is normally the strongest constraint of the genitive alternation (Rosenbach 2014), so one would expect learners to be more sensitive to the constraint if anything, as suggested by prescriptivists (see Murphy 2012; cited in Heller, Szmrecsanyi & Grafmiller 2017: 10). Additionally, the constraint should be very reliable and therefore a priori highly learnable (see Yoshimura & MacWhinney 2010; MacWhinney 2014; Gries & Ellis 2015), yet the analysis revealed that learners struggle with the constraint until they are advanced.

What makes the constraint so challenging to master is the fact that animacy is encoded by the noun itself rather than a recurrent linguistic element, so learners have to properly classify the animacy of thousands of different nouns in order to approximate native speakers. This raised the

question as to whether nouns from certain animacy categories are more problematic than others. In this respect, I consider collective and locative categories as the main suspects for two reasons. For one, these animacy categories are less distinctly associated with the genitive variants, so they are less reliable features of the variant constructions, which makes them more difficult to learn (e.g., Gries & Ellis 2015). Secondly, nouns like *Italy*, *province*, *city* etc. can refer to either a location or a group of people depending on the context (see (184)), so learners might misclassify these nouns. Since locatives tend to take the *of*-genitive whereas collectives prefer the *s*-genitive, this would lead to non-native choices of genitive variant.

However, this theory could not be confirmed because collective possessors were conflated with animate possessors and locative possessors were combined with inanimate and temporal possessors in the corpus-based regression model due to data sparsity issues. Hence, I further investigated this finding with the rating task experiment without conflating animacy levels to answer the following research question:

RQ 1: How does possessor animacy affect the intuitions of learners at different proficiency levels in a rating task experiment compared to correlations between animacy and genitive variant in spoken EFL?

Given the results from the corpus-based investigation, I expect that both native speakers and EFL learners prefer the *s*-genitive with animate and collective possessors, whereas they prefer the *of*-genitive with inanimate, locative and temporal possessors. However, B1 and B2 learners should find the *s*-genitive more natural with locative possessors compared to native speakers, and they should find the *of*-genitive more natural with collective possessors. Methodologically, this research question was explored with a mixed-effects linear regression model that focused on the correlation between the ratings of the participants and the animacy of the possessor in the experimental items.

The second aim of the rating task experiment was to cater to some of the conceptual limitations of the corpus-based regression model by studying the genitive alternation from a complementary perspective. For one, corpus data cannot give insight into the process whereby speakers assess the suitability of eligible variant constructions in a choice context. Instead, corpus data only contains a record of the variant that was produced as the result of this process. By comparison, the slider bar in the rating task experiment essentially offers the participants a visual means to distribute

many points across both variants, which provides a more nuanced look at the competition between the variants.

More importantly, corpus-based regression models cannot provide a direct reflection of speakers' implicit grammatical knowledge regarding the variant constructions. Rather, corpus-based models primarily capture patterns of language use because the results of the models are determined by the frequency of the variants in a given context. While usage-based approaches certainly assume a strong correlation between language use and implicit grammatical knowledge, the correlation is not perfect because a variant that is not often used in a given context is not necessarily deemed unacceptable (Kempen & Harbusch 2005: 344; Arppe & Järvikivi 2007: 149–150; Divjak, Dąbrowska & Arppe 2016: 26–27). In other words, corpus-based models “reflect the characteristics of mental processes and structures yielding usage, even though we do not know the exact form of these mental representations” (Divjak & Arppe 2013: 230), or in more simple terms, corpus-based models show what speakers say, but not necessarily what they know (Ellis 2017: 46; see also Schönefeld 2011: 13; Kinne 2020: 173). Whereas experimental data is typically less valid ecologically than corpus data, rating task experiments do not concern language production and instead tap into the intuitions of speakers regarding the acceptability of the structures (Arppe & Järvikivi 2007: 132; Gilquin & Gries 2009; Klavan & Divjak 2016: 361), which arguably provides a better representation of speakers' implicit knowledge of the probabilistic constraints.

Given the complementary nature of corpus data and experimental data, the second research question is as follows:

RQ 2: How does the relationship between experimental ratings and corpus data vary by the proficiency level of participants and speakers?

I hypothesize that the intuitions of the participants correlate with the language use of the speakers in the corpus. This would provide additional evidence that significant effects reported by the corpus-based model indeed derive from the implicit grammatical knowledge of speakers as suggested by Bresnan and Ford (2010) (see also Klavan & Divjak 2016), rather than simply capturing frequent patterns of language use.

On the methodological plane, the experimental items stem from the 2,302 genitive observations collected from the TLC (Gablasova, Brezina & McEnery 2019). This way, it was possible to directly compare the intuitions of the participants reflected by their acceptability ratings with the

predictions from the corpus-based regression model for the same experimental items. That is, if the corpus-based model predicts that an *s*-genitive is highly likely, then the acceptability ratings for the *s*-genitive should be high as well. The correlation between the rating data and the predictions from the corpus-based model was analyzed through a second mixed-effects linear regression model.

Finally, I used the rating data from the experiment to investigate the individual behavior of speakers. In principle, both linear regression models I used to analyze the rating data generalize across the ratings of individual speakers. Likewise, the set of significant constraints discovered by the corpus-based model cover all learners in the data, although not every learner is necessarily sensitive to each constraint. To illustrate, say a model reports significant effects for constraints A, B, C, D and E. It is possible that the probabilistic grammar of one learner consists of constraints A, B, C and D, whereas another learner only pays attention to constraints B, C, D and E (see Divjak et al. 2016: 28). In this regard, Kinne's (2020: 128–138) found that end-weight almost categorically determined particle placement in phrasal verbs for certain learners, whereas for other learners the constraint was not relevant at all. In the present context, the question arises as to whether all B1 and B2 learners are less sensitive to possessor animacy as a group, or whether this pattern is only representative of the strategy of several individual learners within these proficiency levels. Such a question is especially relevant in the context of learner language considering that learners exhibit more individual variability than native speakers (Jäschke & Plag 2016: 511–512; Kinne 2020: 128).

Unfortunately, the corpus data is not suitable to explore individual differences between speakers, mainly because each speaker contributes only a limited number of observations: out of the 514 speakers in the corpus data, only 157 (31%) produce 5 or more genitive constructions. Moreover, these genitives are probably not distributed across a wide range of contexts, so they would not offer a comprehensive picture of a learner's behavior. By contrast, the rating task experiment fits both of these criteria. For one, each participant contributed a larger number of ratings. Because the experiment was designed to test the correlation between the ratings and the animacy of the possessor as well as the overall predictions of the corpus-based model, the experiment consisted of a balanced selection covering a wide range of authentic choice contexts. Hence, I conducted an analysis of the individual variation between speakers with the following research question in mind:

RQ 3: How do the ratings of individual participants compare to the aggregated results regarding possessor animacy and the relation between intuitions and language use?

This question was explored mostly through monofactorial statistics and by-participant visualizations of the ratings.⁸

The remainder of this chapter is structured as follows. In the next section, I present the data and methodology of the experiment, including the selection of materials, the recruitment of participants, the design and procedure as well as the method of statistical analysis. The results are presented in Section 7.3 and discussed in light of the three research questions in Section 7.4.

7.2 Data and Methods

Participants were shown an original corpus excerpt with both possible genitive variants as in Figure 7.1. Using the slider bar, participants could indicate which variant they preferred. The position of the slider represented a rating from 0 to 100 for the variant on the right side of the slider (*the film's name* in Figure 7.1). Higher scores indicated that the variant on the right side was more natural, with a score of 100 indicating that this was the only acceptable variant in this context. Inversely, a score of 0 indicated that this variant was not acceptable at all.

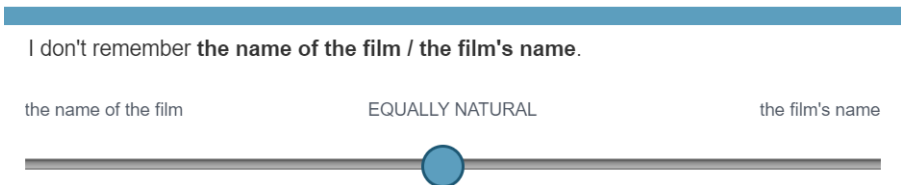


Figure 7.1: An example of a target experimental item presented to participants.

In order to shed more light on the animacy constraint, the rating task experiment was designed to test the effect on the ratings of all five animacy levels for which the genitive data was annotated based on Zaenen et al. (2004; see also Wolk et al. 2013), namely animate, collective, inanimate, locative and temporal possessors (see Section 6.2.3.6). In short, animate possessors mostly concern humans or entities represented as humans and animals (e.g., *god's plan*). Collective possessors cover collective nouns like *government* or *company* and groups of animates who display an identity as

⁸All data and code are available in an online repository at <https://osf.io/6ng7x/>.

a group (e.g., *the values of the parents*). Locatives refer to locations that are fit for humans, as in *the south of France*. Possessors that were marked as temporal refer to a time period or a specific point in time (e.g., *next week*).

Importantly, the experimental items stem from the 2,302 genitive observations I collected from the TLC (Gablasova, Brezina & McEnery 2019) for the corpus-based analysis of the genitive. This is relevant for the second aim of the experiment, which is to compare the predictions from the corpus-based model reflecting the patterns of language use of the speakers in the corpus with the ratings of the participants in the experiment, which capture their intuitions regarding the naturalness of the variants.

7.2.1 Materials

The rating task experiment consists of filler items and target experimental items as in Figure 7.1.

7.2.1.1 Target experimental items

The set of target experimental items was selected from the genitive observations collected from the TLC on the basis of a variety of criteria. This includes not only native speaker observations but also learner observations, which is justified as long as there is no indication in the context or through ungrammaticalities that the speaker is an EFL learner, as this could otherwise affect the perceived naturalness of the experimental items. From a methodological standpoint, including learner data makes it easier to find experimental items to suit the goals of the experiment, since the data mostly consists of learner observations.

Controlling for confounding variables

Since I am interested in the animacy constraint specifically, it was necessary to control for any other effects that could potentially influence the ratings. According to the corpus-based model presented in Section 6.3.2, the choice of genitive variant is also influenced by structural persistence, possessor definiteness and final sibilancy, possessor and possessum length, the lexical density of the surrounding context as well as lexical effects.

To account for these confounding variables, I only selected observations that had definite possessors and no final sibilant, mostly because the combination of these variables accounted for the most observations. For structural persistence, the corpus-based model revealed that the *s*-genitive is less likely if an examiner produced an *of*-genitive, so I did not consider observations preceded by examiner *of*-genitives.

To control for the influence of the lexical effects, the genitive observations could only feature possessor and possessum heads that

exhibited relatively minimal bias for either variant. This meant more concretely that I only included observations for which the possessor and possessum head lemma had an intercept adjustment in the corpus-based model that did not exceed 1 or -1 (see Section 6.3.1).⁹

Controlling for the length of the constituents is more challenging because it is unlikely one can find enough observations of similar lengths, not to mention the effect of constituent length is harder to interpret in the corpus-based model because it was standardized. I attempted to control for their effect on the genitive variants in various ways, but ultimately these methods would have restricted the pool of potential experimental items considerably. Therefore, the constraint was controlled for during statistical analysis by entering it as a fixed effect in the mixed-effects linear regressions models used to analyze the ratings.

In the corpus-based model, the lexical density of the surrounding context was operationalized as the type-token ratio in the 100 words surrounding the genitive observation (see Section 6.2.3.5). Including this much context would make the experiment rather long (Dörnyei & Taguchi 2009: 71) and possibly disadvantage low-proficiency learners, so the surrounding context was reduced for the experimental items. As a consequence, the predictions from the corpus-based model would no longer be completely accurate for the corresponding corpus observation. This was nonetheless important for the second research question on the correlation between the predictions from the corpus-based model and the ratings. Therefore, I fitted a new mixed-effects logistic regression model on the genitive data starting from the beyond optimal model described in Section 6.2.6, except that I did not include type-token ratio to provide more accurate predictions for the experimental items without as much context. The results from this reduced model remain almost exactly the same as in the original corpus-based model. The model has an R^2_{marginal} of .37, an $R^2_{\text{conditional}}$ of .75 and a C -value of .95, which indicated an excellent fit of the data (Baayen 2008). Like the original corpus-based model, this updated model without type-token ratio predicted the likelihood of the *s*-genitive on a probability range from 0% to 100% for the 2,302 genitive observations collected from the TLC.

Additionally, I did not consider corpus observations where the genitive variant does not match the prediction of the corpus-based model. The model might have predicted the alternative genitive variant because it

⁹ I experimented with lower thresholds for the intercept adjustments, but this ultimately forced me to discard many observations, which would have made the item selection process even more stringent.

did not take into account an unknown source of variation, which I did not want to influence participants' ratings. Since the accuracy of the corpus-based model is 88%, this condition removes 12% of the total number of genitive observations. Following the logic of Gries et al. (2020) in the context of MuPDAR (see also Section 6.3.3.2), model predictions between 0.4 to 0.6 are considered correct by default since they allow for both variants.

Out of the original 2,302 genitive observations, 563 fit these selection criteria. 103 (18.3%) were produced by native speakers and the remaining 460 (81.7%) came from the learners.

Additional criteria

From the pool of 563 eligible observations, I made sure to select observations that are plausible and interpretable by the participants (Stowe & Kaan 2006). For example, if an observation includes a proper noun that is unknown to the participant, the participant may not understand the concept and therefore not know whether the possessor refers to a person, a place etc., which makes the rating unreliable. Hence, only proper nouns referring to well-known country names like *India* or *Greece* for example could figure as possessors in the experimental items.

Although the experimental items did not include much context, they should preferably deal with a single topic so that the use of the genitive observation is somewhat contextualized. For example, an observation like (186) would be excluded because it is not entirely clear how *the main of the manuscript* fits in the surrounding context. Since both native speaker and learner observations were eligible for the experiment, I removed observations containing any indications as to the role of the speaker in the exam. For learners specifically, this means that I did not consider observation with clear ungrammaticalities, as in (187), where *alphabetic* is used as a noun. In order to prevent structural persistence effects from occurring within experimental items, observations that feature other genitive observations in the surrounding context were removed as well (Szmrecsanyi 2006; Engel et al. 2022). Also excluded were observations where the genitive variant is or has an embedded genitive, as in (188), and so were cases with *all* as in (189) due to the ambiguity they entail (see also Section 6.2.2.2).

(186) because as you know er there were many kind of entertainments
and that I know so er the main of the manuscript was theatre er
however er the artist in that time history were not servants any
more (TLC:2_SP_21: C2 Spanish)

- (187) some dictionaries with the the alpha-alphabetic erm of Japan and the the pronunciation of Japanese (TLC:7_ME_1, B2 Spanish)
- (188) Dolly was created by replacing the DNA of the one sheep's egg with the DNA of another sheep's (TLC:2_7_RU_4, B2 Russian)
- (189) he understood er like other poets that er all the erm erm principle ideas of the French Revolution were doom to be defeated (TLC:2_IT_3, C1 Italian)

Simplifying the items

Since the experimental items come from native speakers and learners at any proficiency level, it is possible that there are lexical items in the observations produced by the more proficient speakers that are not familiar to low-proficiency learners participating in the experiment. Therefore, I simplified the observations by either removing or replacing the lexical items that do not figure in the New General Service List (New-GSL, Brezina & Gablasova 2015), which contains a stable vocabulary of 2,122 highly frequent English words that even low-proficiency learners are most likely to have encountered. Since the predictors in the model often relate to the features of the possessor and possessum, removing or replacing these constituents would again affect the accuracy of the corpus-based predictions. Hence, if either the possessor or the possessum includes any words that do not figure in the New-GSL, the observation was discarded. Although country names are not included in the list, even low-proficiency learners are most likely familiar with them, so these possessors were retained. To make sure that the observations do not incorporate too much syntactic complexity that could disadvantage low-proficiency learners, subclauses were also transformed into main clauses.

Balancing animacy and corpus-based predictions

In order to replicate the finding from the corpus-based model that B1 and B2 learners are less sensitive to the animacy constraint, the items included in the experiment had to be balanced with regard to possessor animacy. At the same time, the experiment also investigated whether the ratings of the participants reflect the predictions from the corpus-based model: if the corpus-based model predicts that the *s*-genitive is highly likely in a certain context, then the ratings for the *s*-genitive should be high as well. In this respect, previous research found that acceptability ratings do not correlate in a linear fashion with the predictions from regression models, especially when the model predictions indicate that the probability of a given variant is low (Kempen & Harbusch 2005: 330; Arppe & Järvikivi 2007: 149–150; Divjak, Dąbrowska & Arppe 2016: 26–27), so the correlation between the

ratings and the predictions of the corpus-based model may not be as strong across the whole range of corpus-based predictions. To properly quantify the overall correlation between the ratings and the corpus-based predictions, the experimental items should therefore also be balanced with respect to the predictions of the corpus-based model. In other words, the observations selected for the experiment should be distributed equally across the entire range of probabilities predicted by the corpus-based model. Since the model predicts the probability of the *s*-genitive from 0% to 100% for each observation, I selected the experimental items from five probability bins of 20% each. The first bin contains the genitive observations for which the model predicts a probability between 0 and 20% for the *s*-genitive, the second bin covers the range from 20% to 40% etc. To illustrate, the probability of the *s*-genitive for the experimental item in Figure 7.1 is 72.18% according to the corpus-based model, so it belongs to the fourth bin covering the range from 60% to 80%.

Ideally, the experiment should feature five items from each of these bins, one for each level of possessor animacy. However, the experiment would then consist of 25 target items, which would make the experiment rather long considering that I included filler items as well. More importantly, not every bin contains observations that fit all selection criteria while featuring the right type of possessor to boot. For example, there are no eligible observations with temporal possessors in the fifth bin, where the *s*-genitive is highly likely. It is not surprising that certain combinations of probability and possessor animacy are infrequent considering that the animacy constraint is perhaps the constraint that exerts the most influence on the probability of the variants to begin with. Animate possessors, for example, strongly favor the *s*-genitive, so there are very few observations with animate possessors for which the model predicts an *of*-genitive.

Due to such missing observations, the experiment featured only three animacy levels in each probability bin instead of five. To ensure that the experiment remained balanced, I made sure that each animacy level is equally frequent in the experiment and figures in bins across the entire probability range. Additionally, this made the experiment shorter, since each participant saw only 15 target items. The distribution of the animacy levels across probability bins is shown in Table 7.1.

Table 7.1: Distribution of the target items across probability bins and animacy levels.

Probability of the s-genitive	Possessor animacy in the experimental items		
0-20%	Temporal	Animate	Locative
20-40%	Temporal	Inanimate	Collective
40-60%	Animate	Inanimate	Locative
60-80%	Temporal	Inanimate	Collective
80-100%	Animate	Collective	Locative

For example, the three target items from the first probability bin contain a temporal possessor, an animate possessor and a locative possessor respectively. Target items with animate possessors also figure in the third bin and the fifth bin so that animate possessors are distributed equally across the entire probability range.

While the participants only got to see 15 target items, I collected up to three different observations that fit the selection criteria for each of these combinations of animacy level and probability bin in Table 7.1. For the temporal possessors in the first bin for example, I extracted the following three observations:

- (190) it arrived somewhere in the middle no at the beginning of the nineteenth century (TLC:2_8_RU_6, B2 Russian)
- (191) teachers probably spend the most crucial hours of the day with us (TLC:2_IN_24, NS)
- (192) what are the differences between the movies of the past and the present (TLC:2_6_IN_58, NS)

These three observations would be used in three different versions of the experiment: target item (190) was used in the first version, (191) in the second version and (192) in the third version. By including multiple target items for each category in Table 7.1, the results from the statistical analysis are more reliable as they are aggregated across multiple items. However, because of the stringent selection procedure, it was not always possible to find three suitable observations, in which case multiple versions of the experiment contained the same experimental item.

In total, there are 37 target experimental items across the three versions of the experiment. An overview of these experimental items is given in Table 7.2 with their original genitive variant in the corpus data.

Table 7.2: Overview of the experimental items with the genitive variant they originally featured in the TLC data.

Target item	Probability	Possessor head	Possessor animacy
This trend began somewhere at the beginning of the nineteenth century	0-20%	century	temporal
Teachers probably spend the most crucial hours of the day with us	0-20%	day	temporal
It was too late and the right side of the ship hit the rocks	0-20%	ship	locative
There will always be someone who wants the advice and the knowledge of the travel agent	0-20%	agent	animate
What are the differences between the movies of the past and the present?	0-20%	present	temporal
I want to meet the president of America	0-20%	America	locative
What is the name of your city?	0-20%	city	locative
I am interested in what you said about the importance of the role model	0-20%	model	animate
Parents are very worried about many things, so they forget about the education of their children	0-20%	child	animate
Mexico has a lot of people, almost double the population of the United Kingdom	20-40%	United Kingdom	collective
he saw that the name of the author of the guide was "James Bond"	20-40%	guide	inanimate
You should be careful because hot drinks can hurt the roof of your mouth	20-40%	mouth	inanimate
We want to talk about the time of history in which Elizabeth lived	20-40%	history	temporal
It was created after the foundation of the football association	20-40%	association	collective
The most famous reward is given to the best player of the year	20-40%	year	temporal
This is the story of your life	20-40%	life	inanimate
Scientists believe in helping each other, so they look at the goals of the community	20-40%	community	collective
This flower is my grandpa's birthday gift	40-60%	grandpa	animate
It is necessary to know some details about the customer's private life	40-60%	customer	animate
The tsunami hit the coast of India	40-60%	India	locative
I think you have talked about the game's pattern	40-60%	game	inanimate
We must be sure that the equipment's motors are turned down	40-60%	equipment	inanimate
School is a way to create the future of the children	40-60%	child	animate

I propose that students prepare for the next day's class at home	60-80%	day	temporal
What are people saying about the government's response?	60-80%	government	collective
I don't remember the film's name	60-80%	film	inanimate
The world's population is expected to grow a lot	60-80%	world	collective
What is your idea about the world's future?	60-80%	world	collective
The products do not stay good for long, so we have to check the bottle's date	60-80%	bottle	inanimate
I think it is very important to support your country's player	80-100%	country	collective
What happened if you did not follow the principal's rules?	80-100%	principal	animate
The former president made this period in China's history difficult	80-100%	China	collective
When he was twelve years old he started playing for Barcelona's team	80-100%	Barcelona	locative
You should into the facts rather than the media's version	80-100%	medium	collective
The boy's mother picks the girl, so it is an arranged marriage	80-100%	boy	animate
Beijing's air is very bad	80-100%	Beijing	locative
What has changed since your grandmother's days?	80-100%	grandmother	animate

7.2.1.2 Filler items

The filler items in the experiment had a dual purpose. On the one hand, they should distract the participants from the goal of the experiment (see Schütze & Sprouse 2013: 39), which was to investigate the genitive alternation and more specifically the effect of animacy on the perceived naturalness of the variants. More specifically, the experiment aimed to replicate the finding that B1 and B2 learners are less sensitive to the animacy of the possessor than native speakers, so the second purpose of the filler items was to assess the proficiency level of the participants. Hence, the filler items are the questions from the Oxford Quick Placement Test, a short proficiency test consisting of 60 questions of various types, most of which require the participant to fill in the blank in a sentence or text by selecting the right word among several options (Allan 2001). The score participants received on the test from 0 to 60 was used to determine their CEFR proficiency level from A1 to C2 (Council of Europe 2001; see also Crosthwaite, Bailey & Meeker 2015).

In total, the experiment consisted of 15 target experimental items of the type exemplified in Figure 7.1 and 60 filler questions stemming from the

Oxford Quick Placement test. This means that there are four times more fillers than there are target items, which is well above the rule of thumb stating that there should at least be twice as many fillers as target items (Keating & Jegerski 2015: 16–17).

7.2.2 Participants

Statistical power analysis

To ensure the validity and reliability of the results, it is desirable to have an idea of the statistical power of the experiment design. Statistical power depends on the statistical test, the number of observations, the size of the effect under investigation, the variability in the data and the number of conditions to be compared (Gelman & Carlin 2014). The smaller the effect to be discovered, the more statistical power is needed. Likewise, datasets that incorporate much variability require more statistical power. Finally, studying how five animacy levels influence the ratings across five proficiency levels requires more statistical power than investigating the relation between the ratings and a single binary variable for example.

Most of these factors cannot be controlled for. For example, the effect size is already determined since I want to replicate the corpus-based finding that B1 and B2 learners are less sensitive to possessor animacy than native speakers. The number of comparisons could be adapted, so one could conflate the five animacy levels into a binary variable to improve statistical power, but this would eliminate important distinctions in the data and potentially make the experiment less balanced to boot. Therefore, the best way to ensure sufficient statistical power in this case was simply to collect enough rating data. Given that each participant rated no more than 15 experimental items, the only way to collect more data was to recruit more participants.

In the literature, an experiment is considered well-powered when the statistical test has an 80% chance of finding the desired effect when this effect is indeed present in the data (Cohen 1992; Brysbaert & Stevens 2018). For many statistical tests, statistical power can be determined rather easily by looking at Cohen (1992). This is far less straightforward for the statistical method I used, namely mixed-effects linear regression (Brysbaert & Stevens 2018; Kumle, Vö & Draschkow 2021), especially prior to data collection rather than afterwards (see Gelman & Carlin 2014). Nonetheless, this could be done through simulations with an artificial dataset, as outlined by Kumle, Vö and Draschkow (2021).

The power analysis was done in R (R Core Team 2022) using the *simr* (Green & MacLeod 2016) and *mixedpower* (Kumle, Vö & Draschkow 2021)

packages. First, I generated a dataset with artificial data from 150 participants equally distributed across the five proficiency levels, viz., B1, B2, C1, C2 and native. Then, I entered the relevant parameters relating to statistical power in an artificial mixed-effects linear regression model, including the expected effect size, the variability in the data and the predictor levels to be compared. To ensure that the design would be able to replicate the corpus-based findings, the effect sizes I chose were rather small for the interaction between possessor animacy and proficiency level as well as the other fixed effects that would figure in the model.¹⁰ However, I had no a priori idea about the variability in the dataset since the power simulation was performed before the analysis, in which case it is recommended to consult similar designs if available. In this respect, Engel et al. (2022) have investigated the dative alternation with a very similar rating task as a complement to a corpus-based investigation. Hence, the variance in their model informed the parameters of the artificial regression model, including the variance explained by the random effects and the unexplained variance. With this artificial model, I was able to simulate statistical power for each of the fixed effects for different numbers of participants. The result was then visualized in plots as exemplified in Figure 7.2.

¹⁰ To build the artificial model, I indicated that on the rating scale from 0 to 100, native speakers would give 15 points more to an animate possessor than an inanimate possessor, whereas B1 and B2 learners would only give 10 points more to animate possessors, so there would be a small difference of 5 points between native speakers and B1/B2 learners.

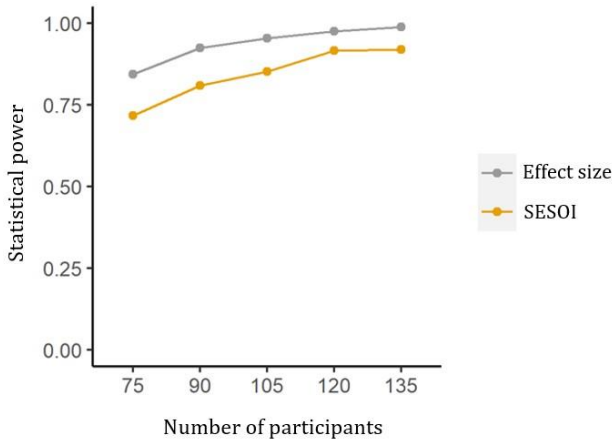


Figure 7.2: Visualization of statistical power for a change from an animate to an inanimate possessor at the B1 level compared to native speakers. The different number of participants is shown on the x-axis and the statistical power is shown on the y-axis. The grey line is for the effect size and the yellow line is for the smallest effect size of interest (SES01), which tests statistical power for an effect size that is 15% smaller.

The plot in Figure 7.2 visualizes the statistical power for one of the effects found in the corpus-based model, namely that B1 learners are less sensitive to the animacy constraint than native speakers. In the experiment, this effect would result from B1 learners differentiating less between animate and inanimate possessors in their ratings compared to native speakers. Since the experiment was designed to replicate this effect, it is desirable that the artificial model reaches a statistical power of at least 0.8. The plot shows that for the rather small effect size I entered in the artificial model, collecting data from 75 participants would suffice. However, at least 90 participants would be required for the smallest effect size of interest, which measures statistical power for an effect size that is 15% smaller. The plots for the other fixed effects in the simulation also showed that at least 90 participants would be needed. Therefore, I decided to collect ratings from 105 participants to ensure that the model would not lack power. Importantly, this simulation assumes that the participants are distributed equally in each of the five proficiency levels, viz., B1, B2, C1, C2 and native. Hence, collecting data from 105 participants implies collecting data from 21 (105/5) participants from each of these proficiency groups.

