

Beyond Anglo-Norman: the Lexical Influence of Old French Dialects on Middle English

Beth Beattie

University of Glasgow

Abstract. This project examines lexical borrowings between the Old French (OF) of Champagne and Picardy and Middle English, identified through spelling variations, and what these borrowings reveal about cultural links between English and France. The methodology consisted of using the Middle English Dictionary (MED) and the Linguistic Atlas of Early Middle English (LAEME) to establish the frequency of variant spellings of open-class Old French words, thus indicating the strength of the potential borrowing of an Old French word or feature. An examination of the context of texts in which a high concentration of dialectal forms were found was used to determine the cultural background that facilitated such transmission. The different source materials for the MED and LAEME resulted in a marked culling of forms to those found in both datasets, but the variant spellings given for those that remained proved ample. The analysis of the texts containing these variant spellings produced fewer results than expected; The majority of the manuscripts featured few words of note, with the exception of Arundel 57, which provided a glimpse into OF usage that was decidedly more varied than solely Anglo-Norman or Central French. This lack of final data, in combination with the wide distribution of the lexical items across manuscripts about which not much is known, made it impossible to focus on a particular lexical source and examine the reasons behind lexical transmission. The possibility to undertake such research remains, with further extended examination required.

Keywords: Old French; Middle English; loanwords; spelling; historical dialectology

1 Introduction

Due to rigorous examination and centuries of study, the Dark Ages can no longer be accurately named ‘dark’. The term is now very seldom used outside of popular culture, yet the Middle Ages in Western Europe still hold a certain fascination, especially to historical linguists. The role of Old French (OF) in the development of Middle English (ME) – especially the Anglo-Norman (AN) and Central French (CF) dialects – has been the recipient of an inordinate amount of attention in uncovering the history of English, yet they were not the only OF dialects to exist during this period. Spanning the width of the langue d’oïl was a continuum of around twelve major dialects, each with their own distinct features that differentiated them from the incipient CF standard. Given the proximity and cultural ties between England and France between the years of 1100 and 1400, it seems likely that other dialects played a role in contributing to ME vocabulary, even if a fleeting one. Such discoveries would be instrumental in detailing the picture of cross-Channel links, both on a linguistic and a societal level.

1.1 Research questions

The research questions this dissertation aims to answer are:

- (1) Is it possible to identify loanwords into Middle English from Old French dialects other than Anglo-Norman and Central French?
- (2) Can the origins of these loanwords be determined through spellings?
- (3) What can these borrowings reveal about cultural links between England and France?

1.2 Method

First, differentiating features of Old French dialects that result in spelling variation will be established, providing examples of these features in literature to confirm usage. The MED shall then be used to search for open-class words known to be of Old Northern French origin, and examine the variant forms given in each entry to identify possible features found in a specific Old French dialect. These potentially relevant forms will form a basis for a LAEME search to establish the frequency of these forms, thus indicating the strength of the potential borrowing of an Old French word or feature. If particular texts feature a high volume of Old French dialectal forms, those texts will then be examined more in-depth to identify the reason for such concentrations, most likely due to scribal influences or – more interestingly for my research – cultural links with a particular region of France.

1.3 Old French Dialect Features

The most suitable elaboration of the distinguishing features between the various OF dialects can be found in Einhorn's 1974 handbook on Old French, and it is their table and description of these features that forms the basis for each stage of analysis.

Table 1: *Old French dialectal features, based on Einhorn, 1974, p. 137¹.*

Characteristics	S	SW	W	N	AN	P	Wn	Ch	L	FC	B
1. [o:] > <i>ou, o</i>	xx	xx	xx	xx			xx	x	xx	xx	xx
2. <i>ei</i> stays, or > <i>e</i>	xx	xx	xx	xx	xx						
3. <i>ie</i> > <i>e</i>	x	xx	xx	x	xx						
4. Cl.1 impf. in - <i>oue, -oe</i>	x	x	x	x	x						
5. Tonic <i>el, eu</i> > <i>al, au</i>		xx	x								

¹ References to these features will be labelled numerically according to this chart. For example, '16' refers to '16. Use of *w*'

6. 1 st p.pl. - <i>om(s)</i> , - <i>on</i>		xx	xx	xx	x						
7. [ɛ] ($\tilde{+} n$) stays [ɛ̃]			xx	xx	xx	xx	xx				
8. Graphy <i>ai</i> for <i>ei</i>			xx								
9. 1 st p.pl. - <i>um(s)</i> , - <i>un(s)</i>				x	xx						
10. [k], [g] for <i>ch</i> , <i>j</i>				x	x	xx					
11. <i>c (+e, i) > ch</i>				x	x	xx	x	x			
12. <i>eau > iau</i>				x		xx	x				
13. [e:] > <i>ei</i>				x			xx	x	xx	x	
14. [o:], [o] > <i>u</i>					xx						
15. Final <i>z > s</i> 12 th c.					x	xx					
16. Use of <i>w</i>					x	xx	xx	x	xx		
17. <i>la > le</i>						xx	xx				
18. Final <i>t</i> remains						xx	xx	x	x		
19. Cl.1.impf. in – (<i>i</i>) <i>eve</i>							x		x	x	x
20. <i>a > ai</i>								x	xx	xx	x
21. Initial <i>e > a</i>									xx	x	
22. <i>lo, lou = le</i>									xx	x	x
23. Tonic [ɛ] > <i>a</i>										xx	x
24. <i>al, able > aul, auble</i>										x	xx
25. <i>ei (+nasal) > oi</i>											xx

The dialects and the areas in which they were spoken are:

- South (S): Baronnais, Nivernais, Berry, Orléans
- South-West (SW): Angoumois, Saintonge, Aunis, Poitou
- West (W): Touraine, Anjou, Maine, Brittany
- Norman (N): Normandy
- Anglo-Norman (AN): England
- Picard (P): Picardy, Artois
- Walloon (Wn): North East (Belgium)
- Lorraine (L)
- Franche-Comté (FC)
- Burgundy (B)
- Champagne (Ch)

Table 1 lists the identifying features and the dialects in which they appear, illustrating a strong presence with two crosses and a weaker presence with one cross. Einhorn elaborates on the condensed versions of the features and gives examples of each:²

1. Tonic (or most stressed) syllables containing [o:] became [ø], spelled <eu>, in Parisian French and Picard, but [u] in other dialects, spelled <u> in Anglo-Norman and <ou> or <o> elsewhere.

- *nevou* (*neveu*)

- *seignor* (*seigneur*)

2. The diphthong /ei/, instead of becoming <oi> during the twelfth century, was lowered through [ɛi] to [ɛ], spelled <ei> or <e>.

- *saveir* (*savoir*)

- *la metié* (*la moitié*)

- *le rei* (*le roi*)

- *la vee* (*la voie*)

- *aveit* (*avait*)

3. Tonic syllables containing [iɛ] became [ɛ:], spelled <e>.

- *chevaler* (*chevalier*)

- *manere* (*maniere*)

- *ben* (*bien*)

- *la pere* (*la pierre*)

- *cel* (*ciel*)

4. In the thirteenth century, Class 1 verbs (where the infinitive ends in <-er>) used <-oue>, and later <-oe>, imperfect conjugations alongside the standard endings.

Standard OF Imperfect Conjugations	Dialectal Imperfect Conjugations A
--	------------------------------------

² Einhorn, 1974, pp. 135–140; own italics); The forms in brackets are the equivalent Parisian French forms of the dialectal examples. The IPA symbols have also been updated to correspond with current usage.

-eie	-o(u)e
-eies	-o(u)es
-eit	-o(u)t
-iiens	-iiens
-iiez	-iiez
-eient	-o(u)ent

5. Tonic syllables containing <el> and <eu> appeared as <al> and <au>.

- *quaus* (*quels*)
- *tau* (*tel*)
- *corporau* (*corporel*)

6. The endings <-om(s)> and <-on> were used instead of <-ons> in the 1st person plural.

- *aloms* ! (*allons* !)
- *cum nos disiom* (*comme nous disons*)
- *nous voulon* (*nous voulons*)

7. [ɛ] followed by <n> remained [̃ ɛ] and was not lowered to [̃ ă].

- *prent* does not rhyme with *avant*

8. The spellings <ei> and <e> (see the sound change in feature 2) were at times spelled <ai>.

- *trais* (*trois*)
- *monaie* (*monoie*)
- *saient* (*soient*)
- *otraierent* (*otroierent*)

9. The endings <-um(s)> and <-un(s)> were used instead of <-ons> in the 1st person plural.

- *donum* (*donons*)
- *volums* (*voulons*)
- *nus volun et comanduns* (*nous voulons et comandons*)

10. [k], spelled <c>, <k> and [g], spelled <g>, replaced <ch> and <j> respectively.

- *camp* (*champ*)
- *castel* (*chastel*)
- *cose* (*chose*)
- *escaper* (*eschaper*)
- *gardin* (*jardin*)
- *goie* (*joie*)

11. <c> followed by <e> or <i> became <ch>.

- chité (*cite*)
- merchi (*merci*)
- grache (*grace*)

12. The triphthong <eau> became <iau>.

- biaus sire (*beaus sire*)
- les oisiaux (*les oiseaux*)

13. Tonic syllables containing [e:] became the diphthong [ei].

- teil (*tel*)
- doneir (*doner*)
- son peire (*son pere*)
- sa bontey (*sa bonté*)
- l'assembleie (*l'assemblée*)

14. Tonic syllables containing [o:] and [o], and initial [o], became [u], spelled <u> and later <ou>.

- duner (*doner*)
- sun seignur (*son seignor*)
- pur sue amur (*por soe amor*)

15. Final [ts], spelled <z>, soon became [s] (In Picard, the grapheme <z> was rare, with 2nd person plural using <-(i)és>).

- assés (*assez*)
- vaillans (*vaillanz*)
- se vos volés (*se vos volez*)
- vos disiés (*vos disez*)

16. Germanic initial <w> was retained instead of becoming <g> or <gu>; <w> could replace initial <v> or <vu> and was sometimes used as an intervocalic glide.

- warder (*garder*)
- wages (*gages*)
- Willaume (*Guillaume*)
- ju wel (*je vuel*)
- il lowent (*il loent*)
- awoust (*aoust*)

17. The feminine article and pronoun “la” became “le”, which was not contracted after “a” or “de”. “Li” could replace it in the nominative.

- le contesse (*la contesse*)
- a lequele (*a laquele/auquele*)
- li vostre amie (*la vostre amie*)

18. Final <t> was retained. The ending <-eit> was common.

- volentet (*volenté*)
- portet (*porté*)
- tenut (*tenu*)
- la veriteit (*la verité*)

19. <-er> verbs used <-eve> and <-ieve> endings in the imperfect.

Focusing on Champagne, the region played an integral role in trade, politics, and culture, which had a strong impact on its language. Ayres-Bennett observes that ‘while Champenois is essentially a central dialect, which therefore has much in common with Francien, it also shares features with its neighbouring dialects to the north-east (Walloon), east (Lorraine), and south-east (Burgundian)’ (1996, p. 69). Such sharing of features is typical of dialects in close proximity; however, the prestige of Champagne maintains the distinctiveness of the dialect and makes it an interesting dialect to analyse.

1.5 Benefits

This research aims to move beyond the giants of Anglo-Norman and Parisian French to offer a wider linguistic analysis of the influence of Old French lexis on Middle English. It is very easy to fall into the trap that contacts between languages during the medieval period was between two homogenous masses, yet this could not be further from the truth. Dialectal diversity is known to be significant during this period with French not having the single dominant standard that it does today, and the history books do not seem to take this into account. A more nuanced picture of language contact between Old French and Middle English is required, which must by definition take into account the variation within each language.

