

DIATHESIS AND VALENCY CHANGES IN PRONOMINAL-VERBAL CONSTRUCTIONS IN FRENCH

Résumé: Le présent article poursuit deux objectifs: passer en revue les classifications existantes de la classe grammaticale des verbes pronominaux français, ainsi que proposer une nouvelle typologie de ces verbes basée sur le concept de diathèse et validée par une analyse de corpus. L'étude de 33 structures prototypiques des verbes pronominaux associée à l'annotation de 1232 structures dans plusieurs corpora nous a permis de distinguer 8 catégories de diathèses applicables aux verbes pronominaux. Les principaux avantages de cette nouvelle typologie résident dans l'importance accordée à la polysémie des verbes, ainsi qu'à l'utilité du concept de diathèse dans l'analyse simultanée de la syntaxe et de la sémantique sur un échantillon important. Cet article fournira également des clés pour comprendre la réduction de valence s'opérant dans la pronominalisation, avec pour effet d'offrir des outils supplémentaires pour appréhender ce phénomène complexe en français.

Abstract: The aims of this study are twofold: to give an overview of existing classifications for the grammatical class of pronominal verbs in French and to present a new classification of these verbs based on the concept of diathesis and validated through a corpus analysis. Based on the thorough analysis of 33 prototypical pronominal sentences along with the corpus annotation of 1232 structures, we were able to distinguish 8 patterns of diatheses relevant to pronominal verbs. The main tenets of this typology lie in an increased focus on the polysemy of the pronominal verbs, as well as the systematic analysis of syntactic and semantic layers of a large sample. Our findings also confirm that French pronominal verbs induce a valency reduction compared to their neutral diathesis counterpart, which sheds a new light on our understanding of an otherwise complex phenomenon.

Keywords: French, Reflexive Verbs, Pronominal Verbs, Corpus Analysis, Diathesis, Valency

1. Introduction

Providing a clear and exhaustive description of grammatical concepts is generally a serious challenge for linguists, although such descriptions have proven very useful in teaching situations, whether for native learners or for second and foreign language (L2) learners.¹ This challenge is of particular interest for the class of pronominal verbs in French, since this grammatical phenomenon is widely acknowledged as being especially complex to describe.²³ Pedagogical grammars found in L2 textbooks – one of the most accessible resources containing classifications concerning French pronominals – have indeed been considered fragmentary and misleading regarding the way these verbs should be categorized.⁴⁵ It is therefore not very surprising that pronominal verbs remain a major source of errors for L2 learners,⁶⁷ even though they are commonly used in all facets of our daily lives and conversations.

One widespread criticism of the grammatical descriptions of pronominal verbs is their lack of systematicity in the linguistic criteria used.⁸ While some categorizations of pronominal verbs emphasize morphosyntactic differences, such as the presence of a direct object, others are focused on semantic roles found in pronominal sentences.⁹ The difference in criteria implies that classifications of pronominal verbs can be difficult to compare. Moreover, several classifications tend to only categorize pronominal verbs at lemma level, disregarding their different meanings or

1 B. Combettes, 'Quelle(s) description(s) grammaticale(s) pour l'enseignement?', *Repères*, 39 (2009): pp. 41-56.

2 D. Gaatone, 'Réflexions sur les verbes pronominaux réfléchis et réciproques', *Folia Linguistica*, 8 (1975): pp. 199-222.

3 J. Bień, 'Sur la nature des confusions notionnelles au sein du verbe pronominal: Diathèse', *Lublin Studies in Modern Languages and Literature*, 31 (2009): pp. 128-135.

4 K. Connors & B. Ouellette, 'Describing the meanings of French pronominal-verbal constructions for students of French ←→ English translation.', *Language sciences*, 18 (1996): pp. 213-226.

5 F. Delbarre, 'Nécessité de contextualisation des manuels publiés au Japon - Une étude de cas: l'expression de l'état', *Revue japonaise de didactique du français*, 5 (2010): pp. 146-166.

6 V. Kraus, 'L'apprentissage du français langue seconde chez les immigrants chinois, coréens et japonais à Montréal', *UWSpace*, 1997, <<http://hdl.handle.net/10012/14293>>.

7 K. Nakamura, *Problèmes typiques des apprenants japonophones de français* (Tokyo: Suguradai-Shuppansha, 2004).

8 Connors & Ouellette, 'Describing the meanings of French pronominal', pp. 213-226.

9 Bień, 'Sur la nature des confusions notionnelles', pp. 128-135.

uses, even though these verbs rarely have monolithic uses.¹⁰ These findings call for more thorough or more consistent classifications of French pronominal verbs, as this should facilitate their acquisition.

The present paper first aims to provide an overview of the linguistic classifications that have been designed to date, and further discuss their advantages and limits. We will then investigate a new type of classification of pronominal verbs based on the concept of diathesis. This concept, defined as the mapping of the constitutive syntactic functions onto the semantic roles in a verbal process, can be used methodologically as an anchor to understand and classify linguistic structures, including pronominal verbs, in a more efficient way. Systematic use of diathesis as a tool for analysis has the additional advantage of clarifying the relationship between the pronominal verb and its non-pronominal counterpart. Indeed, several authors have postulated that the underlying coreferentiality within pronominal-verbal constructions reduces the valency¹¹ of the verb,¹²¹³¹⁴¹⁵ which means that the presence of *se* would necessarily have an impact on the diathetic mapping of the process when compared with its non-pronominal variant. As this hypothesis has not yet been empirically proven, this work will be the opportunity to verify it. Finally, approaching pronominal verbs as verbal structures based on a combination of semantic, syntactic, and valency analyses will also highlight their different uses in context, allowing us to systematically cover the polysemy of pronominal verbs.

2. Overview of pronominal verbs: definitions and previous classifications

¹⁰ Gaatone, 'Réflexions sur les verbes pronominaux', pp. 199-222.

¹¹ Valency is the term used to describe the number of arguments a verb can potentially take; for instance, a divalent verb predicates a subject and a direct object. Valency can also be applied to semantic roles (see J. Allerton, *Valency and the English Verb* (London: Academic Press, 1982).)

¹² S. C. Dik, 'Valentie en valentie operaties in Funktionele Grammatika. Valentie in Functionele Grammatica', *Interdisciplinair Tijdschrift voor Taal-en Tekstwetenschap*, 5 (1985): pp. 95-114.

¹³ Bień, 'Sur la nature des confusions notionnelles', pp. 128-135.

¹⁴ D. Creissels, 'Réflexivisation, transitivité et agent affecté', in *L'énoncé réfléchi*, ed. by A. Rousseau & others (Rennes: Presses Universitaires de Rennes, 2007): pp. 83-106.

¹⁵ G. Corminboeuf & F. Heyna, 'Nominalisation et diathèse', *Le Français moderne*, 83 (2015): pp. 34-53.

Before moving on to the review of existing classifications of pronominal verbs in Section 2.2, we will first discuss some defining concepts about pronominal verbs in the next section.

2.1. What is a pronominal verb?

According to Bouard,¹⁶ the category of French "pronominal verbs" did not exist as such in grammar manuals before the early XVIIIth century. This category used to be seen as a transformation of the active voice thanks to the clitic *se* and the auxiliary *être*, or as a particular case of neutral case. However, XVIIIth century grammarians started to consider pronominal verbs as a distinct structure and attempted to define their properties, as their use was increasingly shifting away from their original reflexive meaning. This divergence might originate from the fact that in Old and Middle French, the so-called middle¹⁷ meaning (intransitive use of a verb describing an internal process) and the reflexive meaning (formed with *se* + *verb*) that were originally separated merged into a single form, as illustrated in the example in Table 1.¹⁸ This created greater ambiguity in modern French, as the same form might correspond to either the middle or the reflexive meaning.

	Middle meaning	Reflexive meaning
Old French	Il lieve [He wakes up]	Il se lieve [He stands up]
Modern French	Il se lève [He wakes up/He stands up]	

Table 1: Fusion of middle and reflexive meanings

In modern French, a pronominal verb is made up, from the syntactic point of view, of a personal pronoun (*se*) in agreement with the grammatical subject, and a verb in the active voice.^{19,20}

In French, this formal structure is a prerequisite for the verb to be considered a pronominal verb.

¹⁶ B. Bouard, 'Verbe pronominal et voix pronominale dans les grammaires françaises des 18e et 19e siècles', *Colloque annuel de la Société d'Histoire et d'Epistémologie des Sciences du Langage* (Paris, 2011): pp. 733-757.

¹⁷ Middle meaning derivate from the Latin *media-tantum* forms, which are active intransitive verbs with a passive form.

¹⁸ L. Schøsler, 'Réflexions sur l'optionnalité des compléments d'objet direct en latin, en ancien français, en moyen français et en français moderne', *La langue, les signes et les Êtres: actes du colloque de l'Institut d'Etudes Romanes de l'Université de Copenhague* (Copenhague, 1999): pp. 9-28.

¹⁹ Connors & Ouellette, 'Describing the meanings of French pronominal', pp. 213-226.

²⁰ M. D. Joffre, 'La synonymie partielle entre deux signifiants: un moteur d'évolution et de création. Le cas de la genèse, en latin, du pronominal français.', *Orbis Linguarum*, 50 (2018): pp. 103-114.

From the semantic point of view, coreferentiality between the agent (A) and the patient (P) is sometimes seen as inherent to pronominal verbs.²¹ This implies that A and P should have no intrinsic semantic difference in order to acknowledge coreferentiality between the grammatical subject and the pronoun *se*, so that they could be exchanged without affecting the meaning of the sentence.²² However, we can see that this is not systematic with verbs having more than one reading, such as *Je me bats*, which is commonly understood as *I fight* rather than its direct reflexive meaning *I beat myself*.