Recruiting participants

The participants for the study were recruited on Prolific (Prolific 2023), an online platform where individuals can register to complete paid online

experiments. These individuals fill in information about themselves upon registration to Prolific, which then allows researchers to choose who can participate in their experiment via prescreening. For the purposes of the study, it was important that the participants in the experiment resemble the speakers included in the TLC data of the corpus-based model as closely as possible.

The questionnaire was administered to 25 native speakers of English (13 female, 12 male, age range: 19-67, $M = 36$, $IQR = 24$). Through the selection of the appropriate prescreeners on Prolific, I made sure that these are monolingual native speakers of British English, who grew up and currently live in the UK and hold British nationality. For the learner participants, the prescreeners made sure that they were raised monolingually with either Chinese, Hindi, Urdu, Spanish, Italian, Hungarian, Polish or Russian as their L1, which covers the most common L1 backgrounds featured in the TLC data used for the corpus-based model. Since the proficiency level of the learner participants was determined by their score on the filler items from the Oxford Quick Placement Test (see Section 7.2.1.2), the questionnaire was administered to learners until a satisfactory number of participants was reached for the B1, B2, C1 and C2 levels in accordance with the results from the power simulation. It was particularly challenging to obtain enough data from B1 learners, so I conflated the data from 18 B1 learners and 7 A2 learners. In total, the questionnaire was distributed to 101 EFL learners (54 female, 47 male, age range: 19-60, $M = 27$, $IQR = 15$). Table 7.3 presents an overview of the distribution of the EFL learners by their L1 and proficiency level.

Table 7.3: Distribution of the learner participants by L1 background and proficiency level. Percentages indicate the contribution of the L1 background to a given proficiency level. For the column total, the percentages indicate the contribution of the proficiency level to the total participants.

	A2	B1	B2	C1	C2	Row total
Chinese	0	0	3 (10%)	10 (35.8%)	3 (16.7%)	16 (15.8%)
Hindi	0	0	3 (10%)	0	1 (5.6%)	4 (4%)
Urdu	0	0	4 (13.3%)	3 (10.7%)	2 (11.1%)	9 (8.9%)
Hungarian	0	3 (16.7%)	5 (16.7%)	1 (3.6%)	4 (22.2%)	13 (12.9%)
Italian	1 (14.3%)	2 (11.1%)	5 (16.7%)	6 (21.4%)	3 (16.7%)	17 (16.8%)
Polish	1 (14.3%)	5 (27.8%)	4 (13.3%)	3 (10.7%)	4 (22.2%)	17 (16.8%)
Russian	0	3 (16.7%)	2 (6.7%)	2 (7.1%)	0	7 (6.9%)
Spanish	5 (71.4%)	5 (27.8%)	4 (13.3%)	3 (10.7%)	1 (5.6%)	18 (17.8%)
Column Total	7 (6.9%)	18 (17.8%)	30 (29.7%)	28 (27.8)	18 (17.8%)	101

Table 7.3 shows that the A2/B1 level features participants from five different L1 backgrounds, namely Hungarian, Italian, Polish, Russian and Spanish. At the B2 level, all eight L1 backgrounds are represented. The C1 and C2 level include participants from seven L1 backgrounds each. Each participant contributed 15 ratings, so the dataset consists of 1,515 learner ratings and 375 native speaker ratings for a total of 1,890 ratings.

Removing data

The duration of the experiment varied considerably across participants, both among native speakers (range in minutes: 6:46-24:16, $M = 13:55$, $IQR = 5:18$) and learners (range in minutes: 11:57-94:42, $M = 24:44$, $IQR = 15:19$). Nonetheless, no submissions were excluded from further analysis because all participants performed well on the simpler questions from the Oxford Quick Placement Test aimed at beginners and intermediate learners, so it is unlikely that they answered the questions randomly.

7.2.3 Design and procedure

General overview

The survey was built in Qualtrics (Qualtrics 2023) and administered online to participants recruited on Prolific. Before accessing the Qualtrics survey, the learner and native speaker participants on Prolific saw a slightly

different description of the survey. Since the filler items stemmed from a proficiency test, namely the Oxford Quick Placement Test (see Section 7.2.1.2), the description for the learner participants stated that I am interested in different ways of measuring how well speakers of different languages know English as follows:

“Thank you for your interest in this survey! As part of my PhD thesis, I am interested in different ways of measuring how well speakers of different languages know English. After a few questions about the language(s) you speak, you will see different types of language questions (55 questions in total) with some instructions on how to answer them.”

The filler items from the proficiency test might seem particularly easy for native speakers, so I explained that their data would be used to test the accuracy of these measurements:

“Thank you for your interest in this survey! As part of my PhD thesis, I am interested in different ways of measuring the English proficiency level of language learners from different mother tongue backgrounds. To see how well these different measurements work, we would like to test them on native speakers of English as well. After a few questions about your background as a native speaker of English, you will see different types of language questions (55 questions in total) with some instructions on how to answer them.”

This brief description also included the approximate duration of the survey as well as the payment participants would receive upon completion. From there, they could access the Qualtrics survey and give their consent. At the beginning of the survey, the prescreening questions were repeated. If the participants' answers were inconsistent with the information they reported on Prolific, they were automatically removed from the survey.

Because the experiment featured different types of questions, alternating filler items and target items could potentially have confused the participants, so questions of the same type were presented together instead. Hence, the participants answered the first 25 filler items from the Oxford Quick Placement Test. Then, they had to rate the 15 the target experimental items before answering the remaining 35 filler items. At the end of the survey, participants could leave comments or feedback about the study. The participants were then automatically returned to Prolific and their survey was submitted for approval and payment.

To make sure that the instructions were clear and that there were no technical issues, the survey was first piloted on five learners and six native speakers, who were made aware that they would participate in a pilot study before they started.

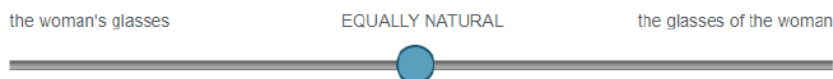
Presentation of the target items

The target items were preceded by instructions and an example where participants could move the slider for practice as shown in Figure 7.3 below.

Below you will see sentences with two options to say the same thing, which are presented in a random order. Your task is to read each sentence carefully and choose which option sounds more natural to you. To do so, you can drag the slider towards the option that sounds more natural to you. The closer you place the slider to one end, the more natural that option seems to you, and the less natural the other option. If both seem equally natural, you can leave the slider in the middle (you will still have to use the slider to go the next question). Below follows an example.

Example:

They picked up **the woman's glasses** / **the glasses of the woman**.



For, example you find the option on the left side of the slider (**the woman's glasses**) to be much more natural than the option on the right side of the slider (**the glasses of the woman**), but the option on the right side is not completely unnatural to you. In this case, you would move the slider to the option on the left side (**the woman's glasses**), but not completely, as shown below:

Example:

They picked up **the woman's glasses** / **the glasses of the woman**.

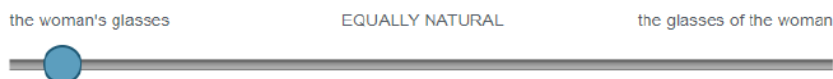


Figure 7.3: Instructions shown to participants for the target experimental items with a concrete example.

The pilot study revealed that native speakers were more inclined to use the extreme ends of the slider compared to learners, so their instructions were slightly rewritten to emphasize that they should read the instructions carefully and that they should feel free to use the entire range of the slider.

As stated in Section 7.2.1.1, each participant saw only one observation for each combination of probability bin and possessor animacy shown in Table 7.1, so each participant rated 15 experimental items in total. Instead of relying on a single experimental item for each combination of possessor

animacy and probability bin, I tried to collect up to three observations and ended up with the 37 experimental items shown in Table 7.2. These three observations were distributed across three different versions of the experiment. While distributing the 37 experimental items in this way, I made sure each version did not contain more than two instances of the same possessor or possessum head noun so as to avoid lexical priming effects (Szmrecsanyi 2006; Hartsuiker et al. 2008). Each participant was randomly assigned to one of the three versions.

As for the order in which the 15 experimental items were presented, I wanted to make sure that items from the same probability bin or with the same possessor animacy were perfectly spread out across the experiment. As suggested by Stowe and Kaan (2006: 54), I therefore presented the items in pseudorandom order as shown in Table 7.4

Table 7.4: Pseudorandomized order of the 15 experimental items from Table 7.1.

Target item	Probability bin	Animacy
Item 1	0-20%	Animate
Item 2	40-60%	Locative
Item 3	80-100%	Collective
Item 4	20-40%	Inanimate
Item 5	60-80%	Temporal
Item 6	0-20%	Locative
Item 7	40-60%	Animate
Item 8	80-100%	Locative
Item 9	20-40%	Collective
Item 10	60-80%	Inanimate
Item 11	0-20%	Temporal
Item 12	40-60%	Inanimate
Item 13	80-100%	Animate
Item 14	20-40%	Temporal
Item 15	60-80%	Collective

It is possible that participants give slightly different ratings at the end of the experiment compared to the beginning, either because they are bored or tired, or because they became better at rating the naturalness of the variants. As a result, there would be a systematic difference between the ratings from the first few items in Table 7.4 compared to the later items. To control for such practice and fatigue effects (Abbuhl, Gass & Mackey 2013: 120), it should not always be the same experimental items that figure at the beginning or the end of the experiment. This is achieved by changing the

order of the items for different participants, so that each item is ideally equally frequent at the beginning and at the end of the experiment across the surveys of individual participants (Abbuhl, Gass & Mackey 2013: 120).

Concretely, the items in Table 7.4 were first put in three groups of five: Group 1 contained the first five items, Group 2 items 6 to 10 and Group 3 the last five items. Then, the order of these groups was altered in the experiment for individual participants, so that certain participants first rated the items from the first group, whereas others first saw the items from the third group etc. Specifically, participants were randomly assigned to one of three group orders, namely Group 1 - Group 2 - Group 3, Group 3 - Group 1 - Group 2 or Group 2 - Group 3 - Group 1. Regardless of the order, I made sure that the position of the genitive variants on the rating scale was altered each time. This means that if the *s*-genitive figured on the right side of the slider as in *the film's name* in Figure 7.1, the *s*-genitive would figure on the left in the following item (see also Engel 2022: 92).

7.2.4 Statistical analysis

The ratings were analyzed as the response variable in two mixed-effects linear regression models using the *lme4* package (Bates et al. 2015) in R (version 4.2.2, R Core Team 2022). The first model looked into the effect of possessor animacy on the ratings and the second model investigated the correlation between the ratings from the participants and the predictions from the corpus-based model. In both models, I employed treatment coding to compare how the effects of the predictors influence the ratings for native speaker participants compared to learner participants at different proficiency levels. Since the position of the variants on the rating scale alternated with each experimental item (see Section 7.2.3), the *s*-genitive was presented on the left side of the slider in half the items. In this case, the participants' ratings corresponded to a score from 0 to 100 for the *of*-genitive rather than the *s*-genitive. Because the corpus-based model predicted the likelihood of the *s*-genitive, these ratings were transformed to also reflect a score for the *s*-genitive before analyzing the data. For the analysis, the data from the A2 and the B1 learners was conflated together to ensure sufficient statistical power (see Section 7.2.2).

The advantages of mixed-effects linear regression

Using mixed-effects linear regression models to analyze the rating data has two main advantages. For one, the method offers the possibility "to simultaneously consider all factors that potentially contribute to the understanding of the structure of the data" (Baayen, Davidson & Bates 2008: 409). That is, such models can take into account the individual behavior of

participants, who might not all use the range of the rating scale in the same way (see Schütze & Sprouse 2013: 39). Additionally, although the experimental items have been selected carefully, it is possible that certain individual experimental items are generally rated differently from others. Hence, the recommendation for such experiments is to include crossed random effects for the participant and the experimental item (Baayen, Davidson & Bates 2008). This refers to random intercepts adjustments as well as random slopes by participant to account for the fact that fixed effects might not affect individual participants in the same way, or by item to account for specific proficiency levels rating items differently (Linck & Cummings 2015). Including random effects in this way ensures that the results from the model regarding the relevant fixed-effects are robust and generalizable (see Gries 2015).

An additional benefit of linear regression models is that it is easy to control for potential confounding variables (Linck & Cummings 2015). For example, it was not feasible to balance the experiment with respect to the length of the possessor and the possessum while designing the experiment (see Section 7.2.1.1), so the ratings are likely to be influenced by constituent length in addition to the animacy of the possessor. As a result, it would be unclear whether dissimilarities in the ratings of the participants are strictly due to the variable of interest, namely possessor animacy, or whether constituent length played a role as well. An easy way to disentangle the effect of both variables is to add the length of the constituents as a fixed effect in the model.

The animacy model

The first linear regression model pertained to the first goal of the experiment, which is to test whether and how the five animacy levels influence the ratings of native speakers compared to learners at different proficiency levels. Hence, the first model included an interaction between the proficiency level of the participants and the animacy of the possessors in the experimental items. Collective possessors figured as the reference animacy level because this made it easier to test the hypothesis that learners primarily struggle with collective and locative possessors. If B1 and B2 learners are prone to misclassifying possessors between these two categories, then they should give more similar ratings to possessors from these categories compared to native speakers. Since it was not possible to balance the experiment for the length of the constituents, these variables were entered as fixed effects in the model. The length of the constituents is not normally distributed, so these variables were standardized by subtracting the median rating (see Heumann, Schomaker & Shalabh 2016:

42) and dividing the resulting value by two standard deviations as suggested by Gelman (2008). The final model was generated by the following formula:

Animacy model $\leftarrow$ lmer(*s*-genitive rating ~ proficiency + possessor animacy + possessor length + possessum length + proficiency:possessor animacy + (1|target item) + (1|participant), data = rating data)

As suggested by Linck and Cummings (2015) and Baayen et al. (2008), I experimented with adding random slopes, but the model ran into convergence issues, so the final model only features intercept adjustments for participant and target item. The final model is significantly better than the null model ($p < 0.001$) and has an R^2_{marginal} of 0.15 and an $R^2_{\text{conditional}}$ of 0.35.

The corpus prediction model

The second model dealt with the second goal of the experiment, which is to assess the overall relation between the language use of the speakers in the corpus data and the intuitions of the participants in the experiment. To do so, the model investigated to what extent the ratings of the native speakers and the learners correlates with the predictions from the corpus-based model for the experimental items under study. Specifically, this model featured the probability of the *s*-genitive according to the corpus-based model as a predictor, which was allowed to interact with the proficiency level of the participants to reveal whether the strength of the correlation varies according to proficiency level. The corpus-based prediction was centered based on the median and then standardized following Gelman (2008). The final model was generated by the following formula:

Corpus prediction model $\leftarrow$ lmer(*s*-genitive rating ~ proficiency + corpus-based prediction + proficiency:corpus-based prediction + (1|target item) + (1|participant), data = rating data)

For this model, I also tried to incorporate random slope adjustments, but this led to convergence issues, so the model again only contains intercept adjustments for participant and experimental item. The final model is significantly better than the null model ($p < 0.001$) and has an R^2_{marginal} of 0.14 and an $R^2_{\text{conditional}}$ of 0.34.

Importantly, the probability of the *s*-genitive was calculated based on the effect of all predictors present in the corpus-based model. Therefore, it would be redundant to include any other predictors in this model. At the same time, this means that the prediction is influenced by the interaction

between proficiency level and possessor animacy in the corpus-based model. Hence, if the experimental item was produced by a B1 learner in the original corpus data, the corpus-based prediction would reflect the behavior of B1 learners specifically. Since B1 and B2 learners are supposedly less sensitive to the animacy of the possessor than native speakers, the corpus-based prediction for an experimental item produced by a B1 or B2 learner might not be entirely accurate for a native speaker participant and vice versa. While this is admittedly not ideal, the experiment features target items from all proficiency levels in the original corpus data, so all participants had to rate items produced by speakers with a different proficiency level. Therefore, I maintain that the correlation between the predictions of the corpus-based model and the ratings from the participants provides an accurate representation of the relation between the language use and intuitions of speakers across proficiency levels. For the first model on the effect of possessor animacy, this is not problematic considering that I am mostly interested in whether the ratings for the *s*-genitive are higher or lower depending on the proficiency level of the participant.

Individual differences

It is important to keep in mind that the results from the corpus-based models represent a generalization across the individual probabilistic grammars of the speakers in the corpus. As explained in the introduction to this chapter, it is possible that the probabilistic grammar of an individual speaker does not perfectly overlap with the constraints the model deems significant as shown by Divjak et al. (2016: 28). This is all the more relevant in the context of learner language since it features even more individual variability compared to native speakers (Jäschke & Plag 2016: 511–512; Kinne 2020: 128). In this regard, Kinne (2020: 128–138) found that the effect of end-weight differed across learners for the particle placement alternation and the verb-PP alternation: some learners rely on the constraint almost categorically, whereas others barely take the constraint into account. In the present case, it is for example unclear whether B1 and B2 learners are less sensitive to possessor animacy as a homogeneous group, or whether there are any other strategies at play.

In the corpus data, I was not able to investigate individual differences between speakers because very few speakers contribute enough observations, which are not balanced across a wide range of contexts to boot (see Section 7.1). By comparison, each participant in the experiment rated 15 items, equally distributed across probability bins and animacy levels. With this in mind, investigating the ratings of individual participants can provide interesting insights regarding the inter-speaker variability for

the animacy constraint. At the same time, looking into the correlation between the corpus-based predictions and the ratings of individual speakers could shed light on the extent to which the probabilistic grammar of individuals overlaps with the significant set of constraints in the corpus-based model.

Inspired by Kinne (2020), I therefore calculated Cronbach’s alpha on the 15 ratings of each participant to get an idea of how much individual variability there is at each proficiency level. Cronbach’s alpha is a measurement of intra-class correlation mostly used to gauge the internal consistency of instruments with multiple items. If all items in a given instrument capture the same result as desired, then the items are internally consistent as reflected by a high Cronbach’s alpha (Gliem & Gliem 2003). In this case, Cronbach’s alpha measures the consistency of the participants’ ratings at each proficiency level. Moreover, I visualized and compared the ratings of individual participants against the aggregate results from both mixed-effects linear regression models.

7.3 Results

This section reports the results from the analysis of the rating data. First, I present the results from the animacy model and then the corpus-prediction model. The third subsection deals with individual differences between participants.

7.3.1 The animacy model

The random effects structure of the model is shown in Table 7.5.

Table 7.5: Overview of the random effects in the animacy model, their estimated variance and their standard deviation.

Random intercept	Variance	St. dev.
Target item	200.74	14.168
Participant	69.24	8.321

Both intercept adjustments contributed to the quality of the model according to likelihood ratio tests as implemented in the `anova()` function, with the target item explaining the most variance.

Table 7.6 presents the results for the fixed effects in the model.

Table 7.6: Effects of the individual predictors on the ratings in the animacy model. Positive coefficients indicate an increase in ratings for the *s*-genitive. (Marginally) significant *p*-values are written in bold.

Predictor	Coefficient	Std. error	<i>t</i> -value	<i>p</i> -value
(Intercept)	63.503	6.27	10.128	<0.001
Possessor length	-1.214	5.287	-0.230	0.82
Possessum length	-15.878	5.212	-3.046	0.005
Possessor animacy				
Collective → animate	16.916	8.391	2.016	0.049
Collective → inanimate	-16.456	8.889	-1.851	0.07
Collective → locative	-17.109	9.378	-1.824	0.074
Collective → temporal	-20.857	8.986	-2.321	0.024
Proficiency				
Native → A2/B1	1.959	5.363	0.365	0.715
Native → B2	-11.019	5.146	-2.141	0.033
Native → C1	-16.116	5.218	-3.089	0.002
Native → C2	-11.402	5.929	-1.923	0.055
Proficiency x possessor animacy				
A2/B1 + animate	-14.191	6.814	-2.083	0.037
A2/B1 + inanimate	-6.64	6.813	-0.975	0.33
A2/B1 + locative	-5.071	6.82	-0.744	0.457
A2/B1 + temporal	-7.592	6.82	-1.113	0.266
B2 + animate	4.291	6.538	0.656	0.512
B2 + inanimate	13.391	6.535	2.049	0.041
B2 + locative	5.489	6.534	0.840	0.401
B2 + temporal	0.934	6.534	0.143	0.886
C1 + animate	9.226	6.629	1.392	0.164
C1 + inanimate	8.113	6.627	1.224	0.221
C1 + locative	7.001	6.63	1.056	0.291
C1 + temporal	8.148	6.630	1.229	0.219
C2 + animate	5.105	7.539	0.677	0.498
C2 + inanimate	3.288	7.526	0.437	0.662
C2 + locative	7.905	7.508	1.053	0.293
C2 + temporal	4.496	7.508	0.599	0.549

Whereas the length of the possessor does not influence the ratings, the model shows that the ratings for the *s*-genitive decrease when the length of the possessum increases, which echoes the finding from the corpus-based model that native speakers and learners alike tend to prefer the *of*-genitive when the possessum is long in spoken language production, in conflict with the short-before-long principle (Behaghel 1909) or Easy First (MacDonald 2013).

Since animacy is part of an interaction with proficiency level, the main effects for animacy only pertain to the reference proficiency level, namely native speakers. These effects show that native speakers give higher ratings for the *s*-genitive when the possessor is animate rather than collective.

Collective possessors in turn receive higher ratings than inanimate, locative and temporal possessors. This difference is marginally significant for inanimate and locative possessors and significant for temporal possessors, which appear to induce the lowest ratings for the *s*-genitive. The main effects for proficiency level indicate that B2 and C1 learners give significantly lower ratings for the *s*-genitive than native speakers when the possessor is collective. The same effect is marginally significant for C2 learners.

The model features two significant interaction effects. While native speakers find the *s*-genitive more natural when the possessor is animate rather than collective, this effect is much weaker for A2 and B1 learners, who give similar ratings for animate and collective possessors. That is, the main effects show that native speakers rate *s*-genitives approximately 17 points higher on the scale from 0 to 100 when the possessor is animate rather than collective. For the same change from collective to animate, A2 and B1 rate the *s*-genitive 14 points lower than native speakers, so A2 and B1 learners only rate the *s*-genitive 3 points (17-14) higher when the possessor is animate. By contrast, B2 learners do not distinguish between inanimate and collective possessors as much as native speakers. Native speakers prefer the *s*-genitive with collective possessors rather than inanimate possessors, so they rate the *s*-genitive 16 points lower with inanimate possessors. By comparison, the *s*-genitive only receives 3 less points when the possessor is inanimate rather than collective for B2 learners.

The interaction between proficiency level and possessor animacy is visualized in the partial effects plot in Figure 7.4. Partial effects plots summarize the information from the model by changing the value of the relevant constraint, viz., possessor animacy, for each of the proficiency levels while holding the other constraints in the model, namely possessor and possessum length, at their default level (Fox 2003). In this way, the plot shows how a change in possessor animacy affects the ratings for the *s*-genitive for native speakers and learners across proficiency levels.

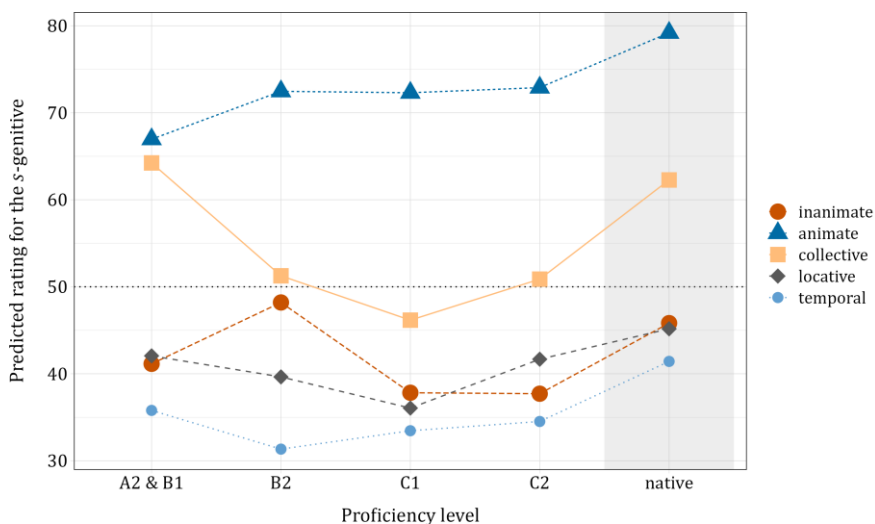


Figure 7.4: Partial effects plot of the interaction between proficiency level and possessor animacy. Predicted ratings on the y-axis are for the s-genitive. The vertical distance between the shapes reflects the effect size of the predictor. The plotted probabilities are calculated with the standardized length of the possessor and possessum at their default level, viz. 0. The native proficiency level (shaded) is the reference level.

As suggested by the interaction effects presented in Table 7.6, Figure 7.4 shows that the ratings for animate possessors and collective possessors are very similar at the A2/B1 level whereas native speakers clearly give higher ratings for the s-genitive when the possessor is animate rather than collective. Specifically, it seems that A2 and B1 learners give slightly higher ratings to collective possessors while at the same time giving lower ratings to animate possessors. Regarding the interaction effect involving B2 learners and native speakers with respect to collective and inanimate possessors, the figure shows that B2 learners give significantly lower ratings for the s-genitive than native speakers when the possessor is collective (see Table 7.6) while at the same time giving slightly higher ratings when the possessor is inanimate. As a result, there is barely a difference between the ratings for collective possessors and inanimate possessors at the B2 level as opposed to native speakers who give higher ratings for the s-genitive when the possessor is collective.

7.3.2 The corpus prediction model

The random effects structure of the corpus prediction model is shown Table 7.7. As in the animacy model, both intercept adjustments contributed to the

quality of the model according to likelihood ratio tests as implemented in the `anova()` function, with the target item explaining the most variance.

Table 7.7: Overview of the random effects in the corpus prediction model, their estimated variance and their standard deviation.

Random intercept	Variance	St. dev.
Target item	198.10	14.075
Participant	69.37	8.329

The parameters of the random effects are very similar in the animacy model, probably because the fixed effects in both models essentially capture the same variability. That is, the corpus-based prediction in the current model is calculated based on all predictors, whereas in the animacy model, these same predictors were either largely controlled for during experiment design (e.g., definiteness, final sibilancy, lexical biases), or they were entered as fixed effects in the model (viz., constituent length, proficiency level and possessor animacy).

Table 7.8 presents the results for the fixed effects in the model.

Table 7.8: Effects of the individual predictors on the ratings in the corpus prediction model. Positive coefficients indicate an increase in ratings for the *s*-genitive. (Marginally) significant *p*-values are written in bold.

Predictor	Coefficient	Std. error	<i>t</i> -value	<i>p</i> -value
(Intercept)	56.525	3.239	17.449	0.047
Corpus prediction	31.751	5.352	5.932	<0.001
Proficiency				
Native → A2/B1	-4.618	3.191	-1.447	0.150
Native → B2	-6.125	3.057	-2.004	0.047
Native → C1	-9.580	3.103	-3.087	0.003
Native → C2	-7.243	3.506	-2.066	0.041
Proficiency x corpus prediction				
A2/B1 + corpus prediction	-18.294	4.315	-4.239	<0.001
B2 + corpus prediction	-8.081	4.129	-1.957	0.050
C1 + corpus prediction	-3.836	4.173	-0.919	0.358
C2 + corpus prediction	-2.534	4.731	-0.536	0.592

The corpus-based prediction is part of an interaction term with proficiency level, so the main effect again pertains to native speakers only. This main effect shows that when the model predicts a high likelihood for the *s*-genitive, native speakers also give higher ratings for the *s*-genitive. The main effects for proficiency level compare how the different proficiency

levels rate the *s*-genitive when the prediction of the corpus-based model is at its default level, namely 0. Since corpus-based prediction was centered on the median, 0 corresponds to the median prediction and therefore should approximate an equal likelihood of the *s*-genitive and the *of*-genitive.¹¹ These effects show that when the corpus-based model predicts that either variant is equally likely, the B2, C1 and C2 learners generally give lower ratings for the *s*-genitive than native speakers, whereas A2 and B1 learners do not differ from native speakers in these contexts.

The model features a significant interaction showing that the corpus-based prediction has a weaker effect on the ratings of A2 and B1 learners compared to native speakers. When the corpus-based model predicts that the *s*-genitive is highly likely, native speakers rate the *s*-genitive 32 points higher compared to a context where either variant is more or less equally likely. By comparison, A2 and B1 learners only give the *s*-genitive 14 more points (32-18) when the corpus-based model predicts that the *s*-genitive is highly likely. The effect is also less pronounced for B2 learners, who give only 24 points more to the *s*-genitive in these cases, but this interaction effect slightly exceeds the threshold for significance ($p = 0.05$).

