

## LANGUAGE IN TIME: ASSESSING MEDIEVAL FRENCH REGISTERS IN A QUEST FOR ACCURACY IN HISTORICAL LINGUISTICS

Historical linguistics can resort to neither recordings nor interviews of native speakers. The only type of material that can be dissected is of a textual nature. This leads to two main issues: first, it has been widely observed in the literature that the choice of a certain register over another shapes the results of the investigation<sup>1</sup>. For instance, the syntax of a poem will offer different perspectives on a language than the syntax of a recipe. Second, and this is particularly relevant to French amongst other languages, the written language often retains archaisms that are not found in the oral tongue. However, building a corpus with non-literary material may lead to consistent findings in diachronic studies. More precisely, the use of legal material is put forward here. This choice stems from other studies investigating different languages and it is discussed here within the French historical context. The present article comes from a doctoral project that recently sought to create a digitalised corpus of unexplored texts in order to investigate the evolution of pronouns in the diachrony of French and subsequently produce a quantitative analysis within the framework of generative linguistics<sup>2</sup>.

Medieval France has produced a sizeable array of texts that allows researchers to carry out studies on the language history. The productivity of medieval scribes can be illustrated with the online corpus *Frantext*<sup>3</sup>, in which twenty-nine texts of considerable lengths are available for the sole period of 1100 to 1199. Unlike other languages that do not enjoy a long tradition of writing, the tangible linguistic heritage French has at its disposal is significant enough to study its evolution. There is no particular obstacle to accessing texts from the Old French period<sup>4</sup>. Nonetheless, the choice of register will greatly influence the findings.

Selecting literary material might indeed not accurately represent a language: effects of register have been shown conclusively for a number of languages, as discussed below. In her discussion of the loss of V2<sup>5</sup> and pro-drop<sup>6</sup> in Medieval French, Adams points out the limits of her study, as literary writers from the fourteenth and fifteenth centuries intertwine the old system, i.e. the norm from the previous centuries, with their contemporary one<sup>7</sup>. This produces occurrences of archaic forms. I also argue that the literary language is dependent on a series of factors that creates a significant distortion not suitable for the study of the evolution of the language.

In the last two decades, a new trend has strengthened in historical linguistics with the creation of non-literary corpora. As pointed out by Balon and Larrivé in their study of the loss of pro-drop in the diachrony of French, literary texts are highly dependent on style and genre<sup>8</sup>. A similar observation is made in the work of Kytö, who discusses the creation of corpora in empirical studies: according to her, data primarily represent the register and genres of a text<sup>9</sup>. She points out that literary genres have a significant impact on the results of a corpus study (e.g. verse texts do not use the same language as prose). Similarly, in her study of Old Russian, Le Feuvre uses data extracted from birch bark manuscripts found in Novgorod, Russia<sup>10</sup>. The documents she uses are

non-literary, which allows her to separate literary constructions from non-literary ones and to contrast her findings with those from earlier research based on chronicles. Along the same lines, Stolk uses non-literary papyri to study case variation in post-Classical Greek, on the grounds that they do not undergo linguistic normalisation and modernisation as literature does<sup>11</sup>. Back to Medieval Romance, the purpose of using non-literary texts can also be illustrated in the study of Díez Del Corral Areta who, in order to counter the influence of stylisation in Medieval Spanish, uses a corpus of witness declarations and letters with the aim of studying the *inmediatez comunicativa*, the ‘immediacy of communication’<sup>12</sup>. The benefit of her selection is that the language studied is closer to what the vernacular was at the time it was set down on paper. The idea that literature is somewhat constructed is also highlighted in Balon and Larrivé’s study of French, where they argue that, unlike the vernacular which is the first linguistic form acquired, the *pratique normée*, ‘standardised practice’, consists of constructions that do not necessarily exist in the spoken language (e.g. the double negation *ne ... pas* is a construction that remains valid in written French, whereas in spoken French, *pas* alone suffices)<sup>13</sup>. They apply this idea to historical linguistics and suppose that the gap between colloquial French and the written norm was already present during the Middle Ages. Indeed, they suggest that the loss of pro-drop is anterior in legal documents: whilst in literary documents, pro-drop becomes optional during the thirteenth century and sporadic in the fifteenth century, their findings show that, in legal documents, pro-drop is in minority during the twelfth century already and disappears during the thirteenth century. Their investigation exposes a three-hundred-year difference in the diachrony of the language, suggesting that pro-drop was not part of the oral grammar anymore when it was still regarded as the written norm. Such findings enforce the necessity to reassess the use of literary documents for historical work.

