Recherches anglaises et nord-américaines

53 | 2020
Internal Variation

A Variationist Approach to the Spread of Emergent Features in Middle English

Michael Percillier

Electronic version
URL: https://journals.openedition.org/ranam/744
DOI: 10.4000/ranam.744
ISSN: 3000-4411

Publisher
Presses universitaires de Strasbourg

Printed version
Date of publication: September 24, 2020
Number of pages: 23-36
ISBN: 9791034400690
ISSN: 0557-6989

Electronic reference
Michael Percillier, “A Variationist Approach to the Spread of Emergent Features in Middle English”, *Recherches anglaises et nord-américaines* [Online], 53 | 2020, Online since 01 February 2024, connection on 07 February 2024. URL: http://journals.openedition.org/ranam/744 ; DOI: https://doi.org/10.4000/ranam.744

The text only may be used under licence CC BY-NC-SA 4.0. All other elements (illustrations, imported files) are “All rights reserved”, unless otherwise stated.

A Variationist Approach to the Spread of Emergent Features in Middle English

Michael Percillier♦

The Middle English (ME) period saw a number of drastic changes in the English language, some of which can be attributed to language contact, whereas the cause of other developments as either language-internal or language-external remains a matter of debate. For instance, the influx of lexical material from Old French (OF) during the language-contact situation following the Norman Invasion (1066–ca 1500) is rather uncontroversial. In contrast, the question whether this incoming lexical material also gave rise to grammatical changes is more difficult to assess. Emergent features that bear similarities to a contact language, such as the rise of argument structures involving prepositional objects headed by *to*, beg the question whether they were influenced by similar constructions in the contact language, in our example French prepositional objects headed by *à*, or rather developed independently. The project *Borrowing of Argument Structure in Contact Situations* (BASICS), in which the research presented in this article is embedded, investigates if verbs of French origin entering ME carried along their argument structure, thus resulting in grammatical changes, and if so, to which extent.[1]

Contact explanations of such changes have been supported by means of a feature's diachronic spread. For example, the spread of *th-* type 3PL personal pronouns at the expense of the older *h-* type occurred from North to South, as can be determined by comparing data from the *Linguistic Atlas of Early Middle English* (LAEME, Laing, 2013) and the *Electronic Version of a Linguistic Atlas of Late Mediaeval English* (eLALME, Benskin et al., 2013). The direction of this spread has been interpreted as support for an Old Norse contact explanation

♦ Michael Percillier, *University of Mannheim*.

1 Funding from the *Deutsche Forschungsgemeinschaft* for the BASICS project (project number 265711632) is gratefully acknowledged.

(Hogg, 1992: 146; Howe, 1996: 154-157; Proffitt, 2018: "they, pron., adj. adv., and n.").

While such a diatopic spread supports a contact-induced explanation, it does so in the capacity of a necessary condition, but not a sufficient one. A case in point is made by Cole (2018), who argues for a Northumbrian dialect origin of *th*-type 3PL personal pronouns, which would also be supported by a North-South spread. In other words, a diatopic spread in the "right direction" is required to support a contact-induced explanation, but further evidence is needed in order to render such an explanation tenable. Conversely, a diatopic spread in the "wrong direction" should be enough to discredit a specific contact- or dialect-related explanation.

Within the research project at hand, this means that features purportedly attributed to contact with OF need to clear a "first hurdle" of not spreading in the unexpected direction, which in this context would be North to South. Although Norman rule included Northern England as well, features emerging due to contact with OF are expected to reach the North later than other regions. Any indication that a feature spread from North to South would definitely make a French contact explanation highly unlikely. In order to test this kind of diatopic spread, existing resources for ME have to be adapted and expanded.

The article is structured as follows: first I describe the motivation for mapping the PPCME2, and the method used for doing so. Following this, I probe the soundness of the method by testing a feature that is evidently due to contact with OF, namely French-based verbs (henceforth, FBVs), as a baseline. I then apply the method to two features, namely the verbal {en-} prefix and the *to* prepositional object construction, to investigate their diatopic spread. The implications of the findings for measuring diatopic spread to probe language contact effects are then discussed.

The need and method for mapping the PPCME2

Resources for ME such as LAEME (Laing, 2013) and eLALME (Benskin et al., 2013) already provide opportunities to track the spread of certain features. However, they employ a feature catalogue system, which means that more complex queries involving syntax, verb classes, verb etymologies, etc., cannot be performed "out of the box". Put differently, one is simply out of luck if the feature to be investigated does not happen to be already covered by LAEME and eLALME. Furthermore, LAEME and eLALME merely indicate presence versus absence of a feature at a given survey point, but not which variant may be dominant there.