In short, while A2/B1 and B2 learners generally still prefer the *s*-genitive when the corpus-based model predicts an *s*-genitive, they are less sensitive to these predictions than native speakers, so the correlation between corpus-based predictions and ratings is weaker for these learners. By extension, the ratings from native speakers correspond more closely to the corpus-based predictions, which implies that their intuitions in the rating task align more closely with the language use of the speakers in the corpus. With this in mind, it is interesting that as the proficiency level of the learners increases, their ratings become more similar to those of the native speakers, and by extension, to the language use of the speakers in the corpus. Hence, the interaction is only marginally significant for B2 learners and the coefficients gradually approach zero at the C1 and C2 levels.

The interaction between proficiency level and the corpus-based predictions for the *s*-genitive is visualized in the partial effects plot in Figure 7.5.

¹¹ The median of the corpus-based prediction across all ratings is 0.48, indicating a very slight preference for the *of*-genitive.

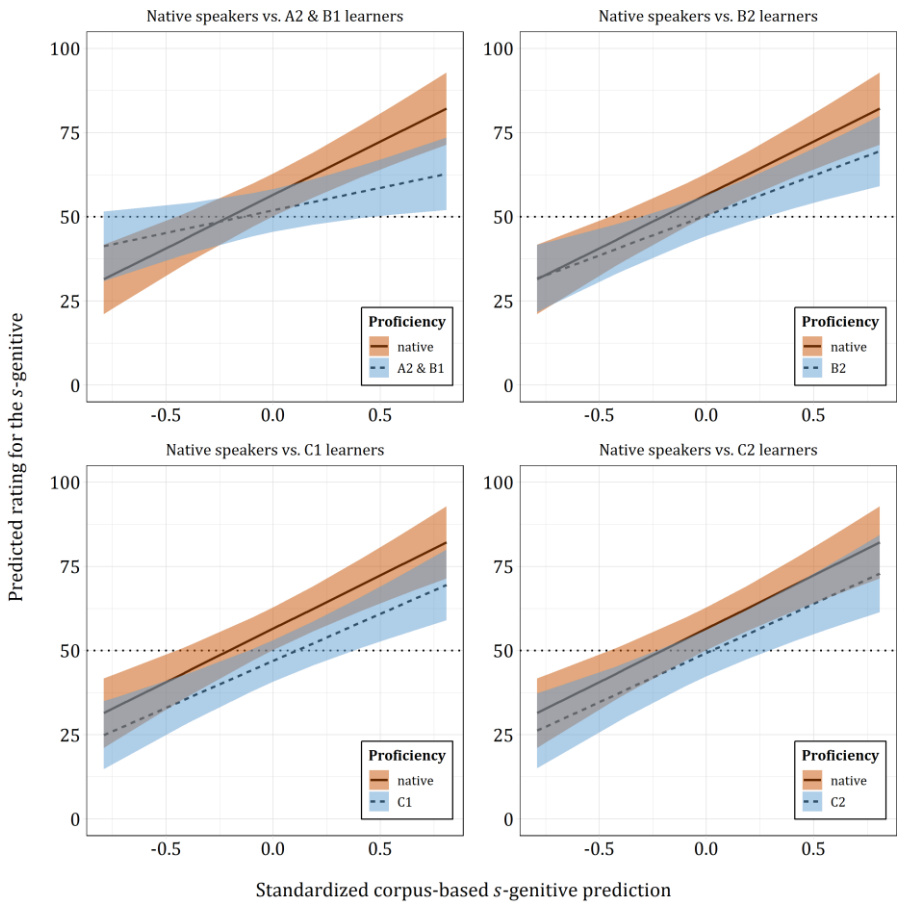


Figure 7.5: Partial effects plot of the interaction effect between proficiency level and the corpus-based prediction for the s-genitive. Predicted ratings on the y-axis are for the s-genitive. The values on the x-axis represent the standardized corpus-based prediction for the s-genitive. The plot shows the effect of the corpus-based prediction on the rating for the s-genitive for native speakers in orange and each learner proficiency level separately in blue.

The figure shows that the ratings from the native speakers are most strongly associated with the predictions of the corpus-based model: their slope is steeper compared to the slopes for the learners, indicating that an increase in the predicted likelihood of the s-genitive leads to a larger increase in ratings for the s-genitive. As their proficiency level increases, the slopes for the learners become steeper as well, so they resemble that of native speakers more.

The upper left panel of Figure 7.5 shows that A2 and B1 learners, who differ significantly from native speakers, exhibit the flattest curve. Specifically, the curve seems to indicate that A2 and B1 learners give lower ratings for the s-genitive than native speakers when the s-genitive is highly

likely according to the corpus-based model, and they also give higher ratings than native speakers when the *s*-genitive is unlikely, i.e., when the corpus-based model predicts a preference for the *of*-genitive. Hence, A2 and B1 learners diverge from native speakers both in contexts of language use that favor the *s*-genitive and in contexts that favor the *of*-genitive. B2 learners also give slightly lower ratings for the *s*-genitive than native speakers when the corpus-based model predicts an *s*-genitive, but they give the similar ratings when the model predicts an *of*-genitive.

The overall repeated measures correlation between the predictions from the corpus-based model and the ratings from all participants is moderate at $r = 0.38$ ($p < 0.001$) (Bakdash & Marusich 2017). If one considers that ratings above 50 represent a preference for the *s*-genitive, it is possible to also examine how often the participants' intuitions and the predictions from the corpus-based model agree on the same genitive variant in a categorical way. If the intuitions of native speakers are more aligned with the language use of the speakers in the corpus, then I expect a higher degree of agreement between their ratings and the corpus-based predictions compared to the other proficiency levels and especially A2 and B1 learners. Hence, for each participant proficiency level, I selected a small, balanced sample consisting of 100 observations for which the corpus-based model predicts an *s*-genitive, and 100 for which it predicts an *of*-genitive. This sample showed that the intuitions of the native speakers correspond to the predicted variant in 80% of the cases. As expected, the agreement is lower for A2 and B1 learners, whose intuitions match the predicted variant in only 68% of the cases. In general, learners' intuitions better approximate language use as they become more proficiency, with an agreement of 73% at the B2 level, 77% at the C1 level and 81% at the C2 level.

7.3.3 Individual differences

Previous research has revealed that there is considerable individual variability among speakers and especially language learners (Jäschke & Plag 2016: 511–512; Kinne 2020: 128), which raises the question as to whether the aggregated results from the linear regression models capture the behavior of individual participants adequately. First, I tested how consistently speakers within each proficiency level rate the experimental items. To do so, I grouped participants with the same proficiency level together and calculated Cronbach's alpha on the 15 ratings of each participant. A higher Cronbach's alpha indicates that the participants with the same proficiency level rate the items in more or less the same way and are therefore homogeneous, whereas lower values indicate that the

participants at a given proficiency level are heterogeneous because they rate the target items differently.

Table 7.9: Cronbach's alpha calculated on the ratings from the participants in each proficiency level as a measure of internal consistency.

Proficiency level	Cronbach's α
A2/B1	0.72
B2	0.9
C1	0.94
C2	0.95
Native speaker	0.95

Table 7.9 shows that native speakers and C2 learners are the most consistent, followed by C1 learners and then B2 learners, all of whom exhibit excellent internal consistency above .9 (Gliem & Gliem 2003). By comparison, A2 and B1 learners are much less consistent with a Cronbach's α of .72. This means that for the same experimental items, A2 and B1 learners give ratings that vary widely, so these learners do not agree on how natural they find the genitive variants with different possessors in various contexts. This is not due to conflating A2 and B1 learners together, since B1 learners alone are even less consistent with Cronbach's α of .63.

In the following subsections, I therefore take a closer look at how the ratings of individual A2/B1 learners relate to the aggregated results from both mixed-effects linear regression models. I will also explore the ratings of individual B2 learners since they are also involved in (marginally) significant interactions in both models.

Possessor animacy

The animacy model revealed that A2 and B1 learners do not distinguish between animate and collective possessors as much as native speakers. If the possessor is animate, native speakers find the *s*-genitive more natural than if the possessor were collective. By comparison, A2 and B1 learners find the *s*-genitive more or less equally natural with both types of possessors. When looking at the ratings of individual participants, I expect the same pattern to appear. That is, each native speaker should give higher ratings on average to the *s*-genitive when the possessor is animate, whereas individual A2 and B1 learners should give similar ratings for both categories.

To compare how animate and collective possessors influenced the ratings of individual participants, I looked at their raw rating data. Since each level of possessor animacy occurred three times in the experiment (see Section 7.2.1.1), I compared the average rating for animate possessors with the average rating for collective possessors for each individual participant. To illustrate, recall that animate possessors occur in probability bin 1, 3 and 5. If a given participant gave the *s*-genitive a rating of 75 for the first item, 25 for the second and 90 for the third, then the average rating of this participant for *s*-genitives with animate possessors is 63.3. Supposedly, the average score for collective possessors should be lower than 63.3 for native speakers. On the other hand, A2 and B1 learners should give more or less the same average rating for both animacy levels.

One might argue that comparing ratings in this way is problematic considering that the ratings are largely determined by the probability bins in which the items occur, as evidenced by the correlation between corpus predictions and ratings (see Section 7.3.2). Nonetheless, this does not prevent the animacy model from finding differences between the animacy levels.¹² Moreover, I made sure that each animacy level is represented in items across the entire probability range, so there should not be a large difference between the average corpus-based predictions for each animacy level.

Figure 7.6 plots the average *s*-genitive rating for animate and collective possessors for A2/B1 learners and native speakers separately.

¹² Because the prediction of the corpus-based model also depends on the strong effect of possessor and possessum length, the inclusion of these constraints as fixed effects in the animacy model admittedly helps the model to determine the effect of possessor animacy in isolation from the corpus-based predictions.

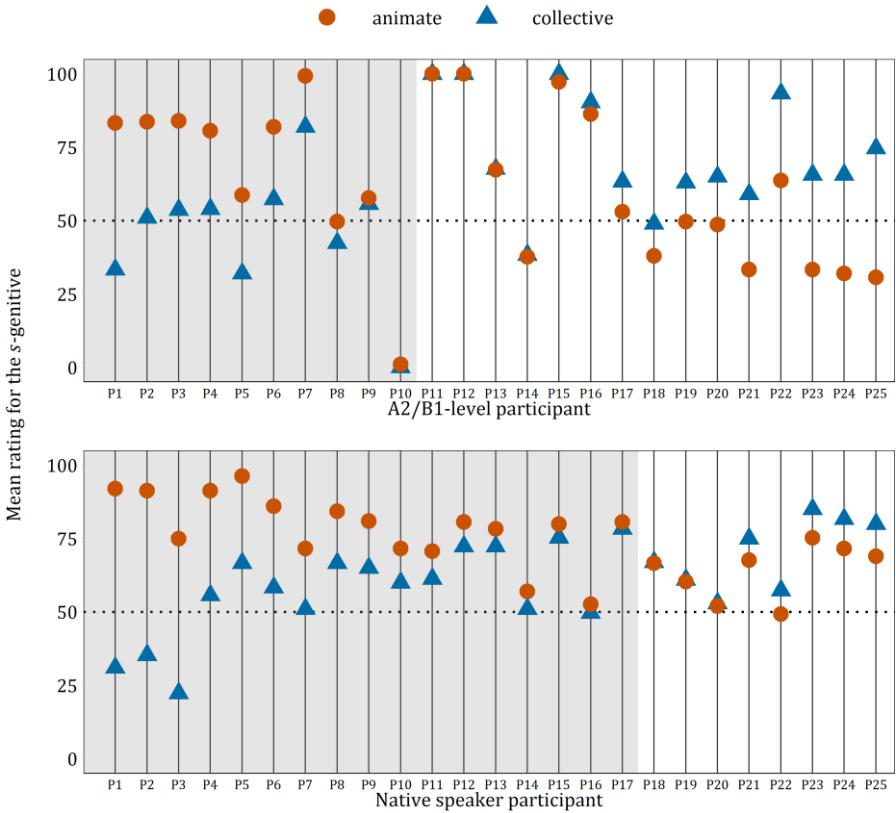


Figure 7.6: Individual mean ratings for animate and collective possessors for A2/B1 learners (top) and native speakers (bottom). The participants are ordered from left to right based on how much higher their average rating is for animate possessors. Participants who give higher ratings to animate possessors are shaded.

Figure 7.6 shows that 17 out of 25 (68%) native speaker participants give higher *s*-genitive ratings on average to animate possessors than to collective possessors (P1 to P17). In the A2 and B1 learner group, 15 out 25 (60%) participants exhibit the opposite pattern whereby they find the *s*-genitive more natural or equally natural when the possessor is collective rather than animate (P11 to P25). This ties in with the overall results from the animacy model, but contrary to what one might expect, neither group is entirely consistent. In the native speaker group, 8 participants gave equal or higher ratings to collective possessors (P18 to P25), much like the majority of A2 and B1 learners. In the A2 and B1 group, 10 learners (2 of whom are A2 learners, namely P2 and P4) gave higher ratings to the *s*-genitive when the

possessor is animate rather than collective, in accordance with the main tendency among native speakers.

The animacy model also revealed that B2 learners do not distinguish as much between collective and inanimate possessors as native speakers. Whereas native speakers clearly prefer the *s*-genitive with a collective possessor rather than an inanimate possessor, B2 learners find the *s*-genitive more or less equally natural in both cases.

Figure 7.7 plots the average *s*-genitive rating for collective and inanimate possessors for B2 learners and native speakers separately.

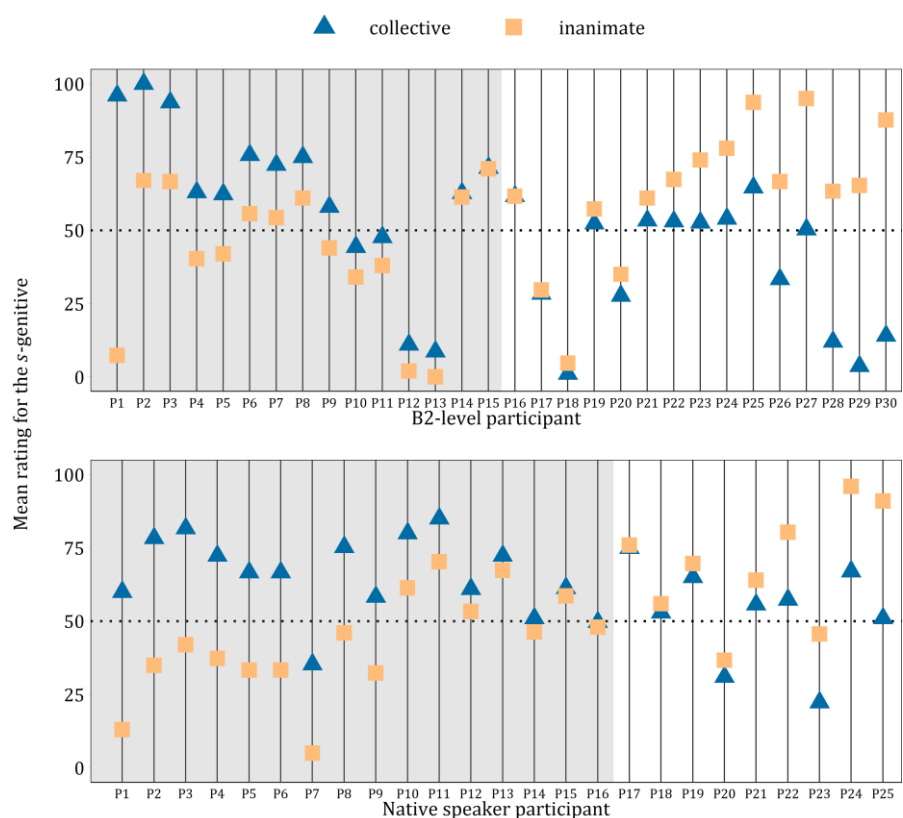


Figure 7.7: Individual mean ratings for collective and inanimate possessors for B2 learners (top) and native speakers (bottom). The participants are ordered from left to right based on how much higher their average rating is for collective possessors. Participants who give higher ratings to collective possessors are shaded.

The figure reveals a similar picture for this interaction effect as well. 16 out of 25 (64%) native speaker participants prefer the *s*-genitive when the possessor is collective rather than inanimate (P1 to P16). Half of the B2

learners do the opposite, so they either find both types of possessors equally natural in *s*-genitives or they prefer *s*-genitives with inanimate possessors (P16 to P30). Again, the 9 remaining native speakers follow the same pattern as these B2 learners (P17 to P25). Inversely, the other half of B2 learners resemble the majority of native speakers in that they also prefer the *s*-genitive with a collective possessor rather than an inanimate possessor (P1 to P15).

Correlations between predictions and ratings

The corpus prediction model found that native speakers give higher ratings to the *s*-genitive when the corpus-based model predicts that the *s*-genitive is more likely. Compared to native speakers, this tendency is less pronounced for A2 and B1 learners and to a lesser extent B2 learners, so the correlation between ratings and corpus-based predictions is supposedly weaker for these learners. The question is again whether this is the case for all relevant learners, so I consulted the raw rating data of the relevant participants and checked how well their 15 ratings correlated with the corpus-based prediction for the same 15 experimental item. Since each of the 37 items included in the experiment has a corresponding corpus-based prediction, this correlation is accurate across different versions of the experiment.

The result of this analysis is portrayed in Figure 7.8 for native speakers (see Jäschke & Plag 2016: 504 for a similar representation of rating data from individual learners).

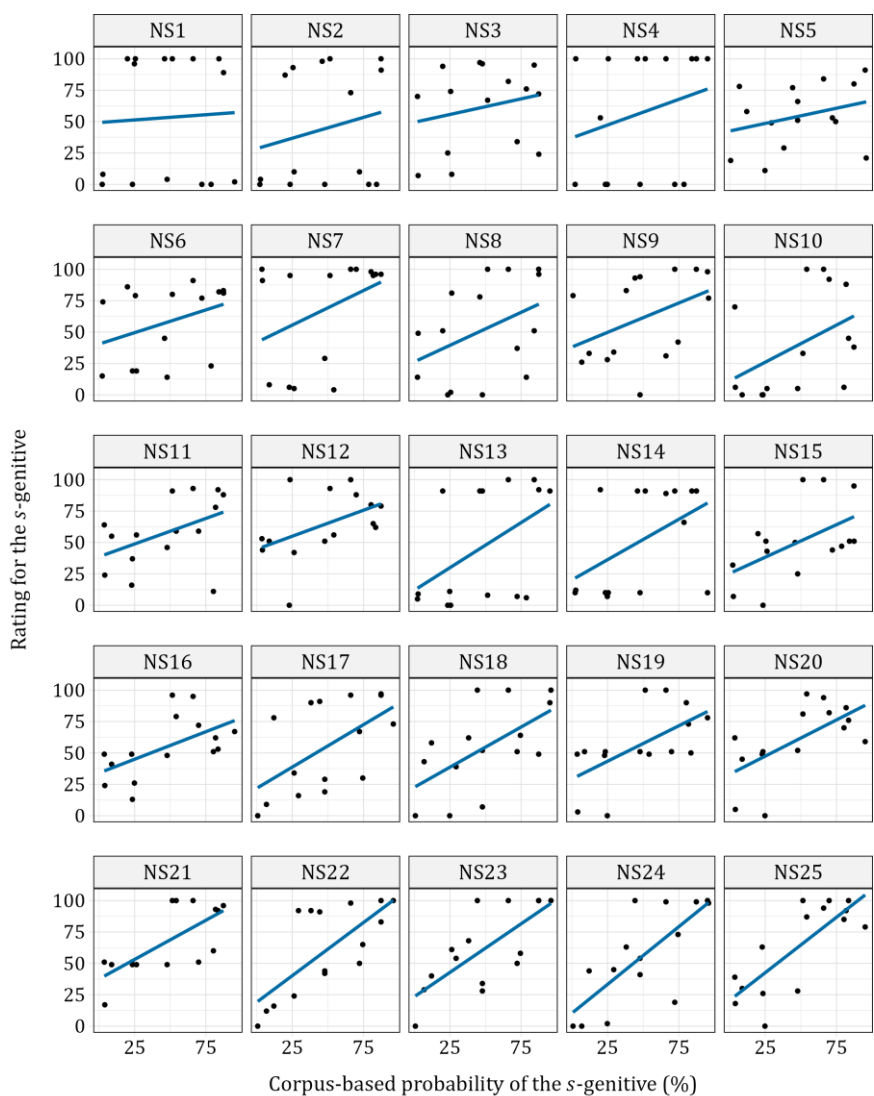


Figure 7.8: Individual native speaker participants' ratings for the *s*-genitive plotted against the corpus-based probability of the *s*-genitive (in %). Regression lines for participants whose ratings correlate positively with the corpus-based predictions are blue, whereas negative correlations are orange. The participants are ordered by the degree to which their ratings correlate positively with the corpus-based predictions.

The figure shows that the ratings of all native speakers correlate positively with the predictions from the corpus-based model, as expected.

A similar visualization is presented for A2 and B1 learners in Figure 7.9.

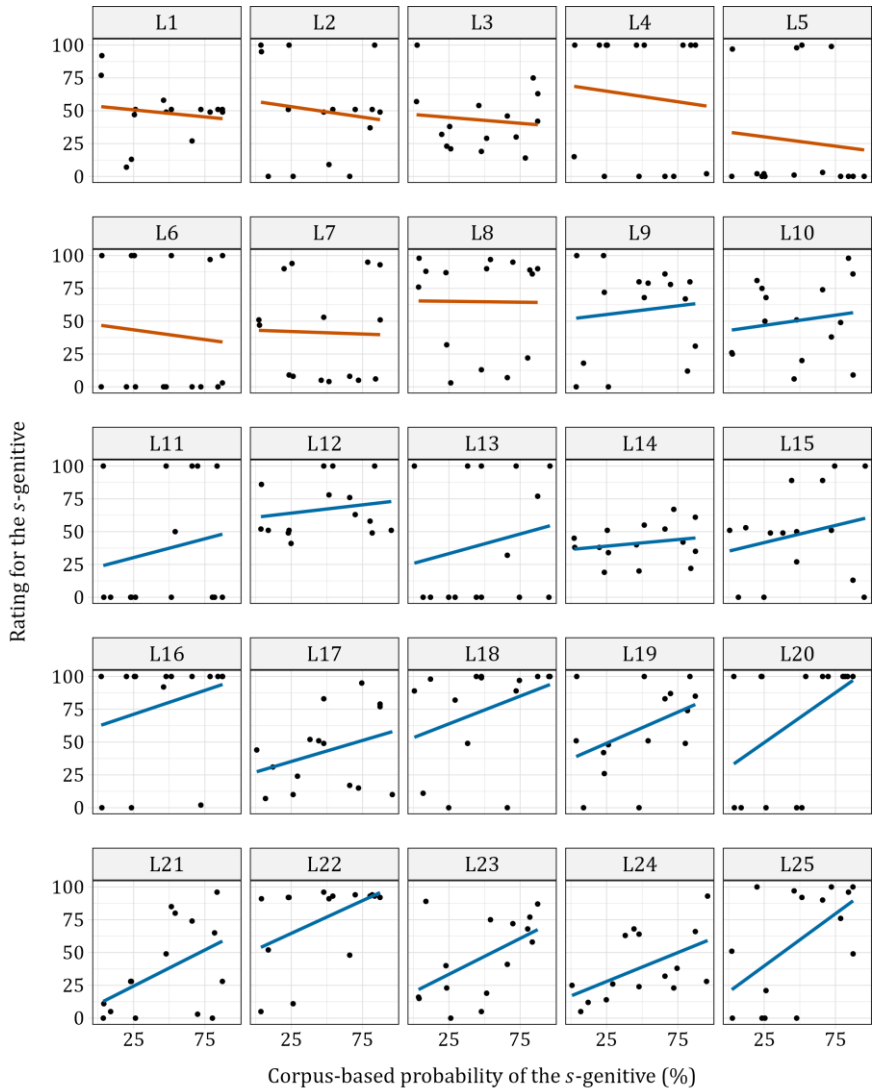


Figure 7.9: Individual A2/B1-level participants' ratings for the *s*-genitive plotted against the corpus-based probability of the *s*-genitive (in %). Regression lines for participants whose ratings correlate positively with the corpus-based predictions are **blue**, whereas negative correlations are **orange**. The participants are ordered by the degree to which their ratings correlate positively with the corpus-based predictions.

The figure shows that there is a positive correlation between ratings and corpus-based predictions for most A2/B1 learners (L9 to L25). This is not surprising considering that the overall correlation between ratings and corpus-based predictions for A2 and B1 is positive despite being weaker than for native speakers. However, the ratings from 8 out of 25 (32%) A2/B1 learners correlate negatively with the corpus-based predictions (L1

to L8), three of whom are A2 learners (L2, L4, L7). Therefore, it seems that some A2/B1 learners deviate considerably from the other learners, which might have influenced the aggregated results of the model. A similar visualization of the B2 data showed that the ratings from only one B2 learner correlate negatively with the corpus-based predictions.

7.4 Discussion

In this chapter, I conducted a rating task experiment designed to complement the corpus-based investigation of the genitive alternation in three ways. First, the experiment shed more light on the finding that B1 and B2 learners are less sensitive to the animacy of the possessor than native speakers. In the corpus-based investigation, the animacy constraint was entered as a binary distinction between animate and inanimate possessors because there was not enough data to include the five-way distinction between animate, collective, inanimate, locative and temporal possessors for which the data was annotated (see Section 6.2.3.6). With this binary distinction, it was not possible to determine whether B1 and B2 learners struggle with specific animacy categories. By contrast, the experiment was designed to test the effect of the five animacy categories on the ratings. Secondly, the corpus-based investigation mostly captures patterns of language use, whereas rating task experiments tap into the intuitions of speakers (Arppe & Järvikivi 2007: 132; Gilquin & Gries 2009; Klavan & Divjak 2016: 361), which arguably provide a better reflection of speakers' implicit grammatical knowledge. Finally, previous research has shown that there is considerable individual variability among speakers and especially learners (see Jäschke & Plag 2016; Kinne 2020). This variability could not be explored with the corpus data, mainly because speakers only contributed a few genitive observations each. By comparison, the rating data is more suitable as each participant rated 15 experimental items, which are balanced for possessor animacy and the probability of the *s*-genitive to boot.

On the methodological plane, the ratings from 25 native speakers and 101 EFL learners from eight different L1 backgrounds with a proficiency level ranging from A2 to C2 were analyzed with two separate mixed-effects linear regression models. The first model focused on the first research question presented in the introduction to this chapter regarding the effect of possessor animacy on the ratings of native speakers and learners compared to their language use. The second model is concerned with the second research question, which investigates how the intuitions from the participants in the experiment relate to the patterns of language use in general. Therefore, the second model tested whether the predictions from

the corpus-based model, which reflect spoken language use, correlate with the ratings of the participants in the experiment as this would provide additional evidence that the native speakers and learners are generally sensitive to the probabilistic constraints driving the choice of variant (Bresnan & Ford 2010). For the third research question, I also explored the individual variability among participants by calculating the consistency of the ratings across participants with the same proficiency level and comparing the ratings of individual participants with the aggregated results from both linear regression models.

In the remainder of this discussion section, I address each of the three research questions separately in light of the findings from the experiment. Given these findings, I then refine the theory-based account of the acquisition of the genitive probabilistic grammar.

7.4.1 How does possessor animacy affect the intuitions of learners at different proficiency levels in a rating task experiment compared to correlations between animacy and genitive variant in spoken EFL?

As expected, the corpus-based investigation of the genitive alternation revealed that native speakers prefer the *s*-genitive when the possessor is animate whereas inanimate possessors take the *of*-genitive. Surprisingly, this tendency is weaker for B1 and B2 learners. Since possessor animacy was conflated into a binary constraint for statistical analysis, the question arises as to whether certain underlying animate and inanimate possessors are particularly challenging for learners. Therefore, the experiment was designed to test the effect of possessor animacy using a more fine-grained distinction between animate, collective, inanimate, locative and temporal possessors.

Results from a mixed-effects linear regression model showed that native speakers find the *s*-genitive most natural when the possessor is animate, followed by collective, inanimate, locative and temporal possessors. This pattern aligns with previous research stating that the *s*-genitive is preferred with referents that are higher on the animacy scale (Rosenbach 2014: 232). By contrast, A2 and B1 learners find the *s*-genitive more or less equally natural when the possessor is animate or collective. Figure 7.4 suggests that this is due to A2 and B1 learners giving slightly higher ratings for collective possessors while at the same time giving lower ratings for animate possessors. B2 learners, on the other hand, find the *s*-genitive equally natural when the possessor is collective or inanimate, whereas native speakers prefer the *s*-genitive with collective possessor.

