

New insights into the typology of motion in the history of French: Evidence from the manner verb lexicon

Mariapaola Piccione* and Thomas Rainsford

Universität Stuttgart

*mariapaola.piccione@ling.uni-stuttgart.de

Résumé. Notre étude vise à étudier l'expression de déplacement selon la typologie de Talmy en diachronie. Nous examinons l'hypothèse de Slobin (1997, 2004) selon laquelle la proportion de verbes de manière est plus importante dans les langues à cadrage satellitaire et l'hypothèse de Schøsler (2008) selon laquelle la différence entre les langues à cadrage satellitaire et les langues à cadrage verbal se trouve dans l'emploi des verbes de manière et dépend aussi du type de texte. Notre étude est innovante du point de vue méthodologique car elle cherche à vérifier ces hypothèses sur un très grand ensemble de données en s'appuyant sur des mesures adaptées. Nos résultats montrent une augmentation de la saillance de la manière (au sens de Slobin 1997, 2004), contrairement à ce que nous attendions. Cette augmentation de la proportion des verbes de manière pourrait s'expliquer par une augmentation de la diversité lexicale dans le domaine de l'expression du déplacement en général. De plus, nos résultats soutiennent l'hypothèse de Schøsler, car les facteurs textuels influent sur l'emploi des verbes de manière.

Abstract. Our study aims to investigate the Talmyan typology of motion encoding in the history of French focusing on testing Slobin's (1997, 2004) hypothesis stating that the proportion of manner verbs is greater in s-framed languages and Schøsler's (2008) hypothesis stating that the difference between s-framed and v-framed languages is to be found in the use of manner verbs and that the nature of the texts might play a role. Our study is methodologically innovative since it tests these hypotheses against a very large dataset using tailored measures. Our findings show an increase in manner saliency (as defined in Slobin 1997, 2004), contrary to what we would expect. This increase in the proportion of manner verbs might be explained by a more general increase in the overall lexical diversity of motion expression in French. Moreover, our results support Schøsler's hypothesis, as textual factors affect the use of manner verbs.

1 The typological shift from an s-framed to a v-framed language

In 1972, Talmy first published his typology concerning the study of different lexicalization patterns based on the analysis of the verb root and the satellite. The patterns examined by Talmy were grouped into the broader set of "motion events", which describe constructions involving location or movement. The motion event comprises different components such as the Figure, which is the object that moves, the Ground, the object with respect to which the Figure moves, the Path, described as the path followed by the Figure, and the Movement, which can indicate either movement or location (Talmy 1985). Following this typology, languages can be divided into two categories: languages with a verbal frame (henceforth called v-framed languages), if the Path is conflated in the verb root, and languages with a satellite frame (henceforth called s-framed languages), if it is conflated in the satellite (Talmy 1991, 2000). The two types of languages also differ in the conflation of Manner, which describes how a movement is performed; s-framed languages such as English conflate Manner in the verb root (1) while v-framed languages such as French typically encode it in a secondary clause (2), as shown by the well-known Talmyan example:

(1) English

The bottle [floated]_{MANNER} [into]_{PATH} the cave

(2) French

La bouteille [entra]_{PATH} dans la grotte [en flottant]_{MANNER}

In this framework, Romance languages are classified as v-framed, despite deriving from Latin which is considered s-framed. In the evolution from Latin to Romance a typological shift has occurred, possibly due to changes in word order (Buridant, 2000), in the syntax (e.g., Dufresne et al. 2000, 2001, 2003; Acedo-Matellán & Mateu 2008, 2013), and/or in the lexicon (Stolova 2015).

How the shift from Latin to French occurred has been the subject of discussion in the literature. Many scholars agree in considering the shift as a gradual change (Acedo-Matellán 2010; Acedo-Matellán & Mateu 2013; Talmy 2000), including hybrid intermediate stages where both s-framed and v-framed characteristics persisted in the language (e.g., Iacobini & Fagard 2011; Iacobini 2012; Slobin 2004). Contradicting this view, Troberg and Burnett (2017) claimed the shift to be abrupt causing the rise of a new s-framed system during the Middle Ages. In a recent study, Troberg and Leung (2021) observed the loss of four particles (i.e., *jus* 'down', *fors/hors* 'out', *arrière* 'back', *avant* 'forward') in Old French and concluded that Old French should be considered as a "weak" v-framed language compared to Modern French, a "strong" v-framed language. Despite the divergences on how the shift has evolved, it is widely acknowledged that Old French still displays s-framed characteristics such as the use of prefixes with a directional and aspectual meaning (e.g., Dufresne, Dupuis, & Tremblay 2003; Kopecka, 2009), the adjectival resultative constructions (e.g., Buridant 2000; Troberg & Burnett 2017), the verb particle constructions (e.g., Burnett & Tremblay 2009; Troberg & Burnett 2017) and the use of telic Path with manner verbs (Troberg & Burnett 2017). While many studies have focused on investigating different s-framed characteristics of Old French, to our knowledge, no study has systematically investigated another typical s-framed characteristic: the manner verb lexicon. For this reason, we decided to investigate the development of French manner verbs to have a more complete overview of the shift and the status of Old French.

1.1 The lexicon of manner verbs

S-framed languages are characterised by a larger manner verb lexicon and a higher frequency of use of manner verbs. By comparing narratives in three v-framed languages (Spanish, Hebrew and Turkish) and two s-framed languages (English and German) Berman and Slobin (1994) observed that the s-framed languages allow for a more detailed specification of Manner and suggested that this was probably due to a richer manner verb lexicon. The same conclusion was reached by Slobin (1997) when conducting a study examining the translations of *The Hobbit* into s-framed and v-framed languages. The comparison of translations demonstrated that v-framed languages show a preference for expressing directionality in contrast to s-framed languages which favour the expression of Manner. This led Slobin (1997) to hypothesize that s-framed languages have at their disposal a larger inventory of manner verbs compared to v-framed languages, most of all when considering those verbs encoding a very specific and detailed Manner, and that the differences in motion encoding strategies correlate with differences in rhetorical styles. Phylogenetic approaches also confirm the correlation between the size of the lexicon of manner and path verbs and motion encoding strategies. In particular, it has been shown that Proto-Indo-European had a tendency towards s-framed patterns of motion encoding and that this could be correlated with a larger set of manner verbs compared to the set of path verbs (Verkerk 2014).