Although today’s legal material may require a certain level of literacy, the use of French in the Middle Ages was justified by the need to be understood by all, and particularly by the people lacking literacy. The *Serments de Strasbourg* in 842 are a good example of the necessity to use the vernacular language over Latin - the latter remains indeed the written language even after it is not effectively spoken anymore. This document is part of a larger book written by Nithard, one of Charlemagne’s grandsons, who narrates the oaths taken by his two cousins, also grandsons of Charlemagne. The book, *Histoire des fils de Louis le Pieux*, is written in Latin. Nithard relates the political history of his grandfather, uncle and cousins. He exposes the day Louis II the German and Charles II the Bald promised each other to form an alliance against their brother Lothair I. He ensures that, despite his Latin narration, the oath taken by Louis II the German is transcribed into *romana lingua*, i.e. the ‘roman tongue’. The contrast between Latin and the *romana lingua* illustrates that, as early as 842, the population of Northern France does not speak Latin anymore. Thus, the first French text identified is a legal document. A few more centuries will nonetheless be necessary for the legislation to abandon Latin completely.

Gradually from the mid-twelfth century, the French language is implemented in legal documents: sporadically at first, until it becomes the official language in the mid-sixteenth century

when, in 1539, François I enacts the *Ordonnance de Villers-Cotterêts*; an order that gives French its status of official language for public documents in the Kingdom of France.

Although it is not before the sixteenth century that French becomes the official language, Normandy in the Middle Ages is a dynamic region and proves to be precursory, as it sees its first legal text written in French in 1150. There is, however, and to the best of my knowledge, a gap between the *Serments de Strasbourg* in 842 and the *Leis Willelme* in 1150: during the three centuries that separate the two texts, there is no legal material available in French.

In spite of the lack of material for periods anterior to 1150, legal texts are a valuable source for studying the history of the language. As Latin has become unintelligible, the laws established by the duchy necessitate some update. From the twelfth century onwards, the vernacular language begins to infiltrate official documents. The use of *françoys*, 'French', is explicitly justified in the Norman text *Le Grant Coustumier du pays et duché de Normendie* written in 1534 by Guillaume Le Rouillé, a legal scholar. The monograph starts with the following:

Le grant Coustumier du pays et duche de Normendie trefutile et profittable a tous praticiens. Lequel est le texte diceluy en francoys proportionne a lequipollent de la glofe ordinaire et familialeire.

My translation:

The great Customs of the country and duchy of Normandy are very useful and profitable to all professionals. The aforementioned text is in French, proportional to the equivalence of the ordinary and familiar language.

The objectives of the document are clear: the language must be intelligible and understood by all, therefore the vernacular tongue shall be used. The latter is explicitly referred to as *francoys*; unsurprisingly, this is only five years before François I enacts the *Ordonnance de Villers-Cotterêts*. The relevance of using this type of material is evidenced here.

Evidently, there is room to object as for the veracity of legal French as the language of the street: there must be a degree of stylisation in this register too, as there were no native speakers of legal Medieval French. The choice of deeds over literature is motivated by the fact that the former is stylised to a much lesser degree than the latter, on the grounds that its objectives are to be efficient and straightforward.

In the last ten years, the subfield of Laws and Corpus Linguistics has gained in popularity. Examples for the English language include *Lawcorpus*<sup>14</sup>, a set of both synchronic and diachronic corpora developed by Brigham Young University; and an example of epistolary set of data, the *Corpus of Early English Correspondence*<sup>15</sup>, compiled by the University of Helsinki. For French, the undergoing creation of a legal corpus is led by the Laboratoire CRISCO at the Université de Caen, with the *Projet ConDé*<sup>16</sup>. The latter reunites legal documents from Normandy. The region of Poitou also has a corpus of legal texts, the *Actes royaux du Poitou*<sup>17</sup>, which is assembled by the École Nationale de Chartres with the support of the Consortium Sources Médiévales.

One could add other corpora to the list: the nascent interest for this type of material is motivated by the taking in consideration of the importance of intra and extra linguistic features of

a text. According to Kytö, “[m]ost researchers are aware that in empirical studies the generalizability of their results largely depends on the register(s) or genres their data represent”<sup>18</sup>. She continues her argumentation and points out that there are two possibilities to account for recurring patterns in historical linguistics: either setting up a group of texts from several registers or analysing a group of texts from the same register and sub registers. The former method expects to represent the entire language and it undoubtedly evidences general trends of development. On the other hand, the single-register method offers a more consistent glimpse of the language, as it takes into consideration factors specific to the register which may impact the results.

The analysis of a corpus of legal texts to investigate linguistic changes within Medieval French brings in a novel perspective as it seeks to grasp transitions and innovations in the structure of the language with precision. Fragments from the past anachronise the literary language and need to be eliminated. It is through lessening this obstruction that historical linguistics has entered a revolution that seeks to accurately locate the vernacular language in time.