Figure 7.4 suggests that this is again due to a combination of B2 learners giving slightly lower ratings for collective possessors (see Table 7.6) and higher ratings for inanimate possessors.

Comparison with the corpus-based model

At first glance, the fact that A2/B1 and B2 learners do not clearly distinguish between certain animacy levels in their ratings compared to native speakers seems to replicate the corpus-based finding that B1 and B2 learners are less sensitive to the distinction between animate and inanimate possessors. Moreover, the ratings of native speakers and learners diverge regarding animate, collective and inanimate possessors, implying that there are indeed certain underlying animacy levels learners struggle with in particular. The more proficient C level learners no longer deviate from native speakers with regard to the effect of possessor animacy, which also ties in with the results from the corpus-based model.

However, a closer look at the results from both types of data reveals some important differences. To explain the results from the corpus-based model, I posited that the B1 and B2 learners experience difficulties with so-called “borderline cases” (Yamamoto 1999: 139–146), viz. collective and locative possessors. For one, locative and collective possessors are less distinctly associated with the genitive variants than animate, inanimate and temporal possessors, which affects cue reliability and complicates associative statistical learning (Gries & Ellis 2015: 236). Additionally, many nouns such as *region*, *Italy*, *village* etc. can be used to refer to either a place or a group of people depending on the context, so it can be difficult to categorize such possessors accurately. By extension, B1 and B2 learners who struggle with these categories are expected to use the *s*-genitive more often with locative possessors and the *of*-genitive more often with collective possessors than native speakers. In terms of ratings, this would be reflected by learners giving more similar ratings to collective and locative possessors than native speakers.

In this regard, the experiment certainly suggests that B2 learners find the *s*-genitive less natural than native speakers when the possessor is collective up until C1 level (see Table 7.6). This could indeed result from B2 learners misclassifying collective possessors like *region*, *Italy* as locative possessors, which prefer the *of*-genitive. However, this explanation does not account for the ratings of A2 and B1 learners as they do not deviate from native speakers regarding the effect of collective possessors. Interestingly, the participants also do not appear to experience difficulties with locative possessors as expected. Instead, A2/B1 and B2 learners struggle with animate and inanimate possessors respectively. This is surprising,

considering that the association between on the one hand animate possessors and the *s*-genitive and on the other hand inanimate possessors and the *of*-genitive is very strong: 82.8% of animate possessors in the data take the *s*-genitive and 95.8% of inanimate possessors take the *of*-genitive (see Table 6.7). Hence, these categories are very reliable and should therefore be highly learnable (Yoshimura & MacWhinney 2010; MacWhinney 2014; Gries & Ellis 2015), to the extent that it would be more advantageous to simply always using the *s*-genitive when the possessor is animate and the *of*-genitive when the possessor is inanimate, as suggested by prescriptivists (Murphy 2012; cited in Heller, Szmrecsanyi & Grafmiller 2017: 10). Furthermore, the visualization of the animacy effect in the partial effects plot in Figure 6.16 suggests that B1 learners tend to use the *s*-genitive with inanimate possessors. The same finding cannot readily be inferred from the rating data of A2 and B1 learners in the experiment, who do not appear to give particularly high ratings to the *s*-genitive with either inanimate, locative or temporal possessors (see Figure 7.4). If anything, Figure 7.4 points to the opposite pattern whereby A2 and B1 learners tend to favor the *of*-genitive more often with animate possessors.

Additionally, while a trade-off between locative and collective possessors seems plausible, it is unclear how learners would fail to differentiate between animate and collective possessors or collective and inanimate possessors, unless there are important distinctions beyond the five-level annotation of animacy that I did not take into account. In general, animacy is difficult to operationalize in all its facets, since the linguistic animacy distinctions are language-specific rather than biological and they are not diachronically stable (Zaenen et al. 2004; Wolk et al. 2013). Rather than a trade-off, the more likely explanation is that learners struggle with each of the categories independently.

The acquisition of independent categories

In principle, the effect of the animacy constraint in a given choice context is the result of a two-step process: speakers first determine the animacy of the possessor, and then choose the preferred genitive variant for this animacy level. For a possessor like *dog* for example, learners would first have to determine that the referent in the context is animate. Then, they would be inclined to produce this possessor in an *s*-genitive, because animate possessors prefer this variant.

Both steps could be at the origin of the dissimilarities between native speakers and learners. That is, learners might not yet have fine-tuned how the animacy categories affect the choice of variant, so they might not know that collectives prefer the *s*-genitive, but not as much as animate possessors

for example. Alternatively, they simply do not have enough experience with the linguistic elements, viz. the individual possessor noun phrases, to determine their animacy in a consistent way to begin with.

The latter explanation seems more plausible for two reasons. For one, what makes the animacy constraint more complicated than other constraints is not the fact that five distinct animacy levels are associated with the genitive variants; it is the fact that animacy is encoded by the head noun of the possessor, which can be realized by thousands of different lemmas. Secondly, low-proficiency learners might not even consider the overall animacy of the possessor when choosing between the genitive variants, because learners do not start out knowing which variant any of these categories prefer. Instead, they are exposed to genitive exemplars with thousands of different nouns, each of which has its own preferred genitive variant. Once they learn how many of these nouns pattern with the genitive variants via associative statistical learning, they may recognize that some of the nouns behaving the same way have their animacy as a common feature. However, abstracting over exemplars to generate prototypical features in this way is gradual and emergent (Abbot-Smith & Tomasello 2006; Ellis & Wulff 2019: 44). Hence, A2/B1 and B2 learners might choose a genitive variant based on what they know about the distribution of the noun in particular without considering its animacy level. In such an account, the animacy effect would be more akin to a lexical effect in the early stages of acquisition (see Section 6.4.3), and developing an understanding of the animacy categories might take a long time, as suggested by Divjak et al. (2021: 78).

If low-proficiency learners primarily struggle with individual nouns, then the most plausible explanation for the differences between native speakers and learners regarding the animate, inanimate and collective possessors is that the nouns included in the experiment for these categories are somehow problematic. According to usage-based approaches, language learning is driven by input (e.g., Ellis & Wulff 2019), so nouns that are encountered more frequently in the input should be easier to learn. To compare the frequency of the possessor nouns across animacy levels, I checked their ranking in the New General Service List (New-GSL, Brezina & Gablasova 2015). The expectation is that the animate, collective and inanimate possessors feature less frequent nouns on average than locative and temporal possessors, so they should be found lower on the list. The average ranking of animate (593), collective (688) and inanimate (970) possessors in the list is indeed lower on average than temporal possessors (289). Most locative possessors are country names, which are not included in the New-GSL, but it is highly likely that learners are also very familiar

with words like *America* and *India*, so overall it seems that temporal and locative possessors were represented by easier nouns in the experiment. Another difficulty with collective possessors in the experiment is that they mostly feature singular collective nouns with plural reference like *association*, *community* and *government*, which are potentially confusing for learners (e.g., Dziemianko 2008). Nonetheless, one would need to adopt a more sophisticated approach to collect more conclusive evidence considering that the overall frequency of a noun does not correlate perfectly with how frequently it is encountered as a possessor in genitive exemplars.

Regardless of the challenges involved with learning the animacy constraint, there are no longer any dissimilarities between native speakers and learners at the C1 and the C2 level. According to the theory-based account, these more proficient learners have been exposed to more input from the target language, and by extension, more genitive exemplars. As a result, they have encountered more nouns as possessors in genitive exemplars and they have learned which variant these nouns prefer through associative statistical learning. Over time, they managed to converge on the same animacy distinctions as native speakers by abstracting across many possessor nouns.

7.4.2 How does the relationship between experimental ratings and corpus data vary by the proficiency level of participants and speakers?

The corpus prediction model showed that there is a positive correlation between the predictions from the corpus-based model and the ratings in the experiment: when the corpus-based model predicts a high likelihood for the *s*-genitive, the participants in the experiment also give higher ratings for the *s*-genitive. These results suggest that the significant constraints included in the corpus-based model, i.e., the constraints that drive language use, are indeed part of speakers' implicit linguistic knowledge as posited by the Probabilistic Grammar framework (Bresnan & Ford 2010; Engel et al. 2022). However, this correlation is far from perfect ($r = 0.38$ ($p < 0.001$) (Bakdash & Marusich 2017)), which means that there is much variability in the ratings that is not explained by the corpus-based predictions.

On the one hand, this could mean that some of the constraints driving language use are not relevant during introspective tasks like the rating task experiment. Unfortunately, it is not possible to test the relevance of each constraint because the corpus-based prediction is calculated based on the effect of all significant constraints simultaneously, not to mention all constraints except for possessor animacy are controlled for. On the other

hand, this divergence between corpus-based predictions and ratings is to be expected given that introspection yields slightly different results than corpus data (Arppe & Järvikivi 2007: 152). Moreover, the strength of the correlation is comparable to recent studies that adopt the same approach for native speakers of English exclusively (see Engel 2022: 176; Engel & Szmrecsanyi 2023: 370).

The correlation between the corpus-based predictions and the ratings is weaker for the A2/B1 learners and to a lesser extent the B2 learners compared to native speakers. This difference could be explained in two ways. On the one hand, the linear regression model might have inferred that the ratings from the A2/B1 and B2 learners correspond less well to the corpus-based predictions because these learners tended to stick to the middle range of the rating scale, as implied by their flatter curve in Figure 7.5. One reason for this might for example be that A2/B1 and B2 learners are less confident about their choices than native speakers. In this case, their intuitions about the naturalness of the variants would in principle be similar to those of native speakers, but they did not want to give ratings on the extreme ends of the slider since a high rating for one variant implies that the other variant is much less natural given the design of the experiment. In other words, although learners might have a clear preference for one variant that aligns with the intuition of native speakers, they might not have wanted to exclude the other variant completely. To test this possibility, I compared to what extent the ratings of A2/B1, B2 and native speakers deviated from the middle point of the rating scale, namely 50. For example, if a learner gave a rating of 75 for the *s*-genitive, the deviation from 50 is 25. If A2/B1 and B2 learners stick more to the middle range of the rating scale, their deviations from 50 should on average be smaller than those of the native speakers. Wilcoxon tests revealed that neither A2/B1 learners ($p = 0.99$) nor B2 learners ($p = 0.34$) gave ratings that are significantly closer to the middle point of the scale than native speakers. Hence, the interaction effect is probably not due to these learners sticking to the middle range of the rating scale.

This supports the second possible explanation, which is that the ratings from A2/B1 and B2 learners correspond less well to the corpus-based predictions compared to native speakers because these learners do not always share the same intuitions as native speakers. More specifically, the intuitions from the native speakers in the experiment appear to correspond more closely to how native speakers and learners use the genitive variants in spoken language production. Put differently, it seems that A2/B1 and B2 learners are more likely than native speakers to produce a genitive variant that they do not find the most natural in a given context.

Specifically, the partial effects plot in Figure 7.5 shows that the intuitions of A2 and B1 learners diverge from those of native speakers both in contexts where speakers would use the *s*-genitive and in contexts where they would use the *of*-genitive in language production. In *s*-genitive contexts, A2 and B1 learners find the *of*-genitive more natural than native speakers whereas A2 and B1 learners find the *s*-genitive more natural than native speakers in *of*-genitive contexts. B2 learners share the same intuitions as native speakers in *of*-genitive contexts, but they differ from native speakers in *s*-genitive contexts, where they find the *of*-genitive slightly more natural.

The discrepancy between intuitions and language use in low-proficiency learners

The question arises as to why A2/B1 and B2 learners would produce genitive variants that they do not find the most natural in the rating task experiment. To some extent, this result can be explained by the fact that low-proficiency participants are less consistent in their ratings, which could attenuate the relevant statistical association in the model, and by extension, the regression slope in Figure 7.5. This is also partially due to the effect of the animacy constraint. In accord with previous literature, the corpus-based model found that the animacy of the possessor is one of the strongest determinants of genitive choice in spoken language production (Rosenbach 2014). Hence, the corpus-based prediction was largely determined by the effects of possessor animacy, with animate and collective possessors strongly favoring the *s*-genitive and inanimate, locative and temporal possessors taking the *of*-genitive. In the previous section, I explained that the effect of the animacy constraint influences the language use and the intuitions of the A2/B1 and B2 learners slightly differently, which can partially explain why the intuitions of these learners correspond less well to the corpus-based predictions as they are largely based on the way animacy operates in language use.

However, the effect of the animacy constraint does not overlap with the findings from the corpus prediction model (see Section 7.3.2) well enough to account for this finding in its entirety. For example, the corpus-prediction model showed that A2 and B1 learners find the *s*-genitive more natural in *of*-genitive contexts compared to native speakers. If this were due to how possessor animacy affects the ratings, then I would expect A2 and B1 learners to give high ratings for the *s*-genitive when the possessor is either inanimate, locative or temporal, which strongly favor the *of*-genitive in language use. Yet, the visualization of the animacy effect in Figure 7.4 does not indicate that these learners give particularly high ratings for the *s*-genitive in these contexts.

Therefore, another explanation is that spoken language involves various kinds of time-sensitive, processing-related factors that influence language production and might cause a speaker to produce a variant that they find less natural but that is easier to process in a given context. These restrictions did not apply during the experiment, so the participants could always choose the variant they found more natural. This led to differences between the intuitions of the participants and their language use, especially for low-proficiency learners for whom processing-related considerations during spoken language production are arguably more important given that their language production is in all likelihood less efficient and automatized than that of more advanced learners and native speakers (Segalowitz 2003).

In reality however, most processing-related constraints are controlled for during the experiment and previous research including the corpus-based model showed that even low-intermediate learners are surprisingly similar to native speakers when it comes to the actual use of the genitive variants in spoken language production (see Gries & Wulff 2013). Therefore, one cannot exclude the alternative possibility that low-proficiency learners differ from native speakers because of factors that are specific to the rating task experiment. As stated above, the ratings feature much unexplained variability that does not follow from the patterns of spoken language use, especially at the lower-proficiency levels. Since rating task experiments offer the participant more freedom than forced choice tasks for example, learners might have a different interpretation of what sounds natural to them as they complete the experiment. Future research should therefore investigate what these ratings capture exactly at low-proficiency levels compared to higher proficiency levels and native speakers. In any case, learners no longer differ from native speakers in any context as they become more proficient and reach the C proficiency levels, either because their intuitions and language use converge, or because they complete the rating task experiment in the same way as native speakers.

7.4.3 How do the ratings of individual participants compare to the aggregated results regarding possessor animacy and the relation between intuitions and language use?

First, an investigation of the internal consistency of the ratings at each proficiency level showed that all participants exhibit excellent internal consistency, except for the A2 and B1 learners who are more heterogeneous, meaning that they do not share the same intuitions about the naturalness of the genitive variants in the experimental items. This supports the assumption that learners exhibit much individual variability (Jäschke &

Plag 2016: 511–512; Kinne 2020: 128) and further suggests that the aggregated results from both mixed-effects linear regression models on the ratings might not adequately capture the behavior of all individual participants.

The aggregated results from the animacy model showed that A2 and B1 learners distinguish less clearly between animate and collective possessors than native speakers. Whereas native speakers find the *s*-genitive more natural when the possessor is animate rather than collective, A2 and B1 learners find the *s*-genitive more or less equally natural with both types of possessors. Likewise, native speakers prefer to use the *s*-genitive with collective possessors rather than inanimate possessors, but B2 learners again find both types of possessors equally fitting for *s*-genitives.

An exploration of individual participants' ratings indicated that most participants align with the results of the model, so most A2 and B1 learners either give similar ratings for animate and collective possessors, or they might even find the *s*-genitive more natural with a collective possessor. Likewise, half of the B2 learners rate the *s*-genitive the same or higher when the possessor is inanimate rather than collective. However, among both groups of learners there are individuals who approximate the behavior of native speakers in the model: more than a third of the A2/B1 learners favor the *s*-genitive when the possessor is animate rather than collective, and half of the B2 learners find the *s*-genitive more natural with collective rather than inanimate possessors. If the acquisition of the animacy constraint is indeed largely a matter of lexical knowledge at first, then the learners whose ratings approximate those of native speakers are supposedly more familiar with the lexical items that figured in their version of the experiment than the learners who deviate from native speakers.

Importantly, native speakers are also not perfectly consistent in this regard since a few native speakers rated the *s*-genitive higher with collective possessors than animate possessors, much like A2 and B1 learners. Like B2 learners, a third of the native speakers found the *s*-genitive equally natural with either inanimate or collective possessors.

As for the corpus prediction model, results showed that all speakers find the variant predicted by the corpus-based model more natural in the rating task experiment, but this correlation is weaker for A2/B1 and to a lesser extent B2 learners compared to native speakers. A closer look at the correlations between ratings and corpus-based predictions for each participant showed that there is considerable inter-speaker variability. Firstly, there are many A2 and B1 participants whose ratings exhibit a rather weak correlation with the corpus-based predictions, as expected from the model results. For other participants, the ratings correlate more

strongly with the corpus-based predictions. Like the native speaker participants, the intuitions of these learner participants overlap more closely with the language use of the speakers in the corpus. Contrary to the model results, some A2 and B1 participants' ratings even correlate negatively with the corpus-based predictions, implying that the intuitions of these learners go against the patterns of language use. By comparison, B2 learners appear to align more closely with the results from the model, which again indicates that inter-speaker variability diminishes with proficiency level.

In general, this investigation into the ratings of individual participants showed that there is much inter-speaker variability that cannot be interpreted from the aggregated results of the model to the extent that some results cannot readily be generalized to all learners, especially at the lower proficiency levels. Provided learners build an interlanguage grammar that accommodates their language input in a rational way (Ellis 2006: 103), the learners' behavior ultimately depends on the exemplars they have encountered in their previous language experience (Ellis & Wulff 2019). Because low-proficiency learners have less experience with the target language, their interlanguage grammar is built on fewer exemplars that can be categorized in various ways, which naturally leads to more variability across individual learners. Overall, the findings suggest that the model results partially depend on the distribution of these specific strategies among the relevant speakers. Nonetheless, these conclusions are based on rather simple comparisons of the raw ratings across different versions of the experiment, so they remain rather tentative.

7.4.4 Revisiting the genitive alternation

A limitation of many studies on alternation phenomena in EFL including the study of Gries and Wulff (2013) on the genitive alternation is that they exclusively rely on corpus data. Yet, there are many well-documented benefits to including behavioral data from experiments, some of which have become apparent in this analysis of the genitive alternation with a rating task experiment.

First of all, the experiment helped disentangle multiple possible interpretations of the corpus-based finding that B1 and B2 learners are less sensitive to possessor animacy than native speakers (see Section 6.3.2.2). Supposedly, the animacy constraint is challenging to learn via associative statistical learning because animacy is encoded by the highly variable possessor noun phrase. Specifically, I argued that locative and collective possessors are particularly problematic because certain nouns can fit both

categories depending on the context. However, this could not be confirmed with the corpus-based analysis because animacy was conflated into a binary variable, contrary to experiment, which was designed to test the effect of five animacy levels on the ratings. The analysis of the experiment data revealed that while collective possessors are indeed problematic for EFL learners, locative possessors are not. Instead, it is animate and inanimate possessors, supposedly the most learnable categories, which led to non-native like ratings from A2/B1 and B2 learners. Given that it is unlikely that learners somehow confuse inanimate and collective nouns, I argued that learners do not necessarily struggle with classifying nouns in the right animacy category, simply because low-proficiency learners might not be aware that the animacy of the noun is associated with the choice of genitive variant to begin with. Instead, learners first acquire the association between concrete possessor nouns and the genitive variants before gradually realizing that animacy is an important underlying factor the nouns have in common (see Abbot-Smith & Tomasello 2006). Overall, this interpretation of the experiment results further suggests that the acquisition of the animacy constraint is largely a matter of lexicogrammatical knowledge at the early stages.

Secondly, the experiment catered to some of the limitations of corpus-based models, most notably that corpus data cannot reveal what speakers know about the language (Ellis 2017: 46; see also Schönefeld 2011: 13; Kinne 2020: 173). By contrast, the experiment is based on introspection and therefore provides another perspective on the genitive alternation (Arppe & Järvikivi 2007: 152) that is arguably more representative of speakers' implicit grammatical knowledge. The experiment showed that the predictions from the corpus-based model correlate with the intuitions of the participants in the experiment, providing additional evidence that the constraints operating on speakers' language use derive from their implicit grammatical knowledge (Bresnan & Ford 2010). Moreover, converging evidence from both types of data increases the robustness of the corpus-based findings (e.g., Hoffmann 2006: 169; Schönefeld 2011: 20–24). Nonetheless, the correlation between participants' intuitions and the language use of the speakers in the corpus is far from perfect, especially for low-proficiency learners. This raised the question as to whether these learners approached the rating task experiment differently than native speakers and more advanced learners. Additionally, these learners might also be more strongly affected by prescriptivism. To gain more insight into these questions, it would have been useful to ask participants at the end of the experiment when they think it is appropriate to use the *s*-genitive and the *of*-genitive.

Finally, the experiment data was more suitable for an exploration of the individual differences between participants. While this part of the analysis relied on rather simple descriptive statistics and visualizations, results point to considerable inter-speaker variability both in terms of individual constraints like possessor animacy and in terms of the overall correspondence between intuitions and language use. That is, there are multiple learner strategies within each proficiency level including native speakers, so that it is entirely possible for individual learners to approximate native speakers even though the model finds dissimilarities at the group level. From this perspective, one might argue that the model results partially depend on the distribution of these strategies at a given proficiency level.

8 General discussion

This chapter presents a general discussion of the findings from both alternations. First, I revisit the three research questions of the project. This revealed certain discrepancies between both alternations, which I interpret in Section 8.2. Given the insights gathered from the analysis of both alternations, I then revisit the theory-based account of the acquisition of probabilistic grammars in Section 8.3. In Section 8.4, I discuss some pedagogical implications of the findings and I conclude the chapter in Section 8.5 with an evaluation of the methodology I adopted to address the research questions of the thesis.

8.1 The research questions

8.1.1 How does the set of constraints of learners compare to that of native speakers at different proficiency levels?

Hypothesis 1 *As proficiency increases, learners better approximate the target probabilistic grammar.*

Learners are able to reach a higher proficiency level on official language tests like the GESE exam featured in the TLC (Gablasova, Brezina & McEnery 2019) because they have developed better communicative competence in English, which requires language learning. According to usage-based approaches to SLA (e.g., Ellis & Wulff 2019), language learning of any kind is driven by input from the target language, so more proficient learners have received more input. By extension, they have been exposed to more exemplars of the future markers and the genitive variants whose construction features can be inferred via associative statistical learning (e.g., Romberg & Saffran 2010). Probabilistic constraints encompass one or more of these construction features. For the case studies, I argued that if the probabilistic grammar of a learner does not differ from that of native speakers with regard to any of the relevant constraints, then the learner approximates the target probabilistic grammar of native speakers. With this in mind, the hypothesis implies that lower-proficiency learners should deviate from native speakers with regard to more constraints than more proficient learners.

The number of significant differences between native speakers and learners at different proficiency levels for the future marker alternation and the genitive alternation is shown in Table 8.1.

Table 8.1: Number of significant interaction regressors between native speakers and learners across proficiency levels for the future marker alternation (Table 5.17) and the genitive alternation (Table 6.16).

<i>Proficiency</i>	<i>Future marker alternation</i>	<i>Genitive alternation</i>
B1	4	2
B2	7	1
C1	4	0
C2	2	0

Table 8.1 shows that this hypothesis was largely confirmed in both case studies, so learners who are less proficient typically diverge from native speakers with regard to more constraints than more advanced learners. One potential counterexample comes from the B1 learners for the future marker alternation, who deviate from native speakers with respect to fewer constraints than the more proficient B2 learners. Nonetheless, B2 learners use *be going to* more often than B1 learners and therefore better approximate the distribution of the variants in the target language: native speakers use *be going to* in slightly more than one third of the cases, whereas B1 learners only use *be going to* one in ten times (see Table 5.20). While the present study is mainly concerned with the probabilistic conditioning of the variants, results strongly suggest that the acquisition of the probabilistic grammar associated with the future markers is linked to such an increase in the distribution of *be going to*. This, in turn, implies that learners mainly try to figure out when *be going to* is used when they already have *will* at their availability (see Bardovi-Harlig 2005).

Table 8.1 also highlights one of the important dissimilarities between both alternations, namely that the future marker alternation is apparently harder to master than the genitive alternation, given that learners at the C1 level still exhibit important differences with native speakers. Another important contrast, discussed in more detail in the next section, is that learners have to remove certain constraints to approximate the target probabilistic grammar associated with the future markers, whereas mastering the genitive alternation requires learners to acquire additional constraints or strengthen existing ones. Ultimately, C2 learners reach the target probabilistic grammar with all its intricacies, except for some persistence effects in the future marker alternation.

Hypothesis 2 *Learners are only sensitive to constraints that are part of the target probabilistic grammar.*

According to usage-based approaches to SLA such as the CREED model, the interlanguage of learners accommodates the target language input they receive in a rational way (Ellis 2006: 102–103). This means that if the language input of learners exposes the tendency for *be going to* to co-occur with second-person subjects and learners implicitly pick up on this, they will not go against this input by consistently using *will* with second-person subjects instead. By extension, I hypothesized that they should only consider constraints that are also present in the language input they receive. Supposedly, the learners want to learn British English given that they participate in a British English language exam, so the learners should only be sensitive to constraints that also affect the language use of the examiners in the corpus, who are native speakers of British English. Because processing-related constraints are not acquired according to the theory-based account, this hypothesis only concerns feature-based constraints.

The analysis of the genitive alternation supports this hypothesis: learners only consider constraints that form part of the probabilistic grammar of native speakers. When it comes to the future markers, B1, B2 and C1 learners are all influenced by constraints that do not condition the choice of variant for native speakers, namely the presence of temporal adverbs, the type of verb and the temporal proximity of the future event. Since the language system of second language learners is tuned to their L1 (Ellis 2006: 109), these constraints could in principle be the result of transfer from learners' L1 (Rosenbach 2017). Given the typological diversity of the L1s included in the data, I argued that it is unlikely that transfer effects from a single L1 background could produce significant effects for these constraints, except for Spanish L1, which accounts for a large chunk of the data. Yet, supplementary analyses showed that these effects could also not be attributed to Spanish EFL learners. At the same time, the typological diversity of the L1s makes it unlikely that they would induce the same transfer effects (see Section 5.4.2). Therefore, I proposed two other explanations for this surprising finding.

First, I argued that these constraints result from learners restricting the usage of *be going to* in particular. When learners acquire *be going to*, they already have at their availability the future marker *will*, which is not only more frequent, it is arguably easier to process as well since it does not require learners to inflect the auxiliary *be* (Bardovi-Harlig 2002; Bardovi-Harlig 2005). Adding *be going to* with largely the same function as *will* would be redundant and fill a primarily expressive need (see Ellis 1999:

470). This expressive need might not be the priority of learners, who instead decide to use *be going to* specifically for future events that are relatively certain to happen in the near future. The reason why learners from a wide variety of L1s are able to converge on these additional constraints is because they relate to inherent characteristics of future events. Each future event inherently relates to a point in future time and they always involve a degree of uncertainty because future events have not yet happened, which may influence the occurrence of temporal adverbs.

A second possibility is that learners have inferred these constraints from EFL textbooks. Future-time reference is an important topic of such textbooks, to the extent that Burton (2023) claims it is part of the ‘canon’. These textbooks can provide a large chunk of the target language input for language learners (see Le Foll 2023: 259–260), although the rules they teach often do not correspond to either findings from the field of second language acquisition (Mishan 2022: 497) or authentic corpus data (Burton 2023: 25). In a selection of textbooks, I found that some of these prescriptivist rules are reflected by the probabilistic grammar of learners. In this interpretation of the findings, learners are taught to use *be going to* in specific cases as described by rules like “use *be going to* for future events with present indications”, which has been tied to the tendency for *be going to* to be favored for temporally proximate events (Haegeman 1989; Brisard 2001: 266). As learners become more proficient, they realize through exposure to the target language that native speakers do not follow these rules, so they gradually remove these restrictions on the use of *be going to*. Nonetheless, learners’ language use deviates from these textbooks in important respects, which calls the impact of these prescriptivist rules into question.

Hypothesis 3 *Lower-proficiency learners are sensitive to fewer constraints than more advanced learners and native speakers.*

If one assumes that language learning is rational (see Ellis 2006: 103) and learners only infer the constraints that are present in the output of native speakers, then it follows that the maximum number of constraints learners can implement is determined by the number of constraints present in their input, viz. the language production of native speakers. According to previous research on cues, learners acquire and adjust a given cue before turning to the next cue (Ellis & Sagarra 2010; Ellis & Sagarra 2011: 601). If the acquisition of probabilistic constraints involves learning construction features that resemble these cues, then I argue that probabilistic constraints are also learned one by one. As a result, low-proficiency learners who have

not yet acquired some constraints will be sensitive to fewer constraints than native speakers. Since the probabilistic grammar of learners features fewer constraints, their probabilistic grammar is simpler (Shin 2014).