A way to circumvent these limitations is to map the texts from the *Penn-Helsinki Parsed Corpus of Middle English, Second Edition* (PPCME2, Kroch

and Taylor, 2000), which contains 1.2 million words and is divided into four sub-periods: M1 (1150–1250), M2 (1250–1350), M3 (1350–1420), and M4 (1420–1500). The corpus is syntactically annotated, which allows for complex queries using the dedicated program *CorpusSearch* (Randall, 2010). Further, the annotation of lexical verbs in the corpus can be enriched with an additional layer which provides information on verb lemmas as well as verb origin, thus distinguishing French-based verbs from other verbs (Percillier, 2016; 2018).

Mapping the PPCME2 entails assigning each corpus text to a geographic location. For texts that are common to the PPCME2 and LAEME/eLALME, this information could be taken over. However, given the idiosyncratic mapping system of LAEME/eLALME, their grid system was converted to real latitude/longitude values so as to be able to generate new maps. Remaining PPCME2 texts were located based on metadata from the PPCME2 website. Textual base, i.e. whether the text is an English original or a text translated from French, is also marked as a variable due to the possibility of translation-induced contact (see Haeberli, 2018).

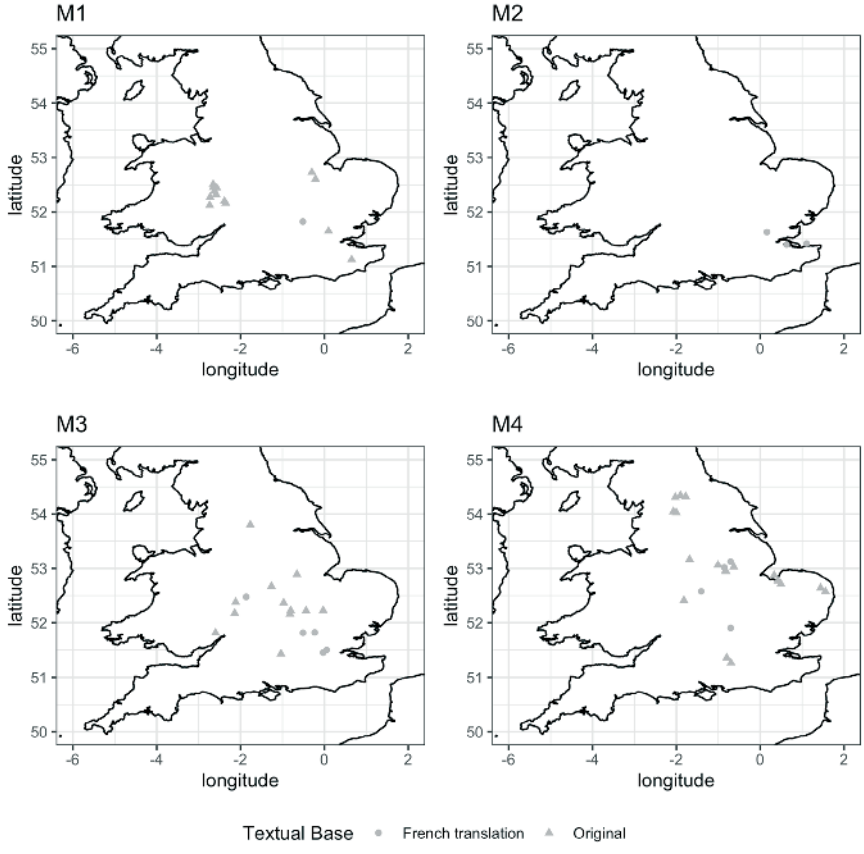


Figure 1: Survey points in the PPCME2, grouped by manuscript sub-period

The survey points thus obtained, shown in Figure 1, reveal multiple coverage gaps, the most obvious of which lies in the M2 sub-period. The presence of only three texts for this period represents an already well-known data gap (Truswell et al., 2019: 21–22). Figure 1 suggests that this gap is worse than previously thought, as M2 not only represents a smaller sample size, but is further restricted to an extremely limited geographical area, namely Kent and the South of the Eastern Midlands, and only consists of texts translated from French.

An alternative view comparing the texts' estimated manuscript dates and latitudes, shown in Figure 2, reveals the extent of the data gap in M2, where the second half of the 13th century is represented by a single text. Additionally, Northern texts only appear in the late 14th century, that is M3, and it is only in the middle of the 15th century, which corresponds to M4, that diatopic coverage is representative.