In addition to the differences in the size of the lexical inventory, the use of manner verbs differs across languages. For instance, v-framed languages have manner verbs at their disposal –even if to a lesser extent– but they are simply used less (Slobin 2004). This difference in use is also found within the family of s-framed languages, as some languages make more use of manner verbs than others. For this reason, Slobin (2004) suggested that the investigation of rhetorical styles does not support a two-way typology (i.e., s-framed vs. v-framed) but rather the idea of a continuum of manner saliency in which, among the languages investigated by Slobin, Spanish and French (v-framed) on the one side and Thai and Russian (s-framed) on the other side occupy the two opposite poles of the continuum. This theory also supports the claim that if the encoding of Manner is easily accessible to speakers, then this would lead to a progressive increase in Manner saliency over time. To test this claim, Fanego (2012) investigated the diachrony of English examining the evolution of Manner saliency from Old English to Modern English and found that it is indeed

the case that the ease of accessibility of Manner in a language (e.g., larger Manner verb lexicon and greater use of Manner expressions) leads to an increase in Manner saliency for that language.

While a number of studies highlight the s-framed properties of Old French, as seen above, only Schøsler (2008) draws attention to the different lexical resources of Old and Modern French, using extracts from *Charrette* for Old French and *Germinal* by Zola for Modern French and comparing these with extracts from Classical (*Bellum Gallicum*) and Late Latin (*Peregrinatio*), and Danish (*Saere Historier*). Schøsler noticed that the main difference between these languages is not found in the size of the inventory (i.e., the number of verb types) but in the frequency of use of manner verbs (i.e., token frequency). In fact, Schøsler affirms that the quantity of manner verbs found in the Old French text appears not to differ so much from the number of manner verbs found in s-framed Danish (*Charrette*, 57.1% manner types vs. *Saere Historier*, 58.3% manner types). But a different picture emerges when comparing the frequency (i.e., the use) of manner verbs across languages, since the Danish text shows a greater use of manner verbs (40.2%) compared to Classical Latin (21.1%), Late Latin (6.8%), Old French (23.6%) and Modern French (10.8%). Schøsler concludes that, when conducting a typological study, it is crucial to distinguish between the lexical resources of a language and their actual use and that the latter can vary according to the nature of the text (2008, p. 127). Nevertheless, as Schøsler mentions, these results need to be corroborated by more data. So, what is missing in the literature is a systematic investigation of the lexicon of motion verbs in the history of French, which constitutes an interesting case since it is claimed to show many s-framed characteristics during the Middle Ages before developing into the v-framed language *par excellence*.

1.2 The lexicon of manner verbs as a typological diagnostic in the diachrony of French

The present study focuses on the lexical resources of French, more precisely, on the inventory and use of manner verbs. The importance given to manner verbs has several justifications: (i) within the framework of Talmy's typology, the verb plays a fundamental role since the whole typology is based on its semantics; (ii) the verb is the only lexical category whose presence is obligatory in a clause and the investigation of the manner component is important since (iii) the verb may lexicalize only one of Manner or Path (Beavers, Levin, & Tham 2010, p. 357).

The goal of the present study is to provide a diachronic investigation of the development of the lexicon of manner verbs using a wide corpus covering the evolution of French starting from Old French texts up until modern times. In particular, we will test Slobin's (1997, 2004) hypothesis concerning the size of the manner verb inventory (types) and Schøsler's (2008) hypothesis concerning their frequency of use (tokens). Throughout the study, we will try to answer the following research questions:

- (3) Does the proportion of manner verbs decrease or increase over time from Old French to Modern French (in types)?
- (4) Does the token frequency of manner verbs decrease from Old to Modern French?

With regard to the first research question, given that Old French shows many s-framed characteristics and that the size of the lexicon of manner verbs is a typological indicator of s-framed languages (Slobin 2004; Beavers, Levin, & Tham 2010; Fanego 2012; Verkerk 2014), then Old French should show a larger number of manner verbs compared to Modern French. As a consequence, we would expect the size of the inventory to decrease over time. As for the second research question, we expect to see the same pattern for the tokens, namely a decrease in the use of manner verbs from Old French to Modern French which parallels the loss of other s-framed characteristics.

2 Methodology

2.1 Overview of the MVIC database

The analysis presented in this paper is based on a database of *Motion Verbs in Context* (MVIC), currently under development as part of the project *Lexical change in motion: Motion verbs and motion lexicalization from medieval to modern Romance* (H1) within the DFG-funded Research Unit FOR 5157 *Structuring the Input in Language Processing, Acquisition and Change* (SILPAC).¹ The database takes the form of an annotated concordance of occurrences of motion verbs extracted from several electronic corpora covering a chronological span encompassing the entire history of French. The database has two main purposes. Firstly, it enables the study of lexical change within the class of motion verbs, since it is possible to calculate the relative frequency of different lexemes in the reference corpus across time. In addition, the annotation provides information on the argument structure of every motion event, which makes it possible to examine changes from s-framed to v-framed characteristics, for example the loss of verb-particle constructions in the 15th and 16th centuries.

The current version of the MVIC database draws on two electronic corpora: the MCVF-PPCHF treebank corpus (Martineau et al. 2021; Kroch et al. 2021) for the Middle Ages and a sub-corpus of texts in the public domain included in the Frantext database² for Middle French and especially the period from the 16th to the 20th century. In total, it is based on a corpus of almost 15 million tokens, including almost 300,000 tokens of motion verbs. The analysis in this paper is limited to the subclass of intransitive motion verbs (class A in our database, see below), of which there are a total of almost 190,000 tokens.

Table 1: Composition of the MVIC database by period

Period	Number of tokens	Tokens of motion verbs	Tokens of class A motion verbs
before 1100	39,478 (0.27%)	1,144 (0.38%)	796 (0.42%)
1100-1199	553,659 (3.73%)	13,607 (4.54%)	9,177 (4.84%)
1200-1299	247,848 (1.67%)	6,100 (2.04%)	4,217 (2.23%)
1300-1399	254,253 (1.71%)	6,253 (2.09%)	4,146 (2.19%)
1400-1499	606,280 (4.09%)	14,046 (4.69%)	9,553 (5.04%)
1500-1599	2,200,730 (14.83%)	44,858 (14.97%)	27,506 (14.51%)
1600-1699	3,035,376 (20.46%)	56,708 (18.93%)	34,718 (18.32%)
1700-1799	3,095,542 (20.87%)	60,009 (20.03%)	36,488 (19.25%)
1800-1899	3,097,477 (20.88%)	63,767 (21.29%)	40,929 (21.60%)
1900-1999	1,705,166 (11.49%)	33,081 (11.04%)	21,972 (11.59%)
Total	14,835,809 (100.00%)	299,573 (100.00%)	189,502 (100.00%)

2.2 Creation of the database

The creation of the database involved two main tasks. First, we created an annotated inventory of French verbs of movement throughout the history of the language using lexicographical resources.³ Second, we extracted the verbs on this list from historical corpora to build the database of motion verbs in context. We will address the details of each of these stages in turn.