Again, this hypothesis can be maintained for the genitive alternation. The least proficient B1 learners are not sensitive to the definiteness of the possessor, contrary to more proficient learners and native speakers, while the other constraints are shared by all speakers. Hence, the choice of genitive variant is conditioned by fewer constraints for B1 learners compared to more advanced learners and native speakers. The same cannot be said of the future marker alternation, because B1, B2 and C1 learners are sensitive to constraints that are not present in the output of native speakers. As a result, the choice of future marker is constrained by more factors in EFL and *be going to* in particular is restricted to a limited number of contexts. In this respect, I showed that the choice of variant is more predictable in EFL than for native speakers, who do not adopt such a pronounced distinction between both variants. While this suggests that learners' probabilistic grammar is more complex (Shin 2014), this more stringent set of constraints does not necessarily mean that learners incur more processing costs. For one, learners simply stick to *will* nine times out of ten because the constraints that conspire in favor of *be going to* rarely co-occur in their language production. Moreover, a stronger distinction between the future marker could facilitate the choice of variant, given that learners tend to deviate from native speakers mostly in contexts where neither variant is clearly preferred (see Gries & Deshors 2015; Wulff & Gries 2019).

8.1.2 How does proficiency level modulate the probabilistic grammar of learners?

Hypothesis *Learners first acquire constraints that facilitate associative statistical learning.*

The learnability of cues and their association with a co-occurring construction via associative statistical learning depends on a number of criteria. A cue that is more frequent, reliable, regular and adjacent to the construction is easier to learn than a cue that is infrequent, unreliable, irregular, and non-adjacent (e.g., MacWhinney 2001; Romberg & Saffran 2010: 2–3; Yoshimura & MacWhinney 2010: 552; Misyak & Christiansen 2012: 304; Yi, Lu & Ma 2017: 520). The construction features covered by the probabilistic constraint share many characteristics with cues, so I assume the same learnability criteria apply. The learnability of the factors

was operationalized as the association strength between the feature-based constraints and the variant constructions in the language production of native speakers according to monofactorial tests conducted in Section 5.2.2 for the future marker alternation and Section 6.2.3 for the genitive alternation.

At first glance, the hypothesis is untenable for both alternations. For the future marker alternation, the type of clause and the temporal proximity of the future event are the probabilistic constraints most strongly associated with the choice of future marker in the language production of the native speakers. While clause type is indeed one of two probabilistic constraints where the effect on learners and native speakers is similar, this is not the case for temporal proximity. Instead, it is the grammatical person of the subject that appears to be more learnable, with both speaker populations preferring *be going to* with second person subjects. For the genitive alternation, the animacy of the possessor and the semantic relation that holds between the possessor and the possessum are most distinctly associated with the genitive variants, yet the only feature-based constraint that operates on the variants in the same way for learners and native speakers is the final sibilancy of the possessor. In fact, possessor animacy is arguably the least learnable constraint given that it is the only constraint learners struggle with until they reach the C level.

Upon closer inspection, I argued that the operationalization of learnability based on association strength as a proxy for the reliability of cues or construction features is probably too simplistic. For one, a monofactorial assessment of the strength of individual constraints in isolation from other factors is not ideal given the inherently multifactorial nature of alternation phenomena. This is evidenced by the fact that neither temporal proximity nor semantic relation was ultimately part of the probabilistic grammar of native speakers according to the mixed-effects logistic regression models, in spite of the fact that these constraints were most distinctly associated with the genitive variants according to monofactorial tests. More importantly, cue reliability or cue-outcome contingency only captures one aspect of learnability.

Rather than rejecting the hypothesis, I argued that a more in-depth assessment of statistical learnability can explain the order of acquisition. For instance, B2 and C1 learners struggling with the influence of negation on the choice of future marker can be explained by the fact that negators are rather infrequent. Besides the frequency of construction features, the most important factor that underlies the learnability of a probabilistic constraint as a whole seems to be its variability, i.e., the number of linguistic elements it encompasses: probabilistic constraints that involve more

linguistic elements whose association with the variants has to be learned are more challenging to master.

For the future marker alternation, this explains why learners differ from native speakers regarding the effect of verb type, temporal proximity and temporal adverbs. There are hundreds of verbs, various temporal adverbs, and many ways to signal the temporal proximity of a future event including contextual cues. At the same time, temporal adverbs and other features related to temporal proximity are often not adjacent to the future marker, which also complicates associative statistical learning (see Gómez 2002). By comparison, the linguistic elements covered by the clause type constraint come from a rather limited set of invariable subordinators. To learn the preferences of apodoses and protases for example, learners only need to pay attention to the subordinator *if*. From this perspective, it is also not surprising that the grammatical person of the subject is highly learnable as the effect in question concerns the second-person pronoun *you*, which is adjacent to the future marker, highly frequent, and most importantly, invariable.

The same rationale accounts for the findings from the genitive alternation. Although possessor animacy is the strongest constraint of the genitive alternation (Rosenbach 2014), animacy is encoded by the noun without any overt markers. If each noun is considered a linguistic element, then learners have to deal with thousands of linguistic elements before they can accurately replicate the target probabilistic grammar for this constraint. Importantly, learners are not aware of the association between the genitive variants and possessor animacy because this is not directly apparent from the language input learners receive. Instead, learners are simply exposed to concrete possessors in genitive exemplars, each with their preferred variant as evidenced by the random effects in the logistic regression model (see Section 6.3.1). For example, learners would notice that a possessor like *dog* prefers the *s*-genitive, whereas a place like *street* prefers the *of*-genitive. Gradually, learners generalize across exemplars to recognize that possessors that follow similar distributions share the same type of animacy (see Abbot-Smith & Tomasello 2006; Ellis & Wulff 2019: 44). In principle, the same process underlies the acquisition of verb type in the future marker alternation: learners first pick up on the preferences of individual verbs (see Section 5.3.1) before applying an abstract distinction between stative and dynamic verbs based on inference across many of these verbs. Compared to possessor animacy, final sibilancy involves a very limited number of sounds as linguistic elements, which facilitates associative statistical learning. Possessor definiteness, on the other hand, was argued

to be difficult to acquire because it is indeed a rather weak constraint (Heller 2018: 104).

In short, a comprehensive assessment of the learnability of constraints via associative statistical learning can account for the order of acquisition for both alternations under study. However, future research should try to find a way to properly operationalize learnability to generate more accurate a priori predictions about the order of acquisition, which would ideally take into account association strength, frequency, adjacency and the number of linguistic elements involved. Moreover, it would be interesting to track the acquisition of the constraints for less proficient learners as well, since B1 learners are already highly similar to native speakers for the genitive alternation.

8.1.3 How do native speakers and learners at different proficiency levels compare regarding processing-related constraints?

Hypothesis *Lower-proficiency learners are more sensitive to processing-related constraints than more advanced learners and native speakers.*

In this study, I adopted a strict definition of processing-related constraints inspired by the criteria Tamminga, MacKenzie and Embick (2016) use to identify psychophysiological constraints. What distinguishes processing-related constraints from feature-based constraints is that they do not involve linguistic elements nearby the choice context, so associative statistical learning cannot take place. Moreover, these constraints relate to features of the brain, such as working memory etc., so learners do not have to learn these constraints. In fact, I hypothesized that they should be more sensitive to these constraints in accordance with findings from previous research (e.g., Wulff, Lester & Martinez-Garcia 2014; Deshors & Gries 2014; Wulff 2016; Wulff, Gries & Lester 2018; Wulff & Gries 2019; Paquot, Grafmiller & Szmrecsanyi 2019) because learners have to juggle between two or more language systems (Kroll & Dussias 2013; cited in Wulff 2016: 133), or because production is less automatized for especially low-proficiency learners compared to native speakers (Segalowitz 2003).

In the genitive alternation, possessor thematicity and discourse status were not relevant, probably because their operationalization is not ideal to capture the kind of short-lived lexical priming effects these constraints could induce (Hartsuiker et al. 2008). As for the length of the constituents

and structural persistence, they do not have a significantly stronger effect on the outcome for learners compared to native speakers.

In the future marker alternation, the effect of clause length, which influenced C1 learners more than native speakers, can be retraced to one particular part of the exam, namely the presentation task. Hence, this finding is not generalizable and probably not representative of spoken EFL given that learners were allowed to prepare their presentation in advance (Gablasova & Brezina 2015: 122; Gablasova et al. 2017: 618–619). The effect of structural persistence, on the other hand, was much more pronounced for especially low-proficiency learners, to the extent that the strongest effect conditioning the choice of variant for B2 learners is their tendency to use *be going to* after an examiner. Nonetheless, the fact that these strong persistence effects for *be going to* quickly diminish at the C1 level raises the question as to whether these results are strictly processing-related, or whether they are partially due to B1 and B2 learners' limited understanding of *be going to* in particular. Given that B1 and B2 learners only use *be going to* in a limited number of cases, hearing the examiner, who is a native speaker of British English, use *be going to* in a much wider range of contexts leads to more surprisal and a higher activation level for *be going to* (Ellis & Wulff 2019: 47–49). As a driver of implicit language learning, these persistence effects could possibly help refine the constraints (see e.g., Bock et al. 2007: 455–456 and Tooley & Bock 2014: 113–114). If the effect of structural persistence is stronger for variant constructions that learners do not master well, then it makes sense that structural persistence does not play such an important role in the genitive alternation, where even low-proficiency B1 learners nearly approximately native speakers' use of the variants.

While neither study provided much support for the hypothesis, I adopted a rather restrictive definition of processing-related constraints compared to previous research, which did not specify which constraints should be considered processing-related. In fact, one could argue that if language processing is incremental (Timmers et al. 2013: 10; Philipp et al. 2017), then virtually all constraints can be said to relate to processing given that they influence the choice between semantically equivalent structures during language production. For example, a referent's animacy can be considered processing-related in the context of alternations (Branigan, Pickering & Tanaka 2008; see Röthlisberger, Grafmiller & Szmrecsanyi 2017), even though animacy is semantic-pragmatic in nature (see Croft 2003: 128–132; Corbett 2010; Rosenbach 2014: 238).

Rather than positing that learners should be *more* sensitive to processing-related effects, the theory-based account licenses a weaker version of this hypothesis:

Hypothesis *Learners are sensitive to the relevant processing-related constraints in the target probabilistic grammar.*

Since processing-related constraints derive from features of the brain, they do not have to be learned contrary to feature-based constraints. Instead, these universal effects apply to learners by default, so they should at least be sensitive to the same processing-related constraints as native speakers. This weaker version of the hypothesis holds true for both alternations. In the genitive alternation, all learners are sensitive to the processing-related constraints that also influence native speakers, namely structural persistence, the length of the constituents and the lexical density of the environment surrounding the choice context. Similarly, the only processing-related constraint affecting native speakers' choice of future marker, viz. structural persistence, also plays a role for even the least proficient learners in the data.

8.2 Discrepancies between the alternations

The development of the probabilistic grammars underlying both alternations in spoken EFL is similar in many regards. In both cases, learners better approximate the target probabilistic grammar as they become more proficient, which can be attributed to the fact that more proficient learners have received more input filtered through general learning mechanisms (e.g., Ellis & Wulff 2015: 76; Ellis & Ogden 2017: 607; Yi, Lu & Ma 2017: 520; Wulff & Ellis 2018: 37; Ellis & Wulff 2019: 41). Moreover, learners are sensitive to the same processing-related constraints that also play a role in the target probabilistic grammar for both alternations. Finally, the alternations have in common that learners first acquire the feature-based constraints that are easy to learn via associative statistical learning. Specifically, one of the most important factors impacting the learnability of a constraint besides its reliability is the number of linguistic elements it involves: constraints that encompass fewer linguistic elements are easier to learn than constraints that are realized by many different linguistic elements.

Nonetheless, the studies also uncovered some important dissimilarities between the alternations. In the genitive alternation, learners are only sensitive to the constraints that form part of the target

probabilistic grammar of native speakers. This ties in with the hypothesis that language learning is rational (Ellis 2006: 102–103): the inferences learners make about the use of the variants accommodates the input they receive in a rational way. Moreover, low-proficiency learners in the genitive alternation also adhere to the hypothesis that learners are sensitive to fewer constraints than the target probabilistic grammar. This follows from the assumption that the construction features are learned one-by-one like cues (see Ellis & Sagarra 2010; Ellis & Sagarra 2011), so in particular low-proficiency learners are not yet sensitive to all the probabilistic constraints present in their input. In short, the use of the genitive variants is not exactly the same for learners and native speakers, but these dissimilarities are mostly predictable.

By comparison, the development of the probabilistic grammar regulating the choice between the future markers deviates from the hypotheses in major aspects. Most importantly, learners are sensitive to constraints that are not present in the target probabilistic grammar, which calls into question the hypothesis that their probabilistic grammar is inferred rationally from the output of native speakers. By extension, they were found to rely on more constraints than native speakers, which also runs counter to the predictions of the theory-based acquisition account. A third difference is that learners somehow take longer to master the future markers compared to the genitive variants, which implies that the future markers represent a more challenging aspect of the English language for learners (see also Haegeman 1989: 292; citing Wekker 1976: 123). How one accounts for these discrepancies between both alternations depends on one's explanation for the findings from the future marker alternation.

***Be going to* for specialized future contexts**

The first explanation is that learners differentiated more strongly between *will* and *be going to* than native speakers because *be going to* is arguably a redundant addition to learners' repertoire of future forms, which already contains the more frequent, semantically equivalent and invariant form *will* (see Bardovi-Harlig 2005). Instead of inferring patterns of language use from their target language input, which would have primarily catered to an expressive need to have multiple options to convey future tense (see Ellis 1999: 470), learners opted for a functional distinction between the future markers based on inherent characteristics of future events. The question is why learners do not discern more clearly between the genitive variants as well. For example, one might have expected learners to distinguish between the variants based on the semantic relation that is conveyed (see Section 6.2.3.7).

In this regard, I argue that the genitive alternation offers much less flexibility than the future marker alternation because it is a permutation alternation, meaning that the variants involve a change of word order (De Troij, Grondelaers & Speelman 2023). Word order is closely related to principles like Easy First (MacDonald 2013), which states that in language production, linguistic elements that are easy to process are preferably produced before more demanding elements. From this perspective, all constraints that impact the processing demands of the constituents could in principle be considered processing-related. This concerns even semantic-pragmatic features of referents like animacy (see Croft 2003: 128–132; Corbett 2010; Rosenbach 2014: 238), because animacy influences the accessibility of referents and by extension their processing demands (Branigan, Pickering & Tanaka 2008). Hence, the main function of the variants in such permutation alternations is to accommodate language processing by offering multiple word orders that allow the speaker to satisfy these processing principles. By extension, the genitive variants ensure production fluency, which could explain why learners are so quick to approximate the target probabilistic grammar.

In the future marker alternation, the influence of processing factors is less apparent, even if one adopts a looser definition than the one used here (see Section 3.2.1). For one, it is unclear which variant is easier to process: while *will* is arguably simpler because it does not involve the inflection of the auxiliary *be* (Bardovi-Harlig 2002: 196), *be going to* is longer and could buy the speaker time to process upcoming syntactically complex patterns (Szmrecsanyi 2003). Moreover, negation, which is supposedly linked to processing, is not relevant for the low-proficiency learners, who are also not particularly sensitive to a processing-related constraint like the length of the future marker clause. Hence, the use of the future markers is not primarily driven by processing considerations, which gives learners more freedom to differentiate between the variants functionally. By extension, the variants do not enhance production fluency, so the incentive to reach the target probabilistic grammar is weaker.

Prescriptivist influences on *be going to*

The second possibility is that learners are influenced by language input from global EFL textbooks on future time reference. These textbooks have been slow to adapt to recent insights from the field of SLA (Mishan 2022: 497), so they tend to stick to prescriptivist rules that do not reflect actual language usage (see Section 5.4.6) (Burton 2023: 25; see also Winter & Le Foll 2022; Le Foll 2023). Learners then follow these rules that do not operate in the target probabilistic grammar, resulting in additional

constraints. As learners become more proficient, they realize that these rules do not correspond to the output of native speakers, so they gradually unlearn the constraints via associative statistical learning.

From this perspective, there is essentially no difference between the acquisition of the future marker and the genitive constraints as both are inferred rationally from learners' input. The main difference is simply that the genitive alternation is less influenced by prescriptivism as evidenced by the fact that the variants are barely mentioned in the sample of popular EFL textbook series I consulted (see Section 5.4.6).

8.3 The acquisition account revisited

For the most part, the results from both studies confirm the hypotheses in support of the theory-based acquisition account, whose main purpose was to explain how EFL learners acquire and refine the probabilistic constraints to reach the target probabilistic grammar of (British) English native speakers. Most importantly, this account distinguished between two types of probabilistic constraints, namely feature-based constraints and processing-related constraints.

Processing-related constraints like structural persistence derive from the characteristics of the brain shared by learners and native speakers. Hence, these constraints should be part of learners' probabilistic grammar by default, which was confirmed in both studies (see Section 8.1.3). In fact, one could argue that these processing-related effects should be stronger for learners because they have to juggle between multiple language systems (Kroll & Dussias 2013), not to mention their language production is in all likelihood less automatized (Segalowitz 2003), but this strong version of the hypothesis was refuted. Nonetheless, the constraints generally behaved as expected, so this aspect of the theory-based account does not have to be refined.

As the name indicates, feature-based constraints involve linguistic features that are statistically associated with the variant constructions. One of the defining characteristics of construction features is that they are realized overtly by concrete linguistic elements nearby the choice context. To illustrate, second-person subjects as a construction feature of *be going to* is realized by the adjacent second-person pronoun *you*. Through associative statistical learning across many exemplars, learners realize that second-person subjects are associated with *be going to*, which gradually becomes a prototypical feature of the construction (Ellis & Wulff 2019: 44). Through a two-step process, the prototypical feature helps determine the choice of variant in language production. First, learners have to determine

which abstract construction feature a given linguistic element belongs to, and then they have to determine the variant construction that is prototypically associated with said construction feature. For example, if learners are to produce a future marker with *you* as its subject, they first categorize it as a second-person subject. Then, they would recognize that second-person subjects are a prototypical feature of *be going to*, so this is the variant they would produce.

From this perspective, possessor animacy represents a much more challenging constraint to learn because the animacy of a referent is encoded by the lemma of the noun without any overt marking. This makes it much harder to properly categorize the nouns in the right animacy category. Once the animacy of the noun has been determined, learners can use the variant that is prototypically associated with it: animate and collective possessors are features of the *s*-genitive, whereas inanimate, locative and temporal possessors prototypically prefer the *of*-genitive (see Abbot-Smith & Tomasello 2006: 284). In general, the idea that construction features drive language production aligns with the implicit assumption that probabilistic constraints impact language production. However, following the findings from the studies, this theory-based acquisition account has to be refined in two ways.

8.3.1 Two modifications

First linguistic elements, then construction features

Most importantly, this original account is not representative of the earliest stages of language acquisition. When the learners first encounter exemplars of the future markers and genitive variants, they are not aware that second-person pronouns prefer *be going to*, or that animate possessors prefer the *s*-genitive. Therefore, construction features cannot realistically drive language production. Instead, beginner learners can only rely on the concrete linguistic elements in their input. Via associative statistical learning, they would realize that certain linguistic elements co-occur with a given variant construction. For example, learners would find out that the possessor *today* or *someone* prefers the *s*-genitive, whereas *god* favors the *of*-genitive (see Section 6.3.1). It is only when learners have figured out how many individual nouns pattern with the variants that they might realize that nouns with a similar distribution share the same animacy. Since this process can take a long time, it is not the abstract construction features but rather the concrete linguistic elements themselves that drive language production for beginner learners (see Divjak et al. 2021). Once learners have managed to recognize the construction features at a later stage, these

features could become productive and drive language production as stipulated in the previous section (see Abbot-Smith & Tomasello 2006: 284; Azazil 2020: 420).

However, not every construction feature is going to be equally productive. If the above process is applied to clause type for example, learners would have to realize that individual subordinators are often associated with *be going to* and then they would gradually generate subordination as an abstract construction feature of *be going to*. However, there is only a limited set of subordinators, meaning that learners have probably encountered the subordinators often enough to keep relying on the subordinators' individual preferences during language production. In fact, constraints that are represented by a very limited set of linguistic elements might not even generalize to such abstract construction features to begin with. For example, second-person subjects are always encoded by *you*, so this is the only linguistic element that underlies the second-person subjects as a construction feature of *be going to*. As a result, it is likely that learners simply associate *you* with *be going to* directly instead of generating a redundant construction feature.

By contrast, construction features that encompass a myriad of linguistic elements like possessor animacy for example are more productive, considering that learners are bound to produce nouns they have not encountered frequently as possessors, in which case they can rely on the prototypical features to choose the more acceptable variant construction (see Azazil 2020: 446). The question arises as to how learners decide which variant to pick with new linguistic elements before they have developed such abstract construction features. In this respect, I argue that learners might rely on specific concrete, highly frequent exemplars as the standard of comparison (see Azazil 2020: 420). In principle, the same applies for constraints such as verb type, presence of temporal adverbs and temporal proximity of the future event.

This first revision of the theory-based account better captures the link between the learning mechanisms that operate on the raw linguistic input learners receive, namely associative statistical learning, and the mental representation of the constructions, which subsequently plays a role in language production. Specifically, it highlights the fact that linguistic elements drive the formation and productivity of abstract construction features.

Accommodating the future markers

An important assumption of the theory-based account is that language learning is rational, meaning that learners should only be sensitive to

constraints they can infer from the target language input. However, the results from the future marker alternation showed that learners are sensitive to the type of verb, the presence of temporal adverbs and the temporal proximity of future events, which do form part of the probabilistic grammar of the native speakers in the corpus. Ideally, the theory-based account should somehow accommodate these results as well.

In principle, EFL learners could be sensitive to these additional constraints due to transfer from their L1 (Rosenbach 2017), but this is unlikely given the diversity of L1s included in the data and the results from supplementary analyses on the future marker data. A more likely possibility is that learners acquired these additional constraints via exposure to EFL textbooks. The question is then whether the constraints were inferred rationally from the input learners received from these textbooks or whether they reflect prescriptivist rules that were explicitly formulated in the textbooks. In the latter case, learners apply these rules top-down without first learning the association between the relevant linguistic elements and the variant constructions. Instead, they would have to determine how concrete linguistic elements fit these rules to decide which variant to use. In the former case, the acquisition of probabilistic constraints is exactly the same as for the genitive alternation, except that the input is not an accurate representation of authentic language use (Burton 2023: 25).

In the present case, I argue that explicit prescriptivist rules are more likely the cause of the additional constraints because this account can better explain how low-proficiency B1 and B2 learners are sensitive to so many constraints. Supposedly, acquiring these constraints via associative statistical learning would have taken much input and time, as evidenced by the fact that learners gradually unlearned these constraints via associative statistical learning until the C2 level. This is rather unlikely given the proficiency level of learners and their rather infrequent use of *be going to*, whereas explicit rules can indirectly give rise to such constraints as soon as they are implemented.

The second possibility is that learners implemented these additional constraints to restrict *be going to* to certain, nearby future events. Specifically, I argued that the constraints pertain to inherent characteristics of future events that all learners can recognize without any exposure to language input at all, so these constraints can also be applied in a top-down fashion. In order for the theory-based account to accommodate this finding, I argue that learners are able to come up with some of their own functional distinctions between variants in alternations that offer them this flexibility, like substitution alternations where processing considerations play a minor role (see Section 8.2).

8.3.2 The acquisition of constraints in five stages

With the implementation of these revisions, I offer a step-by-step account of the acquisition of probabilistic constraints according to the theory-based account, starting from the concrete linguistic elements present in the linguistic input of EFL learners. Admittedly, since the data does not feature beginner learners and because I cannot directly access the mental representations of learners (see Section 9.2), more empirical data is needed to establish the early stages of acquisition in particular.

Stage 1

Learners are first exposed to the exemplars of the relevant constructions in their input. Provided exemplars are analyzed as chunks, learners are not aware of the various relevant linguistic elements in the construction (see Myles, Hooper & Mitchell 1998; Myles, Mitchell & Hooper 1999). By extension, they have no knowledge of how any of the surrounding linguistic elements associate with the constructions. Admittedly, it is highly likely that all learners under study have already progressed through this stage.

Stage 2

By comparing exemplars, learners have started identifying the relevant linguistic elements of the construction (see Myles, Hooper & Mitchell 1998; Myles, Mitchell & Hooper 1999), like for example the infinitive verb following the future markers, or the possessor noun in genitive exemplars. Learners notice that the constructions are statistically associated with these linguistic elements in the surrounding context via associative statistical learning (e.g., Romberg & Saffran 2010; Rebuschat & Williams 2011b; Misyak & Christiansen 2012). For example, they realize that clauses introduced by a subordinator like *because* often take *be going to*, or that a possessor like *dog* typically co-occurs with the *s*-genitive. Learners first pick up on the associations concerning linguistic elements that are highly frequent, invariant in form, adjacent to the construction and strongly associated with it (Romberg & Saffran 2010: 2–3; Yoshimura & MacWhinney 2010: 552; O'Shannessy & Meakins 2012: 380; Misyak & Christiansen 2012: 304; Gries & Ellis 2015). At this stage, learners could in principle also implement constraints from similar constructions in their L1 via transfer in a top-down fashion (see Rosenbach 2017).

Stage 3

Learners have now picked up on the association between the constructions and many linguistic elements of various types. By comparing exemplars of the variant constructions, they gradually recognize that linguistic elements

that associate with the constructions in the same way have some abstract feature(s) in common, which then become a prototypical feature of the construction (Abbot-Smith & Tomasello 2006).

To illustrate, learners not only understand that *because* is associated with *be going to*, but also for instance *that*, *as*, *before*, *since*, *though* etc. Over time, they notice that these linguistic elements associated with *be going to* share the function of subordinator as a common abstract feature. Hence, learners recognize across exemplars that *be going to* is associated with subordinators in general, which becomes a prototypical feature of the construction. Similarly, learners are aware that possessors like *dog*, *man*, *plumber*, *police agent* etc., often occur in *s*-genitives and gradually realize that they denote animate referents as a common feature, which becomes part of the prototypical features of the construction. The same process applies for the acquisition of the other animacy levels as construction features of the *s*-genitive and the *of*-genitive. For subjects of future markers, only the linguistic element *you* is associated with *be going to*. Hence, there is in all likelihood no such formation of construction features because there are no multiple linguistic elements that share common features. Instead, *you* is associated with the variant directly.

Stage 4

At this stage, the features of constructions are productive (see Azazil 2020: 420). That is, the choice of variant no longer depends solely on learners' understanding of the concrete linguistic elements; they are able to compare the linguistic elements with the prototypical features of the eligible constructions to produce the construction that provides the best fit for the linguistic element.

This is not particularly consequential for linguistic elements that belong to closed lexical classes like subordinators, because learners are unlikely to encounter infrequent elements of this type where the productivity of the construction feature plays a role. This is more relevant for construction features that encompass highly variable linguistic elements, like possessor animacy in the genitive alternation and verb type, temporal adverb and temporal proximity in the future marker alternation. For instance, if learners have to produce a genitive variant with a possessor whose association with the eligible constructions is not entirely clear, they can compare the characteristics of the possessor with the prototypical features of the variant constructions to assess which constructions provides the best fit (Azazil 2020: 446). Hence, if the possessor is animate based on their understanding of the noun, they would conclude that the *s*-genitive is

the more fitting variant because animate possessors represent an important prototypical feature of the construction.

In principle, these construction features can also be a source of non-native like choices of variant if learners' understanding of these features deviates from that of native speakers. At this stage, the construction features exert a similar influence on language production as the top-down effects that operate on the use of the future markers (see Section 8.3.1).

Stage 5

Learners have perfectly approximated the target mental representation of the constructions and their features, and so their use of the variants is native-like. As a result, learners no longer deviate from native speakers with respect to any probabilistic constraint as defined in the relevant literature, which encompass an arbitrary number of these construction features. For a constraint like possessor animacy for example, learners not only have to know that animate possessors represent a prototypical feature of *s*-genitives; they also have to realize that collective possessors represent a slightly less prototypical feature of this variant. Given that collective possessors are notoriously challenging for learners (e.g., Dziemianko 2008), they might take longer to approximate native speakers with respect to this construction feature, which could be reflected by differences on the level of the probabilistic constraint during statistical analysis as in the corpus-based analysis of the genitive alternation (see Section 6.3.2.2). By contrast, a constraint like final sibilancy only concerns a single construction feature, namely that the *s*-genitive avoids final sibilants, which in turn is encoded by a limited number of linguistic elements, so this constraint is acquired much quicker.