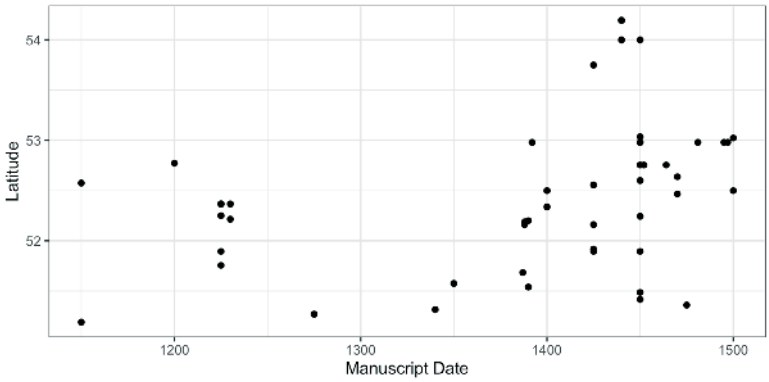


Figure 2: Distribution of texts by manuscript date and latitude in the PPCME2

The transition from M1 to M2 has been described as a period of change for multiple phenomena. To mention just two examples, Percillier (2016: 212) observes a sudden increase in FBVs from M1 to M2 with little fluctuation thereafter, and Zehentner (2018: 158) describes a sudden increase in the proportion of *to* prepositional object constructions (*to*-POC) at the expense of double object constructions (DOC).[2]

A consequence of the M2 data gap is that the studies finding M2 to be a transitional period not only base their conclusions on a small data set, but may also (unknowingly) conflate diachronic development and diatopic variation given that M2 data is limited to the Southeast of England. Further, translation effects are not controlled for given the lack of texts originally written in English in M2, in contrast to the other sub-periods.

The most promising solution to the M2 data gap lies in the adaptation of additional texts to the PPCME2's annotation format. Such projects are already in

2 See Truswell et al. (2019: 22) for a list of further studies establishing M2 as a transitional period.

an advanced stage, most notably the *The Parsed Corpus of Middle English Poetry* (PCMEP, Zimmermann, 2018) and a parsed version of LAEME, the *Parsed Linguistic Atlas of Early Middle English* (PLAEME, Truswell et al., 2018). Efforts by the author to lemmatise and map these additional resources are also well under way.

The reason for which the PPCME2 does not contain the texts included in the PCMEP and the PLAEME lies in its focus on prose, justified by the fact that rhyme and metre in verse may influence syntax. Rather than choosing between coverage gaps or the effect of rhyme and metre as undesirable effects, verse texts can be included under the condition that text type, i.e. prose versus verse, is treated as an extralinguistic variable that may or may not influence a given linguistic feature, akin to how text genre (law, fiction, sermon, etc.) is already treated in the PPCME2.

Pending the completion and annotation of these additional resources, the present article investigates how including the diatopic dimension provides insights into the spread of emergent features. To mitigate the effect of the M2 data gap, the ME period is divided into only two sub-periods, namely Early Middle English (eME, M1–M2) and Late Middle English (lME, M3–M4). This division is similar to the comparisons that can be made between the LAEME and the eLALME. More fine-grained diachronic divisions, for example 50-year intervals as done by Jenset and McGillivray (2017: 182) or a periodization based on a given feature's semantic developments as proposed by Perek and Hilpert (2017), can only be undertaken once texts from the PCMEP and PLAEME have been added. When applying a division between early and late ME, as shown in Figure 3, the data gap is mitigated but not entirely removed, as Northern texts only appear in the later periods M3 and M4. In spite of these drawbacks, this article probes the feasibility of testing a given feature's diatopic spread in order to provide arguments for or against explanations of contact-induced or contact-influenced change.

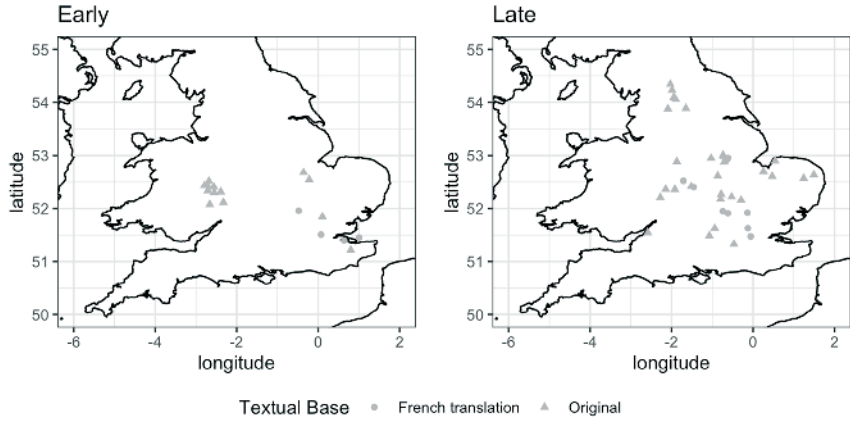


Figure 3: Survey points in the PPCME2, grouped by early (M1–M2) and late (M3–M4) sub-periods

Baseline: The spread of French-based verbs

To test the mapping method, it should be applied to an obvious case such as FBVs before further features are to be investigated. In terms of methodology, the annotation layer marking verb etymology, i.e. French-based versus other, is taken advantage of. A first query was made for all FBVs, followed by a second query for all lexical verbs. Dividing the frequencies from the first query by those from the second query yields a ratio of FBVs per text.