2.2.1 Identification and annotation of motion verbs

Our list of motion verbs is based on data from two principal lexicographical resources: the *Dictionnaire du Moyen français* (DMF 2020), which focuses on the period from 1330-1500, and the *Trésor de la langue française informatisé* (TLFi) for the modern language. Beginning with a short list of core base motion verbs (e.g. *aller* ‘to go’, *se déplacer* ‘to move’), we used a “reverse lookup” technique to find verbs in these two dictionaries which contain a motion verb in their definition. Once all such verbs had been identified, we classified them in two stages. In the first stage, we elected whether or not to include them in the database based on the following semantic definition of a motion verb:

- (5) A motion verb encodes possible (but not obligatory) translocation of the Figure.

The notion of translocation is linked to the expression of Path, and consequently, an important diagnostic for inclusion in our database was the possible use of the verb in contexts in which a Path of motion is clearly expressed. This can be tested by adding one or more locative complements to the sentence in order to express at least two points on the Path of motion. For example, the verb *arriver* ‘to arrive’ encodes the end of a Path, but can also be combined with a locative expressing the Source (6); equally, the verb *partir* ‘to leave’ encodes the start of a Path but can be combined with a locative expressing the Goal (7):

- (6) *Elle est arrivée du Canada.* (Source)
‘She arrived from Canada’
- (7) *Il est parti pour Paris.* (Goal)
‘He left for Paris’

In the same way, verbs of putting (e.g. *mettre* ‘to put’) and verbs of separation (e.g. *quitter* ‘to leave’) can both be used to express respectively the endpoint and the startpoint of a translocation. However, they are not compatible with locatives expressing other components of the Path:

- (8) * *Elle a mis sa tasse d’un placard en un autre.*
‘She put her mug from one cupboard to another’
- (9) * *Il a quitté la salle pour le couloir.*⁴
‘He left the room for the corridor’

In (8), *mettre* is incompatible with the expression of the Source (*d’un placard*), while in (9), *quitter* is incompatible with the expression of a Goal (*pour le couloir*), which shows that these verbs do not express translocation along a Path. These verbs were therefore not included in the database.

The definition of motion reported in (5) also allows us to include in our database the manner verbs which do not obligatorily express a translocation of the Figure but are compatible with it:

- (10) *Il a sauté pendant des heures sur le trampoline.* (No translocation)
‘He jumped for hours on the trampoline’
- (11) *Il a sauté d’une chaise à l’autre.* (Translocation)
‘He jumped from one chair to the other’

While our database includes both intransitive voluntary motion verbs (class A) such as *arriver*, in which the Figure is the subject, and transitive caused motion verbs (class B) such as *jeter* ‘to throw’, in which the Figure is the object, in the present paper we focus on intransitive voluntary motion verbs only (class A). Based on their semantics, (motion) verbs can either lack the manner component (e.g., *to enter*) or encode it in their root (e.g., *to run*) and those verbs encoding Manner are claimed to be based on one, in some cases two, specific and more elementary parameters such as speed (e.g., *to run*), body motion pattern (e.g., *to walk*), the shape of the Path (e.g., *to zigzag*), purposeless (e.g., *to wander*), instrument (e.g., *to ski*), and so on (see Stosic 2009, 2019; Aurnague 2011; Moline & Stosic 2016). For this reason, Stosic (2009, 2019) has suggested that Manner should be treated as a cluster rather than a unitary category. The implications of such categorisation for the development of the expression of Manner over time are unclear and so we consider only the overarching feature of Manner in the present study, although investigating the

development of different types of manner verbs within a more fine-grained classification may provide new insights in future work. It follows that our class of intransitive voluntary motion verbs was further subdivided into two classes depending on whether they encode the Manner of motion in the verb root (class A2) or not (class A1). In total, the verb list contains 121 A1 verbs and 127 A2 verbs.

2.2.2 *Extraction and annotation of the data*

Once the list of motion verbs was established, we extracted all tokens of these verbs from the MCVF-PPCHF and a subcorpus of public domain texts in Frantext. As the MCVF-PPCHF is not lemmatized, we created CorpusSearch queries for each verb using the list of forms for each verb in question provided for the LGeRM lemmatizer.⁵ The Penn-treebank-format parsed files containing the results were converted into a tabular concordance using the Concordance Manager application,⁶ which were manually corrected to ensure that each verb was correctly lemmatized.

For the Frantext texts, we first downloaded the source XML files from the Frantext website and created an offline corpus using TXM (Heiden et al. 2010). On the grounds of feasibility, we defined a 13.2 million word subcorpus, focusing on travel literature, letters, journals and narratives, assuming that these genres would contain more occurrences of motion verbs. As the automatic lemmatization of the texts is not always accurate, we extracted potential tokens of each verb using both the automatic lemma and a regular expression to identify rarer forms which were misidentified by the lemmatizer. As with the MCVF-PPCHF data, the extracted concordances were manually checked to correct the lemmatizations.

To build the final version of the database, the concordances containing the occurrences of motion verbs were enriched with a range of additional data in order to enable statistical analysis. First, we added standardized *textual metadata* (title, author, date of composition, form, text type, etc.) using the documentation of the original corpora. Second, we added *verbal metadata* (verb class, possible argument structures) from the verb list created from lexicographical resources. Third, we added *annotation of the argument structure* for each occurrence (reflexive, presence of direct object, presence of locative complements). For the MCVF-PPCHF, this was based on the syntactic annotation from the original treebank while for the Frantext text, the extracted data was parsed using the HOPS dependency parser (Grobol and Crabbé 2021) and a modern French language model built from the SEQUOIA corpus (Grobol et al. 2022). Although the analysis in this paper does not require the annotated argument structure, it will form an important tool in future research.

3 Results

This section will present the results of our analysis in two parts. First, we focus on the token frequency of manner verbs with respect to non-manner verbs in the MVIC. Second, we will present a method for estimating the size of the manner verb inventory across time from the data in the MVIC and will report the results. All the analyses were run using R Studio (Posit Team 2023). All the models were fitted including date of composition of the text and the type and form of the text as predictors. In the models, we fitted date to check the evolution of manner verbs over time and text type and form to test Schøsler's (2008) theory that the use of manner verbs might depend on the nature of the text. Since type and form are often related (e.g., correspondence is always in prose and poetry is always in verse in our corpus), we grouped these factors into a single predictor (type.form). This combined variable has several levels: *correspondance prose* 'prose correspondence', *didactique prose* 'prose didactic text', *didactique vers* 'verse didactic text', *poésie vers* 'verse poetry', *récit prose* 'prose narrative', *récit vers* 'verse narrative', *récit mixte* 'mixed narrative', *théâtre prose* 'theatre in prose', *théâtre vers* 'theatre in verse'. In all the statistical models, *récit prose* was set as reference level for comparison since most of our data is included within this category.