2 Research Context

2.1 Old French Overview

Before delving into the development of medieval French it must first be established exactly what is meant by the term ‘Old French’. Huchon helpfully gives a grammatical reasoning behind his labelling:

‘La denomination d’ancien français englobe souvent l’état de la langue du IX^e siècle au XVI^e siècle. On préférera toutefois parler pour le XIV^e siècle et le XV^e siècle de moyen français, conservant le terme d’ancien français pour l’époque où le français est une langue à déclinaisons à deux cas’ (Huchon, 2003, p. 53).

[Its southern counterpart had a surprising uniformity across the region: “The denomination ‘Old French’ often encompasses the state of the language from the ninth to the fourteenth century. Nevertheless, for the fourteenth and fifteenth centuries, it is better to use the term ‘Middle French’, reserving the term ‘Old French’ for the period where French distinguished between two grammatical cases] (own translation)

Such grammatical variances are of little importance for our purposes, but the period that Huchon defines through such differentiation is a sufficient base for further elaboration and exploration.

Much like the popular notion of the origins of the French nation, the French language is traditionally believed to originate from the early 800s with the writing of the Oaths of Strasbourg in 842 (Huchon, 2003, p. 27). However, works written in the vernacular were not commonplace until much later, and even then, the first increase in vernacular writings is found at the end of the eleventh century in the langue d’oc in the south of France, rather than in the langue d’oïl in the north (Lodge, 1993, p. 110). It is during the twelfth century that vernacular literary works appear with any frequency in the langue d’oïl, and “only in the thirteenth century is French used in written prose and in non-literary as well as literary texts” (Lodge, 1993,

p. 107). The reason behind this spread in vernacular writing from south to north is due to the growing literary interests of the aristocracy: “They found in the courtly literature which originated in the langue d’oc an expression in the vernacular language of the earthly values of chivalric fame and physical pleasure they were beginning to adopt” (Lodge, 1993, p. 109). This literature that the aristocracy was so fond of was designed for oral performance, however, with specific stylistic features that Latin and other languages could not accurately capture- in other words, “a vernacular performance required a vernacular script” (Lodge, 1993, pp. 108-9).

One of the most marked differences in the use of the vernacular in the langue d’oïl, compared to in the langue d’oc, is that there was a high degree of regional variation from the outset³ (Lodge, 1993, p. 113).

‘Du point de vue linguistique, ce qui frappe dans l’occitan des troubadours, c’est qu’il présente, dès ses premières manifestations, c’est-à-dire dès le XI^e siècle, une assez grande unité : les différences dialectales y sont en effet minimales et sans aucun rapport en général avec la provenance dialectale du troubadour : l’idiome est sensiblement le même du limousin jusqu’à la Méditerranée’ (Bec, 1967, p. 69)

[From the linguistic point of view what is striking in the Occitan of the troubadours is that it presents a reasonable degree of unity from its earliest manifestations onwards, that is from the eleventh century; dialectal differences are in fact minimal and bear no relationship with the dialectal origin of the troubadour; the language is perceptibly the same from the Limousin to the Mediterranean] (trans. Lodge, 1993, p. 111)

Despite the influence of the southern troubadours on the development of the courtly literature in the north that acted as a means of disseminating vernacular writings, this lack of a uniform variety across the region allows us to better examine the specific regional varieties in the langue d’oïl.

In discussing the state of Old French around 1200, Einhorn observes that the dialect of Paris was “only one of many competing dialects in northern France for in the passage from Latin to Old French local differences had developed, slight or more marked, merging or overlapping into neighbouring regions” (1974, p. 135). These varying features are more on the level of phonetics as opposed to lexis or syntax, as it has been observed that a lot of the dialects had a shared vocabulary (Huchon, 2003, p. 60). What is particularly noteworthy is that a number of these regional varieties gained quite high levels of prestige as the wealth and influence of their population centres grew (Lodge, 1993, p. 98). With Paris having not yet grown to the position of cultural and economic dominance that it has today, the status of places like Picardy and Champagne was considerably higher than in more modern times, and this is reflected in textual evidence:

‘The earliest twelfth-century texts contain numerous Norman or western features (reflecting the literary influence of the Plantagenet court). The famous romances of Chrétien de Troyes (c.1180) contain Champenois features. Many literary texts composed in the thirteenth century reflect the linguistic usage of the great Picard towns. Interestingly, it is only in the thirteenth century that we begin to find vernacular texts written in the Paris region’ (Lodge, 1993).

Economics was the reason for the rise in status of the northern towns around Picardy, and it was also the cause of their decline during the fourteenth century. Furthermore, the annexation of Normandy and Anjou at the beginning of the thirteenth century led to their regional features falling out of favour (Lodge, 1993, p. 132; Nezirovic, 1980, pp. 183-5). Paris had been growing in both economic and political influence since the end of the eleventh century, due to its prime location surrounded by excellent agricultural land and the easy access to the city afforded by the Seine (Lodge, 1993). There were other factors beyond its position that contributed to Paris' rise in status: the expansion of royal power, the centralisation of administrative power, the creation of its reputation as an intellectual hub (Huchon, 2003, p. 61). All of these developments helped to establish Paris as the crucible where a more standard French began its construction.

2.2 French Influence on Middle English

The history of French influence on English does not begin in 1066, but well before, due to contact between Norman and Anglo-Saxon rulers; Ethelred II even sought protection in Normandy to escape Viking raiders (Blake, 1992, p. 423). After 1066 is when things step up a gear in terms of greater contact between the two languages. However, post-Conquest, the evidence of English usage is very little, and mostly anecdotal (Blake, 1992, p. 424). An important consideration regarding the Conquest itself is that William's invading force was not a monolithic horde of Norman speakers; other languages and dialects such as Breton and Picard would have been well-represented (Rothwell, 1998, pp. 149-50). Despite the fact that subsequent settlers were mostly Norman, it is important to remember that overgeneralisation of dialectal variation during this period can result in an inaccurate linguistic overview.

A major shift in the usage of French in England comes relatively quickly after the Conquest, as French gradually fell out of usage as a vernacular in favour of English. The decline of Anglo-Norman began soon after the Conquest in certain areas and was only accelerated when Normandy was lost to France during the reign of King John (Blake, 1992, p. 427). However, as the prestige of Anglo-Norman was on the wane, the rise in prestige of Paris and Parisian French grew. It was seen as the 'langue du jour' and was the key to social advancement, as evidenced by the development of grammars and word lists to assist in the acquisition of the fashionable continental dialect (Blake, 1992, p. 423; 427).

Before the influence of Paris raised the prestige of its dialect, Continental French gained a foothold in England through Henry II's marriage to Eleanor of Aquitaine:

'First, the king's marriage to Eleanor of Aquitaine brought a new influx of courtiers just at the time when the descendants of the first Norman French invasion were becoming totally assimilated. Furthermore, her patronage of French literary production and the prestige of the new literary forms evolving in France, the romance and lyric poetry, encourages the use of French in the literate segments of society' (Kibbee, 1991, p. 14).

Such patronage resulted in an explosion of literary compositions that began with Henry II and did not fade until the death of Henry III. Anglo-Norman culture had a thriving native literary tradition, which provided a ready market for the Continental French works. Despite the separate development of the two traditions, both Anglo-Norman and Continental French works shared similar themes and language (Kibbee, 1991, p. 18). Therefore, the incursion of Continental French into Anglo-Norman literature was aided by cultural similarities, resulting in the phasing out of the former leviathan.

Contrasted with the new sophisticated French of the court, the speakers of Anglo-Norman began to feel that their language was inferior, which no doubt contributed to its early decline (Kibbee, 1991, p. 24; Burnley, 2000, pp. 66-7). However, as English began to reclaim its lost territory during the fourteenth century, the effects of all manner of French varieties would be left on the language.

2.3 Loanwords

Early Middle English was a time of great expansion in the lexicon. Of the 60,000 lexemes recorded in the MED between 25% and 30% of them are loanwords, taken from Latin, Old French, and Old Norse for a variety of purposes, such as filling lexical gaps due to the period's technological and societal upheaval (Brinton & Bergs, 2017, p. 173). An important factor to consider when looking at information concerning loans from French and Latin is that it is often challenging (or even impossible) to differentiate Latin and French loans; at times, the Latin and French spellings are identical, whereas at others, the Latin spellings were changed due to Middle English conventions. Brinton and Bergs cite 'allegory' (ME 'allegorie' < Latin *allegoria*) and 'desk' (ME 'deske' < Medieval Latin *deska*) as examples of this (2017, p. 173).

What we do know about French loanwords in English begins prior to the Conquest, with words pertinent to the nobility: 'prūd' (valiant), 'castel' (castle), 'gingifer' (ginger), and 'capun' (capon) (Blake, 1992, p. 429). Subsequent borrowings are traditionally divided into two waves: the first arriving with the Normans, and the second coming from Continental or, more specifically, Parisian French (Blake, 1992, p. 426). More recently, Durkin acknowledges that this view of the period is too narrow, and the English/French/Latin trilingualism of medieval England complicates this rather neat perspective (2014, p. 229). However, earlier borrowings from Norman are noticeably different from later ones, mostly due to spelling differences (Blake, 1992, p. 430), so this distinctiveness of Norman is useful to keep in mind when examining the circumstances of loanwords.

The division between earlier and later loans is also exemplified by an increase in borrowing from 1250 into the fourteenth century, during which time the language of writing in England gradually changed to English. The prestige of the semantic categories of these loans is similar in places to those of the earlier Norman words, but more have appeared over a broader scope:

'These later loans occur in many additional semantic areas where French was prestigious, such as administration, fashion, social life, food, medicine, and learning, but also in core areas of everyday life... French loans are, however, rare in shipping and seafaring, as well as in farming, which possible reflects the lower prestige of these fields' (Brinton & Bergs, 2017, p. 174).

When discussing lexical borrowing it is easy to gloss over the entire country assume that everyone everywhere experienced the same depth of borrowing. Brinton and Bergs remind us that 'not all Middle English dialects nor their speakers experienced language contact to the same extent' (2017, p. 166). This idea links in with the domains in which French loanwords are found; you are more likely to encounter French if you are wealthy and educated, and less likely if you are a member of the peasantry.

Relating to the prestigious use of Continental French, Brinton and Bergs chose to distinguish Law French as a separate variety existing from the thirteenth century dedicated solely to being used in court and in records (2017, pp. 186-7). The strength of their argument for differentiating this usage as a separate

variety is not entirely convincing, but it is important to remember the diversity of roles that French played and the effect that can have on the language.

2.4 Anglo-French Relations

The relationship between England and France underwent many fluctuations throughout the medieval period, swinging from close allies to bitter enemies on multiple occasions. The very nature of this relationship, however, came to be redefined during the centuries in the lead up to the thirteenth century:

‘One feudal and political, the antagonism between [England and France] now became primarily a commercial one, in which the stakes were the wine trade of Gascony and the woollen manufacturers of Flanders, which were dependent of English raw wool. A conditional factor was sea-power or maritime supremacy in the Bay of Biscay and the English Channel, then often called the ‘Narrow Sea’. A glance at the map will show that the bond of connection between England and Flanders and between England and Gascony was the sea’ (Thompson, 1960, p. 55)

There are multiple facets to appreciate, with the most important being the growing mercantile echelon of society acting as a stabilising factor in the intense squabbling between powerful domains. Thompson goes into great detail of the precise nature of these interactions and their impact on the language: the role of English wool was integral in forging trade routes from Yorkshire to London and Dover, before continuing on to Bruges to fuel the renowned Flemish clothmakers. This vital relationship with northern regions of France resulted in a greater presence of Picard forms in treatises from this period. Furthermore, the nobility’s love of wine maintained the importance of the English lands in Gascony, thus the same usage of Gascon forms has been noted in trade documents (Thompson, 1960, pp. 61-80). Gascon is less relevant for our purposes, but the role of Picard in England is extremely promising in looking for any fingerprint it leaves on Middle English.