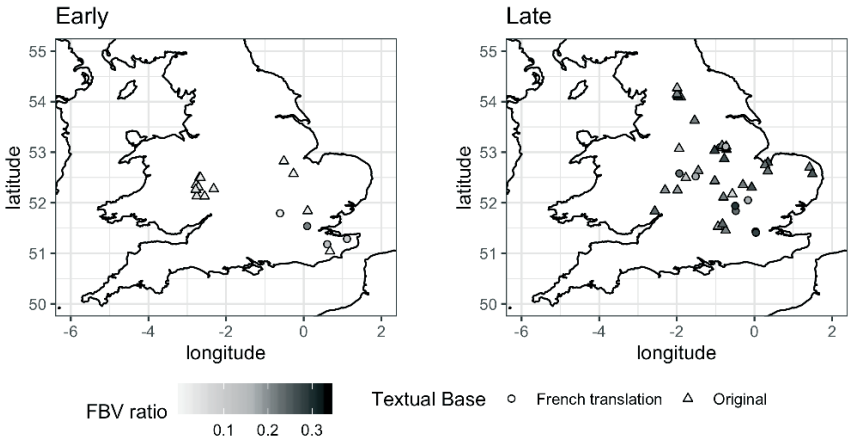


Figure 4: Map showing ratios of French-based verbs per text in early and late Middle English

The map of FBV ratios, shown in Figure 4, displays an obvious increase in FBVs from eME to lME. Within each sub-period, there is a slight tendency for texts further North to display a lower FBV ratio. A more detailed view is offered in Figure 5, where an increase in latitude results in an overall decline of FBV-ratios, however not in a straightforward manner. Textual base appears to exert an effect, as texts translated from French generally exhibit higher FBV ratios in eME. It should be noted that these texts are from M2, meaning that their higher FBV ratios may also be caused by the temporal progression from M1 and M2 as well as their lower latitude, among a host of other possible factors. The effect of textual base is less pronounced in lME. For both periods, an effect of increasing latitude on lower FBV ratios can be perceived, which supports a South to North spread. However, the fact that this correlation is not clear-cut for a feature that is obviously caused by contact with French sets the expectation that other features whose status as contact-induced or contact-influenced is to be probed will not show obvious patterns either.

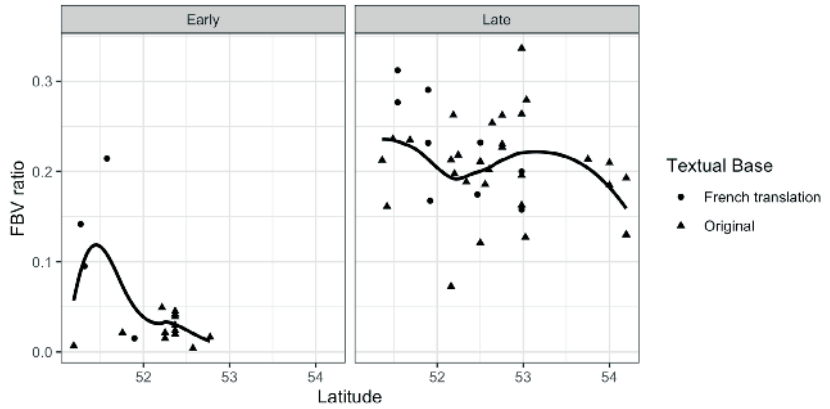


Figure 5: Correlation between latitude and ratios of French-based verbs per text in early and late Middle English

Testing potential contact-induced features

In the present section, two features are probed with regard to their diatopic spread. The two features under investigation are the verbal {en-} prefix, which was copied from OF, and the *to*-POC, which already existed in Old English but may have become more prominent in ME due to contact with French.[3]

The spread of verbal {en-} prefixes

The {en-} prefix was copied from Old French and it first derived FBVs, as shown in Example (1), before being fully integrated and also attaching to native lexical bases, as shown in Example (2). The premise is that just as with FBVs, one should expect a diatopic spread from South to North.