Other than the presence of *se*, identifying common properties within pronominal verbs is not an easy task. Corminboeuf suggests that the clitic *se* has the effect of reducing the valency of the verb,²³ which means that it often prohibits the presence of either the agent (in the case of passive pronominal-verbal constructions) or the patient (in the case of direct reflexive sentences), as can be seen in the following examples:

	Subject	Verb	Object
Neutral active	Il [Agent]	endort	l'enfant. [Patient]
Agentive passive	L'enfant [Patient]	est endormi	par John. [Agent]
Passive pronominal	Un enfant [Patient]	s'endort (facilement).	
Direct reflexive	Il [Agent]	s'endort.	

Table 2: Valency reduction in pronominal sentences²⁴

For instance, we can see that in Table 2, the valency of *s'endormir* (to fall asleep) is reduced in comparison with *endormir* (to put to sleep). This valency reduction operated through the presence of *se* could, according to Corminboeuf, be a better indicator of a pronominal construction. The role of *se* as a valency operator influencing the process in terms of agentivity is also hypothesized by Bień and Creissels,^{25,26} although it has not been empirically proven in a large-scale study.

²¹ Gaatone, 'Réflexions sur les verbes pronominaux', pp. 199-222.

²² Creissels, 'Réflexivisation, transitivité et agent affecté', pp. 83-106.

²³ Corminboeuf & Heyna, 'Nominalisation et diathèse', pp. 34-53.

²⁴ Ibid.

²⁵ Bień, 'Sur la nature des confusions notionnelles', pp. 128-135.

²⁶ Creissels, 'Réflexivisation, transitivité et agent affecté', pp. 83-106.

The observations made in this section allow us to define pronominal verbs as verbal structures taking the form *NP "se" V*, composed of a reflexive pronoun ("*se*") agreeing with the grammatical subject (*NP*) and followed by a verb (*V*).²⁷ In the present work, we will attempt to prove, through a systematic methodology, that the presence of the reflexive pronoun does cause a reduction in valency of the process by one unit, for instance by dismissing the patient, the agent or the dative of a verbal process. The hypothesis that the clitic *se* modifies the verb's valency and, thus, the whole sentence from a diathetic perspective also implies that pronominal verbal constructions could be analysed according to their diathetic value. This analysis of the diatheses applicable through the use of pronominal verbs, also premised by Rivière and then by Dixon and Aikhenvald as promising grounds for analysis,^{28,29} will be covered in Section 4 of this study.

2.2. Review of the literature on classifications of pronominal verbs

As mentioned above, in modern French, pronominal verbs do not always express coreferentiality between the agent and the patient at the semantic level. This semantic confusion has so far prevented researchers from agreeing on a shared classification of pronominal verbs. This section reviews several classifications that have been introduced and aims to highlight their strengths and shortcomings.

One of the first classifications of pronominal verbs was proposed by Dangeau (1714-1717) and, more recently, reported and refined by Zribi-Hertz³⁰. The pronominal verbs are here classified according to syntactic criteria into four categories: reflexive (*Pierre se lave*), reciprocal (*Pierre et Marie s'aiment*), passive (*Le vin se boit frais*), and subjective (*Je m'évanouis*). In the reflexive derivation, the clitic represents a direct accusative object to the grammatical subject (*Pierre lave*

²⁷ J. L. Burston, 'The pronominal verb construction in French: An argument against the fortuitous homonymy hypothesis', *Lingua*, 48 (1979): pp. 147-176.

²⁸ N. Rivière, 'Le pronominal, les rôles actanciels et la diathèse', *Cahiers Charles V*, 23 (1997): pp. 7-46.

²⁹ R. M. Dixon & A. Y. Aikhenvald, *Changing valency: Case studies in transitivity* (Cambridge: Cambridge University Press, 2000).

³⁰ A. Zribi-Hertz, 'La réflexivité ergative en français moderne', *Le Français moderne*, 55 (1987): pp. 23-54.

lui-même); in the reciprocal derivation, the presence of the clitic implies that multiple subjects become, in turn, objects performing the process on each other (*Pierre aime Marie et Marie aime Pierre*); in the passive derivation, the clitic allows the subject to act as the object of a general statement in an active derivation (*On boit le vin frais*). This classification remains popular, as it is still used in some modern French textbooks and used by many researchers as a starting point for further analysis, thanks to its simplicity.³¹ Its main shortcoming lies in the heterogeneity of the last category (subjective), which encompasses all pronominal verbs that do not belong to the first three categories.³² Zribi-Hertz therefore attempted to further subdivide this category into the following classes: intransitive with unused non-pronominal counterpart (*s'évanouir*, *se repentir*), intransitive with non-reflexive counterpart (*se douter*, *se jouer*), intransitive with reflexive counterpart (*se mourir*), and intransitive with a similar meaning as the non-pronominal variant (*se casser*, *se gonfler*). While this subdivision provides a clearer view of the subjective category, the classification still has the disadvantage of considering many verbs as belonging to a single category, thus denying their polysemy, as well as not taking lexicalized pronominal-verbal structures into consideration (such as *se prendre la tête*).

By comparing syntactic structures from pronominals and their transitive variants, Boons, Guillet, and Leclère identified six types of relationships: reflexive (*Pierre se chatouille*), reciprocal (*Pierre et Marie s'aiment*), body parts constructions (*Pierre se lave les pieds*), possessive reflexive (*Pierre se dépense à mener à bien ce travail*), intransitive (*Le pont-levis s'est abaissé*), and "phantom agent" constructions (*Le vin se boit frais*).³³ In each case, the structure of the transitive variants of a verb is analysed before they are turned into their pronominal form, for instance with the transitive syntactic structures **NP V N1** (*Je chatouille Marie*), **NP V à N1** (*Je parle à Marie*) all turning into **NP se V** (*Je me chatouille/Je me parle*) to form the "reflexive" category. This

³¹ Bień, 'Sur la nature des confusions notionnelles', pp. 128-135.

³² Ibid.

³³ J.-P. Boons, A. Guillet & C. Leclère, 'La structure des phrases simples en français, constructions intransitives', *Journal of Linguistics*, 15 (1976): pp. 187-192.

classification has the advantage of using a methodology linking the pronominal structures with their transitive equivalent, making the meaning of the structures clearer.³⁴ However, the focus on syntactic arguments at the expense of the semantic dimension is not without flaws, as it would, for instance, make it impossible to determine the structure of verbs that have no record of being used in a non-pronominal equivalent.³⁵

Geniušienė proposed a division between subjective and objective pronominal structures.³⁶ Her classification is based on the recession of syntactic arguments: subjective structures are marked by object recession, while objective structures are marked by subject recession. A good example is the distinction between the autocausative sentence *Je me lève* [I wake up], in which the object is depleted (subjective structure), and the anticausative sentence *La porte s'ouvre* [The door opens], where the subject is depleted (as the action of opening cannot be performed without an external subject). Although this classification does not seem to consider structures where the recession is unclear (such as *Je me lave les dents*), the overall methodological soundness and the relevance of this distinction was praised by Melis,³⁷ who proposed an adapted version of Geniušienė's model.

Desclés and Guentchéva suggested a classification of pronominal verbs based on the interpretation of the roles of the agent and patient, and divided the structures into five categories (reflexive, reciprocal, middle, mediopassive, and passive).³⁸ The breakdown follows the following pattern:

<i>Structure</i>	<i>Example</i>	<i>Agent</i>	<i>Patient</i>
<i>Reflexive</i>	<i>Pierre se rase</i>	Pierre	Pierre
<i>Reciprocal</i>	<i>Les enfants se battent</i>	Les enfants	Les enfants
<i>Middle</i>	<i>Le vautour s'abattit sur le lapin</i>	Le vautour	Le lapin

³⁴ Bień, 'Sur la nature des confusions notionnelles', pp. 128-135.

³⁵ J. François, 'Quelles sont les origines des verbes essentiellement pronominaux du français?' in *Rahmen des Sprechens: Beiträge zur Valenztheorie, Varietätenlinguistik, Kreolistik, Kognitiver und Historischer Semantik*, ed. by S. Dessi Schmidt & others (Peter Koch, 60, 2011): pp. 33-42.

³⁶ E. Geniušienė, *The typology of reflexives* (Berlin: Mouton de Gruyter, 1987).

³⁷ L. Melis, 'Pronominal Verbs in Old and Modern French, or How prototypes can be restructured on the basis of permanent meaning effects', *Belgian Journal of Linguistics*, 5 (1990): pp. 87-108.

³⁸ J.-P. Declés & Z. Guentchéva, 'Le passif dans le système des voix du français', *Langages*, 109 (1993): pp. 73-102.