Trade was not the only reason for contact between England and France during the Middle Ages. Diplomatic missions, the ability to study at French institutions, and also military expeditions increased the likelihood of people picking up words from the continent to bring back home with them (Rothwell, 1998, p. 144). Such ‘cross-Channel traffic’ embodies the image of how contact between countries results in lexical borrowing.

2.5 The Importance of Scribes

When examining medieval texts, especially manuscripts, there are numerous considerations to be aware of. The most relevant for our purposes is the importance of the scribe, as “le personage important... est celui qui a tenu la plume” [the most important person is he who held the pen] (Monfrin, 1968, p. 33). With copies of each text only being possible by writing them out by hand, the scribe is in a unique place to add their own flavour to a text by using their own language and spellings. There are three possible transmission outcomes from exemplar to copy:

‘A: [The scribe] may leave the language more or less unchanged. This appears to happen only somewhat rarely.

B: He may convert it into his own kind of language, making innumerable modifications to the orthography, the morphology and the vocabulary. This happens commonly.

C: He may do something somewhere between A and B. This also happens commonly.’
(Benskin and Laing, 1981, p. 56; cf. McIntosh, 1973, p. 61)

For translated texts, the most relevant outcome is (A) since that would best preserve any original dialect forms from the source text. However, not all of the texts are translations, so this preference is not universally applicable.

The possible variation in transcription is well-understood by those who study manuscripts, but it is important to note that ‘writers and scribes in England did not live in total isolation, fixed in one spot from cradle to grave, cut off from all contacts and influences outside their place of residence’ (Rothwell, 1998, p. 155). Where they are from is not necessarily where they learned to write, and who they learned to write from was potentially from a different place altogether, as well as any other possible contacts and influences that must be considered when examining spellings or other variances in textual transmission.

The effect that this has on loanwords is staggering, especially during the Middle English period. Due to the lack of a standard language form, one word can have any number of spellings for any number of reasons. And such variability is cumulative, resulting in the fact that ‘loanwords do not inevitably or quite so obviously bear the stamp of their originary situation on them’ (Dance, 2014, p. 171). Filtering through these layers can be impossible at times, but they each play a role in charting a word’s history.

2.6 Things to Consider

Whilst it would be ideal to take every possible angle of lexical borrowing into account, this research must limit its focus to specific aspects. However, there are some considerations that must be acknowledged. First, there is the role of the MED and the Anglo-Norman Dictionary (AND) in widening the accessibility of the study of Middle English:

‘Over the last forty years the steady publication of the voluminous MED has profoundly altered the whole landscape of English etymology, whilst the appearance of the first edition of the AND has made possible at least a preliminary understanding of the presence in medieval England of a French that often differs considerable both in form and, more importantly, in semantic content from what were until recently accepted as the norms of standard Central French’ (Rothwell, 1998, p. 145).

Despite the wide-ranging benefits of these resources for enhancing our ability to study such historical languages, the very nature of historical dictionaries means that they are restricted to using literary texts as sources, thus resulting in an underrepresentation of non-literary texts (Brinton & Bergs, 2017). The language represented in these dictionaries could be seen to lack the more down-to earth and less florid language of treatises and trade documents, for example, but literary texts are often easier to find and thus make more readily-available sources for these dictionaries. Nonetheless, the level of language represented in the MED must be remembered when using it as a source of data.

Another element detailed by Rothwell whose importance in lexical borrowing cannot be overstated is semantics. Differing meanings between Anglo-Norman and Continental French words is one thing, but

when these words are borrowing into English, there is often a discrepancy between the spelling and sense, with one having come from a different side of the Channel from the other (Rothwell, 1998). As integral as it may be to examining Middle English loanwords, the semantics of the loans is too much of a wild card to do proper justice to in this research and is worthy of further consideration in relation to spelling in additional work.

Even the very definition of a loanword must be acknowledged in order to clarify the focus of this research. Brinton and Bergs state that ‘there is no unequivocal way of deciding when a lexical item from one language that is used during discourse in another language – whether by a single speaker, or repeatedly in a community – should be considered a loanword’, in addition to distinguishing between loans that fill a semantic gap in the target language from those that become productive (Brinton & Bergs, 2017, p. 111). All of these notions are worth considering but are not entirely relevant for this research. The importance is not placed on the productivity of the forms selected, nor making a significant distinction between whether a form is used by one person, a town, or the entire country. If the relevant form is cited and used in a deliberate way, then it is worth studying, no matter if it only cited once; any and all uses are significant and worth consideration.

3 Stage 1: MED

3.1 About the MED

The MED is one of three legs that form the trivium of the Middle English Compendium, with the other two being the MED Bibliography and the Corpus of Middle English Verse and Prose. The Bibliography lists all of the source texts used in the MED, greatly expanding on the citations given in the dictionary itself. The Corpus is not an actual corpus, but a collection of searchable texts, useful for examining collocations and the context of words and phrases.

The formulation of the MED began in 1925, drawing together around three million quotations from primary sources dating from between 1175 and 1500. It is widely regarded as the most comprehensive evaluation of Middle English vocabulary thanks to its inclusion of all types of evidence for a form or lexeme. The online edition – first published in 2000 and since revised – allowed the easiest access to the dictionary and its data and is the source of the initial data in this research.

3.2 Purpose

This first portion of this research is dedicated to gathering the base set of lexemes known to derive from some form of Old French and their variant spellings that could possibly have originated in Picardy or Champagne. To identify these relevant variant spellings, the descriptions of the dialects’ distinguishing features as given by Einhorn are used to pinpoint spellings that could show evidence of those dialectal features.

3.3 Method and Analysis

The data from the MED was gathered in two steps: the first, by manually recording each entry that fit the search parameters, and the second by searching through this initial data wave for any of the features

belonging to Picard and Champenois as indicated in Table 1. This method was effective for reducing the thousands of Old French-derived lexemes to the fifty-two discussed in this chapter.

The MED includes multiple possible filters through which to examine its entries, giving you the option of limiting your search by subject, part of speech, and source language. Specifying a period was not available nor required due to the specialised nature of the MED (as indicated in the name). It did not seem relevant to restrict the search by subject since semantics had not been selected as a factor, but since open-class words are the most frequently borrowed, focusing on nouns, verbs, and adjective was logical. Furthermore, such constraints provided a broad spectrum of possible loans whilst eliminating closed-class words that were less likely to contain interesting features (Haspelmath & Tadmor, 2009).

The source language filter was more complicated, however. The MED lists 12,063 words as being derived from Old French, which is too many to analyse for this research. By searching for forms derived at least in part from Old Northern French (ONF) and Continental French (CF), the number of solely Anglo-Norman words was reduced, as was influence from Latin, Occitan, and Germanic languages beyond English. As a result of these filters, the total number of potential lexemes became a much more manageable 176.

The second stage of MED analysis consisted of searching the variant spellings of the chosen lexemes to find any evidence of the dialectal features associated with Champenois and Picard. Any dialectal remnants would have been evidenced in spelling, so it was the multitude of variant spellings for each lexeme that were examined, with reference to the etymologies given in the MED for the ‘standard’ Old French equivalents. All potential features were considered, both those identified as weak (one x) and strong (two x). Therefore, the selected features were:

7. [ɛ̃] (+n) stays [ɛ̃]
10. [k], [g] for *ch, j*
11. *c* (+*e, i*) > *ch*
12. *eau* > *iau*
13. [e:] > *ei*
15. Final *z* > *s* 12th c.
16. Use of *w*
18. Final *t* remains
20. *a* > *ai*

Despite ‘17. *la* > *le*’ being a strong feature in Picard, because it refers to the definite article it was excluded from the list.

If the lexeme and/or its variant spellings and etymology indicated the potential presence of one of the above features, they were recorded in a spreadsheet along with the date of first citation as noted in the MED and the possible feature or features evidenced. To facilitate easier application to LAEME search parameters, the present-day English equivalent (or most modern version) of the lexeme was also recorded, using the links to the Oxford English Dictionary provided in each lexeme’s MED page.

An example of this process can be illustrated with ‘sc̃arnen’. The MED gives multiple possible etymologies for this word: OF ‘escharnir’, Anglo-French (AF) ‘charnir’, AF/ONF ‘escarnir’, and ONF ‘eskarnir’. These potential sources do not agree on whether the first consonant is [k] (spelled <c/k>) or [ʃ] (typically spelled <ch>), which is itself evidence for feature 10. This same discrepancy is noted in the variant spellings listed in the MED: ‘scorn(e)’, ‘scoren’, ‘skorn(e(n))’, ‘schorne(n)’, ‘scorni(e)’, ‘skorni’,

‘scarne(n)’, ‘skarnen’, and ‘scoarnen’ are just the possible ways of spelling the infinitive. Based on the feature description given by Einhorn and the etymology given in the MED, it can be determined that the spellings featuring <c/k> were evidence for this feature being borrowed into English and these forms warranted further investigation.

3.4 Data

The effect of both stages of analysis resulted in whittling down the entire MED to a total of fifty-six lexemes, whose variant forms clearly exemplified just three OF dialectal features: numbers 10, 11, and 16⁴. Forms featuring <w> where more standard OF forms use <g> (16) was the most common feature with 39 instances, with orthographic representations of [k] and [g] instead of <ch> and <j> (10) making up the majority of the remainders with 17 occurrences. There are just two instances in this data set of <c> becoming <ch> when followed by <e> or <i> (11), but those familiar with modern French dialects and Dany Boon films would recognise this feature as still being present in the ‘ch’ti’ of Picardy, so it is important to include. Two of the lexemes featured evidence for two features: ‘wiket’ and all its variant forms are examples of both features 10 and 16, as is ‘wāğǫur’.

I also chose to organise the earliest possible dates of first citation of each lexeme in the data set in chronological order. For a large number of manuscript sources, dating can be more of an approximation, hence the earliest possible date of composition is used. The results displayed in figure 1 indicate that the majority of loanwords in this data set were borrowed between 1200 and 1400, with only a few words appearing outwith these boundaries. The absolute peak of this borrowing was during the fourteenth century, which confirms existing research which concludes that the majority of OF loanwords came into English during the switch in bureaucratic language from French and Latin to English.

⁴ The full list of lexemes can be found in Appendix A.

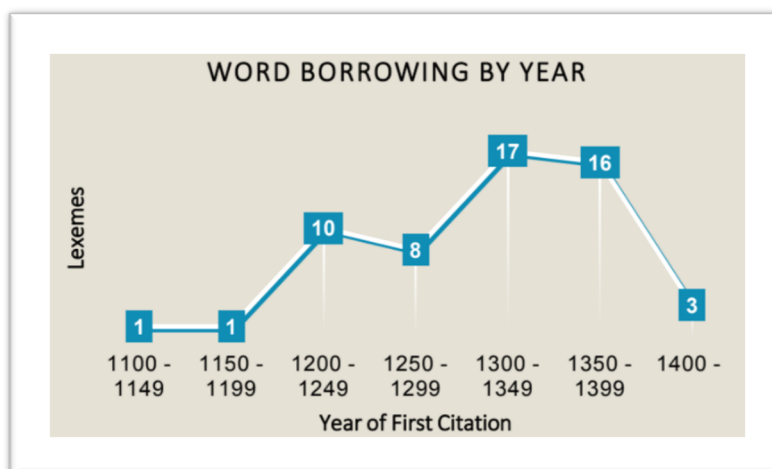


Figure 1: Number of OF loanwords in this dataset borrowed into English between 1100 and 1400.

3.5 Evaluation

This is a challenging period in which to examine etymologies due to the sheer amount of doubt and confusion when it comes to etymologies. ‘Strācūr’ is proposed to have come from an unknown AN or ONF form of the OF ‘estrace’, but because the MED only has one citation of the word from one source text there is very little to go on when trying to uncover a more solid etymology. Other words, like ‘ward(e)’ and ‘waider’ have been determined to be quite resolutely Anglo-Norman in origin, but that does not mean that other forms did not derive from elsewhere. Furthermore, both features 10 and 16 are more well-attested in Picard than in Norman and Anglo-Norman, so it is reasonable to look at Picard origins in these features before examining Norman paths.