3.1 Token frequency of manner verbs

The research question reported in (4) hypothesizes that Old French would make more use of manner verbs in expressing motion events than Modern French, as it is more s-framed. Figure 1 shows date on the x-axis and the proportion of non-manner verbs (A1) vs. manner verbs (A2) on the y-axis. The width of the columns of the spine plot indicates the amount of data, in our case the number of motion events that we have for each period of the language.

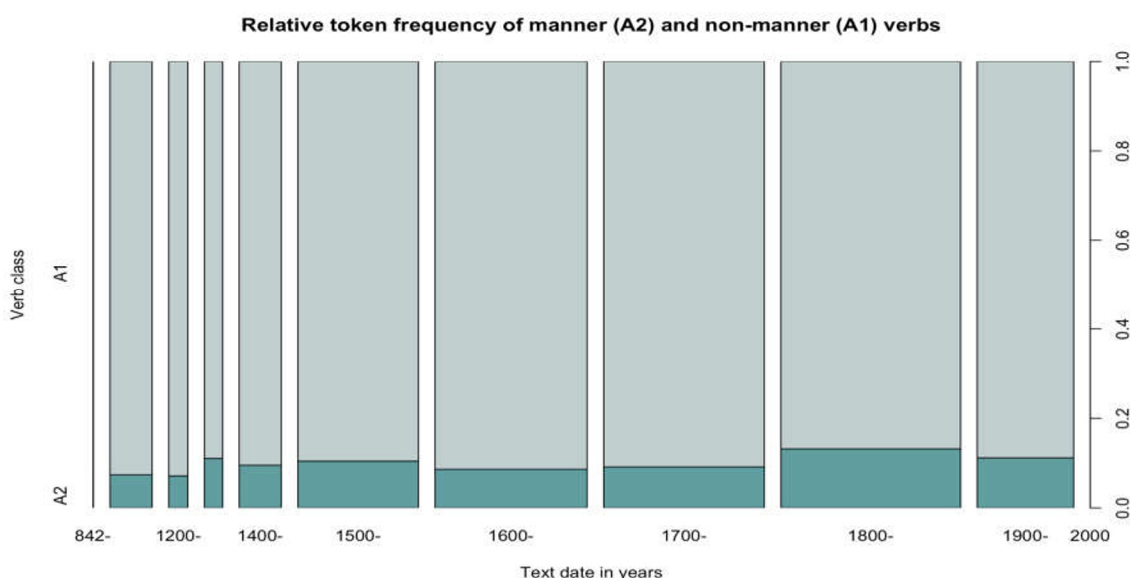


Figure 1. Spine plot depicting the proportion of non-manner (A1) vs. manner verbs (A2) used to express motion events.

Based on our predictions, we expected that manner verbs would be used less over time. The descriptive results do not appear to confirm this trend. Figure 1 shows a slight increase in the proportion of manner verbs, even if it is not uniform. To confirm this tendency, and to test Schøsler's (2008) theory that the difference in use also depends on the nature of the text, we fitted a Generalized Linear Model (*glm*). The model included verb class (A1 vs. A2) as dependent variable and date of composition of the text and type.form of the text as predictors. Considering our set of intransitive motion verbs, the model tried to predict the probabilities of using a manner verb when expressing a motion event according to the predictors that we fitted, namely date and type.form of the text. The results of the model show that the probability of using a manner verb changes according to the date and the type.form of the text. Moreover, we also found significant interactions between type.form and date indicating that, other than an effect of type.form and date alone, the number of manner verbs varies according to a certain type.form and at a given time. Table 2 reports only the significant results of the model.⁷ Since Table 2 reports the output of a *glm*, the coefficient estimates are expressed in log odds. Probabilities of less than 0.5 are represented by an overall negative value and probabilities greater than 0.5 by a positive value. The intercept refers to our reference level (i.e., *récit prose*) when date is at 0; since date was normalized using *z*-scores, date at 0 corresponds to the mean date of the corpus, i.e. 1679. Positive coefficients show that the use of a manner verb is more probable and negative coefficients show that it is less probable than at the reference level. Interactions with date show whether rate at which the probability of using a manner verb changes over time is significantly greater than or less than at the reference level. The table also provides the standard error (SE) for each estimated coefficient, which corresponds to the variability or uncertainty that is associated with the estimates of the model; and the *z*-value obtained by dividing the estimates by their standard errors. Both the SE and the *z*-value are used to compute the *p*-values for each predictor's estimate. Based on the Null and Residual

variance, we can see that the model only explains a small portion of the variance due to the high variability in our data. Nevertheless, the model is significant overall ($p < .001^{***}$).

Table 2. Significant coefficients of the generalized linear model with verb class as dependent variable and date and text type.form as predictors. Dependent variable: probability of using a manner verbs (A2).

<i>Predictors</i>	<i>Estimates</i>	<i>SE</i>	<i>z value</i>	<i>P</i>
(Intercept)	-2.14	0.01	-233.88	<.001***
Date (z-score)	0.18	0.01	-13.72	<.001***
Correspondance.prose	-0.45	0.03	-13.72	<.001***
Didactique.prose	0.12	0.05	2.47	.01*
Didactique.vers	-1.04	0.44	-2.34	.02*
Poésie.vers	0.70	0.07	10.30	<.001***
Théâtre.prose	-0.31	0.06	-5.42	<.001***
Date : Correspondance.prose	0.20	0.07	2.78	.005**
Date : Didactique.vers	-0.59	0.20	-2.96	.003**
Date : Récit.mixte	-0.27	0.11	-2.59	.01*
Date : Récit.vers	-0.10	0.04	-2.52	.01*
Date : Théâtre.prose	-0.23	0.08	-2.93	.004**
Null Deviance	124883 on 188524 (degrees of freedom)			
Residual Deviance	124058 on 188503 (degrees of freedom)			
AIC	124102			
F-statistics	39.274		p-value <.001	

* $p < .05$ ** $p < .001$ *** $p < .0001$

To visualize the effects, Figure 2 plots the raw data together with fitted regression lines for each type.form. Date is on the x-axis and the proportion of A1 vs. A2 verbs is on the y-axis. Each dot of the plot represents the proportion of tokens of manner verbs at that particular date for each text type.form.