<i>Mediopassive</i>	<i>Le mât s'abattit avec un grand fracas</i>	Ø	Le mât
<i>Passive</i>	<i>Les livres se vendent bien</i>	"Les gens"	Les livres

Table 3: Desclés and Guentchéva's classification³⁹

The main strength of this classification is that it sheds light on the identification of specific semantic roles as a major distinctive marker of pronominal structures.⁴⁰

Noticing the huge differences between the nature of criteria chosen to classify pronominals in the previous typologies, Bien mixed together the most prominent distinctive features among them (coreferentiality, syntactic recession, syntactic transitivity, agentivity, and verb lexicality) and suggested a semantic classification of pronominal verbs based on two criteria: agentivity and control over the process, as can be seen in Table 4.⁴¹

<i>Structure</i>	<i>Example</i>	<i>Agentivity</i>	<i>Control</i>
<i>Active</i>	<i>Jean se lave</i>	+	+
<i>Medio-active</i>	<i>Jean s'énervé</i>	+	0
<i>Mediopassive</i>	<i>Le mât s'écroule</i>	0	0
<i>Passive</i>	<i>Le problème se discute</i>	0	-

Table 4: Bien's classification

Bien describes the agentivity feature as the grammatical subject acting on the process as an agent, while the control feature would be the "*capacity to carry out an action and capacity to interrupt it*". This classification enables the well-known concept of agentivity to be divided between two sub-concepts (agentivity and control) that are semantically related, but not equivalent. It also has the merit of allowing lexicalized *se-verbs* (such as *s'évanouir*) to be analysed as medio-active structures. However, other types of structures are indiscriminately mixed. For instance, the structure *Ils se lavent* is only categorized as active, whereas it can have at least two different semantic interpretations: *they wash themselves/they wash each other*.

³⁹ Desclés & Guentchéva, 'Le passif dans le système des voix', pp. 73-102.

⁴⁰ Bien, 'Sur la nature des confusions notionnelles', pp. 128-135.

⁴¹ Ibid.

Following the tenets of construction grammar,⁴² Petersilka classified pronominal structures based on both syntactic and semantic criteria, as well as on whether the root of the pronominal verb is a transitive verb.⁴³ The analysis resulted in a 10 classes distribution: attribute-value (*Tu t'es trahi*), reflexiva tantum (*Elle s'est évanouie*), anticausative (*Le soleil s'est couché*), mediopassive (*Cette tour se voit de loin*), reciprocal (*Pierre et Jean se trompent*), reciprocal reflexiva tantum (*Les gamins se chamaillent sur la place*), reciprocal anticausative (*Nos routes se séparent ici*), completive (*On s'attendait au pire*), attributive (*Ils se pensent souvent drôles*), and attributive with thing-object (*Cette école se veut complémentaire d'autres initiatives*). Methodologically speaking, how this typology takes syntax, semantics, and even etymology into account is commendable; however, the way the sample structures were chosen is not easy to figure out, leading to some degree of confusion. Consequently, it is possible that some existing structures do not fall into any of these categories. Additionally, some categories mainly differentiated by semantics, such as completives (e.g. *On s'attendait au pire*), do not attempt to describe *how* semantics differ with pronominalization (e.g. there is no information on how *On attendait* differs from *On s'attendait*), merely claiming that they take a different meaning. Hence, the classification could benefit from a more thorough examination of the classes it constructed.

By using systematic corpus analysis based on formal criteria, Barque, Candito, and Huyghe built a typology of pronominal verbs.⁴⁴ Through the annotation of pronominal sentences in two corpora, *Sequoia* (318 tokens) and *French Treebank (FTB)* (900 tokens), the researchers were able to identify 15 categories of pronominal structures: intrinsic, autonomous (subdivided into five categories: syntax, selection, aspect, lexicalized, and construction), autocausative, reflexive, reciprocal (including two types: V and seV), mediopassive, anticausative, antipassive, parallel, and

⁴² A. E. Goldberg, *Constructions: A construction grammar approach to argument structure* (Chicago: University of Chicago Press: 1995).

⁴³ C. Petersilka, 'Une approche syntaxique et lexicale des verbes pronominaux', *Studia Romanica Posnaniensia*, 39 (2012): pp. 33-48

⁴⁴ L. Barque, M. Candito & R. Huyghe, 'La classification des verbes réfléchis à l'épreuve d'une annotation en corpus', *Langages*, 216 (2019): pp. 121-137.

autobenefactive. The annotation process showed good results in terms of annotator agreement (Kappa values range from 0.53 – with 13 categories involved – to 0.74 with a broader granularity), although it emphasized how difficult it is to distinguish reflexive interpretations based solely on formal criteria. All in all, the typology appears to be complexified by the large number of categories. However, the use of corpus annotation to certify the typology is particularly innovative.

As can be seen, there have been attempts to classify pronominal verbs according to different principles (syntactic, semantic, or even etymological). However, it sometimes remains difficult to determine whether these authors classify similar verbs in the same way (e.g. whether Zribi-Hertz's passive category is equivalent to Boons, Guillet, and Leclère's “phantom agent” constructions). To have a better idea of how these classifications compare with each other, we sampled 20 verbs from all categories proposed by Zribi-Hertz and attempted to classify the most frequent uses of these verbs under each classification (see Annex 1 available online on request).⁴⁵⁴⁶

The data collected show that while some structures can easily be agreed upon, other verbs such as *s'évanouir* are difficult to classify (it has seven different designations in the seven classifications). Additionally, it appears that earlier typologies rarely take the polysemy of pronominal verbs into account, and the overall tendency to analyse semantic and syntactic parameters separately can lead to many exceptions. Moreover, the sample coverage is limited and rarely supported by corpus analyses (only found with Barque & others), leading to potential methodological weaknesses. To overcome these limitations, it may be useful to suggest a new classification of French pronominal verbs based on the concept of diathesis. By means of a two-phase methodology using both systematic analysis and corpus annotation, we aim to build a typology with enough clarity for potential educational purposes, while remaining precise and thorough for linguists. Before presenting this new typology and the results of the annotation, the next sections will briefly summarize the notion of diathesis and how it applies to pronominal verbs.

⁴⁵ Zribi-Hertz, 'La réflexivité ergative', pp. 23-54.

⁴⁶ Bieñ, 'Sur la nature des confusions notionnelles', pp. 128-135.

3. Application of diathesis to pronominal verbs

3.1. About the concept of diathesis

The concept of diathesis in morphosyntax can be defined as “a pattern of mapping of semantic arguments onto syntactic functions (grammatical relations) [...]”.⁴⁷ According to Kulikov, the most common mapping is "X: S; Y: DO", where the first semantic role (X) is mapped onto the syntactic function of subject, and the second semantic role (Y) is mapped as direct object. This diathesis, the most common one, is often regarded as the active or neutral diathesis.^{48,49} Changing the mapping of the arguments alters the diathesis and can impact the valency of the verb. For instance, promoting the direct object of a neutral diathesis to subject and demoting the original subject to an oblique object (or removing it) will result in a change of diathesis, as seen in the following example:

I see an apple. → *An apple is seen (by me).*

Diatheses are often presented in tabular form for the sake of clarity and convenience. The following table (Table 5) corresponds to the above example of change of diathesis for English:

<i>Example</i>	<i>I</i>	<i>see</i>	<i>an apple</i>
<i>Semantic role</i>	A	V	P
<i>Syntactic function</i>	S	V	DO
↓			
<i>Example</i>	<i>An apple</i>	<i>is seen</i>	<i>(by me)</i>
<i>Semantic role</i>	P	V	A
<i>Syntactic function</i>	S	V	Ø (OBL)

Table 5: Change from neutral to passive (agentless) diathesis

Since changing a syntactic function or a semantic role is a sufficient condition to trigger a switch of diathesis, there are many diatheses in every language. A quick look at the following list of

⁴⁷ L. Kulikov, 'Voice typology' in *The Oxford Handbook of Linguistic Typology*, ed. by J. J. Song (Oxford: Oxford University Press, 2011): pp. 368–398.

⁴⁸ M. Shibatani, 'On the conceptual framework for voice phenomena', *Linguistics*, 44 (2006): pp. 217-269.

⁴⁹ A. Rousseau, 'Propositions pour une description ordonnée des “voix” et des “diathèses”: problématique, statut et conceptualisation du “moyen”', *Langages*, 2 (2012): pp. 21-34.

the main syntactic functions and semantic roles (hierarchically ordered according to Grimshaw)⁵⁰ reveals a high number of potential arguments that could be combined into a diathesis.

Syntactic function	Explanation	Example
Subject (S)	Object predicated by the verb in the clause	<i><u>Jean</u> mange la pomme.</i>
Direct Object (DO)	Object linked directly to a transitive verb	<i>Jean mange <u>la pomme</u>.</i>
Indirect Object (IO)	Object linked by a preposition to a transitive verb	<i><u>Jean</u> offre un cadeau <u>à Marie</u>.</i>
Oblique Object (OBL)	Object linked by a preposition to an intransitive verb	<i>Jean joue <u>au poker</u>.</i>
Dative (DAT)	Object towards which the verb is directed in the clause	<i>Marc écrit <u>pour sa mère</u>.</i>
Genitive (GEN)	Object attributive to another object	<i>Les oreilles <u>du chat</u> sont petites.</i>
Adjunct (ADJ)	Non-mandatory object	<i>Jean mange la pomme <u>sur un banc</u>.</i>
Dummy (∇)	Dummy function (cf. impersonal <i>Il, ça</i>)	<i><u>Il</u> pleut.</i>

⁵⁰ J. Grimshaw, *Argument Structure* (Cambridge: The MIT Press, 1990).

Semantic role	Explanation	Example
Agent (A)	Entity performing the process	<i><u>Jean</u> mange la pomme.</i>
Patient (P)/Undergoer (U)	Entity undergoing the process	<i>Jean mange <u>la pomme</u>.</i>
Beneficiary (B)	Entity benefiting from the process (positively or negatively)	<i>Jean offre un cadeau à <u>Marie</u>.</i>
Experiencer (E)	Entity experiencing the process (state of being, emotion, perception)	<i><u>Jean</u> a peur.</i>
Instigator (I)/Causer (CAUS)	Animate entity causing the process to happen	<i><u>Marc</u> demande à Jean de venir.</i>
Cause (CAUS)/Stimulus (STI)	Inanimate entity causing the process to happen	<i><u>La pluie</u> a fait annuler la représentation.</i>
Circumstant (C) [Locative, Modal, Temporal...]	Entity circumstant to the process	<i>Jean mange la pomme <u>sur un banc</u>.</i>
Dummy (Ø)	Does not refer to any entity	<i><u>Il</u> pleut.</i>

Table 6a & 6b: Main syntactic functions and semantic roles in French

Despite the large number of possible combinations, some diatheses are more prominent in a language. Table 7 shows some major diatheses used in French in Kulikov's typology.