When identifying possible OF dialectal features in the variant spellings, the lack of a standard spelling system was both a blessing and a curse. Consonantal features were easy to pinpoint, but due to the variability of vowels, as well as sound changes that have occurred over the years, it was significantly more challenging to find evidence for vowel-centric features. Furthermore, English has no definite way of marking nasal vowels in nativised words, so feature 7 relating to [ɛ̃] was impossible to identify given the constraints of this projects. Given more time and a larger data pool, greater focus on vowels could have been achieved, but the evidence for the consonantal features previously mentioned was solid enough to be sufficient.

Also, whilst it is reassuring that the data reflects previous research relating to when lexical borrowing was at its peak during Middle English, one must also take the linguistic situation in France into account. The decline of the northern towns and rise in status of Paris occurred during the fourteenth century, so in order to increase the likelihood of minimal Parisian influence on the OF dialects earlier loanwords are preferred. However, that does not mean to say that the words borrowed during the 1300s should be discounted entirely- the growing influence of Paris and the desire to conform to an incipient standard should be taken into account for later loanwords.

4 Stage 2: LAEME Forms and Frequencies

4.1 About LAEME

LAEME is described by its creators Margaret Laing and Roger Lass as the ‘daughter atlas’ of ‘A Linguistic Atlas of Late Middle English’ (LALME) but is quite different in scope and purpose. It is comprised of a corpus of searchable tagged texts and a searchable database of information about the manuscripts and texts used, in addition to numerous explanatory documents. The corpus contains 650,000 words, where each lexical item and derivation and inflectional morpheme is tagged in incredible detail.

The production of LAEME began after the publication of its predecessor ‘A Linguistic Atlas of Late Middle English’ (LALME) in 1987. The initial plan of using a questionnaire-based method of manuscript analysis was abandoned in favour of lexico-grammatical tagging, facilitating the creation of the corpus. The data gathered has facilitated greater research into early ME scribal practises and writing systems, as well as providing a large repository of early ME texts that is a lot easier to work with than searching for each manuscript manually.

4.2 Purpose

This second phase takes the lexemes and variant spellings extracted from the MED and applies them to LAEME in order to gather more detail about those forms. More specifically, they can be used to examine more in-depth the uses and distribution of variant spellings in specific texts. Furthermore, the data subsequently gathered on the frequency of these forms across texts provides a basis for identifying the manuscripts of greatest interest for further examination in the next stage of analysis.

4.3 Method and Analysis

The method used for this stage of data-gathering consisted of searching the ‘Tag Dictionary’ part of LAEME for the present-day English equivalent of each MED lexeme; where possible, the MED gives a link to the Oxford English Dictionary (OED)’s entry on the word. This is not always possible, however, such as with ‘scapelen’ and ‘waider’, so these words had to be omitted from the LAEME search. To further complicate matters, the OED had entries for a number of MED lexemes, such as ‘warisoun’, but LAEME would not accept this as a lexeme — ‘garrison’ would have to be inputted and forms beginning with <w> were extracted from the variant forms manually.

LAEME offers a wide scope of criteria for searching the tag dictionary, such as using suffixes, grammatical words, numbers, and other types of tag as the search string, in addition to restricting the scope of the search to the beginning, middle, end, or entirety of the ‘lexel’. In this instance, using a lexical string and broadening the scope to the entirety of the lexel produced the best results.

Each variant spelling for the lexemes was recorded in a spreadsheet along with their frequency and the number of texts the forms appear in. LAEME is incredibly thorough when it comes to grammatically tagging each form, distinguishing between nouns and nouns that function as direct objects, even going so far as to note which forms appear in a rhyming position. Such a high level of detail can be useful when focusing on specific spellings, but for a wider overview, the grammatical functions were reduced to their broader categories as recorded from the MED. This resulted in some forms being repeated within the wider

word classes, so these duplicates were removed in order to simplify the resulting data.

4.4 Data

A surprisingly large number of words taken from the MED were not recorded in LAEME - just over half of the fifty-six words extracted. The result was a grand total of twenty lexemes identifiable in LAEME, of which four were different grammatical classes of two words: ‘warrant’ exists in this data set as a noun and a verb, and ‘waste’ as a noun and an adjective. Without levelling the duplicate spellings resulting from the broadening of grammatical categories, those twenty lexemes have one hundred and thirty-four forms between them; with levelling, that number reduces to ninety-five. The words with the highest number of variant spellings are ‘wait’ and ‘war’, which is not surprising given the period in question. To limit confusion, the total number of forms without duplicates is used for all further discussion.

Table 2: *Highest frequency spelling variations of lexemes in this dataset across texts in LAEME.*

PDE	Form	Freq.	Texts
war	peorre	14	11
warden	pardeins	11	3
waste	pesste	11	3
snake	snaken	8	7
war	werre	8	4
wain	wa33n	5	1
wait	wayteþ	5	4
warden	pardains	5	2
waste	peste	5	4
waste	peste	5	4
wile	piles	5	4
warnish	warniste	4	3
war	wer	4	4
war	weorre	4	3
war	perre	4	3

In terms of frequency of forms, the vast majority of spelling forms identified are only found once in one text, according to LAEME. Fifty-eight forms have just one citation in one text, with a further seventeen appearing twice across one or two texts. The fifteen highest frequencies are listed in Table 2, with the entire list given in Appendix B.

4.5 Evaluation

First of all, the discrepancy between the number of lexemes identified in the MED and in LAEME is quite dramatic but easily explained. The MED aims to be comprehensive in its study of ME, which by nature involves recording words which possibly have only one citation in one text. By contrast, LAEME's focus is on the texts themselves rather than the breadth of words covered in the corpus, so it is logical that there are words not found in their manuscripts that have evaded tagging. The early ME period is notorious for being hard to quantitatively study due to the availability of material, so the discrepancy between the MED and LAEME's sources is to be expected.

The number of forms with just one citation is also to be expected due to the nature of ME. With the lack of a standard spelling system, countless spelling variations are possible. However, not all forms are made equal. In the steady creep toward standardisation, some spellings are more commonplace than others, as indicated by the high frequency of 'peorre' as a form of 'war' across texts shown in Table 2. Just as some spellings are preferred, others exist but are more peripheral for any number of reasons. It is reasonable to assume that 'uierre' is rare because of its orthography, and the quirky 'peorrre' because of its potentially erroneous triple <r>.

Furthermore, the fact that LAEME distinguishes forms that appear in rhyme position from those that do not could be thought to affect the number of one-time spellings. The manipulation of spelling in order to fit a certain rhyme scheme is not unheard of during the ME period, but it does not seem that the role of rhyme is significant in this instance. Eighteen forms are described as occurring in rhyme position, yet they are relatively evenly distributed between the highest and lowest frequency forms. This indicates that the significance of rhyme in affecting spelling is relatively low for our purposes and can therefore be discounted as a major factor to consider. If the purpose of this research were to be analysing the variant forms identified in the MED and LAEME in and of themselves, these orphan spellings would be discounted. Nevertheless, by grouping these forms by text, they can play a more important part in identifying larger patterns within texts as a whole.

5 Stage 3: LAEME Texts

5.1 Purpose

The final and most lengthy portion of the analysis puts the word and feature lists from the previous chapters into context and examine their citation texts more fully in order to determine the likelihood of these variant spellings deriving from French dialects that are not AN or CF. Through using Einhorn's dialectal feature table, in addition to taking each texts' metadata into account, the aim is to identify the most prominent OF dialect in each manuscript (if any) and examine the possible reasons behind this, whether more general, such as relating to culture or historical events, or more text-specific, such as the idiosyncrasies or personal connections of the scribe or author.

5.2 Method

The initial identification of each form's text(s) involved creating an item list through LAEME, which listed the identification number of each text the feature in question appeared in. By putting all of these numbers for all of the features in numerical order, the texts with the highest total of relevant features were easy to determine. However, each text does not necessarily correspond to an entire manuscript- where a manuscript

is known to have been written in multiple hands, the manuscript is often split up with each text entry in LAEME consisting of the portions of that manuscript written in a specific hand. In order to gain the fullest picture, where a manuscript was divided into multiple searchable texts, each text was examined, regardless of whether that specific text featured one relevant spelling or eight. Grouping together related texts made it significantly clearer how many manuscripts were to be examined and provided the broadest perspective of a longer manuscript within the confines of this research.

As with the previous stages of analysis, only the most relevant texts were selected for further examination; manuscripts (both those consisting of multiple tagged texts and those not) containing two or more interesting spellings made the cut, as did a couple – such as Egerton 613 – whose metadata hinted at noteworthy ties to France. After discounting the surplus texts, forty tagged texts remained, deriving from nineteen manuscripts.

Each tagged text was subject to more direct searching for further indications of OF dialectal spellings using Einhorn’s feature list as a guide. Examples of these indicative spellings include:

- Initial <gu>, to rule out the use of <w>
- Final <ez>, to rule out feature 15
- <che> and <chi>, to test for feature 11
- <eau> and <iau>, to test for feature 12

By examining the number of such spellings and features within each text as a whole, it is possible to attest more solidly whether these spellings indeed derive from Picard or Champenois or another dialect entirely.

5.3 Data

Of the forty texts examined, the vast majority had very little in the way of relevant forms or even French-derived vocabulary. However, five texts featured notable forms and other details that make them worthy of further elaboration.

The first text comes from MS Laud Misc. 108, which was written circa 1300 in Oxfordshire and contains rubrics written in AN and Latin in addition to the ME passages tagged in LAEME. This text is notable because of the clear French influences in the spelling; ‘beau-frere’ features the standard spelling of ‘beau’, indicating a lack of <iau> forms found in Picardy (see feature 12), as well as the clear presence of final <z> further discounting Picard – and possibly AN – as an influential dialect (see feature 15). These terminal <z> spellings are particularly interesting because a number of them are found in the French first-person plural ending <-ez> but applied to English contexts, such as in ‘we ne findez nou3t’ and ‘mani men pinchez’. The first sentence shows the correct person being used for this particular conjugation, however the second shows <-ez> being used with the third-person plural. This could be interpreted as an overextension of <-ez> usage in an attempt to make the language of this text seem more ornate and ‘French’, but such a hypothesis would require further examination. What is clearer, however, is the definite OF influence in this text, despite the inability to pinpoint precisely which dialect was most significant.

The first LAEME text from the MS of *Cursor Mundi* housed at the Royal College of Physicians is an excellent example of a mixed palette of dialectal features that make concrete conclusions challenging. The text is peppered with overtly French words like ‘danais’ and ‘delices’, and the preference for French forms of nationalities and countries is notable. The <ch> in ‘da(n)emarche’ could be read as [k] in other contexts, but given the preference for French forms elsewhere, this spelling should be taken as evidence

against feature 10, thus side-lining Picard and AN. However, the spelling of ‘curtaisi’ muddies the waters; it features the classic AN <u> (see feature 14) but the <ai> is most likely a Walloon spelling (see feature 8). Such a mix of dialects is in itself interesting, however, and definitely indicates a strong French influence in this text.

A further example of a healthy variety of OF dialectal features is the copy of ‘Ancrene Riwe’ found in Cotton Titus D xviii. The only uses of <gu> spellings are in the Latin portions of the texts, suggesting that <w> forms predominated (see feature 16). There is also no evidence for features 11 and 12, which are strong indicators of Picard or AN forms, as is feature 10, which is hinted against thanks to spellings like ‘pa{-}t(r)iarches’. The verb ‘auez’ is present, however, exemplifying feature 15, which further counts against Picard and AN but does indicate a French influence. The word ‘culuert’ could provide a clue, but the MED gives the etymology as just deriving from OF with no further details. The spelling looks to indicate the presence of <u> forms (see feature 14), which would contradict the lack of AN spellings elsewhere in the passage. All in all, it is a rather confusing picture that is painted, but OF does affect at least some of the forms used.