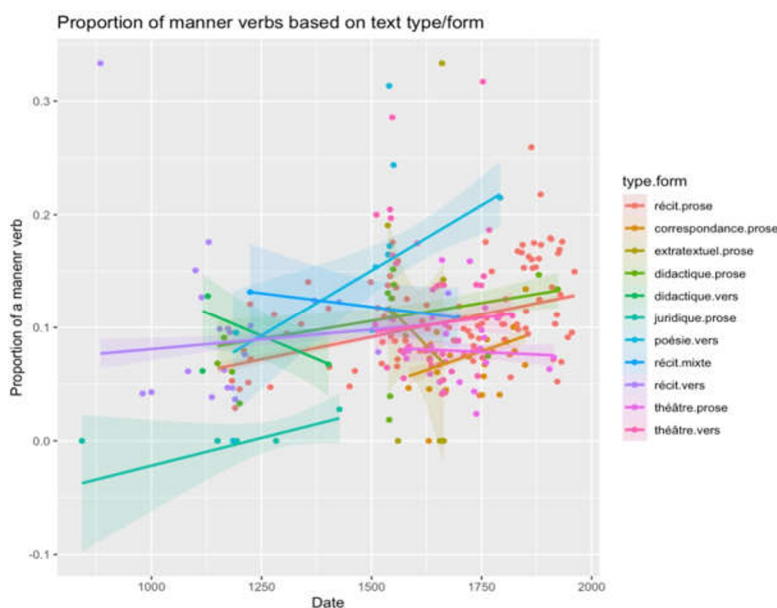


Figure 2. Plot depicting the proportion of manner verbs (A2) for each text type.form over time.

Considering the regression analysis and the plot together, we see that the probability of using a manner verb increases significantly over time at the reference level (*récit prose*). If type.form is *correspondance prose*, the overall probability of using a manner verb is lower, but the increase over time is greater, as we can see from the lower but steeper line in the graph. The estimate for *didactique prose* shows that, for this type.form and independent of date, the probability of using a manner verb is slightly higher compared to *récit prose*, as shown also by the higher line in the plot. The same is valid for *poésie vers*. In the case of *didactique vers* and *théâtre prose*, the probability of using a manner verb is lower overall and actually decreases over time. The effect of *récit mixte* and *récit vers* is only significant in interaction with date meaning that the probability of using manner verbs in these kinds of texts increases less over time compared to the reference level, as we can also see from the graph.

3.2 Size of the manner verb inventory

Calculating the size of the manner verb inventory was less straightforward, as providing an answer to a question such as ‘how many manner verbs existed in the French language in 1760?’ on the basis of a limited corpus is unrealistic. Instead, we opted to calculate the number of types of manner verbs found in samples of 100 motion events and express it as a proportion of the total number of types of motion verbs found in the sample. For example, if a sample of 100 motion events contained 5 different manner verbs and 15 different non-manner verbs, the proportion of manner verbs would be 0.25. If, following Slobin’s theory of manner saliency, the overall number of manner verbs in the language decreases over time as French becomes more v-framed, we predict that the proportion of manner verb types, as opposed to non-manner verb types in our samples, would also decrease.

A feature of this method is that the sample size must remain constant. Larger samples will clearly contain a larger number of verb types; moreover, since the most frequent motion verbs are non-manner verbs (e.g. *aller* and *venir*) and will appear in samples of any size, the proportion of manner verb types, which are generally less frequent, will naturally increase as the sample size increases, therefore distorting the results. We therefore developed a measure that takes inspiration from the Moving Average Type Token Ratio (MATTR) (Covington & McFall 2010). The MATTR is a measure of lexical diversity that is independent of text length. The measure implies the computation of a window that moves along the text and computes

a Type Token Ratio (TTR) at each movement of the window until the end of the text when it computes the average of all the TTRs to give back a lexical diversity measure for each text.

For our measure, we set the size of the window at 100 motion verb occurrences and computed the ratio of manner and non-manner verb types in each sample of 100 verbs. The window then moves forward and the same calculation is repeated. After the computation of the type ratio for each window, a mean value is calculated for each text to obtain a measure that computes the *Moving Average Manner Path Ratio* (MAMPR)⁸. Due to the fixed window size, we include only those texts in our database which contain at least 100 motion verb occurrences.

To test Slobin’s (2004) hypothesis, we used the MAMPR as dependent variable of a Linear Model (*lm*) where date and type.form of the texts were used as predictors. The first model included an interaction between date and type.form. We then conducted an analysis of variance (ANOVA) type III to assess the significance of the predictors of our model. The analysis showed that the interaction was not significant ($F = 1.03, p = .42$) meaning that date and type.form are independent of each other. For this reason, we removed the interaction and fitted both predictors as fixed effects. The results of ANOVA type III showed that both date ($F = 38.85, p < .001^{***}$) and type.form ($F = 5.72, p < .001^{***}$) are significant.

Table 3 reports only the significant results of the model with *récit prose* as reference level. Again, the intercept refers to our reference level (i.e., *récit prose*) when date is at 0. The table provides the standard error (SE) for each estimate and the t-value obtained by dividing the estimates by their standard errors. Based on the Adjusted R squared, we can see that the model only explains 21% of the variance in our data. As for the tokens, this is due to the high variability in our data. Nevertheless, the model is overall significant ($p < .001^{***}$).

Table 3. Significant coefficients of the linear model with MAMPR as dependent variable and date and text type.form as predictors.

Predictors	Estimates	SE	t	P
(Intercept)	0.21	0.00	47.05	<.001***
Date (z-score)	0.03	0.00	6.23	<.001***
Correspondance.prose	-0.04	0.01	-3.76	<.001***
Poésie.vers	0.07	0.02	3.06	.003**
Récit.vers	0.04	0.02	2.38	.02*
Théâtre.prose	-0.04	0.01	-3.06	.003**
Residual standard error	0.04812 on 211 (degrees of freedom)			
Multiple R-squared	0.2432, Adjusted R squared 0.211			
F- statistics	7.535 on 9 and 211 DF			

p-value <.001
 * $p < .05$ ** $p < .001$ *** $p < .0001$

To visualize the direction of the effects, Figure 3 plots the raw data together with fitted regression lines. Date is on the x-axis and MAMPR on the y-axis. The data are plotted based on the text type.form, in this case each dot represents MAMPR for each text at that particular date and the colour represents the type.form.



Figure 3. Plot depicting the average ratio of manner verb types for each text type.form over time.

The results of the model show that the predicted value for MAMPR at the reference level increases over time. Moreover, although the plot shows estimated regression lines with different slopes for each text type, the model shows that there is in fact no significant interaction between type.form and date, and so there is no evidence that the MAMPR changed differently over time in different text types and forms. With regard to the main effects, if type.form is *correspondance prose* or *théâtre prose*, the predicted values of MAMPR are slightly lower overall than *récit prose*, meaning that the inventory of motion verbs contains proportionally fewer manner verbs for these texts as shown by the lower corresponding lines in the graph. The opposite is true for *poésie vers* and *récit vers*, as for these kinds of texts the model predicts higher values of MAMPR compared to *récit prose* which are mirrored by the higher lines in the plot, meaning that the inventory of motion verbs for these type.form contains proportionally more manner verbs.