Diathesis	Example	Syntactic mapping	Semantic mapping
<i>Neutral</i>	Je mange une pomme.	S V DO	A V P
<i>Agentless passive</i>	La pomme est mangée.	S V	P V (A)
<i>Impersonal</i>	Il pleut.	∇ V	Ø V
<i>Canonical reflexive</i>	John se rase.	S V	A = P V
<i>Canonical reciprocal</i>	Pierre et Marie s'embrassent.	S V	A & P V
<i>Anticausative</i>	La branche casse.	S V	P V
<i>Indirect causative</i>	Je lui ai fait manger la	S IO V DO	I A V P

	pomme.		
--	--------	--	--

Table 7: Major diatheses in French

The central role of the neutral diathesis (*Je mange une pomme*) is justified by the fact that many diatheses can be considered extensions of it, such as the indirect causative (*Je lui ai fait manger une pomme* – [I made him eat an apple]).

3.2. Diathesis with pronominal verbs

French pronominal verbs are frequently cited as an example of diathetic change from the neutral use of the verb.⁵¹⁵²⁵³ As we saw in Section 2.1, pronominal verbs have an impact on the valency of the verb, as they affect the semantic and syntactic arguments of the sentence, and they could be represented by one or more diatheses. In this regard, Kulikov's typology of diatheses based on the valency of the arguments provides useful insights into which ones could apply to pronominal verbs. Derivative diatheses (such as various passives) preserve the semantic arguments of agent and patient while demoting them syntactically. On the other hand, operational diatheses (such as reflexive and reciprocal) keep the original meaning of the arguments taken from a neutral derivation but add some operation to them (such as coreferentiality). Finally, diatheses that add or remove semantic arguments include causative, anticausative, applicative, benefactive, and anticausative clauses. While the name given to these diatheses may vary, the patterns they describe remain the same in studies on morphosyntax.⁵⁴⁵⁵⁵⁶⁵⁷

-
- ⁵¹ E. Deronne, 'Des voix aux diathèses', *Article issu d'une intervention au séminaire de l'ATILF* (Nancy, 2006): pp. 2-43.
- ⁵² Bieñ, 'Sur la nature des confusions notionnelles', pp. 128-135.
- ⁵³ W. Weidenbusch, 'La diathèse verbale (voix active, passive, pronominale etc.) comme problème de traduction', *Manuel de traductologie*, 5 (2016): pp. 231-255.
- ⁵⁴ Dixon & Aikhenvald, *Changing valency*.
- ⁵⁵ N. Le Querler, 'Valence et complémentation: L'exemple des verbes trivalents en français contemporain', *Annales de Normandie*, 62 (2012): pp. 175-188.
- ⁵⁶ E. Poole, 'Deconstructing quirky subjects', *Proceedings of the 45th Meeting of the North East Linguistic Society (NELS 45)* (Amherst, 2014): pp. 1-14.
- ⁵⁷ Weidenbusch, 'La diathèse verbale', pp. 231-255.

The interest of considering valency changes when comparing neutral diathesis with other structures, as shown in Kulikov's typology, is that pronominal verbs, for instance, are known to have an impact on valency.⁵⁸ Thus, to determine which diatheses are covered by pronominal verbs, observing how the valency of the verbal process changes could be key. Indeed, although some researchers have hinted at the potential usefulness of the concepts of diathesis and valency regarding pronominal verbs,^{59,60,61} no systematic effort has yet been made to apply these concepts to a classification of pronominal-verbal constructions based on semantic, syntactic, and empirical criteria.

In this paper, we aim to determine whether a diathetic classification of pronominal structures – rather than forms – integrating semantic, syntactic, and empirical analyses all together would be possible, and how valency changes occur within pronominal-verbal constructions. To achieve this aim, Kulikov's typology of diatheses will be used as the main basis of our analysis.

The next section will describe the methodology used to build our classification of pronominal verbs based on diathesis.

4. Two-step analysis

To carry out our classification of pronominal verbs, we applied a two-phase procedure. The objective of the first phase was to determine which diathetic patterns would correspond to a given number of prototypical pronominal structures, while the second phase would attempt to verify that these diatheses could effectively cover all forms found in a large corpus.

4.1. First phase: Identification of diatheses applicable to pronominal verbs

⁵⁸ Corminboeuf & Heyna, 'Nominalisation et diathèse', pp. 34-53.

⁵⁹ Bień, 'Sur la nature des confusions notionnelles', pp. 128-135.

⁶⁰ Creissels, 'Réflexivisation, transitivité et agent affecté', pp. 83-106.

⁶¹ Corminboeuf & Heyna, 'Nominalisation et diathèse', pp. 34-53.

The first phase followed a three-step process: 1) selecting a representative set of verbs that will be subjected to analysis, 2) identifying the major uses for each chosen verb, and 3) determining adequate classification criteria.

Our first step was to determine a set of pronominal verbs representative of the different categories found in the literature (see Section 2.2). We took Zribi-Hertz's classification as a starting point for our analysis,⁶² as it is a modern approach inspired by the authoritative figure of Dangeau's typology in the field of pronominal verbs.⁶³ Indeed, Dangeau's work has been cited as the foundation of the “quatripartition traditionnelle des verbes pronominaux”, which still has prominent impact in grammar textbooks.⁶⁴ Compared to other typologies we've reviewed in this paper, Zribi-Hertz chooses, in her own words, to keep the classical typology untouched apart from delving deeper into understanding of the subjective category. From there, we sampled up to three verbs per category:

Reflexive: *se laver, se regarder, se raser*

Reciprocal: *s'aimer, se voir, se parler*

Passive: *se boire, se vendre, s'utiliser*

Subjective, divided into:

- intransitive with unused non-pronominal counterpart: *s'évanouir, se repentir*
- intransitive with non-reflexive counterpart: *se douter, se jouer*⁶⁵
- intransitive with reflexive counterpart: *se mourir*⁶⁶
- intransitive with similar meaning as the non-pronominal variant: *se casser, se gonfler*.

We also supplemented these 16 verbs with additional pronominal structures in French, which, albeit rare, could not be ignored due to their unique structure, namely the impersonal

⁶² Zribi-Hertz, 'La réflexivité ergative', pp. 23-54.

⁶³ Bieñ, 'Sur la nature des confusions notionnelles', pp. 128-135.

⁶⁴ Melis, 'Pronominal Verbs in Old and Modern French', pp. 87-108.

⁶⁵ *Douter* and *jouer* are both intransitive.

⁶⁶ *Mourir* has a reflexive meaning.

pronominal verbs. This category of pronominal verbs was either absent from the classifications^{67,68,69} or mentioned by several authors to be referred to as part of the passive^{70,71,72} or the subjective.^{73,74} We thus included four additional verbs chosen on the basis of an analysis of impersonal pronominal verbs carried out by Rivière (*s'agir, se passer, se construire, se pouvoir*).⁷⁵ This enabled us to reach a total of 20 sample verbs.

As a second step, we searched for typical uses of each of these 20 verbs in their dictionary entry in the *Trésor de la Langue française informatisé* (TLFi). The TLFi is one of the most exhaustive online dictionaries for French, with over 100,000 entries, and is useful for its historical references. Based on this source, we were able to identify 33 different structures for our 20 verbs, in authentic and historical contexts, which are important given that diathetic mapping should be analysed in context.

Our third step was to determine the classification criteria on the basis of which the various diathetic mappings found in our 33 structures would be described. The four criteria we used for the analysis are based on Kulikov's typology: the syntactic functions of the verb arguments, their semantic role, their cases, and valency changes compared with the neutral diathesis.⁷⁶ Each of these criteria is described in more detail below.

First criterion: syntactic function

⁶⁷ Boons, Guillet & Leclère, 'La structure des phrases simples en français', pp. 187-192.

⁶⁸ Geniušienė, *The typology of reflexives*.

⁶⁹ Bień, 'Sur la nature des confusions notionnelles', pp. 128-135.

⁷⁰ Desclés & Guentchéva, 'Le passif dans le système des voix', pp. 73-102.

⁷¹ J. Albrecht, 'Reflexivkonstruktionen in romanischen Sprachen und germanischen Sprachen' in *Comparatio delectat II, Akten der VII. Internationalen Arbeitstagung zum romanisch-deutschen und innerromanischen Sprachvergleich*, ed. by E. Lavric & W. Pöckl (Innsbruck, 6–8.2, 2012): pp. 445–465.

⁷² Weidenbusch, 'La diathèse verbale', pp. 231-255.

⁷³ A. Zribi-Hertz, 'La construction du français et son statut dans le triangle: moyen-passif-réfléchi', *Linguisticae Investigationes*, 6 (1982): pp. 345-401.

⁷⁴ Zribi-Hertz, 'La réflexivité ergative', pp. 23-54.

⁷⁵ N. Rivière, 'La pronominalisation et les auxiliaires de modalité' in *Morphosyntaxe et sémantique du verbe relations actanciennes, voix, aspect et statut grammatical en français et en arabe*, ed. by J. François J. & others (Caen: Université Caen, 2018): pp. 60-80.

⁷⁶ Kulikov, 'Voice typology', pp. 368–398.