Digby 86 is worth an honourable mention, as it is a manuscript in which half the material is in French. With other manuscripts containing other languages, LAEME’s transcriptions skip over the French (but the Latin is often included), but in this instance some of the French has been transcribed and was searchable as a part of the English portions of the manuscript. There was no evidence in either OF or ME of the <che> and <chi> forms indicative of feature 11, and the presence of both terminal <z> and <eau> forms in the phrase ‘Les diz de seint bernard / comencent Ici tresbeaus’ suggest that the OF of the manuscript is not Picard. Unfortunately, very few of the interesting features of the French carry over to the English passages, but the ability to perform a direct comparison in situ without having to resort to the manuscript facsimile was noteworthy.

5.4 Arundel 57

Despite all these crumbs hinting at the presence of OF dialects, the strongest evidence for their presence comes from Arundel 57, otherwise known as ‘Ayenbite of Inwyt’. The manuscript was written at St Augustine’s in Canterbury and is known to have been completed on 27th October 1340 (Gradon, 1979, p. 1). The text is the only known ME translation of the OF ‘Somme le Roi’ – which was originally composed around 1280 for the children of Philip III of France (‘Laurent D’Orléans, ‘La Somme le Roi,’ n.d.) – and was written (and possibly translated) by Dan Michel, originally a secular clerk who became a priest at St Augustine’s in 1296. The abbey is known to have had two copies of ‘Somme le Roi’, of which one was owned by Michel himself and was his most likely exemplar. This copy has not survived to the present day, although the other is found in Cotton Cleopatra A (Gradon, 1979, pp. 53–4). It has been noted that the language of ‘Ayenbite’ is quite archaic for its period, explained by the fact that Michel was quite an old man when writing it (Gradon, 1979, p. 12). The LAEME transcription of the manuscript also includes many annotations concerning mistranslations, suggesting that OF features would not be direct copies from the exemplar, but in fact subtler influences if not from Michel’s own system.

The ‘Frenchness’ of Michel’s spelling has been well-recorded and discussed, most notably with regard to his vowels. His use of <ou> to represent [u] follows the French system (Gradon, 1979, p. 14), and Wallenberg notes that his <i/y> spellings could represent an analogy with French, in addition to representing [je:] (Wallenberg, 1923, p. 121). However, his spellings do not correspond with AN usage. With regard to Michel’s use of <ie>, Gradon notes that ‘if Dan Michel was familiar, not only with the AN

pronunciation in words such as *chef*, but also with the CF pronunciation, he could just as well have used <ie> as a graph for [ie] as for [e:]_n (Gradon, 1979, p. 32). In addition, Gradon observes that the variation in Michel's use of <o> cannot be explained by aligning these spellings with AN sound change. Instead, CF nasalisation of [a] must be taken into account in order to make sense of some of the uses (Gradon, 1979, p. 40). The result of such confusion is more the ruling out of AN as a source and bringing CF into the fore than providing clear-cut solutions.

The text has clear French influences that move beyond just spelling. An annotation on a use of 'þet' reads 'sic – 'that' for 'than' is a calque on French usage', which implies subtle French grammatical uses. Furthermore, the 'che' in 'che manere' appears to be a variant of *ce*, the French for 'this/that', building on this basic impact of French.

The lack of importance of AN in 'Ayenbite' is corroborated by the lack of the <u> described in feature 14, shown in 'glotounye'. Furthermore, the <ou> spelling corroborates feature 1, which is common to all dialects except AN and Picard. While there may not be a Picard flavour in that particular form, the only example of the <iau> spelling (feature 12) indicative of Picard is found in this text in 'hysiaus'. With further evidence for feature 10 in 'askapie', it cannot be ignored that Picard plays as big a role in the language of this text, if not more.

5.5 Evaluation

Across all of the texts analysed, there was a surprising amount of variation in dialectal forms used; no text favoured solely one dialect. Even with 'Ayenbite', whilst it is possible to conclude that AN was not the majority dialect preferred by Michel, it is not feasible to discern one that was. As evidenced by Gradon's analysis, taking CF into account makes sense of a lot of the vowels, as well as numerous Picard forms being found through this research. Therefore, without more in-depth research, it must be said that the overall use of OF spellings consists of a lot more variation than the historical record suggests.

6 Final Analysis

6.1 Is it possible to identify loanwords into Middle English from Old French dialects other than Anglo-Norman and Central French?

Through the course of this research, it has become clear that the first two questions examined are symbiotic in nature. It is indeed possible to identify variant forms of a lexeme that have been borrowed into English from OF, but such identification is only possible through spelling. Furthermore, these forms appear to be few and far between and when they do exist, they are incredibly difficult to find due to their scattered distribution.

It is impossible to hypothesise a spelling variation's origins from only examining dictionary material, yet it is entirely feasible to take a series of manuscripts or other texts and search for dialectal features without having previously referred to dictionaries. The benefit of having used both the MED and LAEME was that it narrowed down the Brobdingnagian collection of OF loanwords and provided a focus on the forty tagged texts explored. Further research with a previously determined set of texts would not require the initial lexicographical stages.

What is required for identifying loanwords is a reference point of OF dialects, which is where Einhorn's guidelines proved invaluable. It is also useful to have a working knowledge of ME, as certain indicative spellings occur numerous times natively in ME. For example, <chi> was an important feature to search for but occurred most often in 'child' and 'children'. A similar issue was faced with <w>, which is why it became easier to search for instances of <gu> in order to indicate the likelihood of <w> forms. Ultimately, it was the textual context of each potential form that indicates whether it was likely from an OF dialect or a victim of ME's spelling idiosyncrasies.

In terms of identifying a specific dialect of a loanword, however, this was not possible due to the lack of data in the tagged texts. As previously discussed, there was a large amount of variation in features used within texts, with some forms often contradicting each other when it came to suggesting a common dialectal origin. There was greater evidence for Picard forms than those of Champagne, however, most likely due to the greater geographical proximity to Paris resulting in its features being more familiar and widely accepted. Significant further research would need to be undertaken in order to clearly identify such forms' dialect of origin.

6.2 Can the origins of these loanwords be determined through spellings?

This is a challenging period to research because of the general spelling variability on both sides of the Channel. The difference in phonetic inventories between OF and ME means that vowels are particularly susceptible to spelling changes as they make the journey. Variation between scribes, as well as the inherent variability within ME, meant that no overarching conclusions about particular vowel-centric features could be made. It is an area which could provide a great deal of insight with further examination.

As previously mentioned, there were a great deal of spelling features that occurred in ME naturally, but filtering these out was possible, if time-consuming. What posed a greater challenge was the overlap in features, in particular between AN and Picard. While there were features specific to both dialects – such as <iau> for Picard and <u> for AN – they were not always present in the same text. Therefore, being able to distinguish between other northern dialects and AN was not always clear, but reasonable suggestions could be made, particularly in the case of Arundel 57.

6.3 What can these borrowings reveal about cultural links between England and France?

Due to the lack of data resulting from the analyses, there was insufficient evidence to even hint at the reasons behind the forms identified. Those that were present were almost randomly scattered across texts, which combined with a lack of background about the manuscripts and their scribes themselves meant that they revealed nothing about their reason for appearing in their texts. The greater evidence for more Picard forms could be explained by geography, or also the prestige of the region and its proximity to other regional centres and trading hubs, as well as England itself. This is pure speculation, however, with little substantiation to back it up.

Even with 'Ayenbite', where a great deal is known about both the manuscript and its scribe, there were very few conclusions to be drawn. The fact that the origin text 'Somme le Roi' was written for the offspring of the French king makes it unlikely to have ever been written in AN (despite what the British

Library says), yet without the exemplar that Dan Michel used to translate his version, it seems hasty to suggest that his Picard-leaning forms came from this OF manuscript.

Despite the lack of conclusions about why these dialectal forms came into ME, there is still a great deal of potential for further research with other documents whose origins are clearer and more useful.

7 Conclusion

The aim of this research was to bring together multiple resources in order to examine the relationship between medieval France and England on the level of lexical borrowing. The variability of scribal practices during the Middle Ages was the ideal medium through which to identify obscure lexical forms that have otherwise escaped notice. These forms, once unearthed, aimed to provide a gateway into a more nuanced understanding of the nature of links between England and regions of France beyond Normandy and Paris.

The MED proved to be an invaluable resource in narrowing the field to focus on open-class words from specifically the northern portion of France. Its lists of each spelling variation for each ME lexical item and known etymologies were instrumental in creating the initial dataset. Combined with the data from Einhorn concerning the features of each dialect, this dataset was tailored to indicate which features were most likely to be relevant in future stages of analysis.

Once familiarised with, LAEME provided an almost cumbersome amount of detail with regards to frequency of forms and their distribution across texts, building on the MED data. The different sources of the two resources resulted in a marked culling of forms, but the variant spellings given for those that remained proved ample. By organising these forms by text, the analysis of each individual text began, with fewer results than expected. Most manuscripts featured very few words of note, with the exception of Arundel 57, which provided a glimpse into OF usage that was decidedly more varied than solely AN or CF. Still, a large variability in forms remained, which resulted in the inability to draw hard and-fast conclusions.

The ultimate combination of the uncovered variant forms with a cultural reason for their existence was not ultimately possible beyond pure conjecture. A lack of final data, in combination with their wide distribution and the lack of knowledge about manuscripts' origins, made it impossible to focus on a particular lexical source and examine the reasons behind such lexical transmission. The possibility to undertake such research remains, with further extended examination required. The use of a more specific manuscript source – ideally one where the history of the scribe and text itself has survived – or examining non-literary texts would lead to clearer outcomes and be able to further enlighten the ties between England and more peripheral areas of France.

8 References

- Anon (n.d.) Laurent D'Orléans, '*La Somme le Roi*'. Available from <https://www.bl.uk/collection/items/illustration-of-prowess-idleness-david-labour-from-laurent-dorleanss-la-somme-le-roi> [Accessed 30th March 2020].
- Ayres-Bennett, W. (1996) *A History of the French Language through Texts*. London and New York, Routledge.
- Bec, P. (1967) *La langue occitane*. Paris, PUF.
- Benskin, M. and Laing, M. (1981) Translations and Mischsprachen in Middle English manuscripts. In: Benskin, M. and Samuels, M.L. eds. *So meny people longages and tonges: philological essays in*