8.4 Pedagogical implications

The studies on both alternations have revealed that certain constraints impact the choice between variant constructions differently for native speakers and learners across proficiency levels. As a result, learners sometimes do not produce the variant a native speaker would have preferred in a given context, resulting in utterances that may not sound very natural. While learners ultimately converge on the same probabilistic grammar as native speakers, some pedagogical guidance at earlier stages of acquisition could quicken this learning process or even prevent some of these difficulties from occurring altogether. In what follows, I therefore propose to tweak specific aspects of the input learners receive and highlight the need for assistance from teaching materials, instructors or language

partners (Ellis 2006: 110) to draw learners' attention to certain aspects of the variants' use they might have overlooked (see Schmidt 2001).

With respect to the future markers, the main issue of language learners is that they restrict the use of *be going to* to mainly events happening in the nearby future with some certainty. If this is due to learners adhering to recommendations on the use of the future markers from global EFL textbooks that do not reflect authentic native-speaker usage (see Mishan 2022; Winter & Le Foll 2022; Le Foll 2023), then the obvious solution is to do away with these prescriptivist rules. Conversely, if learners restrict *be going to* on their own initiative to differentiate functionally between *will* and *be going to*, two adjustments are possible.

On the one hand, EFL textbooks, which constitute a considerable source of input for some language learners (see Le Foll 2023: 259–260), should emphasize that *will* and *be going to* are interchangeable future markers, rather than presenting either variant as 'the neutral future' as in the *Headway* series for example (Soars & Soars 2019). Indeed, the probabilistic grammar of native speakers shows that both markers are in reality more or less equally acceptable in many contexts given that only clause type, subject grammatical person and negation induce any preference for either variant. This high degree of unconstrained interchangeability is also evidenced by the fact that the model could only predict the right choice of future marker for native speakers in 73% of the cases (see Section 5.4.2).

The second part of the solution would be to expose learners to more exemplars of *be going to* for future events that do not occur with temporal adverbs and that are not temporally proximate. Results from the corpus-based investigation of the alternation suggested that learners might be very sensitive to such exemplars of *be going to* in contexts that contradict their interlanguage as evidenced by the strong effects of structural persistence (see Section 5.4.4).

Compared to the future marker alternation, the case study on the genitive alternation revealed that the probabilistic grammar of even the low-proficiency B1 learners is strikingly similar to that of native speakers (see also Gries & Wulff 2013: 349), with a few exceptions. Concretely, B1 learners are not aware that definite possessors prefer the *s*-genitive, probably because this constraint is rather weak (Heller 2018: 104). A practical way of addressing this issue is to draw attention to the article that accompanies the possessor in the genitive variants. That is, one could expose learners to more *s*-genitive exemplars featuring a possessor with definite article, and *of*-genitive exemplars where the possessor takes an indefinite article. While definiteness is not only determined by the definite

article, it accounts for more than half of the definite possessors in the genitive observations of native speakers (179 out of the 328). Hence, if learners managed to replicate the association between the article and the variant in the language use of native speakers, this would already greatly reduce the number of non-idiomatic genitive variants learners produce as a result of the definiteness constraint. Moreover, the articles are invariant in form, highly frequent and adjacent to the construction, so learners should be quick to pick up on the association between these linguistic elements and the choice of variant via associative statistical learning (e.g., Romberg & Saffran 2010: 2–3; Yoshimura & MacWhinney 2010: 552; Misyak & Christiansen 2012: 304; Yi, Lu & Ma 2017: 520).

Arguably the most important finding of the genitive study is that B1 and B2 learners experience difficulties with the animacy constraint. In view of the results gathered from both corpus data and experimental data, I argued that this stems from insufficient exposure to the individual nouns occurring as possessors in genitive constructions. Indeed, learners can only fully acquire the animacy constraint when they figured out the distribution of many individual possessor nouns, which allows them to recognize that possessors preferring the same variant tend to share the same animacy. As a result, the acquisition of possessor animacy at its early stages is largely a matter of lexical knowledge.

Given that learners are already exposed to all kinds of nouns, it is unclear how this process could be accelerated. Moreover, lexical knowledge cannot be mastered fully at lower proficiency levels and is instead continuously refined throughout language learning. As a result, the acquisition of the animacy constraint is by default a long gradual process, and the same should in principle apply to constraints like verb type in the future marker alternation. One possible suggestion is to expose learners to more lexical elements that are particularly problematic, like collective nouns such as *government*, *community* etc. according to the rating task experiment (see Section 7.4.1, see also Dziemianko 2008).

8.5 Methods evaluation

The research questions of the thesis were mostly answered with the results from mixed-effects logistic and linear regression models, similar to most previous multifactorial analyses of alternation phenomena in EFL (e.g., Gries & Wulff 2013; Wulff, Lester & Martinez-Garcia 2014; Wulff & Gries 2015; Wulff & Gries 2019). Contrary to previous studies, which incorporated L1 as the main learner variable of interest during statistical analysis, a key innovative aspect of the present research was the

introduction of proficiency level in interactions with each of the relevant probabilistic constraints in the models. More specifically, I used treatment coding to compare native speakers against learners, which basically allowed me to track the development of learners' probabilistic grammar towards the target language of native speakers. On the other hand, using the native speakers as the standard of reference comes with the disadvantage that the envelope of variation is restricted to interchangeable contexts for native speakers, which might not match perfectly for learners (see Section 9.2).

With this model design, I was able to discover dissimilarities between native speakers and learners due to general learning mechanisms such as associative statistical learning (see Rebuschat & Williams 2011b), which previous methodological designs could not detect. For the future marker alternation, previous research found that L1 background does not exert much influence on speakers' use of the variants (Bardovi-Harlig 2000: 411), although the distribution of *will* and *be going to* is a known source of difficulties for language learners (Wekker 1976: 123; Haegeman 1989: 292). Admittedly, these previous studies on L1 influence did not take into consideration as many constraints simultaneously, so perhaps a more sophisticated statistical analysis could reveal L1-dependent differences. Nonetheless, the recent study by Bohmann (forthcoming) on future-time reference in ESL also implies that L1 plays a minor role considering that the variety of ESL does not interact with speakers' set of constraints either. By contrast, the logistic regression model with proficiency interactions was able to discover significant differences between native speakers and language learners for six out of nine constraints conditioning the choice of variant. Therefore, I argue that the methodology is well-suited to address the need expressed by previous research on the acquisition of the future markers to investigate the influence of proficiency level (see Bardovi-Harlig 2000: 411–412).

Likewise, the most important finding of the study on the genitive alternation, namely that low-proficiency learners are less sensitive to the animacy of the possessor, did not appear in the study of Gries and Wulff (2013). This is most likely due to the fact that learners from multiple proficiency levels were conflated during statistical analysis, so the difficulties low-proficiency learners experience with the animacy constraint could not be detected when combined with data from more proficient learners who already master the constraint. If proficiency had not been considered, the diverging animacy effects might have been missed completely, or they would have been generalized to all learners regardless of their proficiency level. Therefore, I argue that an accurate and

comprehensive multifactorial account of learners' choice between alternating variants has to control for the influence of proficiency level. Conversely, previous research on the genitive alternation in particular has demonstrated the importance of L1 (see Gries & Wulff 2013), so this learner variable should ideally also be incorporated in the methodological design, which was only partially possible in this case (see Section 9.2).

The genitive alternation was also analyzed with the MuPDAR approach (Gries & Adelman 2014; Gries & Deshors 2014), which also involves regression models but focuses on the constraints that actually lead to non-native choices of variant. However, this analysis was rather unsuccessful, mainly because there are too few genitive observations where learners deviate from native speakers to model the data accurately (see Section 6.3.3.2). The approach might be more productive for alternations that are less similar for learners and native speakers.

Whereas most previous research relied exclusively on such multifactorial corpus-based analyses, the genitive alternation was also studied with a follow-up rating task experiment. Similar experiments were conducted by Jäschke and Plag (2016) for the dative alternation and Kinne (2020) for the effects of end-weight and discourse status on the particle placement alternation and the verb-prepositional alternation. Contrary to Jäschke and Plag (2016) and Kinne (2020), who only included German EFL learners, the rating task experiment presented in the previous chapter included participants from eight L1 backgrounds across six proficiency levels.

This experiment successfully achieved three specific aims. First, the experiment was able to replicate the finding that learners are less sensitive to possessor animacy than native speakers using a more fine-grained operationalization of the constraint, which also shed more light on the acquisition of probabilistic grammars in general. Secondly, the introspective ratings from the participants in the experiment were shown to correlate with the predictions from the corpus-based model on the genitive alternation, providing additional evidence for the cognitive reality of the constraints (Bresnan & Ford 2010; Engel et al. 2022). Although the correlation is moderate, indicating much variability in the ratings that cannot be predicted by the corpus-based model (see Section 9.2), its strength is comparable to other studies with similar designs focusing on native speakers only (Engel 2022: 176; Engel & Szmrecsanyi 2023: 370; see also Verhagen et al. 2020). Finally, the experiment offered the possibility to explore the behavior of individual participants compared to the aggregated results from the regression models.

Overall, the mixed methodology offered robust results regarding the development of the probabilistic grammars for EFL learners. That is, although one cannot directly access the mental representations of the variant constructions in the minds of speakers (see Section 9.2), the corpus-based investigation of the language use of the speakers reflects their implicit grammatical knowledge, and so do the intuitions of the participants in the experiment regarding the naturalness of the variants, both of which provided converging evidence for the cognitive reality of the probabilistic grammars and their development. The trade-off for these robust results is a time-intensive data collection and annotation process, especially in the context of spoken learner language, which is particularly messy.

9 Conclusion

In this final chapter, I first review the main theoretical and methodological contributions of the thesis. Then, I turn to some of its limitations and provide some suggestions for future research.

9.1 Research contributions

While a number of studies have researched alternation phenomena in EFL (e.g., Gries & Wulff 2013; Gries & Deshors 2014; Wulff & Gries 2015; Paquot, Grafmiller & Szmrecsanyi 2019; Kinne 2020), no prior study has explicitly formulated and tested predictions about the development of probabilistic grammars. In this thesis, I have presented a theory-based account of the development and acquisition of probabilistic grammars for language learners inspired by usage-based approaches to SLA (e.g., MacWhinney 2005; Ellis 2006; Ellis & Wulff 2019). This theory-based account was then used to generate predictions regarding the acquisition of the set of constraints guiding the choice between on the one hand *will* and *be going to* and on the other hand the *s*-genitive and the *of*-genitive. These predictions were tested empirically as hypotheses in quantitative corpus-based studies. The corpus-based study of the genitive alternation was also complemented with a rating task experiment.

Overall, the results from the studies largely align with the hypotheses and so lend support to an input-driven, exemplar-based account of the acquisition of probabilistic grammars. Specifically, I argue that probabilistic grammars consist of broadly speaking two types of constraints, namely processing-related and feature-based constraints. Processing-related constraints are part of the probabilistic grammar of learners by default because they stem from features of the brain. Feature-based constraints derive from the fact that each construction has a set of abstract prototypical features it is most often associated with. These features are language specific as they originate from the association between the variant constructions and the concrete linguistic elements in the input of language learners. As a result, these construction features have to be learned, and the general learning mechanism responsible for this is associative statistical learning (e.g., Romberg & Saffran 2010; Rebuschat & Williams 2011b;

Lammertink et al. 2017). The learnability of the statistical association between features and constructions depends on a number of criteria. Most importantly, it depends on the number of concrete linguistic elements the abstract feature encompasses, as well as their reliability, frequency and adjacency to the construction in the language input. Statistical associations that are easier to learn require less language input to master so they are acquired first. However, the study of the future marker showed that learners can develop sensitivity to linguistic elements that are not necessarily associated with the construction in the language use of native speakers, which sheds some doubt on the assumption that language learning is the result of rational inference from language input (Ellis 2006: 102–103). Instead, it appears that lower-proficiency learners can take the initiative to distinguish between the variants in a non-native yet systematic way for added functionality when the variant constructions involved offer them the flexibility to do so.

Methodologically, one of the key innovative aspects of the thesis is that I controlled for learners' proficiency level during statistical analysis, which makes it possible to track the development of their probabilistic grammars throughout language learning. Results from both alternations confirmed that learners struggle with the acquisition of certain constraints at specific proficiency levels due to general learning mechanisms like associative statistical learning, which could not have been detected in a statistical design that does not differentiate between consecutive stages of language learning. As a result, I argue that an accurate and comprehensive multifactorial account of learners' choice between alternating variants must take into account their proficiency level.

Finally, the thesis promotes a mixed methodology combining corpus data with experimental data for the study of learner language. Converging evidence from both types of data increased the generalizability of the findings regarding the development of learners' probabilistic grammars (e.g., Hoffmann 2006: 169; Gilquin & Gries 2009: 9; Schönefeld 2011: 20–24; Divjak & Arppe 2013: 265), and helped shed light on the puzzling effect of the animacy constraint operating on the genitive alternation in particular. This, in turn, highlighted the important distinction between feature-based constraints that encompass a large number of linguistic elements and those that concern only a limited number of elements, which helped refine the theory-based account of the development of probabilistic grammars.

9.2 Limitations and future research

Although the thesis addressed some of the conceptual and methodological gaps of previous research, it suffers from a number of shortcomings that will hopefully inspire future research.

No interactions with L1

One of the main innovative aspects of the thesis is that I controlled for the influence of proficiency level. In this way, the studies can capture how general learning mechanisms shared by all speakers shape the set of constraints at different stages of language learning (see Section 2.1.2.2). However, a shortcoming of the studies is that the influence of learners' L1 is only partially controlled for despite its relevance according to previous research. That is, L1 was included as a random effect, which controls for the overall preference learners from a given L1 might have for either variant, but it cannot control for potential interactions between L1 and the constraints.

Instead, L1 should ideally be included as a fixed effect in the regression models, much like proficiency level. On the one hand, controlling for L1 as a fixed effect would further ensure that the significant interactions with proficiency level in the models are not due to underlying L1s. Instead, I carried out supplementary analyses to exclude this possibility for both the future marker alternation (see Sections 5.3.3) and the genitive alternation (see Section 6.3.3.1). On the other hand, three-way interactions between L1, proficiency level, and each of the probabilistic constraints would bring to light transfer effects at different stages of language learning. As stated in Section 2.1.2.2, previous research on cues found that L1 influence is stronger for low-proficiency learners (Morett & MacWhinney 2013: 134, 149; Wang & Xu 2015: 215), so I would expect low-proficiency learners in particular to transfer patterns from their L1.

Including L1 as a fixed effect was not possible because it would overlap perfectly with proficiency level for native speakers since all speakers with a native proficiency level have English as their L1, which yields errors during the model fitting process. To address this limitation, I tried to fit a supplementary model without native speakers. However, three-way interactions require considerably more data than two-way interactions, so it was ultimately not possible to reach a satisfactory balance between ensuring that there is enough data to fit the regressors and retaining enough distinctions in the predictors to make the results insightful for the research goals of the thesis.

No beginner learners

A second shortcoming of the corpus-based studies is that I could not gain insight into the acquisition of certain feature-based constraints for both alternation phenomena. That is, B1 learners already approximate native speakers with respect to the effect of clause type and subject grammatical person for the future marker alternation and final sibilancy for the genitive alternation. Hence, learners most likely acquire these constraints at earlier stages of language acquisition, namely the A proficiency levels on the CEFR scale (Council of Europe 2001), which are not included in the TLC (Gablasova, Brezina & McEnery 2019).

Prescriptivism?

The corpus-based analysis of the future marker alternation revealed that B1, B2 and C1 learners favor *be going to* for events that are relatively certain to happen in the near future, whereas native speakers find both variants equally acceptable in many contexts. To explain this finding, I argued that differentiating between the future markers in this way benefits learners more than using *be going to* as a less frequent and arguably more difficult functional equivalent to *will* (see Section 5.4.5).

However, another possibility is that learners acquired these additional constraints through exposure to prescriptivist rules from EFL textbooks (e.g., Winter & Le Foll 2022; Le Foll 2023), which do not correspond to authentic language use. This is especially relevant for the future markers since future-time reference is part of the ‘canon’ of EFL textbooks (Burton 2023). Lower-proficiency learners might try to apply these rules in a top-down fashion during language production. As they become more proficient, they gradually weaken these restrictions because they realize through associative statistical learning that these rules do not match authentic language input.

There are arguments in support of both explanations. A closer look at a sample of highly popular, relevant EFL textbooks showed that such prescriptivist rules could account for some of the additional constraints learners are sensitive to (see Section 5.4.6). On the other hand, not all prescriptivist rules in these textbooks appear to make their mark on learners’ use of the future markers. Moreover, not every learner might have been exposed to these EFL textbooks to begin with.

Hence, a limitation of the thesis is that it cannot properly distinguish between these two interpretations. A solution would be to replicate the study with another substitution alternation that is perhaps not discussed in EFL textbooks as much, like the alternation between the indefinite

pronouns (*everybody/everyone, anybody/anyone* etc.) for example (Raumolin-Brunberg & Kahlas-Tarkka 1997).

Differences in language input

A related limitation is that the learners under study are not necessarily exposed to the same language input. That is, learners from one country might mostly encounter English in classroom contexts, whereas learners from another country mostly receive natural language input. Given that English is an institutional language in India, which technically means that the Indian learners in the data are ESL rather than EFL speakers (see Section 4.1.1), it is highly likely that Indian learners of English receive more natural language input than Russian EFL learners for example. The nature of the input learners receive is relevant because it drives the acquisition of the constraints according to the theory-based account. Learners who are primarily exposed to English in a classroom context might be more strongly influenced by prescriptivist rules that do not necessarily represent authentic language use (Burton 2023: 25). While Indian learners of English are probably exposed to more natural language input, this input might incorporate features of Indian English, which is different from the British English variety spoken by the examiners in the corpus.

Opaque experimental ratings

The corpus-based analysis of the genitive alternation was supplemented with a rating task experiment resembling the experiments in Bresnan and Ford (2010), Engel et al. (2022) and Engel and Szmrecsanyi (2023). By virtue of the fact that the experimental items were selected from the genitive observations used for the corpus-based analysis of the genitive alternation, it was possible to directly compare learners' intuitions regarding the naturalness of the variants, captured by the ratings from the participants in the experiment, with the language use of the speakers in the corpus as reflected by the predictions of the corpus-based logistic regression model (Arppe & Järvikivi 2007: 132; Gilquin & Gries 2009; Klavan & Divjak 2016: 361). As expected, results showed that the ratings correlated with the corpus-based predictions, which means that the variant participants deem most natural in a given context also tends to be more frequent in spoken language production. This provided additional evidence that the significant probabilistic constraints in the corpus-based model are indeed part of speakers' implicit grammatical knowledge (Bresnan & Ford 2010).

However, the strength of the correlation is only moderate, so there is much variability in the ratings that cannot be predicted by the corpus-based

model, especially for the participants from the lower proficiency levels. This is partially due to the fact that the highly influential animacy constraint impacted the choice of variant slightly differently for these learners in the experiment and the corpus data, but this cannot fully account for the discrepancies between both types of data (see Section 7.4.2). Yet, no other satisfactory explanation could be found, except that low-proficiency learners might not approach the rating task experiment in the same way as more proficient learners and native speakers. In general, rating tasks leave much to participants' interpretation, and notions like 'acceptability', 'grammaticality' or 'naturalness' are hard to distinguish (Arppe & Järviokivi 2007: 150 citing Penke & Rosenbach 2007: 492). Moreover, participants can vary considerably in their use of the slider bar, with some sticking more to the middle range whereas others are more categorical in their preferences. The instructions probably influence participants as well, as suggested by the fact that the native speaker participants from the pilot study were prone to using the extreme ends of the slider before the instructions were rewritten to emphasize that they could give more gradient ratings using the entire range of the slider (see Section 7.2.3). This raises the question as to whether it is natural for participants to gauge the naturalness or acceptability of the variants on such a fine-grained scale as in the present design to begin with (see Stefanowitsch 2006: 73). Alternatively, this could indicate that native speakers originally confused naturalness with grammaticality, choosing one variant as the only correct option. As for the learner participants, their introspective judgments could also be influenced by what they believe is the correct variant following the teachings of prescriptivist textbooks.

In short, while it is reassuring that the rating data converges with the results from the corpus-based analysis (see Section 8.5), it is not exactly clear what these metalinguistic, introspective ratings capture, especially for low-proficiency learners, so much variability remains unexplained. A better solution might be to opt for experimental techniques that give the participants less freedom of interpretation and instead tap into their implicit grammatical knowledge more indirectly via reaction times for example, as suggested by Engel (2022: 296–297).

Restrictive perspective on syntactic variation

Probabilistic grammars determine which variant construction is most appropriate in contexts where multiple constructions are functionally and semantically eligible. To investigate probabilistic grammars, the dominant approach in the literature is to only include those contexts where the variant constructions of interest are interchangeable. In this case, I only

considered contexts where native speakers allow for both *will* and *be going to* for the future marker alternation and both *s*-genitives and *of*-genitives for the genitive alternation. In Section 4.1.2.3, I argued that it is justified to determine interchangeability according to native speaker standards because the learners in the corpus data are taking an official spoken English language exam, which implies that the language use of native speakers is the target language that learners strive for. It is also more efficient methodologically given that the interchangeable contexts are well-documented for native speakers compared to learners.

While this approach is well-suited to study the development of learners' probabilistic grammars towards the native speaker standard in these specific contexts, it is restrictive in two ways. First, it does not consider that interchangeable contexts for native speakers and learners might not overlap perfectly. By extension, the approach cannot keep track of an important aspect of language learning, which is to understand how the semantic space is divided across potential variants. Hence, an interesting avenue for future research would be to investigate which meanings functionally equivalent variants can express.

More importantly, focusing on two variant constructions in particular cannot provide a comprehensive understanding of future-time reference for EFL learners, nor does it cover the many ways in which learners could express "the relation between the possessor and a possessed entity" (Rosenbach 2014: 221; see also Rosenbach 2002: 27). That is, it remains unclear how the development of *will* and *be going to* interacts with other future-time expressions in English, such as *be to* (Declerck 2010) or the simple present and the progressive (Nesselhauf 2007; Bergs 2010: 217–218). For example, the simple present and the progressive are used for future events that are certain and they often involve adverbs (Bergs 2010: 218), which is reminiscent of the way low-proficiency learners employ *be going to*. With this in mind, learners might associate *be going to* with more certain future events because they do not yet master the use of the simple present and progressive for future events. Likewise, it would be interesting to investigate how compound nouns relate to the *s*-genitive and the *of*-genitive in the language use of EFL learners (see Szmrecsanyi, Biber, et al. 2016).

Bibliography

- Aarts, Bas, Joanne Close & Sean Wallis. 2013. Choices over time: methodological issues in investigating current change. In Bas Aarts, Geoffrey Leech, Joanne Close & Sean Wallis (eds.), *The Verb Phrase in English: Investigating Recent Language Change with Corpora* (Studies in English Language), 14–45. Cambridge: Cambridge University Press.
- Abbot-Smith, Kirsten & Michael Tomasello. 2006. Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review* 23(3). 275–290.
- Abbuhl, Rebekha, Susan Gass & Alison Mackey. 2013. Experimental research design. In Robert J. Podesva & Devjani Sharma (eds.), *Research Methods in Linguistics*, 116–134. New York: Cambridge University Press.
- Aissen, Judith. 1999. Markedness and Subject Choice in Optimality Theory. *Natural Language & Linguistic Theory* 17(4). 673–711.
- Allan, Dave. 2001. *Oxford Placement Test 1*. Oxford, UK: Oxford University Press.
- Allen, Cynthia L. 2003. Deflexion and the development of the genitive in English. *English Language & Linguistics* 7(1). 1–28.
- Arnold, Jennifer E., Anthony Losongco, Thomas Wasow & Ryan Ginstrom. 2000. Heaviness vs. Newness: The Effects of Structural Complexity and Discourse Status on Constituent Ordering. *Language* 76(1). 28–55.
- Arppe, Antti & R. Harald Baayen. 2011. Statistical classification and principles of human learning. In *Proceedings of Quantitative Investigations in Theoretical Linguistics 4*. Humboldt-Universität zu Berlin, Germany.
- Arppe, Antti & Juhani Järvikivi. 2007. Every method counts: Combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory* 3(2). 131–159.
- Azazil, Lina. 2020. Frequency effects in the L2 acquisition of the catenative verb construction – evidence from experimental and corpus data. *Cognitive Linguistics* 31(3). 417–451.
- Baayen, Rolf Harald. 2008. *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Baayen, Rolf Harald, Douglas J. Davidson & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). 390–412.

- Bakdash, Jonathan Z. & Laura R. Marusich. 2017. Repeated Measures Correlation. *Frontiers in Psychology* 8. 456.
- Bardovi-Harlig, Kathleen. 2000. Chapter Seven: Past, Present, and Future. *Language learning* 50(s1). 409–437.
- Bardovi-Harlig, Kathleen. 2002. A new starting point? Investigating formulaic use and input in future expression. *Studies in Second Language Acquisition* 24(2). 189–198.
- Bardovi-Harlig, Kathleen. 2005. The Future of Desire: Lexical Futures and Modality in L2 English Future Expression. In *Proceedings of the 7th Generative Approaches to Second Language Acquisition Conference*, 1–12. Somerville, MA: Cascadilla Proceedings Project.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.
- Bates, Elizabeth & Brian MacWhinney. 1989. Functionalism and the Competition Model. In Brian MacWhinney & Elizabeth Bates (eds.), *The crosslinguistic study of sentence processing*, 3–73. New York, NY: Cambridge University Press.
- Behaghel, Otto. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25. 110–142.
- Bentz, Christian & Bodo Winter. 2013. Languages with More Second Language Learners Tend to Lose Nominal Case. *Language Dynamics and Change* 3(1). 1–27.
- Berglund, Ylva. 2000. Utilising Present-day English corpora: A case study concerning expressions of future. *ICAME Journal* (24). 25–63.
- Bergs, Alexander. 2010. Expressions of futurity in contemporary English: A Construction Grammar perspective. *English Language & Linguistics* 14(2). 217–238.
- Biber, Douglas. 2003. Compressed noun-phrase structure in newspaper discourse: The competing demands of popularization vs. economy. In Jean Aitchinson & Diana M. Lewis (eds.), *New Media Language*, 169–181. London & New York: Longman.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. Harlow: Pearson Education Ltd.
- Bidgood, Amy, Julian M. Pine, Caroline F. Rowland & Ben Ambridge. 2020. Syntactic Representations Are Both Abstract and Semantically Constrained: Evidence From Children's and Adults' Comprehension and Production/Priming of the English Passive. *Cognitive Science* 44(9). e12892.
- Biewer, Carolin. 2011. Modal auxiliaries in second language varieties of English: A learner's perspective. In Joybrato Mukherjee & Marianne Hundt (eds.), *Exploring Second-Language Varieties of English and Learner Englishes: bridging a paradigm gap*, 7–34. Amsterdam: John Benjamins.
- Bloem, Jelke, Arjen Versloot & Fred Weerman. 2017. Verbal cluster order and processing complexity. *Language Sciences* 60. 94–119.