The syntactic functions of the arguments of the pronominal verb are one of the two essential characteristics in diathetic grammar. The nature of a syntactic function is determined by the function an argument takes in the verbal process. Explanations of these functions can be found in Table 6(a).

Categorical (nominal): subject (S), direct object (DO), indirect object (IO), oblique object (OBL), dative (DAT), genitive (GEN), adjunct (ADJ), verb (V), not applicable (Ø).

Second criterion: semantic role

The semantic role of the arguments is the other essential marking in diathetic mapping. The nature of a semantic role is determined by the meaning an argument takes in the verbal process. Details concerning semantic roles can be found in Table 6(b).

Categorical (nominal): agent (A), patient (P), experiencer (E), beneficiary (B), instigator (I), cause (CAUS), circumstant (C), verb (V), not applicable (Ø).

Third criterion: case-marking

Case-marking is a grammatical feature indicative of the syntactic function of the arguments. This criterion is often coded in conjunction with the syntactic function for the sake of clarity, as it may give a more flexible understanding of syntactic functions⁷⁷.

Categorical (nominal): nominative (NOM), accusative (ACC), dative (DAT), genitive (GEN), ablative (ABL), verb (V), not applicable (Ø).

Fourth criterion: valency change in comparison with the neutral diathesis

The valency change compared with the neutral diathesis is essential to determine internal properties of pronominal verbs. The transformation into neutral diathesis will be provided by turning the arguments into an "X: S; Y: DO" mapping, while keeping the meaning of the arguments (semantic roles) as close to the pronominal variant's arguments as possible. In cases where the pronominal sentence would not have a grammatically correct neutral diathesis counterpart, the clitic

⁷⁷ In the sentence "*Il s'achète un cadeau*", the interpretation of "se" can be either IO, possessive, or even ADJ; however, the case designation of Dative encompasses all these functions in one category.

se will be replaced by the object argument (direct or indirect) *soi-même* [oneself], as it is regarded as both the historical meaning of the reflexive pronoun and the closest substitute for it. The method to determine whether the valency decreased or increased will be based on Grimshaw's argument structure theory.⁷⁸

Categorical (nominal): addition, reduction, promotion, demotion (of an argument), fusion (of arguments)

These criteria will be coded in tabular form, commonly used for diathetic mapping representations. Table 8 is an example of the format of a diathetic table. The complete analysis of the 33 structures can be found in an adjunct document.

	<i>Je</i>	<i>me lave</i>
<i>Syntactic function</i>	S	V
<i>Semantic role</i>	A	V
<i>Case-marking</i>	NOM	V
<i>Valency change</i>	Neutral diathesis: * <i>Je lave moi-même</i> ► Reduction of a syntactic argument (V cannot accept DO) ► Reduction of a semantic argument (V cannot accept P)	

Table 8: Example of diathetic mapping

4.2. A new classification of pronominal verbs based on the concept of diathesis

A general look at the diatheses covered by the sample verbs

As a result of the analysis of the diatheses corresponding to the 33 different structures analysed, our inductive approach identified eight different mappings of diathesis. These are broken down in the following table, along with the three types of properties we analysed, i.e. syntactic, semantic, and empirical.

⁷⁸ Grimshaw, *Argument Structure*.

Diathesis	Syntactic	Semantic	Empirical	Examples
Canonical reflexive	S se V	A se V	A = P Valency decrease ($\Theta\emptyset/P$)	<i>Je me lave</i>
Canonical reciprocal	S se V	A se V	$A\alpha = P\alpha$ Valency decrease ($\Theta\emptyset/P$)	<i>Pierre et Marie s'aiment</i>
Reflexive benefactive	S se V DO	A se V P	A = B Valency decrease ($\Theta\Delta F$)	<i>Le chat se lave l'oreille</i>
Autocausative	S se V	E se V	E = CAUS Valency decrease ($\Theta\emptyset/CAUS$)	<i>Je me suis calmé</i>
Agentless passive	S se V	P se V (A)	A = $\emptyset$ Valency decrease ($\Theta\emptyset$)	<i>Le français se parle</i>
Conversive passive	S se V OBL/ADJ	P se V A	A = OBL/ADJ Valency decrease ($\Theta\emptyset$)	<i>L'avion se remplit avec les passagers</i>
Impersonal passive (with dummy subject)	∇ se V DO/OBL	$\emptyset$ se V P	S/A = $\emptyset$ Valency decrease (S/A)	<i>Il se passe des choses étranges.</i>
Anticausative	S se V	P se V	A/CAUS = ? Valency decrease ($\Theta\emptyset$)	<i>Le verre se cassa</i>

Table 9: Typology of diatheses represented by pronominal verbs

From a syntactic point of view, most of these constructions follow the structure *S se V*, with only a few exceptions (reflexive benefactive, conversive, and impersonal passive). While the structure *S se V* only represents the minimal syntactic mapping of a given structure, this relatively common pattern for pronominal verbs could explain the difficulties in interpreting each type of construction.

Semantically, Table 9 distinguishes between two major categories of diatheses represented in pronominal-verbal constructions:

- Diatheses where the subject of the sentence is preserved, but an external argument is demoted/removed (i.e. canonical reflexive, reciprocal, reflexive benefactive, anticausative).
- Diatheses where a patient is promoted as the main predicate and the agent is demoted/removed (i.e. agentless passive, conversive passive, impersonal passive, anticausative).

Empirical observations will supplement syntactic and semantic mappings by linking them together, as well as describe operations on the valency of the arguments. A quick analysis of the constructions enables us to pinpoint the following specific features:

- The function of the reflexive morpheme in the canonical reflexive is to encode coreferentiality between the agent and the patient (they refer to the exact same entity).
- The function of the reflexive morpheme in the canonical reciprocal is to encode coreferentiality between at least two agents and patients that mutually perform the process.
- The function of the reflexive morpheme in the reflexive benefactive is to encode the addition of a dative nuance to the process that the agent carries out on the patient.⁷⁹
- The function of the reflexive morpheme in the anticausative is to encode a self-caused sensorial process affecting the experiencer.

⁷⁹ While most of the time, this diathesis implies partial coreferentiality between the agent and the patient (such as an agent acting upon a part of itself in body-part constructions like *Je me lave les mains*), this is not a prerequisite condition (see below).

- The function of the reflexive morpheme in the agentless passive is to encode the suppression of the agent active in the process.
- The function of the reflexive morpheme in the conversive passive is to encode the demotion of the agent to a circumstantial position.
- The function of the reflexive morpheme in the impersonal passive is to encode the marginalization of the subject/agent as a dummy subject.
- The function of the reflexive morpheme in the anticausative is to encode the deletion of any cause and/or agent identifiable in the process.

4.3. Second phase: Testing of our typology applicability through corpus analysis of *Sequoia* and *FTB*

As a result of the first phase of our research, we identified eight diathetic patterns, which cover the 33 prototypical reflexive structures found in previous work. To verify the degree of applicability of these patterns and the potential need for additional diathetic categories, we carried out a manual corpus analysis of pronominal structures found in the *Sequoia* (332 occurrences) and the *FTB* (900 occurrences) corpora. *Sequoia* is composed of over 3,200 sentences sourced from Europarl (European Parliament texts), *L'Est Républicain* (a French right-wing newspaper), Wikipedia, as well as European Medicines Agency, while the *FTB* includes 1 million words taken from *Le Monde* (1989-1995), one of the most prominent daily newspapers in France. Occurrences were identified by isolating the reflexive clitic *se* and its derivatives, preceded or followed by a verbal form. Three linguists (A1, A2, A3), among which two of them (A1 & A2) had no prior expertise with the concept of diathesis, were asked to determine the diathesis which each structure belongs to and to detail their reasoning, using a three-step procedure described in an annotation guide: 1) identifying the structure's meaning, 2) constructing a diathetic mapping, and 3) answering binary questions to interpretate the mapping. These steps are further detailed in the annotation guide (Annex 2). The

annotation procedure was carried out in an iterative way: about 100 structures were first annotated by the three linguists (referred to as *annotators* from here on), leading to a discussion to resolve disagreements and produce a gold standard. The annotation guide was then refined, and a new 100-structure annotation phase started. At the end of this second phase, we computed a Cohen's Kappa to assess inter-rater agreement on the 200 structures. The values of κ are 0.37 for A1-A2, 0.44 for A1-A3, and 0.64 for A2-A3. It appeared that A1 had more trouble adapting to the typology, while A2 and A3 reached a fairly high degree of agreement. The 1,032 remaining structures were then separately annotated by the annotators after receiving final feedback. At the end of the annotation process, we defined a gold standard based on a consensus between the annotators.