Although the model only explains a small proportion of the overall variance, the results nevertheless suggest that the size of the inventory of manner verbs increases over time and that text type and form play a role in defining manner saliency, since texts of a certain type.form (i.e., *poésie vers* and *récit vers*) show a wider use of manner verbs than others. In this case, the confidence intervals (CIs), represented by the coloured shadows, are quite large meaning that the slope coefficients cannot be estimated with much confidence and, as a consequence, the degree of certainty that the other slopes significantly differ from the slope of the reference level is not high. This is also confirmed by the regression analysis which did not find any significant interactions between the two predictors. Nevertheless, just as for the relative token frequencies above, we see no evidence of a decrease in manner saliency from the type-based MAMPR measure either.

3.3 What about lexical diversity of motion verbs in French?

As shown in the previous sections, neither the analysis of the tokens nor the analysis of the types showed a decrease in manner saliency for French. This conclusion was unexpected and needed further exploration. For this reason, we adopted the same technique of the moving window (described above) to check whether the overall size of the motion verb inventory increased. Therefore, we formulated an additional research question:

- (12) Does the lexicon of French intransitive motion verbs become more diversified?

To test this research question, we developed a measure of lexical diversity for motion verbs that calculates within every moving window the type-token ratio of motion verbs for that window and then computes the average of all the type-token ratios computed for each text (*Moving Average of Motion Type Token Ratio* or MAMTTR). We then used the MAMTTR to fit an *lm* with date and type.form as predictors. The first model included an interaction between date and type.form. We then conducted an analysis of variance (ANOVA) type III to assess the significance of the predictors of our model. The analysis showed that the interaction was not significant ($F = 1.28, p = .26$) meaning that date and type.form are independent from each other. We then fitted a model with date and type.form as fixed effects. The results of the ANOVA type III showed that both date ($F = 84.17, p < .001^{***}$) and type.form ($F = 7.53, p < .001^{***}$) are significant predictors.

Table 4 reports only the significant results of the model with *récit* as reference level. Table 4 reports predicted values of MAMTTR: again, the intercept refers to our reference level (i.e., *récit prose*) when date is at 0. The table provides the standard error (SE) for each estimate and the t-value. Based on the Adjusted R squared, we can see that the model explains 39% of the variance in our data. 61% of the variance in our data remains unexplained. Nevertheless, the model is significant overall ($p < .001^{***}$).

Table 4. Significant results of the linear model with MAMTTR as dependent variable and date and text type as predictors.

Predictors	Estimates	SE	t	P
(Intercept)	0.29	0.00	77.19	<.001***
Date (z-score)	0.03	0.00	9.17	<.001***
Correspondance.prose	-0.04	0.01	-4.43	<.001***
Poésie.vers	0.07	0.02	3.23	.001**
Théâtre.prose	-0.06	0.01	-5.52	<.001***
Residual standard error	0.04028 on 211 (degrees of freedom)			
Multiple R-squared	0.4244, Adjusted R squared 0.3999			
F- statistics	17.29 on 9 and 211 DF			

p-value <.001
 * $p < .05$ ** $p < .001$ *** $p < .0001$

To visualize the direction of the effects, Figure 4 plots the raw data together with fitted regression lines. Date is on the x-axis and MAMTTR on the y-axis. The data are plotted based on the text type, in this case each dot represents the MAMTTR for each text at that particular date and the colour represents the text type.

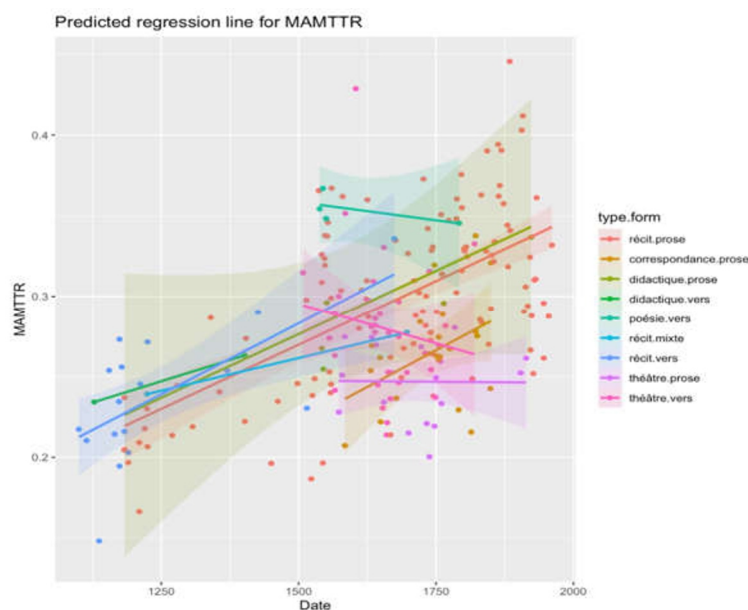


Figure 4. Plot depicting the lexical diversity of motion of each text coded for each text type over time.

The results of the model show that when date increases the predicted value for MAMPR also increases. For *correspondance prose* or *théâtre prose*, the model predicts lower values of MAMTTR compared to *récit prose*, meaning that the inventory of motion verbs is less rich for those texts, irrespective of manner or non-manner verbs, as shown by the lower lines in the plot. The opposite is true for *poésie vers* for which the model predicts higher values of MAMTTR, for this reason, the corresponding line of the plot is higher, meaning a richer lexical diversity of motion.

The results of the model and the plot show that the lexical diversity of the intransitive motion verbs in French increases over time. As was the case for the MAMPR measure, the type and form of the texts play a role in defining motion lexical diversity, for instance, certain kinds of text show a richer vocabulary of motion than others (e.g., *poésie vers*). In this case as well, the CIs are quite large meaning that the slope coefficients cannot be estimated with much confidence and, consequently, this leads to uncertainty in defining whether the other slopes significantly differ from the slope of the reference level. The large CIs explain why the regression analysis did not find any significant interactions between the two predictors. Nevertheless, the plot shows a clear increasing trend. This increase in lexical diversity of motion might have contributed to the increasing use and increasing size of manner verb lexicon since, if the lexicon is more diversified, the chances of using and introducing more infrequent verbs, such as manner verbs, increase.

4 Discussion

Within the Talmyan typological framework, Old French has been considered to be s-framed (Troberg & Burnett 2017) or “weak” v-framed (Troberg & Leung 2021) due to the presence of s-framed features. What also distinguishes s-framed and v-framed languages is the higher proportion and the higher token frequency of manner verbs in s-framed languages (Slobin 1997, 2004; Schösler 2008). So far, no study has investigated the manner verb lexicon in the context of the evolution of French. For this reason, we wanted to test whether the proportion and the use of manner verbs would decrease over the history of French assuming that Old French, which has been shown to display several s-framed characteristics, would also show a higher proportion and use of manner verbs than the more v-framed modern stages of the language. We did so by using the MVIC database that contains motion occurrences starting from the Old French period up until modern times.