- Bock, J. Kathryn. 1982. Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review* 89(1). 1–47.
- Bock, K. & Z.M. Griffin. 2000. The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General* 129(2). 177–192.
- Bock, Kathryn, Gary S. Dell, Franklin Chang & Kristine H. Onishi. 2007. Persistent structural priming from language comprehension to language production. *Cognition* 104(3). 437–458.
- Bohmann, Axel. forthcoming. Future-time reference in World Englishes. *World Englishes*.
- Branigan, Holly P., Martin J. Pickering & Mikihiro Tanaka. 2008. Contributions of animacy to grammatical function assignment and word order during production. *Lingua* 118(2). 172–189.
- Bresnan, Joan. 2007. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In Sam Featherston & Wolfgang Sternefeld (eds.), *Linguistics in search of its evidential base* (Studies in Generative Grammar), 75–96. Berlin: Mouton de Gruyter.
- Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayen. 2007. Predicting the dative alternation. In Gerlof Boume, Irene Krämer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.
- Bresnan, Joan & Marilyn Ford. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language: Journal of the Linguistic Society of America* 86(1). 168–213.
- Bresnan, Joan & Tatiana Nikitina. 2009. The Gradience of the Dative Alternation. In Linda Uyechi & Hee Wee Lian (eds.), *Reality Exploration and Discovery: Pattern Interaction in Language and Life*, 161–184. Stanford: CSLI Publications.
- Brezina, Vaclav & Dana Gablasova. 2015. Is There a Core General Vocabulary? Introducing the New General Service List. *Applied Linguistics* 36(1). 1–22.
- Brisard, Frank. 2001. Be going to: an exercise in grounding. *Journal of Linguistics* 37(2). 251–285.
- Brysbaert, Marc & Michaël Stevens. 2018. Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition* 1(1). 9.
- Burton, Graham. 2023. *Grammar in ELT and ELT Materials: Evaluating its History and Current Practice* (Second Language Acquisition 164). Bristol, Blue Ridge Summit: Multilingual Matters.
- Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4). 711–733.
- Callies, Marcus. 2015. Using learner corpora in language testing and assessment: Current practice and future challenges. In Erik Castello, Katherine Ackerley & Francesca Coccetta (eds.), *Studies in Learner*

- Corpus Linguistics: Research and Applications for Foreign Language Teaching and Assessment*, 21–35. Frankfurt: Peter Lang.
- Claes, Jeroen. 2017. Probabilistic Grammar: The view from Cognitive Sociolinguistics. *Glossa: a journal of general linguistics* 2(1). 62.
- Clarke, Paul. 2008. When can group level clustering be ignored? Multilevel models versus single-level models with sparse data. *Journal of Epidemiology & Community Health* 62(8). 752–758.
- Cleland, Alexandra A. & Martin J. Pickering. 2003. The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language* 49(2). 214–230.
- Cleveland, William S. 1993. *Visualizing data*. Summit, NJ: Hobart Press.
- Clyde, Laura & Ray Parker. 2015. *Pass Trinity Now 5-6 and ISE I*. Black Cat.
- Clyde, Laura & Shannon West. 2015. *Pass Trinity Now 7-8 and ISE II*. Black Cat.
- Cochrane, Stuart. 2015. *Pass Trinity Now 3-4 and ISE Foundation*. Black Cat.
- Cohen, Jacob. 1992. A power primer. *Psychological Bulletin* 112(1). 155–159.
- Comrie, Bernard. 1989. *Language universals and linguistic typology*. 2nd edn. Oxford: Oxford University Press.
- Copley, Bridget. 2009. *The semantics of the future*. New York and London: Routledge.
- Corbett, Greyille G. 2010. Implicational Hierarchies. In Jae Jung Song (ed.), *The Oxford Handbook of Linguistic Typology*, 1–13. Oxford: Oxford University Press.
- Council of Europe. 2001. *Common European Framework of Reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Croft, William. 2003. *Typology and Universals*. Cambridge: Cambridge University Press.
- Crosthwaite, Peter R., Daniel R. Bailey & Ashley Meeker. 2015. Assessing in-class participation for EFL: considerations of effectiveness and fairness for different learning styles. *Language Testing in Asia* 5(1). 1–19.
- Dahl, Östen. 1984. Temporal distance: remoteness distinctions in tense-aspect systems. In Brian Butterworth, Bernard Comrie & Östen Dahl (eds.), *Explanations for Language Universals*, 105–122. Berlin, Boston: De Gruyter Mouton.
- Danchev, Andrei & Merja Kytö. 1994. The construction *be going to* + infinitive in Early Modern English. In Dieter Kastovsky (ed.), *The construction *be going to* + infinitive in Early Modern English* (Topics in English Linguistics 13), 59–78. Berlin, Boston: De Gruyter Mouton.
- Dashti, Abdulmohsen A. 2013. Interacting with domestic workers in Kuwait: grammatical features of foreigner talk. A case study. *International Journal of the Sociology of Language* 2013(224). 63–84.

- Davies, Mark & Robert Fuchs. 2015. Expanding horizons in the study of World Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE). *English World-Wide* 36(1). 1–28.
- De Troij, Robbert, Stefan Grondelaers & Dirk Speelman. 2023. Natiolectal Variation in Dutch Morphosyntax: A Large-Scale, Data-Driven Perspective. *Journal of Germanic Linguistics* 35(1). 1–68.
- Declerck, Renaat. 2010. Future time reference expressed by *be to* in Present-day English. *English Language and Linguistics* 14(2). 271–291.
- Denis, Derek & Sali A. Tagliamonte. 2018. The changing future: competition, specialization and reorganization in the contemporary English future temporal reference system. *English Language and Linguistics* 22(3). 403–430.
- Deshors, Sandra C. 2014. A case for a unified treatment of EFL and ESL: a multifactorial approach. *English World-Wide* 35(3). 277–305.
- Deshors, Sandra C. & Stefan Th. Gries. 2014. A Case for the Multifactorial Assessment of Learner Language: The Uses of *May* and *Can* in French-English Interlanguage. In Dylan Glynn & Justyna A. Robinson (eds.), *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*, 179–204. Amsterdam: John Benjamins.
- Divjak, Dagmar & Antti Arppe. 2013. Extracting prototypes from exemplars: What can corpus data tell us about concept representation? *Cognitive Linguistics* 24(2). 221–274.
- Divjak, Dagmar, Ewa Dąbrowska & Antti Arppe. 2016. Machine Meets Man: Evaluating the Psychological Reality of Corpus-based Probabilistic Models. *Cognitive Linguistics* 27(1). 1–33.
- Divjak, Dagmar, Petar Milin, Adnane Ez-zizi, Jarosław Józefowski & Christian Adam. 2021. What is learned from exposure: an error-driven approach to productivity in language. *Language, Cognition and Neuroscience* 36(1). 60–83.
- Dörnyei, Zoltán & Tatsuya Taguchi. 2009. *Questionnaires in Second Language Research: Construction, Administration, and Processing*. 2nd edn. New York: Routledge.
- Dubois, Tanguy. 2023. The Complexity Principle and lexical complexity in the English and Dutch dative alternation. In Eva Zehentner, Melanie Röthlisberger & Timothy Coleman (eds.), *Ditransitive constructions in Germanic languages: Diachronic and Synchronic Aspects* (Studies in Germanic Linguistics 7), 325–364. Amsterdam: John Benjamins Publishing Company.
- Dubois, Tanguy, Magali Paquot & Benedikt Szendrői. 2023. Alternation phenomena and language proficiency: the genitive alternation in the spoken language of EFL learners. *Corpus Linguistics and Linguistic Theory* 19(3). 427–450.
- Duguine, Isabelle & Barbara Köpke. 2018. Processing strategies used by Basque-French bilingual and Basque monolingual children for the

- production of the subject-agent in Basque. *Linguistic Approaches to Bilingualism* 9(4). 514–541.
- Dziemianko, Anna. 2008. Subject-Verb Concord with Collective Nouns or the Count-Mass Distinction: Which is More Difficult for Polish Learners of English? *Poznań Studies in Contemporary Linguistics* 44(4). 449–468.
- Ellis, Nick C. 2006. Cognitive perspectives on SLA: The associative-cognitive CREED. *AILA Review* 19. 100–121.
- Ellis, Nick C. 2008. The Dynamics of Second Language Emergence: Cycles of Language Use, Language Change, and Language Acquisition. *The Modern Language Journal* 92(2). 232–249.
- Ellis, Nick C. 2017. Cognition, Corpora, and Computing: Triangulating Research in Usage-Based Language Learning. *Language Learning* 67(S1). 40–65.
- Ellis, Nick C., Kausar Hafeez, Katherine I. Martin, Lillian Chen, Julie Boland & Nuria Sagarra. 2014. An eye-tracking study of learned attention in second language acquisition. *Applied Psycholinguistics* 35(3). 547–579.
- Ellis, Nick C., Matthew Brook O'Donnell & Ute Römer. 2014. The processing of verb-argument constructions is sensitive to form, function, frequency, contingency and prototypicality. *Cognitive Linguistics* 25(1). 55–98.
- Ellis, Nick C. & Dave C. Ogden. 2017. Thinking about multiword constructions: Usage-based approaches to acquisition and processing. *Topics in Cognitive Science* 9(3). 604–620.
- Ellis, Nick C. & Nuria Sagarra. 2010. The bounds of adult language acquisition: Blocking and learned attention. *Studies in Second Language Acquisition* 32(4). 553–580.
- Ellis, Nick C. & Nuria Sagarra. 2011. Learned attention in adult language acquisition: A replication and generalization study and meta-analysis. *Studies in Second Language Acquisition* 33(4). 589–624.
- Ellis, Nick C. & Stefanie Wulff. 2015. Second language acquisition. In Eva Dabrowska & Dagmar Divjak (eds.), *Handbook of Cognitive Linguistics*, 409–432. Berlin, Boston: De Gruyter Mouton.
- Ellis, Nick C. & Stefanie Wulff. 2019. Cognitive approaches to Second Language Acquisition. In John W. Schwieter & Alessandro Benati (eds.), *The Cambridge Handbook of Language Learning*, 41–61. Cambridge: Cambridge University Press.
- Ellis, Rod. 1999. Item versus system learning: explaining free variation. *Applied Linguistics* 20(4). 460–480.
- Engel, Alexandra. 2022. *The register-specificity of probabilistic grammars in English and Dutch - Combining corpus analysis and experimentation*. Leuven: KU Leuven.
- Engel, Alexandra, Jason Grafmiller, Laura Rosseel & Benedikt Szmrecsanyi. 2022. Assessing the complexity of lectal competence: the register-specificity of the dative alternation after give. *Cognitive Linguistics* 33(4). 727–766.

- Engel, Alexandra & Benedikt Szmrecsanyi. 2023. Variable grammars are variable across registers: future temporal reference in English. *Language Variation and Change* 34(3). 355–378.
- Fehringer, Carol & Karen P. Corrigan. 2015. The rise of the going to future in Tyneside English: Evidence for further grammaticalisation. *English World-Wide* 36(2). 198–227.
- Ferguson, Charles A. 1975. Toward a Characterization of English Foreigner Talk. *Anthropological Linguistics* 17(1). 1–14.
- Field, Andy P., Jeremy Miles & Zoë Field. 2012. *Discovering statistics using R*. London: Sage.
- Fox, John. 2003. Effect Displays in R for Generalised Linear Models. *Journal of Statistical Software* 8. 1–27.
- Gablasova, Dana & Vaclav Brezina. 2015. Does Speaker Role Affect the Choice of Epistemic Adverbials in L2 Speech? Evidence from the Trinity Lancaster Corpus. In Jesús Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2015: Current Approaches to Discourse and Translation Studies* (Yearbook of Corpus Linguistics and Pragmatics 3), 117–136. Cham: Springer International Publishing.
- Gablasova, Dana, Vaclav Brezina & Tony McEnery. 2019. The Trinity Lancaster Corpus: Development, description and application. *International Journal of Learner Corpus Research* 5(2). 126–158.
- Gablasova, Dana, Vaclav Brezina, Tony McEnery & Elaine Boyd. 2017. Epistemic Stance in Spoken L2 English: The Effect of Task and Speaker Style. *Applied Linguistics* 38(5). 613–637.
- Gardner, Matt Hunt. 2017. *Grammatical variation and change in industrial Cape Breton*. Toronto: University of Toronto.
- Gelman, Andrew. 2008. Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine* 27(15). 2865–2873.
- Gelman, Andrew & John Carlin. 2014. Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives on Psychological Science* 9(6). 641–651.
- Gilquin, Gaëtanelle. 2015. The use of phrasal verbs by French-speaking EFL learners. A constructional and collostructional corpus-based approach. *Corpus Linguistics and Linguistic Theory* 11(1). 51–88.
- Gilquin, Gaëtanelle & Stefan Th. Gries. 2009. Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory* 5(1). 1–26.
- Gliem, Joseph A. & Rosemary R. Gliem. 2003. Calculating, Interpreting, And Reporting Cronbach's Alpha Reliability Coefficient For Likert-Type Scales. In *Proceedings of the 2003 Midwest Research to Practice Conference in Adult, Continuing, and Community Education*, 82–88. Ohio, USA: Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education.

- Goldberg, Adele E. 1992. The inherent semantics of argument structure: The case of the English ditransitive construction. *Cognitive Linguistics* 3(1). 37–74.
- Gómez, Rebecca L. 2002. Variability and Detection of Invariant Structure. *Psychological Science* 13(5). 431–436.
- Götz, Sandra. 2019. Filled pauses across proficiency levels, L1s and learning context variables: A multivariate exploration of the Trinity Lancaster Corpus Sample : *International Journal of Learner Corpus Research* 5(2). 159–180.
- Götz, Sandra, Christoph Wolk & Katja Jäschke. 2022. A contrastive interlanguage analysis across L1s, task types and learning context variables: Contextualizing fluency in advanced spoken learner language. In Agnieszka Leńko-Szymańska & Sandra Götz (eds.), *Complexity, Accuracy and Fluency in Learner Corpus Research* (Studies in Corpus Linguistics 104), 273–298. Amsterdam: John Benjamins Publishing Company.
- Grafmiller, Jason. 2014. Variation in English genitives across modality and genres. *English Language & Linguistics* 18(3). 471–496.
- Grafmiller, Jason & Benedikt Szmrecsanyi. 2018. Mapping out particle placement in Englishes around the world: A study in comparative sociolinguistic analysis. *Language Variation and Change* 30(3). 385–412.
- Granger, Sylviane, Estelle Dagneaux, Fanny Meunier & Magali Paquot. 2009. *International Corpus of Learner English v2*. Louvain-la-Neuve: Presses universitaires de Louvain.
- Green, Peter & Catriona J. MacLeod. 2016. SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution* 7(4). 493–498.
- Gries, Stefan Th. 2003. *Multifactorial analysis in corpus linguistics: a study of particle placement*. London & New York: Continuum Press.
- Gries, Stefan Th. 2015. The most under-used statistical method in corpus linguistics: multi-level (and mixed-effects) models. *Corpora* 10(1). 95–125.
- Gries, Stefan Th. 2017. Syntactic alternation research: Taking stock and some suggestions for the future. *Belgian Journal of Linguistics* 31(1). 8–29.
- Gries, Stefan Th. & Allison S. Adelman. 2014. Subject Realization in Japanese Conversation by Native and Non-native Speakers: Exemplifying a New Paradigm for Learner Corpus Research. In Jesús Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2014* (Yearbook of Corpus Linguistics and Pragmatics 2), 35–54. Cham: Springer International Publishing.
- Gries, Stefan Th. & Sandra C. Deshors. 2014. Using regressions to explore deviations between corpus data and a standard/target: two suggestions. *Corpora* 9(1). 109–136.

- Gries, Stefan Th. & Sandra C. Deshors. 2015. EFL and/vs. ESL? A multi-level regression modeling perspective on bridging the paradigm gap. *International Journal of Learner Corpus Research* 1(1). 130–159.
- Gries, Stefan Th. & Sandra C. Deshors. 2020. There's more to alternations than the main diagonal of a 2×2 confusion matrix: Improvements of MuPDAR and other classificatory alternation studies. *ICAME Journal* 44(1). 69–96.
- Gries, Stefan Th. & Nick C. Ellis. 2015. Statistical Measures for Usage-Based Linguistics. *Language Learning* 65(S1). 228–255.
- Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collocation analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9(1). 97–129.
- Gries, Stefan Th. & Stefanie Wulff. 2005. Do foreign language learners also have constructions? Evidence from priming, sorting, and corpora. *Annual Review of Cognitive Linguistics* 3(1). 182–200.
- Gries, Stefan Th. & Stefanie Wulff. 2013. The genitive alternation in Chinese and German ESL learners: Towards a multifactorial notion of context in learner corpus research. *International Journal of Corpus Linguistics* 18(3). 327–356.
- Grondelaers, Stefan & Dirk Speelman. 2007. A variationist account of constituent ordering in presentative sentences in Belgian Dutch. *Corpus Linguistics and Linguistic Theory* 3(2). 161–193.
- Gullberg, Marianne, Leah Roberts, Christine Dimroth, Kim Veroude & Peter Indefrey. 2010. Adult Language Learning After Minimal Exposure to an Unknown Natural Language. *Language Learning* 60(s2). 5–24.
- Gundel, Jeanette K. 1988. Universals of topic-comment structure. In Michael Hammond, Edith A. Moravcsik & Jessica Wirth (eds.), *Studies in Syntactic Typology*, 209–239. Amsterdam: John Benjamins.
- Guy, Gregory R. & Charles Boberg. 1997. Inherent variability and the obligatory contour principle. *Language Variation and Change* 9(2). 149–164.
- Haegeman, Liliane. 1989. Be going to and will: A pragmatic account. *Journal of Linguistics* 25(2). 291–317.
- Haiman, John. 1980. The Iconicity of Grammar: Isomorphism and Motivation. *Language* 56(3). 515–540.
- Hansen, Tricia. 2015. *Pass Trinity Now 9-10*. Black Cat.
- Hartsuiker, Robert J., Sarah Bernolet, Sofie Schoonbaert, Sara Speybroeck & Dieter Vanderelst. 2008. Syntactic priming persists while the lexical boost decays: Evidence from written and spoken dialogue. *Journal of Memory and Language* 58(2). 214–238.
- Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
- Hay, Jessica F., Bruna Pelucchi, Katharine Graf Estes & Jenny R. Saffran. 2011. Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology* 63(2). 93–106.

- Heller, Benedikt, Tobias Bernaisch & Stefan Th. Gries. 2017. Empirical perspectives on two potential epicenters: The genitive alternation in Asian Englishes. *ICAME Journal* 41(1). 111–144.
- Heller, Benedikt, Benedikt Szmrecsanyi & Jason Grafmiller. 2017. Stability and fluidity in syntactic variation world-wide: the genitive alternation across varieties of English. *Journal of English Linguistics* 45(1). 3–27.
- Heller, Daniel-Benedikt. 2018. *Stability and fluidity in syntactic variation world-wide: The genitive alternation across varieties of English*. Leuven: KU Leuven.
- Heumann, Christian, Michael Schomaker & Michael Shalabh. 2016. *Introduction to Statistics and Data Analysis: With Exercises, Solutions and Applications in R*. Cham: Springer International Publishing.
- Hinrichs, Lars & Benedikt Szmrecsanyi. 2007. Recent changes in the function and frequency of Standard English genitive constructions: a multivariate analysis of tagged corpora. *English Language & Linguistics* 11(3). 437–474.
- Hinrichs, Lars, Benedikt Szmrecsanyi & Axel Bohmann. 2015. Which-hunting and the Standard English relative clause. *Language* 91(4). 806–836.
- Hoffmann, Thomas. 2006. Corpora and introspection as corroborating evidence: The case of preposition placement in English relative clauses. *Corpus Linguistics and Linguistic Theory* 2(2). 165–195.
- Hopper, Paul J. & Elizabeth Closs Traugott. 2003. *Grammaticalization* (Cambridge Textbooks in Linguistics 61). 2nd edn. Cambridge: Cambridge University Press.
- Hosmer, David W. & Stanley Lemeshow. 2000. *Applied Logistic Regression* (Wiley Series in Probability and Statistics). 2nd edn. New York (N.Y.): Wiley.
- Huang, Danqing. 2022. *Diachronic Prototype Semantics of Chinese Radicals*. Leuven: KU Leuven.
- Huddleston, Rodney D. & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Humphries, Jennie. 2017a. *Ready for Trinity GESE Grades 3-4 and ISE Foundation*. ELI Publishing.
- Humphries, Jennie. 2017b. *Ready for Trinity GESE Grades 5-6 and ISE I*. ELI Publishing.
- Hundt, Marianne & Christian Mair. 1999. “Agile” and “Uptight” Genres: The Corpus-based Approach to Language Change in Progress. *International Journal of Corpus Linguistics* 4(2). 221–242.
- Hundt, Marianne & Benedikt Szmrecsanyi. 2012. Animacy in early New Zealand English. *English World-Wide* 33(3). 241–263.
- Ionin, Tania, Marí L. Zubizarreta & Vadim Philippov. 2009. Acquisition of article semantics by child and adult L2-English learners. *Bilingualism: Language and Cognition* 12(3). 337–361.

- Jackson, Carrie N. 2018. Second language structural priming: A critical review and directions for future research. *Second Language Research* 34(4). 539–552.
- Jaeger, T. Florian & Neal E. Snider. 2013. Alignment as a consequence of expectation adaptation: syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition* 127(1). 57–83.
- Jäschke, Katja & Ingo Plag. 2016. The dative alternation in German-English interlanguage. *Studies in Second Language Acquisition* 38(3). 485–521.
- Keating, Gregory D. & Jill Jegerski. 2015. Experimental designs in sentence processing research: A methodological review and user's guide. *Studies in Second Language Acquisition* 37(1). 1–32.
- Kemmerer, David. 2015. *The cognitive neuroscience of language*. New York, NY: Psychology Press.
- Kempen, Gerard & Karin Harbusch. 2005. The Relationship between Grammaticality Ratings and Corpus Frequencies: A Case Study into Word Order Variability in the Midfield of German Clauses. In Stephan Kepser & Marga Reis (eds.), *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives*, 329–350. New York: De Gruyter Mouton.
- Kinne, Alexandra. 2020. *Particle Placement in English L1 and L2 Academic Writing*. Louvain-la-Neuve: Presses universitaires de Louvain.
- Klavan, Jane. 2020. Pitting corpus-based classification models against each other: a case study for predicting constructional choice in written Estonian. *Corpus Linguistics and Linguistic Theory* 16(2). 363–391.
- Klavan, Jane & Dagmar Divjak. 2016. The cognitive plausibility of statistical classification models: Comparing textual and behavioral evidence. *Folia Linguistica* 50(2). 355–384.
- Koptjevskaja-Tamm, Maria. 2001. Adnominal possession. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language Typology and Language Universals* (Handbooks of Linguistics and Communication Science 20), vol. 2, 960–970. Berlin and New York: Mouton de Gruyter.
- Kreyer, Rolf. 2003. Genitive and of-construction in modern written English. Processability and human involvement. *International Journal of Corpus Linguistics* 8(2). 169–207.
- Kroll, Judith F. & Paola E. Dussias. 2013. The comprehension of words and sentences in two languages. In Tej K. Bhatia & William C. Ritchie (eds.), *The Handbook of Bilingualism and Multilingualism*, 216–243. Malden, MA: Wiley-Blackwell Publishers.
- Kumle, Leah, Melissa L.-H. Võ & Dejan Draschkow. 2021. Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods* 2528–2543.
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Philadelphia Press.

- Lammertink, Imme, Paul Boersma, Frank Wijnen & Judith Rispens. 2017. Statistical Learning in Specific Language Impairment: A Meta-Analysis. *Journal of Speech, Language and Hearing Research* 60(12). 3474–3486.
- Landis, J. Richard & Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33(1). 159–174.
- Larsson, Tove, Magali Paquot & Luke Plonsky. 2020. Inter-rater reliability in Learner Corpus Research: Insights from a collaborative study on adverb placement. *International Journal of Learner Corpus Research* 6(2). 237–251.
- Lazaraton, Anne. 1996. Interlocutor support in oral proficiency interviews: the case of CASE. *Language Testing* 13(2). 151–172.
- Le Foll, Elen. 2023. The Potential Impact of EFL Textbook Language on Learner English: A Triangulated Corpus Study. In Kieran Harrington & Patricia Ronan (eds.), *Demystifying Corpus Linguistics for English Language Teaching*, 259–287. Cham: Palgrave Macmillan.
- Lenzo, Kevin. 2014. *The CMU Pronouncing Dictionary*.
- Levin, Beth. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago, IL: University of Chicago Press.
- Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins.
- Li, Yi, Benedikt Szmrecsanyi & Weiwei Zhang. 2023. The theme-recipient alternation in Chinese: tracking syntactic variation across seven centuries. *Corpus Linguistics and Linguistic Theory* 19(2). 207–235.
- Linck, Jared A. & Ian Cunnings. 2015. The Utility and Application of Mixed-Effects Models in Second Language Research. *Language Learning* 65(S1). 185–207.
- MacDonald, Maryellen C. 2013. How language production shapes language form and comprehension. *Frontiers in Psychology* 4(Article 226). 1–16.
- MacWhinney, Brian. 2001. The Competition Model: The input, the context, and the brain. In Peter Robinson (ed.), *Cognition and Second Language Instruction*, 69–90. New York, NY: Cambridge University Press.
- MacWhinney, Brian. 2005. Extending the competition model. *The International Journal of Bilingualism* 9(1). 69–84.
- MacWhinney, Brian. 2014. What we have learned. *Journal of Child Language* 41(S1). 124–131.
- MacWhinney, Brian. 2017. Exposure is not enough. *Bilingualism: Language and Cognition* 20(1). 25–26.
- Marneffe, Marie-Catherine de, Scott Grimm, Inbal Arnon, Susannah Kirby & Joan Bresnan. 2012. A statistical model of the grammatical choices in child production of dative sentences. *Language and Cognitive Processes* 27(1). 25–61.
- Mazzola, Giulia, Bert Cornillie & Malte Rosemeyer. 2022. Asyndetic complementation and referential integration in Spanish: A

- diachronic probabilistic grammar account. *Journal of Historical Linguistics* 12(2). 194–240.
- Milin, Petar, Dagmar Divjak, Strahinja Dimitrijević & R. Harald Baayen. 2016. Towards cognitively plausible data science in language research. *Cognitive Linguistics* 27(4). 507–526.
- Mishan, Freda. 2022. The Global ELT coursebook: A case of Cinderella's slipper? *Language Teaching* 55(4). 490–505.
- Misyak, Jennifer B. & Morten H. Christiansen. 2012. Statistical learning and language: An individual differences study. *Language Learning* 62(1). 302–331.
- Morett, Laura M. & Brian MacWhinney. 2013. Syntactic transfer in English-speaking Spanish learners. *Bilingualism: Language and Cognition* 16(1). 132–151.
- Murphy, Raymond. 2012. *English Grammar in Use. A self-study reference and practice book for intermediate learners of English*. Cambridge: Cambridge University Press.
- Murphy, Raymond. 2019. *English Grammar in Use. A self-study reference and practice book for intermediate learners of English*. Cambridge: Cambridge University Press.
- Myles, Florence, Janet Hooper & Rosamond Mitchell. 1998. Rote or Rule? Exploring the Role of Formulaic Language in Classroom Foreign Language Learning. *Language Learning* 48(3). 323–364.
- Myles, Florence & Rosamond Mitchell. 2014. *Second language learning theories*. London: Routledge.
- Myles, Florence, Rosamond Mitchell & Janet Hooper. 1999. INTERROGATIVE CHUNKS IN FRENCH L2: A Basis for Creative Construction? *Studies in Second Language Acquisition* 21(1). 49–80.
- Nesselhauf, Nadja. 2007. The spread of the progressive and its future use. *English Language and Linguistics* 11(1). 191–207.
- Nesselhauf, Nadja. 2010. The development of future time expressions in Late Modern English: redistribution of forms or change in discourse? *English Language & Linguistics* 14(2). 163–186.
- Nicolle, Steve. 1997. A relevance-theoretic account of be going to. *Journal of Linguistics* 33(2). 355–377.
- O'Shannessy, Carmel & Felicity Meakins. 2012. Comprehension of competing argument marking systems in two Australian mixed languages. *Bilingualism: Language and Cognition* 15(2). 378–396.
- Osselton, Noel E. 1988. Thematic genitives. In Graham Nixon & John Honey (eds.), *An Historic Tongue: Studies in English Linguistics in Memory of Barbara Strang*, 138–145. London: Routledge.
- Oxenden, Clive & Christina Latham-Koenig. 2006. *New English File: Advanced Student's book*. Oxford: Oxford University Press.
- Paquot, Magali, Jason Grafmiller & Benedikt Szmrecsanyi. 2019. Particle placement alternation in EFL learner vs. L1 speech: assessing the similarity of probabilistic grammars. In Andrea Abel, Aivars Glaznieks, Verena Lyding & Lionel Nicolas (eds.), *Widening the*

- Scope of Learner Corpus Research. Selected Papers from the Fourth Learner Corpus Research Conference*, 71–92. Louvain-la-Neuve: Presses universitaires de Louvain.
- Parodi, Teresa, Bonnie D Schwartz & Harald Clahsen. 2004. On the L2 acquisition of the morphosyntax of German nominals. *Linguistics* 42(3). 669–705.
- Patil, Prabahakar B. 1994. Strategies of “Teacher Talk.” *IRAL: International Review of Applied Linguistics in Language Teaching* 32(2). 154–165.
- Penke, Martina & Anette Rosenbach. 2007. An introduction: What counts as evidence in linguistics? In Martina Penke & Anette Rosenbach (eds.), *What Counts as Evidence in Linguistics: The case of innateness* (Benjamins Current Topics 7), 1–49. John Benjamins Publishing Company.
- Philipp, Markus, Tim Graf, Franziska Kretzschmar & Beatrice Primus. 2017. Beyond Verb Meaning: Experimental Evidence for Incremental Processing of Semantic Roles and Event Structure. *Frontiers in Psychology* 8(Article 1806). 1–16.
- Pickering, Martin J. & Victor S. Ferreira. 2008. Structural Priming: A Critical Review. *Psychological bulletin* 134(3). 427–459.
- Pijpops, Dirk, Dirk Speelman, Stefan Grondelaers & Freek Van De Velde. 2018. Comparing explanations for the Complexity Principle: Evidence from argument realization. *Language and Cognition* 10(3). 514–543.
- Poplack, Shana & Sali Tagliamonte. 2000. The grammaticization of going to in (African American) English. *Language Variation and Change* 11(3). 315–342.
- Prolific. 2023. Prolific. London, UK: Prolific. <https://www.prolific.com>.
- Qualtrics. 2023. Qualtrics. Provo, Utah, USA: Qualtrics. <https://www.qualtrics.com>.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London, New York: Longman.
- R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Raumolin-Brunberg, Helena & Leena Kahlas-Tarkka. 1997. Indefinite pronouns with singular human reference. In Matti Rissanen, Merja Kytö & Kirsi Heikkonen (eds.), *Grammaticalization at Work: Studies of Long-term Developments in English*, 17–86. Berlin, Boston: De Gruyter Mouton.
- Rautionaho, Paula, Sandra C. Deshors & Lea Meriläinen. 2018. Revisiting the ENL-ESL-EFL continuum: A multifactorial approach to grammatical aspect in spoken Englishes. *ICAME Journal* 42(1). 41–78.
- Rebuschat, Patrick & John N. Williams. 2011a. Introduction: Statistical Learning and Language Acquisition. In Patrick Rebuschat & John N. Williams (eds.), *Statistical Learning and Language Acquisition*. Boston/Berlin: De Gruyter Mouton.