4.3.1. Distributions of the categories

Figure 1: Diathesis distribution
Sequoia

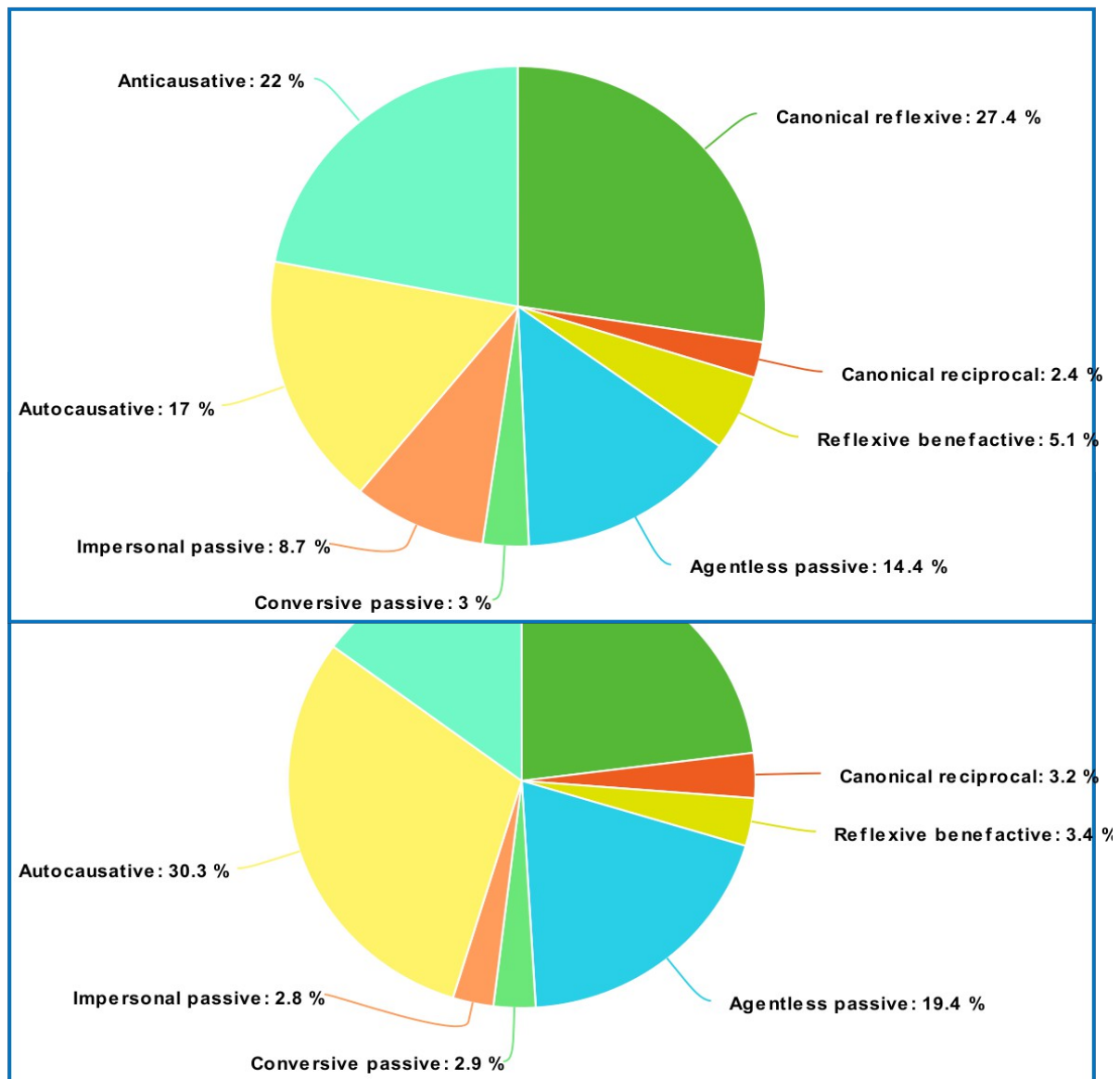
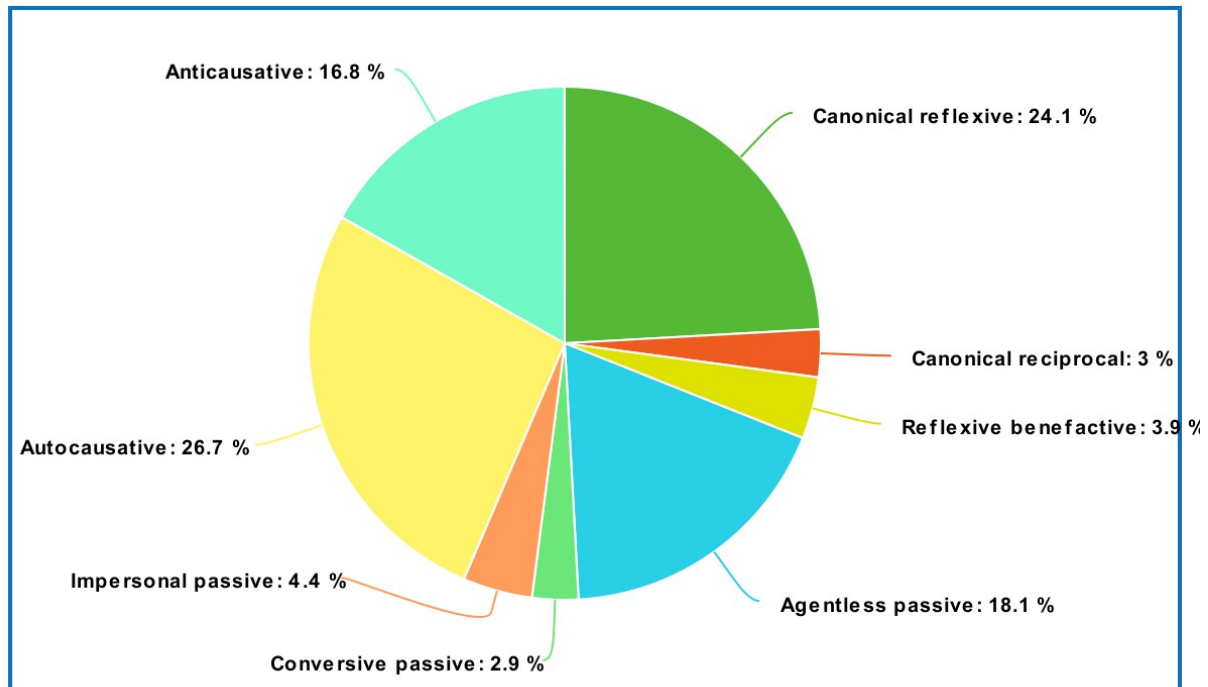


Figure 3: Diathesis distribution
Combined corpora



The first and most important result is that the annotation process and the discussions that followed did not highlight any new structures that did not support our typology. Our typology was then validated by this corpus analysis. The distribution of our eight categories in the corpus is set out in Figure 1 (*Sequoia*), Figure 2 (*FTB*) and Figure 3 (*Combined corpora*).

By adding up numbers from the two corpora, we can observe that four categories were especially well-represented: autocausative (329 structures = 26.7%), canonical reflexive (298 structures = 24.1%), agentless passive (223 structures = 18.1%), and anticausative (207 structures = 16.8%). However, the category distribution in the two corpora is significantly different ($X^2(7) = 51$; $p < 0.001$). Further analysis of these differences through adjusted standardized residuals shows that impersonal passives ($z = 4.53$) and anticausatives ($z = 2.95$) are overrepresented in *Sequoia*, whereas agentless passives ($z = 2.01$) and autocausatives ($z = 4.74$) are overrepresented in *FTB*. The

genre of the texts forming these corpora seems to have considerable impact on the distribution of diathesis types. Autocausatives, for instance, are more common in general newspapers like *Le Monde/FTB* (273 occurrences), whereas medical literature generates a greater amount number of anticausatives, such as the following examples (*Sequoia*):

(1) *L'acide zolédronique ne **se lie** pas fortement aux protéines.*

(2) *Les os deviennent plus fragiles et **se cassent** plus facilement.*

As a general tendency, however, we can see that reciprocals, reflexive benefactives, and conversive passives have low representation, agentless and impersonal passives have low to middle representation, and canonical reflexives, autocausatives, and anticausatives have high representation.

However, these figures are very much related to the definition of each class in our typology. By way of comparison, Table 10 compares our annotation with that of Barque & others,⁸⁰ who divided the majority of the 900 *FTB* tokens into three categories: 372 various autonomous structures (41.3%), 221 anticausative structures (24.6%), and 104 reflexive structures (11.6%):

CATEGORY	REFL	RECI	INTR	AUTO	ANTI	MED	MED- ANTI	AUTO (SEL)	AUTO (SYN)	AUTO (LEX)	AUTO (CON)	AUTO (ASP)
Diathesis												
Canonical reflexive	57	4	5	38	8	5	0	52	27	7	3	5
Canonical reciprocal	2	24	0	0	0	0	0	2	0	1	0	0
Reflexive benefactive	12	2	1	1	2	0	1	2	2	4	0	0
Agentless passive	11	4	0	6	58	27	6	29	22	1	7	4
Conversive passive	2	0	0	1	2	10	0	11	0	0	0	0
Impersonal passive	0	0	0	0	0	0	0	3	22	0	0	0
Autocausative	18	0	25	21	60	4	2	53	85	4	0	1
Anticausative	2	2	4	3	91	3	4	13	3	0	0	9

Table 10: Distribution of pronominal structures in *FTB*, comparison with Barque & others (2019)

⁸⁰ Barque, Candito & Huyghe, 'La classification des verbes réfléchis', pp. 121-137.

As we can see in Table 10, only 70 structures were identified as autocausatives by Barque & others,⁸¹ compared with 273 in our typology. However, a good number of autonomous came under our autocausative category, and we categorized some autonomous in Barque et al. as impersonal passives or agentless passives.⁸² While the definition of autocausative differs between our typologies, the huge gap between categories speaks for the difficulty of identifying some structures. Meanwhile, the number of reciprocals was close enough (36 to 29) to conclude that this category was rather easy to identify. Barque et al. also note that there were no occurrences of autobenefactives, parallels, and or antipassives.⁸³ Indeed, structures we annotated as reflexive benefactive did not correspond to these authors' autobenefactives, which were mostly considered as reflexives.