- Scots and medieval English presented to Angus McIntosh*. Edinburgh, Middle English Dialect Project, pp. 55–106.
- BL Cotton Titus D xviii, fols. 14r-105r: Ancrene Riwe*. In Laing, 2013.
- Blake, N. (1992) *The Cambridge History of the English Language: 1066-1476*. Cambridge, Cambridge University Press.
- Brinton, L. and Bergs, A. eds. (2017) *The History of English: Middle English*. Berlin, De Gruyter Mouton.
- Burnley, D. (1992) *Lexis and Semantics*. Cambridge, Cambridge University Press.
- Burnley, D. (2000) *The History of the English Language: A Source Book*. 2nd ed. Harlow, Pearson Education.
- Dance, R. (2014) Getting a word in: Contact, etymology and English vocabulary in the twelfth century. *Journal of the British Academy*, 2, pp. 153–211.
- Durkin, P. (2014) *Borrowed Words: a history of loanwords in English*. Oxford, Oxford University Press.
- Edinburgh, Royal College of Physicians, MS of Cursor Mundi, entry 1*. In Laing, 2013.
- Einhorn, E. (1974) *Old French: a concise handbook*. London, Cambridge University Press.
- Gradon, P. (1979) *Dan Michel's Ayenbite of Inwyt*. Early English Book Society. Oxford, Oxford University Press.
- Haspelmath, M. and Tadmor, U. eds. (2009) *Loanwords in the world's languages: a comparative handbook*. Berlin, De Gruyter Mouton.
- Huchon, M. (2003) *Histoire de la langue française*. Paris, Librairie Générale Française.
- Kibbee, D. (1991) *For to speke Frenche trewely: the French language in England, 1000-1600: its status, description and instruction*. Amsterdam, Benjamins.
- 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> (28 March 2020)
- Lodge, R. (1993) *French: from dialect to standard*. London, Routledge.
- London, British Library, Arundel 57, fols. 2r-4r; 13r-96v: Ayenbite of Inwyt, Pater Noster etc*. In Laing, 2013.
- McIntosh, A. (1973) Word geography in the lexicography of medieval English. *Annals of the New York Academy of Sciences*, 211, pp. 55–66.
- McSparran, F., et al. 2000-2018. *Middle English Dictionary*. University of Michigan Library. <https://quod.lib.umich.edu/m/middle-english-dictionary/dictionary> (28 March 2020).
- Monfrin, J. (1968) Le mode de tradition des actes écrites et les études de dialectologie. *Revue de linguistique romane*, 32 (125–126), pp. 17–47.
- Nezirovic, M. (1980) Le vocabulaire dans deux versions du Roman de Thèbes. Clermont-Ferrand, *Association des Publications de la Faculté des Lettres de Clermont-Ferrand*.
- OED Online*. Oxford University Press. <https://oed.com> (28 March 2020)
- Oxford, Bodleian Library, Digby 86, all the texts that are suitable for mapping*. In Laing, 2013.
- Oxford, Bodleian Library Laud Misc 108, part 1, fols. 1r-200v, hand A: Life of Christ, Infancy of Christ, South English Legendary*. In Laing, 2013.
- Pope, M. (1966) *From Latin to modern French with especial consideration of Anglo-Norman phonology and morphology*. 2nd ed. Manchester, Manchester University Press.
- Rothwell, W. (1998) Arrivals and Departures: The Adoption of French Terminology into Middle English. *English Studies: A Journal of English Language and Literature*, 79 (2), pp. 144–165.

Thompson, J. (1960) *Economic and social history of Europe in the later Middle Ages (1300-1530*. New York, Ungar.

Wallenberg, J.K. (1923) *The Vocabulary of Dan Miche's Ayenbite of Inwyt*. Uppsala.

9 Appendices

9.1 Appendix One: Full MED Lexemes

Earliest Possible Citation	PDE Equivalent	Lexeme	Part of Speech	Feature
1131	wile	wīle	N	16. Use of w
1160	war	wer(re)	N	16. Use of w
1200	scar	scar	N	10. [k], [g] for <i>ch</i> , <i>j</i>
1200	warrant	warant	N	16. Use of w
1200	ward	ward(e)	N	16. Use of w
1200	warden	wardein	N	16. Use of w
1200	waste	wāst(e)	N	16. Use of w
1200	wimple	wimple	N	16. Use of w
1225	ward	warden	V	16. Use of w
1335	wait	waiten	V	16. Use of w
1225	wicket	wiket	N	16. Use of w
1230	scorn	scōrnen	V	10. [k], [g] for <i>ch</i> , <i>j</i>
1255	wimble	wimble	N	16. Use of w
1273	-	waider	N	16. Use of w
1275	scald	scalden	V	10. [k], [g] for <i>ch</i> , <i>j</i>
1275	ware	wāren	V	16. Use of w
1278	skew	skeu	N	10. [k], [g] for <i>ch</i> , <i>j</i>
1287	-	strācūr	N	10. [k], [g] for <i>ch</i> , <i>j</i>
1295	scoop	scōpe	N	10. [k], [g] for <i>ch</i> , <i>j</i>
1299	snake	snak	N	10. [k], [g] for <i>ch</i> , <i>j</i>
1300	scarce	scārs(e)	N	10. [k], [g] for <i>ch</i> , <i>j</i>
1300	partridge	partrich(e)	Aj	10. [k], [g] for <i>ch</i> , <i>j</i>

1300	archer	archēr, -iēr	N	11. c (+e,i) > ch
1300	warray	werreien	V	16. Use of w
1300	warison	warisōun	N	16. Use of w
1300	warnement	warnement	N	16. Use of w
1300	waste	wāst(e	Aj	16. Use of w
1303	wardecorps	warde-cors	N	16. Use of w
1318	warnestore	warnestōr(e	N	16. Use of w
1328	-	waspail	N	16. Use of w
1330	warnish	warnishen	V	16. Use of w
1330	warish	warishen	V	16. Use of w
1330	waynpain	wain-pain	N	16. Use of w
1333	wain	wain	N	16. Use of w
1335	were	wēr(e	N	16. Use of w
1338	warnison	warnisōun	N	16. Use of w
1338	-	warnisōur	N	16. Use of w
1350	kennet	kenet	N	10. [k], [g] for ch, j
1350	-	werpishen	V	16. Use of w
1350	wager	wāğour	N	16. Use of w
1350	waynoun	wainōun	N	16. Use of w
1359	wode	waid(e	N	16. Use of w
1364	-	scapelen	V	10. [k], [g] for ch, j
1373	warence	warance	N	16. Use of w
1373	wasp	wasp	N	16. Use of w
1375	wallop	walop	N	16. Use of w
1376	wernard	wernard	N	16. Use of w
1378	wardrobe	warde-rōbe	N	16. Use of w
1382	wage	wāğe	N	16. Use of w
1387	scarcity	scārsetē	N	10. [k], [g] for ch, j
1387	botch	bocche	N	11. c (+e,i) > ch
1390	warrant	waranten	V	16. Use of w
1391	strick	strik(e	N	10. [k], [g] for ch, j
1400	tuck	tukken	V	10. [k], [g] for ch, j
1400	warrok	warroken	V	16. Use of w

1404	skellat	skellet	N	10. [k], [g] for <i>ch</i> , <i>j</i>
------	---------	---------	---	--

9.2 Appendix Two: Full LAEME Lexemes

No.	PDE	Form	Freq.	Texts
1	war	peorre	14	11
2	warden	pardeins	11	3
3	waste	pesste	11	3
4	snake	snaken	8	7
5	war	werre	8	4
6	wain	wa33n	5	1
7	wait	wayteþ	5	4
8	warden	pardains	5	2
9	waste	peste	5	4
10	waste	peste	5	4
11	wile	piles	5	4
12	warnish	warniste	4	3
13	war	wer	4	4
14	war	weorre	4	3
15	war	perre	4	3
16	wain	wayne	3	1
17	wait	paiteden	3	2
18	warden	pardein	3	3
19	waste	west	3	3
20	war	porre	3	3
21	scarce	scarse	2	2
22	scorn	scornunge	2	1
23	scorn	scarned	2	2
24	snake	snakes	2	2
25	wain	wayn	2	2
26	wain	waine	2	1
27	wait	waite	2	2
28	wait	waites	2	2
29	warrant	waraunt	2	2

UNIVERSITY OF ABERDEEN
16-18 APRIL 2021

30	warrant	parant	2	2
31	waste	weste	2	2
32	were	pere	2	1
33	war	worre	2	2
34	war	were	2	2
35	war	werres	2	1
36	wimple	wimpel	2	2
37	wimple	pimpel	2	2
38	archer	archer	1	1
39	partridge	pertrich	1	1
40	scald	scalden	1	1
41	scald	scoldeþ	1	1
42	scald	scaldant	1	1
43	scorn	scornige	1	1
44	scorn	scærninge	1	1
45	scorn	schornigis	1	1
46	scorn	schorningis	1	1
47	scorn	scorneþ	1	1
48	scorn	scornes	1	1
49	scorn	scorne	1	1
50	snake	snake	1	1
51	wain	pein	1	1
52	wain	wain	1	1
53	wain	weyn	1	1
54	wain	peines	1	1
55	wait	wait	1	1
56	wait	waiten	1	1
57	wait	wayte	1	1
58	wait	wayten	1	1
59	wait	paitinge	1	1
60	wait	waytinges	1	1
61	wait	waites+*	1	1
62	wait	paited	1	1
63	wait	waiten	1	1
64	wait	paitid	1	1

PROCEEDINGS OF ULAB XI

65	wait	Waitand	1	1
66	warrant	warant	1	1
67	warrant	warand	1	1
68	warrant	warauntie	1	1
69	warden	+wardein*	1	1
70	warden	wardein	1	1
71	warish	warist	1	1
72	warish	i-warisd	1	1
73	garison	wareisun	1	1
74	warnish	warnising	1	1
75	warnish	warnisit	1	1
76	warnish	warnist	1	1
77	waste	Waast	1	1
78	waste	paste	1	1
79	waste	paste	1	1
80	waste	wēste	1	1
81	war	werre	1	1
82	war	uuerre	1	1
83	war	peorrrre	1	1
84	war	weorra	1	1
85	war	pere	1	1
86	war	wers	1	1
87	war	peorren	1	1
88	wile	wyl	1	1
89	wile	wyle	1	1
90	wile	pilis	1	1
91	wile	pil[l]es	1	1
92	wile	wyles	1	1
93	wimple	wimpil	1	1
94	wimple	winpil	1	1
95	wimple	pinpel	1	1

*precedes/follows another noun

Whose Testimony is it? Institutional Influence in the 1641 Depositions

Alex Brownless

Northumbria University

Abstract. The 1641 depositions are oral witness testimonies describing the experiences and losses of (mainly) protestant settlers during the seventeenth-century Irish rebellion (Trinity College Dublin Library, 2010). This study explores the possibility of institutional influence in these witness documents by conducting an authorship attribution analysis using corpus linguistic methods. Building on the notion of idiolect (Coulthard, 2004), this study applies Kredens' (2002) concept of idiolectal style, concerned with the unique ways an individual uses language. This analysis investigates similarities that reoccur through multiple depositions that may suggest the presence of an overarching institutional idiolectal style. To achieve this, a corpus of forty depositions was constructed and analysed on WordSmith Tools (Scott, 2020). The results show similar linguistic constructions in concordances, including identical strings reoccurring throughout multiple depositions. These results suggest that the commissioners influenced the testimonies and indicates the presence of an anonymising institutional narrative. The findings of this study have implications for the credibility of the depositions. Taken together, the analysis of institutional influence has shown an overwhelming tendency of high-frequency structures throughout the corpus. This study argues that these patterns can be attributed to the commissioner's idiolectal style as a co-author of the texts. Therefore, when engaging with these historical documents, this study emphasises that we must consider the broader historical context and the authors' institutional intentions as 'Hidden puppeteers' in the narrative (Goodich, 2006 cited in Johnston, 2010, p. 163).

Keywords: 1641 depositions; authorship analysis; corpus linguistics; idiolect; Irish history; witness testimonies

1 Introduction

The 1641 depositions function as a record of the outbreak of the seventeenth-century Irish rebellion. These depositions consist of oral witness accounts written by a scribe and in the presence of a commissioner (Grund & Walker, 2011, p. 15). They describe the experiences and losses of mainly protestant settlers from all social backgrounds (Trinity College Dublin Library, 2010). Today, transcriptions of the original depositions are fully digitalised and available online to encourage engagement with Irish history and challenge the myths and propaganda of the rebellion. They are further utilised in research as a rich source of information on Early modern Irish life and considered intrinsic to our understanding of the massacre of protestant settlers that ignited the rebellion (Ohlmeyer, 2009, p. 55). Having said this, linguists and historians are beginning to question the credibility of these testimonies, such as the Language and Linguistic Evidence project (2010). Fennell-Clark (2011) asks, 'Can we distinguish directly reported incidents and eyewitness accounts from hearsay, i.e. reports of reports? And can we detect the influence of the clerks and the commissioners in the 'manipulation' of the evidence?' (p. 27).