- Rebuschat, Patrick & John N. Williams (eds.). 2011b. *Statistical Learning and Language Acquisition*. Boston/Berlin: De Gruyter Mouton.
- Redstone, Chris & Gillie Cunningham. 2005a. *face2face elementary*. Cambridge University Press.
- Redstone, Chris & Gillie Cunningham. 2005b. *face2face pre-intermediate*. Cambridge University Press.
- Redstone, Chris & Gillie Cunningham. 2005c. *face2face intermediate*. Cambridge University Press.
- Redstone, Chris & Gillie Cunningham. 2012. *face2face upper-intermediate*. 2nd edn. Cambridge University Press.
- Révész, Andrea. 2012. Coding Second Language Data Validly and Reliably. In Alison Mackey & Susan M. Gass (eds.), *Research Methods in Second Language Acquisition: A Practical Guide*, 201–221. Malden, MA: Wiley-Blackwell Publishers.
- Rohdenburg, Günter. 1996. Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics* 7(2). 149–182.
- Romano, Francesco. 2016. Syntactic planning of English genitives in L1 and L2 production: The Animacy Rule Model. *Lingua* 184. 104–121.
- Romano, Francesco. 2018. The basic continuity hypothesis of L1 to L2 production. *Second Language Research* 34(3). 275–308.
- Romberg, Alexa R. & Jenny R. Saffran. 2010. Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews. Cognitive science* 1(6). 906–914.
- Römer, Ute, Audrey Roberson, Matthew B. O'Donnell & Nick C. Ellis. 2014. Linking learner corpus and experimental data in studying second language learners' knowledge of verb-argument constructions. *ICAME Journal* 38(1). 115–135.
- Rosenbach, Anette. 2002. *Genitive variation in English. Conceptual factors in synchronic and diachronic studies* (Topics in English Linguistics 42). Berlin/New York: Mouton de Gruyter.
- Rosenbach, Anette. 2014. English genitive variation - the state of the art. *English Language and Linguistics* 18(2). 215–262.
- Rosenbach, Anette. 2017. Constraints in contact: Animacy in English and Afrikaans genitive variation – a cross-linguistic perspective. *Glossa* 2(1). 72.
- Röthlisberger, Melanie. 2018. *Regional variation in probabilistic grammars: a multifactorial study of the English dative alternation*. Leuven: KU Leuven.
- Röthlisberger, Melanie, Jason Grafmiller & Benedikt Szmrecsanyi. 2017. Cognitive indigenization effects in the English dative alternation. *Cognitive Linguistics* 18(4). 673–710.
- Royster, Jane F. & John M. Steadman. 1923. The “going-to” future. In *Manly Anniversary Studies in Languages and Literature*, 394–403. Freeport, NY: Books for Libraries Press.

- Sag, Ivan A., Thomas Wasow & Emily M. Bender. 2003. *Syntactic Theory: A Formal Introduction*. 2nd edn. Stanford, California: CSLI Publications.
- Sandlund, Erica, Pia Sundqvist & Lina Nyroos. 2016. Testing L2 talk: A Review of empirical studies on second-language oral proficiency testing. *Language and Linguistics Compass* 10(1). 14–29.
- Schmidt, Richard. 2001. Attention. In Peter Robinson (ed.), *Cognition and second language instruction*, 3–32. Cambridge: Cambridge University Press.
- Schneider, Agnes. 2016. The variation of will vs. be going to: Future time marking in spoken Ghanaian English. In Olga Timofeeva, Anne-Christine Gardner, Alpo Honkapohja & Sarah Chevalier (eds.), *New approaches to English linguistics: building bridges* (Studies in Language Companion Series 177), 141–173. Amsterdam: John Benjamins Publishing Company.
- Schönefeld, Doris. 2011. Introduction: On evidence and the convergence of evidence in linguistic research. In Doris Schönefeld (ed.), *Converging Evidence: Methodological and theoretical issues for linguistic research*, 1–32. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Schütze, Carson T. & Jon Sprouse. 2013. Judgment data. In Robert J. Podesva & Devjani Sharma (eds.), *Research Methods in Linguistics*, 27–50. New York: Cambridge University Press.
- Segalowitz, Norman. 2003. Automaticity and Second Languages. In Catherine J. Doughty & Michael H. Long (eds.), *The Handbook of Second Language Acquisition*, 382–408. Malden, MA: Blackwell Publishing Ltd.
- Selkirk, Elisabeth O. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3. 371–405.
- Shih, Stephanie, Jason Grafmiller, Richard Futrell & Joan Bresnan. 2015. Rhythm's role in genitive construction choice in spoken English. In Ralf Vogel & Ruben Vijver (eds.), *Rhythm in Cognition and Grammar*, 207–234. Berlin, München, Boston: De Gruyter Mouton.
- Shin, Naomi Lapidus. 2014. Grammatical complexification in Spanish in New York: 3sg pronoun expression and verbal ambiguity. *Language Variation and Change* 26(3). 303–330.
- Silverstein, Michael. 1976. Hierarchy of Features and Ergativity. In Robert M. W. Dixon (ed.), *Grammatical Categories in Australian Languages*, 112–171. Australian Institute of Aboriginal Studies.
- Soars, Liz & John Soars. 2005. *Headway pre-intermediate*. 3rd edn. Oxford: Oxford University Press.
- Soars, Liz & John Soars. 2009a. *Headway beginner*. 4th edn. Oxford: Oxford University Press.
- Soars, Liz & John Soars. 2009b. *Headway intermediate*. 4th edn. Oxford: Oxford University Press.
- Soars, Liz & John Soars. 2009c. *Headway advanced*. 4th edn. Oxford: Oxford University Press.

- Soars, Liz & John Soars. 2019. *Headway upper-intermediate*. 5th edn. Oxford: Oxford University Press.
- Song, Sooho. 2018. Second Language Acquisition Theories. In Sooho Song (ed.), *Second Language Acquisition as a Mode-Switching Process: An Empirical Analysis of Korean Learners of English*, 9–36. London: Palgrave Macmillan UK.
- Speelman, Dirk. 2014. Logistic regression: A confirmatory technique for comparisons in corpus linguistics. In Dylan Glynn & Justyna A. Robinson (eds.), *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy* (Human Cognitive Processing: Cognitive Foundations of Language Structure and Use 43), 487–533. Amsterdam: John Benjamins Publishing Company.
- Stanley, Phiona & Marie Stevenson. 2017. Making sense of not making sense: Novice English language teacher talk. *Linguistics and Education* 38. 1–10.
- Stefanowitsch, Anatol. 2006. Negative evidence and the raw frequency fallacy. *Corpus Linguistics and Linguistic Theory* 2(1). 61–77.
- Stowe, Laurie & Edith Kaan. 2006. *Developing an Experiment. Techniques and Design*. Unpublished.
- Swan, Michael. 2005. *Practical English Usage*. 3rd edn. Oxford: Oxford University Press.
- Szmrecsanyi, Benedikt. 2003. Be Going to Versus Will/Shall: Does Syntax Matter? *Journal of English Linguistics* 31(4). 295–323.
- Szmrecsanyi, Benedikt. 2006. *Morphosyntactic persistence in spoken English. A corpus study at the intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis* (Trends in Linguistics. Studies and Monographs 177). Berlin/New York: Mouton de Gruyter.
- Szmrecsanyi, Benedikt. 2013. The great regression: genitive variability in Late Modern English news texts. In Kersti Börjars, David Denison & Alan Scott (eds.), *Morphosyntactic categories and the expression of possession*, 59–88. Amsterdam: Benjamins.
- Szmrecsanyi, Benedikt, Douglas Biber, Jesse Egbert & Karlien Franco. 2016. Toward more accountability: Modeling ternary genitive variation in Late Modern English. *Language Variation and Change* 28(1). 1–29.
- Szmrecsanyi, Benedikt, Jason Grafmiller, Benedikt Heller & Melanie Röthlisberger. 2016. Around the world in three alternations: modeling syntactic variation in varieties of English. *English World-Wide* 37(2). 109–137.
- Szmrecsanyi, Benedikt & Lars Hinrichs. 2008. Probabilistic determinants of genitive variation in spoken and written English: a multivariate comparison across time, space, and genres. In Terttu Nevalainen, Irma Taavitsainen, Päiva Pahta & Minna Korhonen (eds.), *The Dynamics of Linguistic Variation: Corpus Evidence on English Past and Present*, 291–309. Amsterdam: Benjamins.

- Tagliamonte, Sali. 2012. *Variationist sociolinguistics: change, observation, interpretation*. Malden, MA: Wiley-Blackwell.
- Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change* 24(2). 135–178.
- Tagliamonte, Sali A., Mercedes Durham & Jennifer Smith. 2014. Grammaticalization at an early stage: future be going to in conservative British dialects. *English Language and Linguistics* 18(1). 75–108.
- Tamminga, Meredith, Laurel MacKenzie & David Embick. 2016. The dynamics of variation in individuals. *Linguistic variation* 16(2). 300–336.
- Taylor, John R. 1989. *Linguistic categorization: Prototypes in Linguistic Theory*. Oxford: Clarendon Press.
- Timmers, Inge, Francesco Gentile, M. Rubio-Gozalbo & Bernadette Jansma. 2013. Temporal Characteristics of Online Syntactic Sentence Planning: An Event-Related Potential Study. *PLoS One* 8(12). e82884.
- Tooley, Kristen M. & Kathryn Bock. 2014. On the parity of structural persistence in language production and comprehension. *Cognition* 132(2). 101–136.
- Torres-Cacoullous, Rena & James A. Walker. 2009. The Present of the English Future: Grammatical Variation and Collocations in Discourse. *Language* 85(2). 321–354.
- Tyler, Andrea & Hana Jan. 2017. Be going to and will: talking about the future using embodied experience. *Language and Cognition* 9(3). 412–445.
- Ünel, Arzu & Meryem Mirioğlu. 2015. Syntactical modifications in Teacher Talk of native and non-native speakers in EFL classrooms. *Theory and Practice of Second Language Acquisition* 1(1). 125–141.
- Van den Bosch, Antal & Joan Bresnan. 2015. Modeling dative alternations of individual children. In *Proceedings of the Sixth Workshop on Cognitive Aspects of Computational Language Learning*, 103–112. Lisbon, Portugal: Association for Computational Linguistics.
- VanPatten, Bill. 1996. *Input processing and grammar instruction in second language acquisition*. New York: Ablex.
- Verhagen, Véronique, Maria Mos, Joost Schilperoord & Ad Backus. 2020. Variation is information: Analyses of variation across items, participants, time, and methods in metalinguistic judgment data. *Linguistics* 58(1). 37–81.
- Wang, Jia & Caihua Xu. 2015. Cue Competition between Animacy and Word Order: Acquisition of Chinese Notional Passives by L2 Learners. *Open Journal of Modern Linguistics* 5. 213–224.
- Wasow, Thomas. 1997. End-Weight from the Speaker's Perspective. *Journal of Psycholinguistic Research* 26(3). 347–61.

- Weiss, Daniel J., Chip Gerfen & Aaron D. Mitchel. 2010. Colliding cues in word segmentation: The role of cue strength and general cognitive processes. *Language and Cognitive Processes* 25(3). 402–422.
- Wekker, Herman C. 1976. *The expression of future time in contemporary British English: an investigation into the syntax and semantics of five verbal constructions expressing futurity*. (North-Holland Linguistic Series 28). Amsterdam: North-Holland Publishing Company.
- Winter, Tatjana & Elen Le Foll. 2022. Testing the pedagogical norm: Comparing if-conditionals in EFL textbooks, learner writing and English outside the classroom. *International Journal of Learner Corpus Research* 8(1). 31–66.
- Wolk, Christoph, Joan Bresnan, Anette Rosenbach & Benedikt Szmrecsanyi. 2013. Dative and genitive variability in Late Modern English: Exploring cross-constructural variation and change. *Diachronica* 30(3). 382–419.
- Wooldridge, Blue. 2001. ‘Foreigner Talk’: An Important Element in Cross-Cultural Management Education and Training. *International Review of Administrative Sciences* 67(4). 621–634.
- Wulff, Stefanie. 2016. A Friendly Conspiracy of Input, L1, and Processing Demands: That-variation in the Language of German and Spanish Learners of English. In Andrea Tyler, Lourdes Ortega, Hae In Park & Mariko Uno (eds.), *The usage-based study of language learning and multilingualism*. Georgetown: Georgetown University Press.
- Wulff, Stefanie & Nick C. Ellis. 2018. Usage-based approaches to SLA. In David Miller, Fatih Bayram, Jason Rotham & Ludovica Serratrice (eds.), *Bilingual Cognition and Language: The state of the science across its subfields*, 37–56. Amsterdam: John Benjamins.
- Wulff, Stefanie & Stefan Th. Gries. 2015. Prenominal adjective order preferences in Chinese and German L2 English: A multifactorial corpus study. *Linguistic Approaches to Bilingualism* 5(1). 122–150.
- Wulff, Stefanie & Stefan Th. Gries. 2019. Particle Placement in Learner Language. *Language Learning* 69(4). 873–910.
- Wulff, Stefanie, Stefan Th. Gries & Nicholas Lester. 2018. Optional that in complementation by German and Spanish learners. In Andrea Tyler, Lihong Huang & Hana Jan (eds.), *What is applied cognitive linguistics? Answers from current SLA research*, 99–120. Berlin & Boston: De Gruyter Mouton.
- Wulff, Stefanie, Nicholas Lester & Maria T. Martinez-Garcia. 2014. That-variation in German and Spanish L2 English. *Language and Cognition* 6(2). 271–299.
- Yamamoto, Mutsumi. 1999. *Animacy and Reference: A cognitive approach to corpus linguistics* (Studies in Language Companion Series 46). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Yi, Wei, Shiyi Lu & Guojie Ma. 2017. Frequency, contingency and online processing of multiword sequences: An eye-tracking study. *Second Language Research* 33(4). 519–549.

- Yoshimura, Yuki & Brian MacWhinney. 2010. Honorifics: A sociocultural verb agreement cue in Japanese sentence processing. *Applied Psycholinguistics* 31(3). 551–569.
- Zaenen, Annie, Jean Carletta, Gregory Garretson, Joan Bresnan, Andrew Koontz-Garboden, Tatiana Nikitina, M. Catherine O'Connor & Tom Wasow. 2004. Animacy encoding in English: Why and how. In Bonnie Webber & Donna Byron (eds.), *Proceedings of the 2004 ACL Workshop on Discourse Annotation, Barcelona, July 2004*, 118–125. East Stroudsburg, PA: Association for Computational Linguistics.
- Zhang, Xiaoyu & Sang-Gu Kang. 2023. Applying given-before-new principle in L2 English datives development. *Language Acquisition*.
- Zuur, Alain F., Elena N. Ieno & Chris S. Elphick. 2010. A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution* 1(1). 3–14.
- Zuur, Alain F., Elena N. Ieno, Neil J. Walker, Anatoly A. Saveliev & Graham M. Smith. 2009. Mixed effects modelling for nested data. In Alain F. Zuur, Elena N. Ieno, Neil Walker, Anatoly A. Saveliev & Graham M. Smith (eds.), *Mixed effects models and extensions in ecology with R* (Statistics for Biology and Health), 101–142. New York, NY: Springer.

List of Tables

Table 3.1: Classification of the probabilistic constraints included in this study for the future marker alternation and the genitive alternation as either psychophysiological or not, inspired by the criteria of Tamminga, MacKenzie and Embick (2016).	36
Table 4.1: Overview of the coding method for the constraints of the future marker alternation and the genitive alternation.....	52
Table 5.1: Distribution of the potentially interchangeable future markers across native speakers and learners.....	67
Table 5.2: Distribution of the potentially interchangeable future markers across the different learner proficiency levels.....	67
Table 5.3: Distribution of the future markers across proficiency levels and L1 backgrounds.....	69
Table 5.4: Contingency table for sentence type for native speakers ($X^2 = 3.61$, $df = 1$, $p = 0.057$) and learners ($X^2 = 4.16$, $df = 1$, $p = 0.042$, Cramér's $V = 0.04$).	74
Table 5.5: Contingency table for negation for native speakers ($X^2 = 3.31$, $df = 1$, $p = 0.069$) and learners ($X^2 = 16.85$, $df = 1$, $p < 0.001$, Cramér's $V = 0.08$).	76
Table 5.6: Contingency table for clause type for native speakers (Fisher's exact test $p < 0.001$, Cramér's $V = 0.16$) and learners ($X^2 = 56.13$, $df = 3$, $p < 0.001$, Cramér's $V = 0.14$).	80
Table 5.7: Contingency table for persistence for native speakers ($X^2 = 27.27$, $df = 4$, $p < 0.001$, Cramér's $V = 0.19$) and learners ($X^2 = 465.29$, $df = 4$, $p < 0.001$, Cramér's $V = 0.41$).	83
Table 5.8: Contingency table for temporal proximity for native speakers ($X^2 = 10.39$, $df = 1$, $p = 0.001$, Cramér's $V = 0.12$) and learners ($X^2 = 290.17$, $df = 1$, $p < 0.001$, Cramér's $V = 0.32$).....	87
Table 5.9: Contingency table for temporal adverb for native speakers ($X^2 = 2.79$, $df = 1$, $p = 0.095$) and learners ($X^2 = 4.08.17$, $df = 1$, $p = 0.043$, Cramér's $V = 0.04$)	89

Table 5.10: Contingency table for verb type for native speakers ($X^2 = 3.09$, $df = 2$, $p = 0.214$) and learners ($X^2 = 112.92$, $df = 2$, $p < 0.001$, Cramér's $V = 0.2$).....	93
Table 5.11: Contingency table for subject grammatical person and animacy for native speakers ($X^2 = 6.4$, $df = 3$, $p = 0.094$) and learners ($X^2 = 152.39$, $df = 3$, $p < 0.001$, Cramér's $V = 0.23$).....	96
Table 5.12: Overview of the classifying features for each of the future marker constraints.	102
Table 5.13: The feature-based future marker constraints, ranked by the strength of association according to monofactorial tests (Cramér's V)....	102
Table 5.14: Overview of the future marker constraints, as well as the relevant coding distinctions, their expected effect on the choice of variant and their learnability via associative statistical learning.....	103
Table 5.15: Overview of the random effects in the model, their estimated variance and their standard deviation	108
Table 5.16: Main effects of the individual predictors on the choice of future marker.....	112
Table 5.17: Interaction effects of the individual predictors on the choice of future marker	114
Table 5.18: Interaction effects of the individual predictors on the choice of future marker without Mexican EFL learners.....	124
Table 5.19: Interaction effects of the individual predictors on the choice of future marker in the mother tongue model.....	126
Table 5.20: Distribution of the potentially interchangeable future markers for the native speakers and each of the learner proficiency levels.....	133
Table 5.21: The uses of <i>will</i> and <i>be going to</i> according to a sample of popular EFL textbooks and reference grammars	142
Table 6.1: Criteria used to manually determine on a case-by-case basis whether the potential <i>of</i> -genitives and <i>s</i> -genitives identified by the script are interchangeable.....	156
Table 6.2: Distribution of the interchangeable genitive variants across the different proficiency levels.....	163
Table 6.3: Distribution of the genitive observations across proficiency levels and L1 backgrounds.....	164

Table 6.4: Updated distribution of the genitive observations across proficiency levels and L1 backgrounds	166
Table 6.5: Contingency table for possessor discourse status for native speakers ($X^2 = 0.09$, $df = 1$, $p = 0.77$) and learners ($X^2 = 0.42$, $df = 1$, $p = 0.571$)	171
Table 6.6: Contingency table for persistence for native speakers ($X^2 = 19.62$, $df = 4$, $p < 0.001$, Cramér's $V = 0.19$) and learners ($X^2 = 55.27$, $df = 4$, $p < 0.001$, Cramér's $V = 0.18$).....	173
Table 6.7: Contingency table for possessor animacy for native speakers ($X^2 = 247.62$, $df = 4$, $p < 0.001$, Cramér's $V = 0.68$) and learners ($X^2 = 652.82$, $df = 4$, $p < 0.001$, Cramér's $V = 0.61$)	180
Table 6.8: Contingency table for the semantic relation between possessor and possessum for native speakers ($X^2 = 5.42$, $df = 1$, $p = 0.02$, Cramér's $V = 0.10$) and learners ($X^2 = 46.54$, $df = 1$, $p < 0.001$, Cramér's $V = 0.16$)	183
Table 6.9: Contingency table for possessor definiteness for native speakers ($X^2 = 0$, $df = 1$, $p = 0.944$) and learners ($X^2 = 25.60$, $df = 1$, $p < 0.001$, Cramér's $V = 0.12$)	185
Table 6.10: Contingency table for the final sibilancy of the possessor for native speakers ($X^2 = 0.78$, $df = 1$, $p = 0.377$) and learners ($X^2 = 10.05$, $df = 1$, $p = 0.002$, Cramér's $V = 0.08$)	187
Table 6.11: Overview of the classifying features for each of the genitive constraints.....	191
Table 6.12: The feature-based genitive constraints, ranked by the strength of association according to monofactorial tests (Cramér's V)	192
Table 6.13: Overview of the genitive constraints, as well as the relevant coding distinctions, their expected effect on the choice of variant and their learnability via associative statistical learning	193
Table 6.14: Overview of the random effects in the model, their estimated variance and their standard deviation.....	197
Table 6.15: Main effects of the individual predictors on the choice of genitive variant.....	202
Table 6.16: Interaction effects of the individual predictors on the choice of genitive variant.	204
Table 6.17: Main and interaction effects of the individual predictors on the choice of genitive variant in the mother tongue model.....	210

Table 6.18: Effects of the individual predictors on the choice of genitive variant in the native speaker data.....	212
Table 6.19: Main and interaction effects of the individual predictors for non-native choices in the learner data.....	215
Table 7.1: Distribution of the target items across probability bins and animacy levels.....	241
Table 7.2: Overview of the experimental items with the genitive variant they originally featured in the TLC data.	242
Table 7.3: Distribution of the learner participants by L1 background and proficiency level.....	248
Table 7.4: Pseudorandomized order of the 15 experimental items from Table 7.1.....	251
Table 7.5: Overview of the random effects in the animacy model, their estimated variance and their standard deviation.....	256
Table 7.6: Effects of the individual predictors on the ratings in the animacy model.....	257
Table 7.7: Overview of the random effects in the corpus prediction model, their estimated variance and their standard deviation.....	260
Table 7.8: Effects of the individual predictors on the ratings in the corpus prediction model.....	260
Table 7.9: Cronbach's alpha calculated on the ratings from the participants in each proficiency level as a measure of internal consistency.....	264
Table 8.1: Number of significant regressors between native speakers and learners across proficiency levels for the future marker alternation (Table 5.17) and the genitive alternation (Table 6.16).	286

List of Figures

Figure 4.1: An example of a target experimental item presented to participants.....	56
Figure 5.1: Distribution of the tokens across proficiency levels.....	70
Figure 5.2: Distribution of <i>will</i> and <i>be going to</i> across levels of sentence type for native speakers and learners.....	75
Figure 5.3: Distribution of <i>will</i> and <i>be going to</i> across levels of negation for native speakers and learners	77
Figure 5.4: Distribution of <i>will</i> and <i>be going to</i> across levels of clause type for native speakers and learners.....	81
Figure 5.5: Distribution of <i>will</i> and <i>be going to</i> across levels of persistence for native speakers and learners.....	84
Figure 5.6: Distribution of <i>will</i> and <i>be going to</i> across levels of temporal proximity for native speakers and learners	88
Figure 5.7: Distribution of <i>will</i> and <i>be going to</i> across levels of temporal adverb for native speakers and learners	90
Figure 5.8: Distribution of <i>will</i> and <i>be going to</i> across levels of verb type for native speakers and learners	93
Figure 5.9: Distribution of <i>will</i> and <i>be going to</i> across levels of subject grammatical person and animacy for native speakers and learners.....	97
Figure 5.10: Smoothed conditional means of the proportion of <i>be going to</i> by the letter length of the clause for native speakers and learners.....	99
Figure 5.11: Intercept adjustments for the country where the exam was taken	109
Figure 5.12: Intercept adjustments for the verb lemmas	110
Figure 5.13: Partial effects plot of the interaction between proficiency level and negation.....	115
Figure 5.14: Partial effects plot of the interaction between proficiency level and persistence.	116

Figure 5.15: Partial effects plot of the interaction between proficiency level and temporal proximity.....	117
Figure 5.16: Partial effects plot of the interaction between proficiency level and temporal adverb.....	119
Figure 5.17: Partial effects plot of the interaction between proficiency level and verb type.....	120
Figure 5.18: Partial effects plot of the interaction between proficiency level and clause length.....	122
Figure 6.1: Schematic representation of the noun-phrase structure of the <i>of</i> -genitive.....	160
Figure 6.2: Schematic representation of the noun-phrase structure of the <i>s</i> -genitive.....	161
Figure 6.3: Distribution of the tokens across proficiency levels.....	165
Figure 6.4: Smoothed conditional means of the proportion of <i>s</i> -genitives by the letter length of the possessor and the possessum for native speakers and learners.....	169
Figure 6.5: Distribution of the <i>s</i> -genitive and the <i>of</i> -genitive across levels of possessor discourse status for native speakers and learners.....	171
Figure 6.6: Distribution of the <i>s</i> -genitive and the <i>of</i> -genitive across levels of persistence for native speakers and learners.....	174
Figure 6.7: Smoothed conditional means of the proportion of <i>s</i> -genitives by the thematicity of the possessor for native speakers and learners.....	175
Figure 6.8: Smoothed conditional means of the proportion of <i>s</i> -genitives by the type-token ratio of the surrounding context for native speakers and learners.....	177
Figure 6.9: Distribution of the <i>s</i> -genitive and the <i>of</i> -genitive across levels of possessor animacy for native speakers and learners.....	181
Figure 6.10: Distribution of the <i>s</i> -genitive and the <i>of</i> -genitive across levels of semantic relation for native speakers and learners.....	183
Figure 6.11: Distribution of the <i>s</i> -genitive and the <i>of</i> -genitive across levels of possessor definiteness for native speakers and learners.....	185
Figure 6.12: Distribution of the <i>s</i> -genitive and the <i>of</i> -genitive across levels of possessor final sibilancy for native speakers and learners.....	188
Figure 6.13: Intercept adjustments for the heads of the possessors.....	198

Figure 6.14: Intercept adjustments for the heads of the possessums	200
Figure 6.15: Partial effects plot of the interaction between proficiency level and possessor definiteness	205
Figure 6.16: Partial effects plot of the interaction between proficiency level and possessor animacy	207
Figure 7.1: An example of a target experimental item presented to participants	235
Figure 7.2: Visualization of statistical power for a change from an animate to an inanimate possessor at the B1 level compared to native speakers	246
Figure 7.3: Instructions shown to participants for the target experimental items with a concrete example	250
Figure 7.4: Partial effects plot of the interaction between proficiency level and possessor animacy on the ratings	259
Figure 7.5: Partial effects plot of the interaction effect between proficiency level and the corpus-based prediction for the <i>s</i> -genitive	262
Figure 7.6: Individual mean ratings for animate and collective possessors for A2/B1 learners and native speakers	266
Figure 7.7: Individual mean ratings for collective and inanimate for B2 learners and native speakers	267
Figure 7.8: Individual native speaker participants' ratings for the <i>s</i> -genitive plotted against the corpus-based probability of the <i>s</i> -genitive	269
Figure 7.9: Individual A2/B1-level participants' ratings for the <i>s</i> -genitive plotted against the corpus-based probability of the <i>s</i> -genitive	270