(1) *Take Maye buttre & hony & þe white of an egge & menge to-gedir and* ***anoynte*** *thyn eghne with all.* (CMTHORN,9.151-CMTHORN,9.153)
"Take unsalted butter, honey, and an egg white, mix them together and anoint your eyes with it all (= apply it to your eyes)."

(2) *and nou he haþ* ***en-he□ed*** *myn heued vp alle myn enemis.* (CMEARLPS,29.1225)
"and now he has raised/honoured my head (= made me eminent/ powerful) above all my enemies"

3 The term *copying* is preferred over the more traditional term *borrowing* due to conceptual problems with the latter: the donor language is not deprived of any linguistic material it supposedly "lends", nor is this material necessarily identical in the donor and recipient languages, whereas the concept of "copying" allows for non-identicality of original and copied material (see Johanson, 2002: 287–288).

In terms of methodology, a corpus query was based on a manually verified list of verbs prefixed with {en-}.[4] The map of normalised frequencies per text, shown in Figure 6, suggests a South to North spread of the feature, albeit with the highly conspicuous exception of the text with the highest frequency being in the North. A more detailed view, given in Figure 7, reveals that this outlier upends the trend of declining frequencies with increasing latitude. The two texts with the highest frequencies are medical texts, which suggests a strong effect of genre, even more so due to the fact that both are English originals rather than French translations. The prominence of {en-} in these texts is largely due to the high frequency of *enointen* ("to anoint"), which is already shown in Example (1).

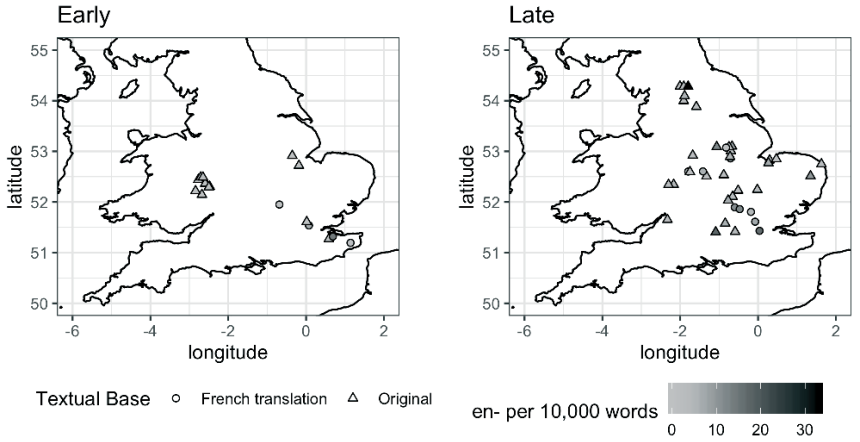


Figure 6: Map of normalised frequencies of the {en-} verbal prefix in Early and Late Middle English

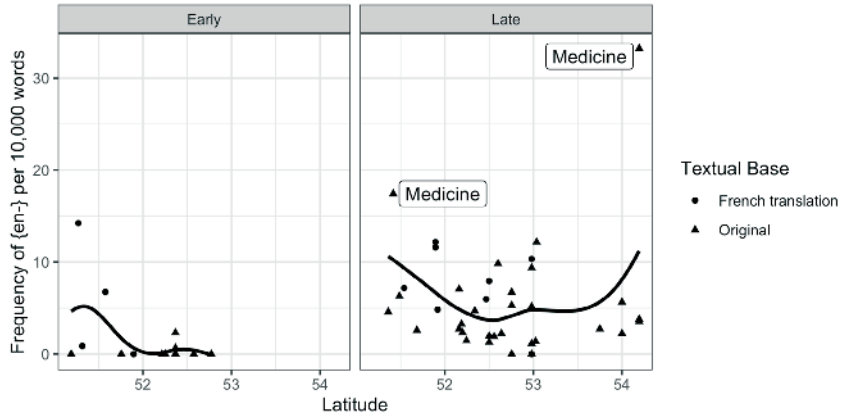


Figure 7: Correlation between latitude and normalised frequencies of verbs prefixed with {en-} per text in early and late Middle English

4 This includes the allomorphs *am-*, *an-*, *em-*.

The spread of *to* prepositional object constructions

The *to* prepositional object construction (*to*-POC), shown in (3), already existed in Old English (see De Cuypere, 2015), and rose to prominence in ME, overtaking its double object construction (DOC) variant (Zehentner, 2018: 158), which is shown in (4). Therefore, the question is not whether the *to*-POC was copied from French, but whether the influx of FBVs, many of which possessed argument structures involving *à*-PP arguments, as in (5), may have acted as a catalyst for the *to*-POC. A diatopic spread from South to North would lend support to this explanation.