The present essay aims to explore these questions proposed above by Fennell-Clark (2011) through using corpus linguistic methods. Section 2 will provide an overview of key literature and concepts,

including the historical context, multi-levelled speech reporting, and authorship analysis. Developing on this, Section 3 will then outline the methodological framework of the following analysis. Finally, in Section 4, this essay will conduct an empirical study to investigate the institutional influence of a deponent's testimony. The analysis is divided into two categories: explicit institutional influences in Section 4.1 and implicit institutional influences in 4.2. Overall, the essay will conclude that there is an abundance of institutional influence in the testimonies, as shown through the analysis in Section 4. These findings have significant implications for the reliability of the depositions.

2 Literature review

2.1 Historical context

The depositions are quasi-legal witness documents concerning the outbreak of the Ulster Rising in 1641 and subsequent civil war throughout Ireland. A commission of eight Church of Ireland clergymen, headed by Henry Jones, were appointed to gather the depositions with two main objectives. Firstly, to compensate those dispossessed by the Irish rebels, and secondly to illustrate the 'great cruelties' endured by the Protestant community (Darcy, 2013, p. 85). Edited segments of these testimonies were then published through English printers and disseminated in order to solicit financial and military aid from the English parliament (Darcy, 2013, p. 85).

This selective editing positioned protestant settlers as victims of the rising and distorted their experiences of the rebellion (Cope, 2001, p. 370). As a result, the 1641 depositions became a highly contested source of evidence of the events (Darcy, 2013, p. 85). Catholic commentators argued that the colonial administration and the deposition commission intended their findings to be utilised as anti-Irish and anti-Catholic propaganda (Darcy, 2013, p. 3). Despite the apparent political, social, and economic grievances of the Catholic gentry, the commission portrayed the rebellion as a religiously motivated event and not as a result of improper colonial governance (Darcy, 2013, p. 101).

2.2 Speech reporting: Hearsay evidence

Reported speech is a representation of earlier discourse that may come from the deponents themselves or other participants at a previous speech event. Any reported speech that the deponent recounts to the commission defined as hearsay evidence (Section 114 (1) CJA, 2003). Yet, considering that the retelling of an event or experience is a vital aspect of the genre of legal testimonies (Walker & Grund, 2017, pp. 1-2), the depositions primarily consist of multi-layered discourse. As a result, the published deposition is a product of several layers of reported speech, including hearsay (Kytö et al., 2007, pp. 68-69).

To consider the reliability of the depositions, we must first consider the validity of this hearsay evidence. This is a difficult endeavour, as marking evidentiality, by stating the source of information being reported is not an obligation in the English language. Subsequently, speech reporters only vaguely indicate the source of their information or expect it to be inferred from the statement's context (Palmer, 1986, p.85 in Grund, 2012a, p. 1). Macleod (2012) found an overwhelming tendency of the hearsay marker 'informed' to be written in a passive construction such as 'this deponent was credibly informed' (Fennell-Clark, 2011, p. 28). These constructions omit the indication of evidentiality entirely by omitting the agent of the speech act, the 'informer'.

The omission of the informer is counteracted by the modifying adverb ‘credibly’ suggesting that the commission is concerned about the reception of hearsay evidence. This is most likely because hearsay, although mostly accepted in seventeenth-century courts, was deemed a weaker form of evidence (Wilson & Walker, 2015, p. 253). Hearsay evidence continued to be considered into the mid-eighteenth century until a negative view began to emerge (Langbein, 2003, in Grund, 2012a, p. 37). Having said this, hearsay evidence is currently admissible in witness testimonies and also in some circumstances of English courts, revealing that this form of evidence is still considered valuable today (Wilson & Walker, 2015, p. 253).

2.3 Speech reporting: Authorship

When considering authorship of historical texts, Lass (2004) establishes that documents presented today are a construction encompassing several co-authors (Grund, 2012b, p. 17). This essay will focus on the three explicit co-authors of the depositions: the deponent, the commissioner, and the scribe. These individuals have specific roles; the deponent provides the oral testimony. The scribe converts the utterance to written form, and the commissioner guides the interaction to produce a legal text.

As discussed above, authorship issues are likely to arise because of ambiguity in reportedness, making it difficult to determine the boundaries between these three authors in retrospect (Walker & Grund, 2017, p. 4). One strategy to overtly signal reported speech is the inclusion of a non-narrative tag such as the *verba dicendi* ‘saith’ (Collins, 2001, p. 5). The Language and Linguistic Evidence in the Depositions project (2010) suggest that an utterance beginning with ‘saith’ [that] may be an attempt to attribute authorship or introduce reported speech (Language and Linguistic Evidence in the Depositions, 2010).

Despite evidence of these subtle linguistic cues, it may not be possible to attribute authors to their roles systematically. Goodich (2006) acknowledges that ‘[historians] may often be painfully aware of [the court personnel’s] presence, functioning almost as hidden puppeteers’ (Johnson, 2014, p. 140). Consequently, in addition to considering the historical context discussed in Section 2.1, we must also recognise the limitations of the depositions as a source of evidence because of the ambiguity of authorship and intentions of the commission.

2.4 Authorship analysis and corpora

Forensic authorship attribution aims to identify authors of disputed or anonymous documents through the analysis of identifiable linguistic cues. Coulthard (2004) states that ‘the linguist approaches the problem of questioned authorship from the theoretical position that every native speaker has their own distinct and individual version of the language they speak and write, their own idiolect’ (p. 433). This suggests that being able to identify the language patterns associated with a document can inform us of the document’s author.

The theory of idiolect is controversial as it has yet to receive empirical support. Building on the notion of idiolect, Kredens (2002) suggests the notion of ‘idiolectal style’ as a less-idealised construct (Kredens, 2002; Turell, 2010 in Wright, 2013, p. 46). Idiolect style is not concerned with the individual’s language system itself but concerned with how the individual uses that system in unique ways. This is influenced by the context and genre and provided an analysis more relevant to authorship attributions as it recognises that the discourse is shaped by the legal genre (Wright, 2013, p. 46). This is more relevant to authorship analysis in depositions as this theory recognises that the discourse is affected by the legal genre.

Corpus methods are significantly beneficial to authorship analysis, as they provide an empirical quantitative approach to language study, in contrast to the traditional qualitative approaches. Because of this, corpus methods can avoid common criticisms associated with discourse analysis on the grounds of potential researcher bias (Baker, 2006 in Wright, 2013, p. 46). They can perform large scale comparisons and find patterns in word frequency, co-occurrence, and collocations fast and efficiently (Durant & Leung, 2016, p. 157).

Having said this, it is important to note that no method of analysis can guarantee authorship attribution or elimination (Cotterill, 2010, p. 578). Corpus methods do not bring complete objectivity to analysis, and researcher intuition remains present, such as deciding which features to investigate how to interpret their findings (Stubbs, 1994 in MacLeod, 2012, p. 116).

3 Methodology

For this study, we constructed two corpora using forty depositions from The Online Depositions Website (<https://1641.tcd.ie> (See appendix A)). Co-commissioners Henry Rugg and Philip Bisse collected all the depositions used for this study. The commissioners were kept as a controlled variable to avoid variations in idiolectal style. Furthermore, the date (1642-1643) and topic (apostasy) were also controlled to limit the common pitfalls occurring in corpus-based authorship analysis from temporal and contextual variations, as recognised in Cotterill (2010, p. 578).

Section 4 comprises of a two-part analysis. Firstly, in Section 4.1, we constructed a corpus from the forty depositions and transferred this to WordSmith Tools software (Scott, 2020). This enabled us to find WordLists lexis throughout the corpus that explicitly indicate the institutional presents of the commissioner and scribe.

Secondly, in Section 4.2, this study constructed a sub-corpus consisting of utterances beginning with the verba dicendi ‘saith’ from the forty depositions⁵. The sub-corpus aimed to investigate language and linguistic features that may be institutively attributed to the deponent with the tag ‘saith’ suggesting that what follows is free direct speech or reported speech.

4 Analysis and discussion

4.1 Explicit influence

This Section aims to identify markers of explicit institution influence. The three markers we have identified, codeswitching, binomials, modifiers, and others not stated below, represent legal discourse and indicate the legal genre in action. As legal discourse is goal orientated, we expect to find that these markers coincide with the commission's agenda to create a clear narrative of the events.

4.1.1 Latin influence

⁵ For the present study, a sub-corpus was not required to conduct the analysis below. However, the ‘saith’ sub-corpus was created to allow for further analysis such as standard-type token ratio comparisons in future study.

By constructing a WordList on WordSmith Tools, we were able to identify frequently used Latin lexis throughout the corpus, as shown in figure 1. Although there was a significant reduction in the influence of Latin in the sixteenth and seventeenth centuries as a result of British colonialism, its historical legacy remained in legal texts (Durant & Leung, 2016, p. 35).

N	Word	Freq.	%	Texts	%	Dispersion	Lemmas
1	FOL	117	0.52	40	100.00	0.90	
3	VIZT	50	0.22	n/a	n/a	n/a	VIZT[28] VZT[22]
4	CORAM	41	0.18	40	100.00	0.83	
2	JURAT	40	0.18	39	97.50	0.83	
5	ANNUM	31	0.14	19	47.50	0.75	
6	COIBUS	14	0.06	12	30.00	0.68	

Figure 1: Latin lexis in the corpus.

The three highest occurring Latin tokens are FOL (folio), meaning ‘a unit for measuring the length of a legal document’, VIZT (videlicet), meaning “namely” and CORAM, meaning “in the presence of”. It is evident from these definitions that Latin serves a legal function in the text. The presence of these reoccurring throughout all the depositions suggests that they are a Legal marker that holds formulaic importance in the texts. This is corroborated by the concordance of ‘VIZT’ that shows its use is highly formulaic and constrained to the specific purpose of stating value.

N	Concordance	File
1	goods & Chattles to the seuerall values following virt worth 1097 li. sterling Of Cowes heiffers horses	Nicholas Roberts
2	goods & chattells to the seuerall values following virt worth 102 li. 0.0 Of Cowes, heifers, bulls, steer	Richard Winchester
3	oods & Chattles to the seuerall values following virt worth Caxviij li. Of Cowes one bull heiffers year	William Hodgins
4	ds and Chattles to the seuerall values following virt worth 765 li. Of Cowes heiffers mares horses Coult	Thomas Browne
5	goods and Chattles to the seuerall values following virt worth 54 li. 10 s. Of horses to the value of four	Thomas Jones
6	ods and Chattles to the seuerall values following virt worth 4969 li. Of Cowes oxen heiffers mares horse	Peeter Mainell
7	les to to the value of seuerall values following virt value of 729 li. Of Cowes heiffers steeres yearli	Peeter Peacocke
8	nd Chattles]] to the seuerall values following virt value of 27 li. 10 s. Of Cattell a s Cowes yearl	Thomas Turner
9	ds and Chattles to the seuerall values following virt value of 64 li. 12 s. Of Cowes heiffers bulls cal	William Tomlinson
10	ods & Chattles to the seuerall values following virt Of steeres one Come & an heifer yearlinge sheep &	Richard Charing
11	ods and Chattles to the seuerall values following virt Of household stufte to the value of forty shillia	Henry Briggs
12	Chattles to the se u erall values following virt virt Of cowes steeres oxen heif f ers mares horses & c	Ann Graham
13	rth 86 li. 5 s. to the seuerall values following virt Of Cattell as Cowes and yearlings and heiffers to	Richard Canes
14	ods and Chattles to the seuerall values following virt 196 li. Of Cowes horses and sheepe to the value o	William Kingswell
15	and Chattles to the se u erall values following virt Of cowes steeres oxen heif f ers mares horse	Ann Graham
16	ds and Chattles to the seuerall values following virt value of 70 li. 10 s. Of Cowes heiffers calves on	Thomas Canes
17	ds and Chattles to the seuerall values following virt the value of 203 li. Of Cowes mares and coults to	William Harrish
18	ds & chattells, to the seuerall values following, virt worth 760 li., Part consisting of debts owing by	Thomas Whiteby
19	ods & Chattles to the seuerall [values] following virt worthe 110 li. Part consisting of debts due by Re	William ward
20	ds and Chattles to the seuerall values following virt value of 89 li., Part consisting of a debt due by	Thomas Wills
21	goods & chattells to the seuerall values following virt worth 950 li. partly by the loss Of the one hal	Margery Hazard
22	s & chattells, to the seuerall values follow i ag virt worth 92 li.	Leonard Webber
23	oods & chattells to the seuerall values following virt: worth 685 li.	Edward Clare
24	oods & chattles to the seuerall values following virt worth 376 li.	Wilyam Haynes
25	goods and Chattles to seuerall values followinge virt worth 314 li.	John Symson
26	goods & Chattles to the seuerall values following virt worth 157 li.	Danyell Spratt
27	ds & Chattles to the seuerall values following virt worth 153 li. 19 s. 8 d.	John Bradish
28	s and Chattles to the seuerall values following virt worth 142 li.	Elizabeth Cooke
29	goods and Chattles to the seuerall values followinge virt worth 725 li.	Ambose Martin
30	goods and Chattles to the seuerall values followinge virt. worth 690 li.	Augustine Kingsmill

Figure 2: Concordance of ‘VIZT’.