4.3.2. Discussion about ambiguous categories

Now that we have an idea of the proportion of each diathesis takes within two French corpora, we need to look at the difficulties that occurred during the annotation process in order to assert the reliability and simplicity of our typology. As such, we computed the recall or sensitivity between each annotator and the gold standard (GS) (a value of 1 for a given diathesis means that 100% of GS annotations were correctly identified by the annotator), and the true positive rate or positive prediction value (a value of 1 for a given diathesis means that all occurrences assigned to this diathesis were correct). The recall rate and true positive rate of each diathesis between the annotators (A1, A2, and A3) and the gold standard (GS) in *Sequoia* (1-200) are as follows:

⁸¹ Barque, Candito & Huyghe, 'La classification des verbes réfléchis', pp. 121-137.

⁸² Ibid.

⁸³ Ibid.

Diathesis	Recall rate (A1 + GS)	True pos rate (A1 + GS)	Recall rate (A2 + GS)	True pos rate (A2 + GS)	Recall rate (A3 + GS)	True pos rate (A3 + GS)
Canonical reflexive	0.8333	0.5195	0.8125	0.7500	0.9167	0.8302
Canonical reciprocal	0.8333	0.5556	0.8333	1	0.8333	0.5556
Reflexive benefactive	0.1429	0.2222	1	0.8750	1	0.9333
Agentless passive	0.4348	0.3704	0.9130	0.7500	0.7826	0.8571
Conversive passive	0.2857	0.2000	0.4286	0.5000	0.8571	1
Impersonal passive	1	1	1	0.8947	1	1
Autocausative	0.4324	0.4000	0.5135	0.7600	0.9459	0.8140
Anticausative	0.08333	0.36364	0.8542	0.8367	0.7083	0.9444

Table 11: Confusion matrix recall and true positive rates compared with the gold standard, *Sequoia*

1-200

As can be seen, some categories were more prone to mistakes than others. A1 and A2 both had trouble with the conversive passive, given their low to average scores in both recall and true positive rates. The autocausative was also difficult to identify for A1 & A2, although A2 rarely assigned this diathesis incorrectly. Individually, A1 had difficulties in both identifying and distinguishing the reflexive benefactive and the anticausative diatheses. On the bright side, the canonical reflexive, reciprocal, and impersonal passive diatheses obtained very high rates for each annotator. The agentless passive was also fairly well defined.

Consequently, four diatheses, namely the reflexive benefactive, the conversive passive, the autocausative, and the anticausative diatheses were seemingly not clear enough for annotators, especially A1. We will hereafter describe the difficulties relating to their interpretation and the rationale for our classification.

The case of reflexive benefactive

This diathesis is defined by two major characteristics: the necessary presence of an object (most of the time, a DO/P) and a valency reduction operated through the suppression of a dative. Evidence of this dative can be found in the neutral diathesis through the addition of an IO, a possessive, or a genitive, which all take the semantic role of beneficiary:

(3) *Il découvre un lion qui se frotte le ventre.*

→ *Un lion qui frotte le ventre à/pour lui-même (IO)*

→ *Un lion qui frotte son ventre (possessive)*

→ *Un lion qui frotte le ventre de lui-même (genitive)*

In each sentence, the semantic interpretation is that of a process aimed at the agent, which implies it also becomes a beneficiary.

This interpretation of the agent/beneficiary duality is supported by Creissels and can also be applied to a sentence such as *Je me mange tout le gâteau*.⁸⁴ Here, the agent performs the process (*manger*) on the patient (*le gâteau*) in order to produce an effect on himself (emphasized by the clitic). Typically, eating involves the participation of the agent in the process to a higher degree than other actions, since it is most often motivated by a goal (satiety, for instance). With this in mind, it is not difficult to understand why the agent would also be a beneficiary in a sentence such as *Je me mange tout le gâteau*. However, this effect can be involuntary and detrimental, as can be seen in the sentence *Je me prends un coup*. *Se prendre* should be understood here as meaning *to take*. A lot of words can be combined with this meaning of *se prendre* and will always take either a positive, voluntary nuance (known as real benefactives), or a negative, involuntary nuance (known as malefactives⁸⁵):

⁸⁴ Creissels, 'Réflexivisation, transitivité et agent affecté', pp. 83-106.

⁸⁵ Zúñiga & Kittilä offer a detailed perspective about the nuances of benefactive and malefactive in their reports, especially in sentences with *give* and *take* (see F. Zúñiga & S. Kittilä (eds.). *Benefactives and*

(4) *Je me suis pris un hamburger. (benefactive)*

(5) *Je me suis pris une boisson. (benefactive)*

(6) *Je me suis pris une amende. (malefactive)*

(7) *Je me suis pris une gifle. (malefactive)*

Although the identity of the agent in such sentences could be debatable due to the semantic construction, from a syntactic point of view, the action of *se prendre* is still carried out by the subject *Je*, which is why we would categorize these sentences as reflexive benefactive. Indeed, they do share the same diathetic mapping (A se V P - S se V DO), and we were also able to recover these benefactive/malefactive nuances in other types of sentences categorized under this diathesis:

(8) *Je me lave les mains. (benefactive)*

(9) *Les patients qui se sont fracturé la hanche. (malefactive)*

(10) *Je me tape les bouchons. (malefactive)*

(11) *Je me réserve le droit de prendre le volant. (benefactive)*

All these sentences belong to the category of *dative turns* as defined by Evrard,⁸⁶ since they revolve around the same principle of an agent being treated as a beneficiary/maleficiary. Although there were few occurrences of reflexive benefactive in the corpus analysis (respectively 17 & 31 tokens, respectively, totaling 3.8%), it should be easy to identify thanks to its peculiar structure. According to the feedback from our annotators, the confusion with other types came from misidentifying the dative meaning in specific sentences.

The case of conversive passive

Conversive passives, much like the agentless passive, are diatheses placing the patient in subject position. However, unlike the agentless passive, this diathesis allows the neutral diathesis agent to be expressed, although in a marginal position (adjunct or oblique complement):

*malefactive*s: *Typological perspectives and case studies* (John Benjamins Publishing, 92, 2010).

⁸⁶ I. J. M. S. Evrard, 'La diathèse en français: un essai de synthèse', *Actas del XXIII Congreso Internacional de Lingüística y Filología Románica* (Salamanca, 2003): pp. 303-316.

(12) *L'avion s'est rempli avec les passagers.*

(13) *La finale se jouera entre Utrecht et Soleuvre.*

There were very few occurrences of this diathesis in the corpus (respectively 10 & 26 tokens, respectively, totaling 2.9%), but it was often confused with aspectual anticausatives, such as (14).

(14) *La séance s'est terminée par des applaudissements.*

The difference between conversive passives and aspectual anticausatives is mainly found in the semantic nature of the adjunct/oblique complement. For instance, unlike *les passagers* towards *remplir* in (12), *des applaudissements* cannot be the active agent of *terminer*.

The case of autocausative

Unlike reflexives which licence dynamic processes controlled by an agent, autocausatives are defined as events in which a conscious being is subject to self-induced sensorial experience.⁸⁷ As such, prototypical examples of the autocausative diathesis can be found within emotion verbs (e.g. *s'énerv*, *s'attrister*, *se fâcher*, *se réjouir*, *s'ennuyer*, *s'effrayer*, *s'étonner*, *s'honorer*, etc.). Pronominal structures are particularly suitable for expressing emotions, as the semantic role of experiencer is prominent in such constructions.⁸⁸ In a sentence such as *Je m'énerv*, *je* is hardly an agent; otherwise, it would imply that *je* is in total control of the process. The fact that autocausatives license experiencers instead of agents/patients is also reflected in tests of patient emphasis such as *Je m'énerv moi-même**, which is less acceptable than the emphasized canonical reflexive *Je me lave moi-même*.⁸⁹

⁸⁷ M. Parry, 'The reinterpretation of the reflexive in Piedmontese: impersonal' SE constructions', *Transactions of the Philological Society*, 96 (1998): pp. 63–116.

⁸⁸ W. Croft, *Syntactic categories and grammatical relations: The cognitive organization of information* (Chicago: University of Chicago Press, 1991).

⁸⁹ K. de Schepper, 'The space between one and two', *Transitivity: Form, meaning, acquisition, and processing*, 166 (2010): 191p.

Similarly, Kailuwet writes that emotion verbs in French turn the subject into an experiencer, especially because the degree of controlled intentionality is lower than in action-focused verbal processes.⁹⁰ The same analysis was made by Bieñ,⁹¹ who treats these verbs as medio-active pronominal structures (agentivity +, control 0). Meanwhile, the diminished control over the process (due to the nature of inchoative verbs) does not prevent the experience from happening willingly either. As such, in (15) and (16), the subject could be considered an experiencer because the verb describes a voluntarily self-induced change of state:

(15) *Le titulaire de l'AMM s'engage à fournir des études.*

(16) *Le Parlement s'est efforcé d'en tenir compte [...]*

Other inchoative verbs potentially characterized as autocausatives would include physical reactions (*se réveiller, se fatiguer, s'épuiser, se mourir*) or perceptive and mental attitudes (*se repentir, s'apercevoir, se douter*), due to the fact that these verbs feature change of state/perceptive events within a conscious experiencer, rather than controlled actions.^{92/93/94} The variety of verbs which can be considered as autocausatives is reflected by the high number of occurrences found in the corpora (329 tokens, 26.7%).

The case of anticausative

Similarly to the agentless passive, the anticausative diathesis is marked by the promotion of a patient into a subject position, while suppressing the agent and/or cause of the process. However, there is a semantic nuance between these diatheses coming from the fact that anticausatives typically dismiss the intervention of an actor with consciousness and volition.⁹⁵ An agentless passive

⁹⁰ R. Kailuwet, 'La classe P dans les verbes français et les verbes de sentiment', *Langue française*, 1 (2007): pp. 33-39.