## References

- Adams, Marianne 1987. ‘From Old French to the theory of Pro-Drop’, *Natural Language and Linguistic Theory*, 5:1, 1-32.
- Balon, Laurent and Pierre Larrivée 2016. ‘L’ancien français n’est déjà plus une langue à sujet nul – nouveau témoignage des textes légaux’, *Journal of French Language Studies*, 26 :2, 221-237.
- Diez Del Corral Areta, Elena 2011. ‘Sobre el Modo, Manera, Suerte, Forma y Arte de conectar’, *Res Diachronicae Virtual* 9, 33-50.
- Koch, Peter and Wulf Oesterreicher 1985. ‘Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte’, *Romanistisches Jahrbuch* 36, 15-43.
- Kytö, Merja 2019. ‘Register in historical linguistics’, *Register Studies*, 1:1, 136–167.
- Le Feuvre, Claire 2008. ‘Le futur périphrastique dans les écorces de boulot de Novgorod’, in *Construire le temps : études offertes à Jean-Paul Sémon*, ed. Jean Breuillard and Serge Aslanoff (Institut d’Études Slaves), pp. 381-394.
- Olivier, Marc to appear. ‘Diachronie de la proclise et de l’enclise avec l’infinitif en français médiéval (12<sup>e</sup>-15<sup>e</sup> s.)’, *Studia Linguistica Romanica*.
- Joanne Vera Stolk, ‘Case variation in Greek papyri: retracing dative case syncretism in the language of the Greek documentary papyri and ostraca from Egypt (300 BCE–800 CE)’ (unpublished doctoral thesis, University of Oslo, 2015), pp. 44-52.

---

<sup>1</sup> Peter Koch and Wulf Oesterreicher, ‘Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte’, *Romanistisches Jahrbuch* 36 (1985), pp. 15-43.

- 
- <sup>2</sup> Marc Olivier, 'Diachronie de la proclise et de l'enclise avec l'infinitif en français médiéval (12<sup>e</sup>-15<sup>e</sup> s.)', *Studia Linguistica Romanica* (to appear).
- <sup>3</sup> Frantext: <<https://www.frantext.fr/information>> [last accessed 12 October 2020].
- <sup>4</sup> The Old French period lasts from the ninth century to the mid-fourteenth century.
- <sup>5</sup> The V2 ordering is observed in languages in which the verb is systematically placed in second position, e.g. Swedish.
- <sup>6</sup> Pro-drop is observed in languages that omit the pronominal subject without altering the grammaticality of a clause, e.g. Italian or Spanish.
- <sup>7</sup> Marianne Adams, 'From Old French to the theory of Pro-Drop', *Natural Language and Linguistic Theory* 5.1 (1987), pp. 1-32.
- <sup>8</sup> Laurent Balon and Pierre Larrivée. 'L'ancien français n'est déjà plus une langue à sujet nul – nouveau témoignage des textes légaux', *Journal of French Language Studies* 26.2 (2016), pp. 221-237 (p. 222).
- <sup>9</sup> Merja Kytö, 'Register in historical linguistics', *Register Studies* 1.1 (2019), pp. 136–167 (p.137).
- <sup>10</sup> Claire Le Feuvre, 'Le futur périphrastique dans les écorces de boulot de Novgorod', in *Construire le temps : études offertes à Jean-Paul Sémon*, ed. Jean Breuillard and Serge Aslanoff (Institut d'Études Slaves) (2008), pp. 381-394.
- <sup>11</sup> Joanne Vera Stolk, 'Case variation in Greek papyri: retracing dative case syncretism in the language of the Greek documentary papyri and ostraca from Egypt (300 BCE–800 CE)' (unpublished doctoral thesis, University of Oslo, 2015), pp. 44-52.
- <sup>12</sup> Elena Diez Del Corral Areta, 'Sobre el Modo, Manera, Suerte, Forma y Arte de conectar', *Res Diachronicae Virtual* 9 (2011), pp. 33-50 (p. 40).
- <sup>13</sup> Balon and Larrivée, p. 225.
- <sup>14</sup> Lawcorpus: <<https://lawcorpus.byu.edu>> [last accessed 26 October 2020].
- <sup>15</sup> Corpus of Early English Correspondence: <<https://www.helsinki.fi/en/researchgroups/varieng/corpus-of-early-english-correspondence>> [last accessed 12 October].
- <sup>16</sup> Projet ConDÉ: <<https://conde.hypotheses.org>> [last accessed 12 October 2020].
- <sup>17</sup> Actes Royaux du Poitou: <<http://corpus.enc.sorbonne.fr/actesroyauxdupoitou/>> [last accessed 13 October 2020].
- <sup>18</sup> Kytö, p. 137.