The results of our analysis do not support any of the hypotheses formulated. First, the analysis in section 3.1 shows that the probability of using a manner verb rather than a non-manner verb to encode a single motion event actually *increases* over the history of French in most text types, with the possible exception of theatre in prose, mixed-form narrative and didactic verse, although since our corpus does not contain particularly extensive samples of the latter of these two types, this would need to be verified for a wider sample. Regardless of change over time, the probability of using a manner verb is highest in poetry and lowest in didactic verse and correspondence. This perhaps indicates that manner verbs are more frequently used in more literary text styles. Second, when we modelled changes in the overall size of the manner verb inventory using the MAMPR measure in section 3.2, the same pattern emerged: within samples of 100 motion events, later texts contain a higher proportion of manner verb types as opposed to non-manner verb types. Poetry shows the highest proportion of manner verb types while correspondence shows the lowest, once again suggesting that more literary genres favour manner verbs. In section 3.3, using the MAMTTR we saw that the overall lexical diversity of our samples of 100 motion events also increased over time. Later texts show a greater variety of motion verbs of all kinds and, once again, the diversity is greatest in poetry and lowest in correspondence.

Our data clearly show an increase over time in the proportion of manner verbs as opposed to non-manner verbs used, both with regard to number of tokens and the number of types. If Old French was to be treated as a new s-framed stage of the language as suggested by Troberg and Burnett (2017), then we should have seen a change pointing to a decrease over time with Modern French showing a smaller manner verb lexicon and less frequent use of manner verbs. In typological terms, our findings seem to favour the classification of Old French proposed by Troberg and Leung (2021), namely as a “weak” v-framed language. In fact, this classification allows us not to ignore the other s-framed characteristics (e.g., particles, adjectival resultative constructions) that are present in Old French but, at the same time, it fits better with the patterns observed in our results. In addition, similar findings, namely a rather restricted use of manner verbs, were reported by Iacobini and Corona (2016) who examined motion encoding in Caesar’s *De bello gallico* and Ovid’s *Metamorphoses*. Their results suggest that Classical Latin already showed v-framed tendencies that then evolved into the v-framed Romance languages. These v-framed tendencies found in Latin may help to explain our findings with regard to manner saliency in the French lexicon of motion verbs.

In addition, our findings support Schøsler’s (2008) theory that the nature of the text might play a role in the use of manner verbs. Based on our data and our (albeit simple) classification of type and form of the texts, we can confirm that type and form not only play a role in the frequency of use of manner verbs, as already suggested by Schøsler, but also in the size of the inventory of manner verbs and the lexical diversity (i.e., richness) of the motion lexicon in French. In particular, poetry appears to show the highest relative frequency of use of manner verbs as well as the richest vocabulary of intransitive motion verbs, while text types like correspondence and theatre in prose show the lowest proportion of manner verbs and the less diversified motion vocabulary. These differences might be explained by differences in style. For instance, poetry is known to be a literary type in which every word is carefully chosen to create specific stylistic effects and/or to observe metrical constraints; this results in the use of more infrequent, high-register words. This stylistic feature could potentially account for the greater size and use of manner verbs in poetry, being manner verbs usually infrequent in v-framed languages, compared to correspondence and theatre in prose which, among the text types, are typically closer to the spoken language favouring less complex stylistic choices and the use of more frequent words. However, this hypothesis would need to be supported by a more detailed classification of the texts.

To conclude, our study is methodologically innovative. It is the first study that investigates changes in the lexicon of motion verbs in French and it does so by quantitatively analysing a very large dataset. Working with such an amount of data allows us to have a pretty good picture of the different stages of French and draw more robust conclusions from the patterns observed in the data. Some topics remain open, such as determining the precise role of each text type and form, accounting for differences in style and register used by certain texts or certain authors or qualitatively looking at the lexicon for each stage of French separately to account for the change. We leave these topics to be addressed in future research, ideally by using an equal or even greater amount of data than the one used in this study. Nevertheless, the present study

contributes to shed a new light on the discussion about the typological trajectory and status of French throughout its history.

References

- Acedo-Matellán, V. (2006). Prefixes in Latin and Romance and the satellite- vs. verb-framed distinction. In Universitat de Barcelona (ed.), *Actes del VII Congrés de Lingüística General*, CD-ROM, Barcelona: Universitat de Barcelona.
- Acedo-Matellán, V. (2010). *Argument Structure and the Syntax-Morphology Interface. A Case Study in Latin and other Languages*. Barcelona: Universitat de Barcelona [PhD thesis]. Downloadable at www.tdx.cat/bitstream/10803/21788/1/VAM_PhD_THESIS.pdf (accessed 8 November 2012).
- Acedo-Matellán, V., & Mateu, J. (2008). The path from satellite-framed Indo-European to verb-framed Romance: A lexical-syntactic account. Paper presented at the *10th Diachronic Issues in Generative Syntax conference (DIGS X)*, Cornell University, Ithaca, 7-9 August.
- Acedo-Matellán, V., & Mateu, J. (2013). Satellite-Framed Latin vs. Verb-Framed Romance: A Syntactic Approach. *Probus*, 25, 227-265.
- Aurnague, M. 2011. How motion verbs are spatial: The spatial foundations of intransitive motion verbs in French. *Linguisticae Investigationes*, 34(1), 1-34.
- Beavers, J., Levin, B., & Tham, W. S. (2010). The typology of motion expressions revisited. *Journal of Linguistics*, 46(2), 331-377.
- Berman, R. A., & Slobin, D. I. (1994). *Relating Events in Narrative: A Crosslinguistic Developmental Study*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Buridant, C. (1987). Les particules séparées en ancien français. In C. Buridant (ed.) *Romanistique-Germanistique: une confrontation*. Strasbourg: Presses universitaires de Strasbourg, 167-204.
- Buridant, C. (2000). *Grammaire nouvelle de l'ancien français*. Paris: Sedes.
- Burnett, H. & Tremblay, M. (2009). Variable-Behavior Ps and the Location of PATH in Old French. In E. O. Aboh, E. van der Linden, J. Quer & P. Sleeman (eds.) *Romance Languages and Linguistic Theory: Selected papers from 'Going Romance' Amsterdam 2007*. Amsterdam & Philadelphia: John Benjamins, 25-50.
- Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian Knot: The Moving-Average Type–Token Ratio (MATTR). *Journal of Quantitative Linguistics*, 17(2), 94-100.
- DMF (2020). *Dictionnaire du Moyen Français*, version 2020. ATILF - CNRS & Université de Lorraine. <<http://www.atilf.fr/dmf>> [accessed 20 December 2023].
- Dufresne, M., Dupuis, F. & Longtin, C. (2001). Un changement dans la diachronie du français : la perte de la préfixation aspectuelle en -a. *Revue Québécoise de Linguistique*, 30, 33-54.
- Dufresne, M., Dupuis, F., & Tremblay, M. (2000). The role of syntactic features in historical change. In S. N. Dworkin & D. Wanner (eds.), *New Approaches to Old Problems. Issues in Romance Linguistics*. Amsterdam: John Benjamins, 129-149.
- Dufresne, M., Dupuis, F., & Tremblay, M. (2003). Preverbs and particles in Old French. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 2003*. Dordrecht: Kluwer, 33-60.
- Fanego, T. (2012). Motion events in English: The emergence and diachrony of manner salience from Old English to Late Modern English. *Folia Linguistica Historica*, 46 (Historica-vol-33), 29-85.
- Grobol, L. & Crabbé, B. (2021). Analyse en dépendances du français avec des plongements contextualisés. In *Actes de la 28ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN-RÉCITAL 2021)*, Lille, France. < <https://hal.archives-ouvertes.fr/hal-03223424>> [accessed 20 December 2023].
- Grobol, L., Regnault, M., Ortiz Suarez, P., Sagot, B., Romary, L. & Crabbé, B. (2022). BERTrade: Using Contextual Embeddings to Parse Old French. In *13th Language Resources and Evaluation Conference*,