⁹¹ Bieñ, 'Sur la nature des confusions notionnelles', pp. 128-135.

⁹² A. Culioli, *Pour une linguistique de l'énonciation: opérations et représentations* (Ed. Ophrys, 1, 1990).

⁹³ Y. Soga, 'Pour une meilleure didactique des verbes cognitifs en construction pronominale (Dossier: Où en est la "subjectivité du locuteur" dans l'enseignement du français?)', *Enseignement du français au Japon*, 31 (2003): pp. 51-60.

⁹⁴ de Schepper, 'The space between one and two'.

⁹⁵ C. Fellbaum & A. Zribi-Hertz, 'La construction moyenne en français et en anglais: étude de syntaxe et de sémantique comparées', *Recherches linguistiques de Vincennes*, 18 (1989): pp. 19-57.

sentence such as *Le français se parle ici* hints that some agent is responsible of the process, even if not mentioned; meanwhile, an anticausative such as *Le verre se casse* seems to indicate a process happening spontaneously. There is no causal effect that can be recollected in the sentence, hence the denomination anticausative.⁹⁶

Historically, the cause in anticausative sentences could be expressed as an external argument *par*-phrase, in the same way that reflexive passives licensed expression of the agent:⁹⁷

(17) *Tout vin donc ne s'aigrist pas par la chaleur.*

(18) *Le contraire **se montre** par les guerres.*

While it is no longer possible to express causes in this manner in **all** types of anticausative sentences, some Canadian dialects still allow their use, testifying historical use of *par*-phrases with pronominal verbs.⁹⁸ Furthermore, many verbs accept causes as external arguments in standard Modern French as adjuncts in preposition phrases, similarly to conversive passives, turning the clause into a *marked anticausative*:⁹⁹

(19) *Le fromage s'est **moisi** rapidement avec l'humidité.*

Typically, anticausatives licence inanimate patients undergoing a self-explanatory process:

(19) *Les os [...] **se cassent** plus facilement.*

(20) *L'héritage des droits d'auteur **se cristallise** dans les traités.*

(21) *La seule lueur d'espoir qui **se détache** actuellement de la situation anarchique [...]*

⁹⁶ M. Labelle & E. Doron, 'Anticausative derivations (and other valency alternations) in French', *Probus*, 22 (2010): pp. 303-316.

⁹⁷ S. Heidinger & F. Schäfer, 'The French Reflexive Passive and Anticausative' in *Evolutions en français. Etude de linguistique diachronique*, ed. by B. Fagard & others (Berlin, Peter Lang, 2008): pp. 135-152.

⁹⁸ J. M. Authier & L. Reed, 'On the Canadian French middle', *Linguistic Inquiry*, 1996: pp. 513-523.

⁹⁹ S. Heidinger, 'Causalness and the encoding of the causative–anticausative alternation in French and Spanish', *Journal of Linguistics*, 51 (2015): pp. 562-594.

Other events characterized as anticausative include aspectual verbs (*se terminer*, *se prolonger*, *se maintenir*) and descriptive verbs (*se définir*, *s'avérer*, *se révéler*).¹⁰⁰ Anticausatives made up for 16.8% of annotated diatheses in the corpora (208 tokens).

5. Discussion

The results we obtained from our analysis of the diathetic mappings based on the most common uses of our initial sample of 20 representative pronominal verbs, as well as the corpus annotation of 1,232 structures, enabled us to classify pronominal-verbal constructions under eight potential diatheses. The use of Kulikov's typology of diatheses helped us find out how each mapping could be interpreted and how these verbs could be compared with other verbs at a more general linguistic level.¹⁰¹ The view of the diathesis as a systematic pattern of arguments enabled us to “dissect” complex structures, in the way that construction grammar encodes syntactic structures with their semantic layer, argument by argument.¹⁰² However, we found diathesis to be a more convenient way to build up a classification, since construction grammar is often criticized for its overgeneration of information.¹⁰³

The eight diathetic mappings we sketched, while varying in terms of complexity, seem consistent enough to make a case for a potential classification of pronominal-verbal constructions.

We believe that a classification based on diathetic structures has the following additional advantages:

- **Methodological consistency:** It interconnects semantic, syntactic, and valency analyses to deliver consistent argument mappings, in the same fashion as Goldberg's construction grammar,¹⁰⁴ but with the aim of generalizing patterns into a typology.

¹⁰⁰ Heidinger & Schäfer, 'The French Reflexive Passive', pp. 135-152.

¹⁰¹ Kulikov, 'Voice typology', pp. 368–398.

¹⁰² Goldberg, *Constructions: A construction grammar approach*.

¹⁰³ R. Van Trijp, 'Cognitive vs. generative construction grammar: The case of coercion and argument structure', *Cognitive Linguistics*, 26 (2015): pp. 613-632.

¹⁰⁴ Goldberg, *Constructions: A construction grammar approach*.

- **Accounting of lexicalized structures:** As a consequence of this method, any verb can have its pronominal form explained, since a diathetic mapping operates at several linguistic levels. Lexicalized (or semi-lexicalized) pronominal verbs (such as *se rendre compte*, *se prendre un coup*), in particular, could be explained on the basis of an empirical analysis when purely semantic or purely syntactic analysis fails to deliver a satisfying explanation.¹⁰⁵
- **Accounting of polysemic verbs:** In particular, the focus on uses rather than forms allows us to connect several verbs that would otherwise be treated as separate cases (such as *se casser*, usually classified as subjective because of a focused view on *Le verre se casse*, whereas *se casser la jambe* exists along with *se laver le visage* as reflexive benefactives).

Our analysis was also the opportunity to confirm that these verbs, which are seemingly heterogeneous, share a common property: a reduction in valency of the verbal process. Even in the case of the reflexive benefactive diathesis, the original dative meaning is actually suppressed when using the *se* variant. Further insights into this issue can be obtained by comparing the diatheses we found for French pronominals with Kulikov's definition of the middle voice,¹⁰⁶ which includes a variety of diatheses often accompanied by an intransitivization of the neutral structure or, in other words, a valency reduction. The researcher writes that the middle voice encompasses the **passive, conversive, anticausative, reflexive, reciprocal, antipassive, and reflexive benefactive** diatheses. It is not fortuitous that these diatheses are, for the most part, represented by the same morphological mark in Russian (reflexive *-sja/-s'*), but also by French pronominal-verbal structures, as shown in the classification we have built up in the present study. This serves as further evidence that pronominal-verbal constructions are characterized by a valency decrease by one degree compared with their neutral diathesis variant. The fact that this valency reduction can apply to different semantic roles – agent, patient or dative – appears in our work as one of the most prominent properties of the pronominal-verbal constructions. Although this observation was already

¹⁰⁵ For the rationales behind lexicalized pronominal verbs, see Gaatone, 'Réflexions sur les verbes pronominaux': pp. 199-222; and Burston, 'The pronominal verb construction in French': pp. 147-176.

¹⁰⁶ Kulikov, 'Voice typology', pp. 368–398.

hypothesized by several authors,¹⁰⁷¹⁰⁸¹⁰⁹¹¹⁰ no attempt had previously been made to confirm this valency reduction based on a systematic analysis.

Limitations

Our diathetic classification and its eight categories reveal recurring patterns that are general enough to be used for a quantity of verbs, while being brought together by the valency operator *se*. However, even though this classification has its merits, it still should be tempered by several facts.

Firstly, the degree of complexity of diathetic mappings implies some degree of interpretation, which means it involves subjectivity. As such, a thorough understanding of notable uses of pronominal verbs is often necessary in order to determine, for instance, the semantic roles of the arguments.

Secondly, while we paid attention to consider as many pronominal-verbal structures as possible, especially in the corpus annotation task, it is possible that certain pronominal-verbal uses might not perfectly fit one of the categories we built up in the present paper. In such a case, it would be interesting to figure out to what degree the said structure is frequent, and how its diathetic mapping could be constructed in order to be incorporated into our typology.

6. Conclusion

The present research had two goals. First, to review the major classifications of French pronominal verbs up until now. The theme choice of pronominal verbs as our subject was motivated by the ever-growing need for a thorough and precise understanding of grammatical concepts, but also by the lack of consistency in the descriptions made over this important category of verbs. While our review of the existing classifications showed that they all had their merits, we wished to explore

¹⁰⁷ Dik, 'Valentie en valentie operaties', pp. 95-114.

¹⁰⁸ Bień, 'Sur la nature des confusions notionnelles', pp. 128-135.

¹⁰⁹ Creissels, 'Réflexivisation, transitivité et agent affecté', pp. 83-106.

¹¹⁰ Corminboeuf & Heyna, 'Nominalisation et diathèse', pp. 34-53.

whether the concept of diathesis could be applied to analyse pronominal verbs in order to bring new keys to their understanding.

Our second objective was to build a new typology, based on the concept of diathesis. The results of our thorough analysis of a sample of 20 pronominal verbs and their 33 major pronominal uses, followed by a corpus annotation of 1,232 structures, allowed us to identify eight categories of diatheses that should be applicable to the great majority of pronominal verbs. The utility of this new conceptualization not only lies in the flexibility of tabular diathetic mappings, but also in the systematic integration of the polysemy of pronominal verbs. In addition, this work was an opportunity for us to confirm that French pronominal verbs always induce a valency reduction compared with their neutral diathesis counterpart, which means that a semantic and/or syntactic element is eluded or deleted in the process of pronominalization.

Our findings might help in enhancing the understanding of the complex phenomenon of French pronominal verbs, as well as in offering a complementary tool to conduct further research in on the matter. We hope that the present work study could lead to more accurate methodologies in the teaching of these verbs, so that practical utility could assist theoretical findings.