(3) *And* ***to þeme*** *he yafe monny ryalle ornamenttus* (PPCME2, CMSIEGE,92.701) "And to them he gave many real ornaments"

(4) *and giaf* ***hem*** *his blescinge* (PPCME2, CMTRINIT,113.1540) "and [he] gave them his blessing"

(5) *ceo adescertes qe jeo ai, ceo doigne* ***a toy*** (AND, Trotter, 2006: "doner[1]", Actes (A) 362d) "What I certainly own, that I give to you"

In terms of methodology, corpus queries of *to*-POCs cannot be undertaken on purely syntactic grounds given that the corpus annotation does not distinguish adjunct *to*-PPs such as spatial adverbials and argument *to*-PPs such as prepositional objects. PPs dominating a pronoun can be assumed to most likely be arguments, whereas this case cannot be made for PPs dominating a noun, hence the need for manual sorting.

To reduce the amount of data in need of manual verification, nouns within *to*-PPs in clauses that have a constituent marked as "OB1" and no constituent marked as "OB2" were annotated for animacy, the idea being that animate nouns are more likely to be part of a prepositional object, e.g. *to the king*, whereas inanimate nouns are more likely to be part of an adjunct, e.g. *to the castle*. Performing a first query for *to*-POCs in which the *to*-PP dominates a personal pronoun, and adding the frequencies to those obtained from a second query for *to*-POCs in which the *to*-PP dominates an animate noun, yields an estimate of *to*-POCs in the corpus.

In contrast, DOCs can be queried in the corpus syntactically based on the annotation of direct objects as "OB1" and indirect objects as "OB2" for noun phrase constituents. Dividing the frequencies of *to*-POCs by the combined frequencies of *to*-POCs and DOCs in each corpus text yields a *to*-POC ratio per text.

The map of *to*-POC ratios, shown in Figure 8, suggests that the texts with the highest *to*-POC ratios are Southern rather than Northern. However, this correlation does not appear to be straightforward, given that the southernmost texts display lower ratios than those in the East Midlands. A more detailed view, given in Figure 9, reveals an overall correlation between increasing latitude and

decreasing *to*-POC ratios, which speaks in favour of a South to North spread. However, the numerous outliers and the genres they belong to indicate that diatopic spread is only one of multiple factors at play. This is the case for lME in particular, where texts belonging to religious genres such as "Bible", "Sermon", and "Religious Treatise" show extremely high *to*-POC ratios, whereas scientific genres such as "Astronomy" and "Medicine" display very low *to*-POC ratios.

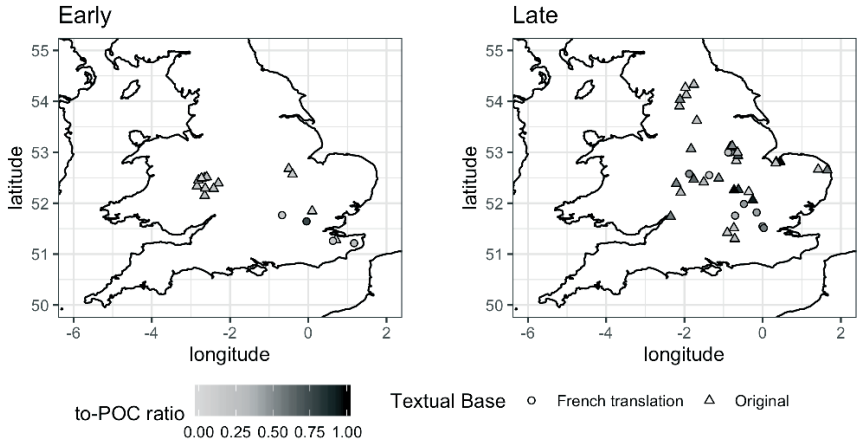


Figure 8: Map showing ratios of to prepositional object constructions per text in early and late Middle English