- European Language Resources Association*, Marseille, France. <<https://hal.science/hal-03736840>> [accessed 20 December 2023].
- Heiden, S., Magué, J.-P., Pincemin, B. (2010). TXM: Une plateforme logicielle open-source pour la textométrie - conception et développement. In S. Bolasco, I. Chiari & L. Giuliano (eds) *10th International Conference on the Statistical Analysis of Textual Data - JADT 2010*, Jun 2010, Rome, Italy. *Edizioni Universitarie di Lettere Economia Diritto*, 2 (3), 1021-1032 <halshs-00549779> [accessed 20 December 2023].
- Iacobini, C., & Fagard, B. (2011). A diachronic approach to variation and change in the typology of motion event expression. *Cahiers de Faits de langue*, 3, 151-171.
- Iacobini, C. (2012). Grammaticalization and innovation in the encoding of motion events. *Folia Linguistica*, 46, 359-386.
- Iacobini, C. & Corona, L. (2016). “Romanes eunt domus”: where you can go with Latin morphology. Variation in motion expression between system and usage. In *MMM 10 Online Proceedings*, 73-87. <<https://doi.org/10.26220/mmm.2726>> [accessed 2 April 2024].
- Kopecka, A. (2009). Continuity and change in the representation of motion events in French. In J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura, & S. Özçalışkan (eds.), *Crosslinguistic Approaches to the Psychology of Language: Research in the Tradition of Dan Isaac Slobin*. New York: Psychology Press, 415-426.
- Kroch, A., & Santorini, B. (eds.) (2021). *Penn-BFM Parsed Corpus of Historical French, version 1.0*. <<https://github.com/beatrice57/mcvf-plus-ppchf>> [accessed 20 December 2023].
- Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. (2017). *lmerTest* Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1-26.
- Moline, E., & Stosic, D. 2016. *L'expression de la manière en français*. Paris: Ophrys.
- Posit team (2023). *RStudio: Integrated Development Environment for R*. Posit Software, PBC, Boston, MA. URL <http://www.posit.co/>
- Schøsler, L. (2008). L'expression des traits manière et direction des verbes de mouvement : Perspectives diachroniques et typologiques. In E. Stark, R. Schmidt, & E. Stoll (eds.), *Romanische Syntax im Wandel: [Festgabe zum 65. Geburtstag von Wulf Oesterreicher]*. Tübingen: Narr Francke Attempo Verlag, 113-132.
- Slobin, D. I. (1997). Mind, Code and Text. In J. L. Bybee, J. Haiman, & S. A. Thompson (eds.), *Essays on Language Function and Language Type: Dedicated to T. Givón*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 437-467.
- Slobin, D. I. (2004). The many ways to search for a frog: Linguistic typology and the expression of motion events. In S. Strömquist & L. Verhoeven (eds.), *Relating Events in Narrative: Vol. 2. Typological and Contextual Perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates, 219-257.
- Stolova, N. (2015). *Cognitive Linguistics and Lexical Change: Motion Verbs from Latin to Romance*. Amsterdam: John Benjamins.
- Stosic, D. (2009). La notion de « manière » dans la sémantique de l'espace. *Langages*, 175, 103-121.
- Stosic, D. (2019). Manner as a cluster concept: What does lexical coding of manner of motion tell us about manner?. In M. Aurnague & D. Stosic (eds), *The Semantics of Dynamic Space in French Descriptive, Experimental and Formal Studies on Motion Expression*. Amsterdam: John Benjamins, 141-178.
- Talmy, L. (1972). *Semantic Structures in English and Atsugewi*. [PhD Dissertation] Department of Linguistics, University of California, Berkeley.
- Talmy, L. (1985). *Lexicalization Patterns: Semantic Structure in Lexical Forms*. In T. Shopen (ed.) *Language Typology and Syntactic Description, vol. III: Grammatical Categories and the Lexicon*. Cambridge: Cambridge University Press, 57-149.
- Talmy, L. (1991). Path to realization: a typology of event integration. *Buffalo Working Papers in Linguistics*, 91(1), 147-187.
- Talmy, L. (2000). *Toward a Cognitive Semantics*. Cambridge, Mass.: MIT Press.

- Troberg, M., & Burnett, H. (2017). From Latin to Modern French: A Punctuated Shift. In E. Mathieu & R. Truswell (eds.), *Micro-Change and Macro-Change in Diachronic Syntax*. Oxford: Oxford University Press.
- Verkerk, A. (2014). *The Evolutionary Dynamics of Motion Event Encoding*. [PhD Dissertation] Max Planck Institute for Psycholinguistics.

¹ <https://silpac.uni-mannheim.de/> [accessed 2 April 2024]

² <https://www.frantext.fr/> [accessed 2 April 2024]

³ The *DinaVmov* database (Stosic et al. 2023) was released after the data for this paper had been prepared and we intend to incorporate it into future versions of our database.

⁴ Note that there are cases in which the verb *quitter* is combined with a Goal complement building grammatical sentences like: “*Il a quitté la campagne pour Paris*” (‘He left the countryside for Paris’). Nevertheless, in such constructions where the Source and the Goal are expressed, the meaning of the verb *quitter* seems to indicate more of a change of state in the sense of “not living there anymore, change of residence, moving out” rather than a movement encoding Path.

⁵ Distributed by ATILF, <http://www.atilf.fr/LGeRM> [consulted 20 December 2023]

⁶ <https://github.com/rainsfordtm/conman> [accessed 2 April 2024]

⁷ The tables showing the results of the statistical models are based on the output of the *summary()* function in R (Kuznetsova, Brockhoff & Christensen 2017).

⁸ For simplicity, in this case we refer as “path verbs” the verbs that do not encode Manner.