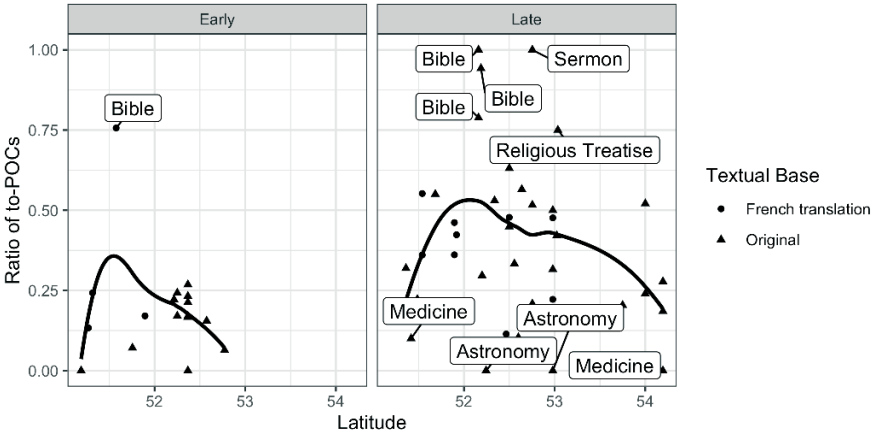


Figure 9: Correlation between latitude and ratios of to prepositional object constructions per text in early and late Middle English

The influence of translation effects may be downplayed for two reasons. Firstly, the distinction between French translations and an "Other" category, which mostly corresponds to English original texts but also includes texts

translated from languages other than French, may not reflect translation effects from these other languages, most notably Latin for Bible texts, which are often translated from the Vulgate. The categorisation of textual base could therefore be more fine-grained. Secondly, English original texts which are thematically linked to translated texts, such as sermons and religious treatises discussing or influenced by Bible texts, may be affected by translation effects as well.

Implications

The analyses performed in the present article have revealed that although a South to North diatopic spread could be observed for the features under investigation, the picture was muddled by the effects of further variables such as textual base and genre. As such, investigating the diatopic spread of potentially contact-induced or contact-influenced features is off to a promising start as a first test method to probe the plausibility of such explanations of change, with the condition that other possible factors are taken into consideration. Multivariate statistical analysis is the obvious method to account for this. However, such an approach could not be used in the present article, as the corpus in its current state does not allow it due to the aforementioned data gaps, for example the smaller sample size and the lack of English original texts in M2, the lack of Northern texts prior to M3, or the general lack of balance between genres and periods.

As mentioned earlier, the addition of further texts to fill the data gaps is the solution to these problems. While the inclusion of texts from the PCMEP and PLAEME may fill the M2 data gap, it introduces a new variable, namely text type (prose versus verse), which would then not be balanced across ME sub-periods. As these additional texts are mostly verse, the sub-period M2 would consist primarily of verse texts, whereas the other sub-periods would be dominated by prose texts. Adding verse texts to the other sub-periods is therefore necessary to achieve a balance of prose and verse texts across the corpus data available for ME in order to statistically test the effect of text type. Further, the relevant metadata of these additional texts, such as genre, need to be harmonised with the metadata of the PPCME2 in order to be usable for statistical purposes.

Another benefit of a more balanced coverage lies in the ability to look at linguistic change in a more fine-grained manner. The present article contrasted eME and lME rather than the four sub-periods defined by the corpus creators. Filling the data gaps may not only allow the use of multivariate analysis with this original subdivision, but may also enable detailed diachronic divisions as those proposed by Jenset and McGillivray (2017: 182) and Perek and Hilpert (2017).

Conclusion

The present article outlined the mapping of texts from the PPCME2 in order to trace the diatopic spread of emergent features. In this context, this applies to potential contact-induced or contact-influenced features in contact with OF. The advantage of the method lies in the ability to investigate any feature that can be queried in the corpus, rather than being limited to a fixed set of features as is the case for previous resources such as LAEME and eLALME.

The spread of the features under investigation, namely FBVs, the verbal {en-} prefix, and the *to*-POC, all followed a trend that indicated a South to North spread. However, it should be noted that this spread is far from unambiguous given the interference from further variables such as text genre and textual origin, i.e. whether a text was an original English language production or the result of a translation from French. It should be noted that such a South to North spread does not constitute a sufficient condition for an explanation of French contact, but merely a necessary condition. As such, the explanation of influence from contact with French is not ruled out, but requires further evidence in order to be corroborated.

The mapping of corpus texts from the PPCME2 and the investigation of the diatopic spread of features have revealed data gaps that go beyond the well-known M2 data gap. Given the effects of text genre and textual origin in addition to regional variation, there is an apparent need to complete the PPCME2 with additional texts. Projects such as the PCMEP and PLAEME fill the M2 data gap, but introduce a portion of verse texts leading to an imbalance of prose and verse across the ME sub-periods. By specifically adding further texts to correct this disparity, applying lemmatisation procedures as well as the mapping procedure described in the present article, a balanced multivariate analysis of ME can be made possible.

References

• BENSKIN, M., LAING, M., KARAISKOS, V., WILLIAMSON, K. (2013): *An Electronic Version of A Linguistic Atlas of Late Mediaeval English*, Edinburgh, The University of Edinburgh. http://www.lel.ed.ac.uk/ihd/elalme/elalme.html (accessed 16 September 2019).

• COLE, M. (2018): "A native origin for Present-Day English they, their, them", *Diachronica* 35, p. 165-209. https://doi.org/10.1075/dia.16026.col.

• DE CUYPERE, L. (2015): "The Old English *to*-dative construction", *English Language and Linguistics* 19, p. 1-26. https://doi.org/10.1017/S1360674314000276.

• HAEBERLI, E. (2018): "Syntactic effects of contact in translations: Evidence from object pronoun placement in Middle English", *English Language and Linguistics* 22, p. 301-321. https://doi.org/10.1017/S1360674318000151.
• HOGG, R.M. (1992): "Phonology and Morphology", in HOGG, R.M. (ed.), *The Cambridge History of the English Language*. Cambridge, Cambridge University Press, p. 67-167. https://doi.org/10.1017/CHOL9780521264747.004.
• HOWE, S. (1996): *The Personal Pronouns in the Germanic Languages: A Study of Personal Pronoun Morphology and Change in the Germanic Languages from the First Records to the Present Day*, Berlin, De Gruyter. https://doi.org/10.1515/9783110819205.
• JENSET, G.B. & MCGILLIVRAY, B. (2017): *Quantitative historical linguistics: A corpus framework*, Oxford, Oxford University Press.
• JOHANSON, L. (2002): "Contact-induced change in a code-copying framework", in JONES, M.C. & ESCH, E. (Eds.), *Language Change*, Berlin, De Gruyter Mouton, p. 285-313. https://doi.org/10.1515/9783110892598.285.
• KROCH, A. & TAYLOR, A. (2000): *The Penn-Helsinki Parsed Corpus of Middle English, Second Edition (PPCME2), release 3*, Philadelphia, University of Pennsylvania. http://www.ling.upenn.edu/hist-corpora/ (accessed 16 September 2019).
• LAING, M. (2013): *A Linguistic Atlas of Early Middle English, 1150–1325, Version 3.2*. Edinburgh, The University of Edinburgh. http://www.lel.ed.ac.uk/ihd/laeme2/laeme2.html (accessed 16 September 2019).
• PERCILLIER, M. (2016): "Verb lemmatization and semantic verb classes in a Middle English corpus", *Proceedings of the 13th Conference on Natural Language Processing (KONVENS 2016)*. p. 209-214. https://www.linguistics.rub.de/konvens16/pub/26_konvensproc.pdf (accessed 16 September 2019).
• PERCILLIER, M. (2018): "A Toolkit for Lemmatising, Analysing, and Visualising Middle English Data", *Proceedings of the Second Workshop on Corpus-Based Research in the Humanities CRH-2*. p. 153-160. https://www.oeaw.ac.at/fileadmin/subsites/academiaecorpora/PDF/CRH2.pdf (accessed 17 January 2020).
• PEREK, F. & HILPERT, M. (2017): "A distributional semantic approach to the periodization of change in the productivity of constructions", *International Journal of Corpus Linguistics* 22, p. 490–520. https://doi.org/10.1075/ijcl.16128.per.
• PROFFITT, M. (ed.) (2018): *Oxford English Dictionary*, 3rd ed., Oxford, Oxford University Press. http://www.oed.com/ (accessed 16 September 2019).
• RANDALL, B. (2010): *CorpusSearch (Version 2.003.00)*[Computer Software]. http://corpussearch.sourceforge.net (accessed 16 September 2019).
• TROTTER, D. (2006): *Anglo-Norman Dictionary 2 Online edition*, London, Modern Humanities Research Association. http://www.anglo-norman.net (accessed 16 September 2019).

• TRUSWELL, R., ALCORN, R., DONALDSON, J. & WALLENBERG, J. (2019): "A Parsed Linguistic Atlas of Early Middle English", in ALCORN, R., KOPACZYK, J., LOS, B. & MOLINEAUX, B. (eds.), *Historical Dialectology in the Digital Age*, Edinburgh, Edinburgh University Press, p. 19-38.
• TRUSWELL, R., ALCORN, R., DONALDSON, J. & WALLENBERG, J. (2018): *A Parsed Linguistic Atlas of Early Middle English*. University of Edinburgh. https://doi.org/10.7488/ds/2310.
• ZEHENTNER, E. (2018): "Ditransitives in Middle English: On semantic specialisation and the rise of the dative alternation", *English Language and Linguistics* 22, p. 149-175. https://doi.org/10.1017/S1360674316000447.
• ZIMMERMANN, R. (2018): *The Parsed Corpus of Middle English Poetry (PCMEP)*. http://pcmep.net/ (accessed 16 September 2019).