4.1.2 Binomials

Examining the concordances of the coordinating conjunction ‘and’ identified an abundance of binomial phrases throughout the corpus. A primary motivation of legal texts to link two synonymic concepts is to avoid ambiguity. Binomials increase the precision and all-inclusiveness of a concept and may have a stylistic function as a significant feature of the legal genre (Bhatia, 1993 in Lehto, 2017, p. 261). This could explain the use of the binomial pair ‘deposeth and saith’ shown in Figure 3. Firstly, it indicates the legality of the discourse, and secondly, it correlates with the trend in Early modern legal English to become increasingly verbose (Hiltunen, 1990 in Lehto, 2017, p. 261).

N	Concordance	File
1	retendants in the Province of Munster &c deposeth and saith That vpon the 25th day of December 1641 & s	Peeter Mainsell
2	sworne & examined before vs by vertue &c deposeth and saith deposeth That on or aboute the 28th of Octo	Thomas Browne
3	rne & examined before vs by vertue of &c deposeth and saith. That on or aboute the 22th of december las	William Tomlinson
4	e shi(linges) He being further examined deposeth and saith That aboute the tin(e) aboute mentioned the	John Ward
5	by vertue of his Maiesties Comission & c deposeth and saith. That on or aboute the 24th of October 1641	Ann Graham
6	rne & examined before vs by vertue of &c deposeth and saith That on or about Candlemas last & since the	William Hodkins
7	n and examined before vs by vertue of &c deposeth and saith That on or aboute the 12th of December 1641	Thomas Cases
8	e and examined before vs by vertue of &c deposeth and saith That on or about Christmas last & since the	Isaak Graneere
9	uly sworne and examined by vertue of &c deposeth and saith that on or aboute the 23th day of December	Elizabeth Cooke
10	starts within the Province of Munster &c deposeth and saith deposeth That on or aboute the 12th of Janu	John Ward
11	e Castle of Clare This examinat likewise deposeth and saith, That within one weecke after the surrender	John Ward
12	rne & examined before vs by vertue of &c deposeth and saith That on or aboute the 4th of december last	Thomas Jones
13	one hundred and forty pounds He further deposeth and saith And that Thomas og fitz Gerrald of Mornane	Peeter Mainsell
14	the true religion The deponent further deposeth and saith That aboute the sixteenth of Aprill last OI	Peeter Mainsell
15	etendants in the Province of Munster &c. deposeth and saith. That vpon the 25th of december 1641 or the	Augustine Kingesmill
16	orne and examined before vs by vertue of &c deposeth and saith That on or aboute the 12th of December 1641	William Harrish
17	e and examined before vs by vertue of &c deposeth and saith That on or about the 25th day of December 1	John Harte
18	d examined bee before vs by vertue of &c deposeth and saith that on or about the first of January 1641	William Kingwell
19	e and examined before vs by vertue of &c deposeth and saith that on or about the 29th of Nouember 1641	Thomas Turner
20	his Maiesties Comission bearing date &c deposeth and saith that on or aboute the 25th day of december	Joane Illavan
21	e and examined before vs by vertue of &c deposeth and saith that on or aboute the first of January las	John Symon
22	e & examined before vs by vertue of &c deposeth and saith that on or aboute the tenth December las	Robert Coyne
23	ne and examined before vs by vertue of &c deposeth and saith that on or about the 16th of December 1641	Richard Cases
24	to one hundred eight & Twenty pounds The deponent And saith that vpon the sixt of June last or therabetu	William Hodkins

Figure 3: Concordance of ‘deposeth and saith’.

Another reason for the inclusion of binomials is to represent a concept in two or more languages. This often occurs in English legal discourse due to historical developments that have left their legacy on the language of the law. An example being ‘Robbed and forcibly dispolued’ and ‘goods and chattels’ as shown in Figure 4, which are a mixture of Old English and Old French constructions.

N	Concordance	File
1	lost was Robbed and forcibly dispoiled of hir goods and Chattells to	Joane fflavan
2	lost was robbed and forceably dispoiled of his goods and Chattles to	Richard White
3	lost was robbed and forceably dispoiled of his goods and Chattles to	Peeter Mainsell
4	lost was robbed and forceably dispoiled of his goods and Chattles to	Augustine Kingesmill
5	lost was robbed and forceably dispoiled of his goods and Chattles to	William Kingwell
6	lost was robbed and forceably dispoiled of his goods and Chattles to	Ambrose Martin
7	lost was robbed and forceably dispoiled of his goods and Chattles to	Richard Smith
8	lost was robbed and forceably dispoiled of hir goodes and Chattells to	Elizabeth Cooke
9	lost was Robbed and forceably dispoiled of his goods and Chattles to	John Symson
10	lost was robbed and forceably dispoiled of his goods & Chattles to to	Peeter Peacocke
11	lost was robbed and forceably dispoiled of his goods and Chattles to	Dermod Grady
12	lost was robbed and forceably dispoiled of his goods and Chattles to	William Harrish
13	lost was robbed and forceably dispoiled of his goods and Chattles to	Thomas Canes
14	lost was robbed and forceably dispoiled of his goods and Chattles to	Thomas Wills
15	lost was robbed and forceably dispoiled of his goods & Chattles to th	William Hodkins
16	lost was robbed and forceably dispoiled of his goods and Chattles to	John Harte
17	lost was robbed and forceably dispoiled of his goods and Chattles wor	Richard Canes
18	lost was robbed and forceably dispoiled of his goods and Chattles to	Thomas Browne
19	lost was robbed and forceably dispoiled of his goods and Chattles to	John Ward
20	lost was robbed and forceably dispoiled of his goods and Chattles to	Ann Graham
21	lost he was robbed and forceably dispoiled of his goods & chattells, to	Thomas Whiteby
22	lost was Robbed and forceably dispoiled of his goods and Chattles []	Thomas Turner
23	lost was Robbed and forceably dispoiled of his goods and Chattles to	Robert Coyne

Figure 4: *Concordance of ‘robbed and forcibly dispoiled’ and ‘goods and chattells’.*

4.1.3 Modifiers

Finally, the last explicit legal markers that this essay will discuss is the use of legal modifiers such as ‘the same’, ‘they said’ and ‘the aforementioned’. The use of such modifiers is distinctively associated with the legal domain, as other discourse communities tend to favour less formal anaphoric references such as the third person pronoun ‘he/she/they’.

These modifiers show institutional intervention, most likely to disambiguate the references made by the deponent and in the hopes of achieving precision of reference. As shown in Figure 5, the first and second term to the right (R1 and R2) in our corpus shows that the modifier ‘said’ was often paired with a person or place. In many of these utterances, it may have been more effective to omit anaphoric reference and use the nominal instead, suggesting that modifiers are also a stylistic marker of these depositions. For example, ‘The said castle’ may be considered to be more ambiguous than ‘Cullen Castle’.

Centre	R1	R2
SAID	COUNTY	GENTLEMAN
	CASTLE	YEOMAN
	COUNTIE	ESQUIRE
	RICHARD	GENT
	PARTIES	WERE
	EARLE	WHERIN
	CATTLE	HUSBANDMAN
	PEETER	
	LEASE	
	LAND	
	CITTIE	
	CASTLES	
	LORD	
	HICKES	

Figure 5: R1 and R2 of modifier ‘said’.

4.2 Implicit influence

Now we have identified the explicit markers of institutional influence, we will now investigate potential implicit influences on the depositions. This will be achieved through analysis of the linguistic patterns in the sub-corpora constructed from utterances beginning with ‘saith’ and its variants shown in Figure 6.

N	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	VALUE	VALUE	WORTH	POUNDS	POUNDS	SAITH	THAT	ABOUTE	MEANES	ABOUTE	THIS
2		VIZT	SAID	KNOWETH	STER	SAYTH		ABOUT		ABOUT	
3		HUNDRED			GENTLEMAN	SAIT		THAT		EXPILLED	
4		THIS				SAIETH		VPON			
5						SAYT					
6						SAIT					

Figure 6: Concordance patterns of ‘saith’ L5 to R5.

4.2.1 ‘Saith’ in the sub-corpora

In Section 2.3, we identified that ‘saith’ may indicate reported speech based on Collins (2001) and The Language and Linguistic Evidence in the Depositions project (2010). However, in this Section we will demonstrate repetition of formulaic structures through analysis of the patterns in the sub-corpora’s concordances. These patterns in the narratives are potentially prompted by specific questions asked by the commissioner to achieve his institutional goals. Churches (1996) notes that ‘[i]n a response to a list of precisely worded queries of a quite technical nature, many will simply answer by rehearsing the words of the interrogative’ (p. 220 in Grund & Walker, 2011, p. 50). This would somewhat explain the parallels; however, if so, the commissioner has omitted all references to his involvement in the discourse. One indication towards hidden interrogatives is the low-frequency verb ‘answered (that)’ which suggests that the deponent is responding to a previous utterance.

4.2.2 Patterns in concordance

As proposed above, reoccurring linguistic patterns in concordance highlight the goal-oriented nature of the legal-lay discourse. This Section will evaluate high occurring concordances outside of that the explicitly understood legal features mentioned in 4.1. As Coulthard’s (2004) theory of idiolect states, no two utterances are the same as no two speakers speak exactly alike. This is the basis of our analysis as considering this corpus comprises of 40 individual testimonies, we would expect to see highly variable idiolectal style to represent each deponent’s contribution as an author.

The concordance of satisfaction (Figure 7) reveals a parallel construct being repeated 15 times in the sub-corpus, ‘therefore this deponent cannot get satisfaction from them. The total of his losses amounts to...’. The term satisfaction exclusively occurs in relation to money, most likely the result of an interrogative about losses produced by the commissioner and then written in a template format by the scribe.

N	N	Concordance	File
1	4	uall rebellion therefore this deponent cannot gett satisfaction from them The totall of his losses amounts to sea	Ambrose Martin
2	5	uall rebellion Therefore this deponent cannot gett satisfaction from them The totall of his losses amounts to one	Dermod Grady
3	9	uall rebellion therefore this deponent cannot gett satisfaction from them The totall of his losses amounts to sea	Peeter Peacocke
4	6	e alie or dead soe that this deponent cannot gett satisfaction from them therefore this deponent cannot gett sati	John Symson
5	7	tion from them therefore this deponent cannot gett satisfaction from them The totall of his losses amounts to thr	John Symson
6	10	erished by meanes of the rebellion he cannot gett satisfaction from anie of them The totall of his losses amount	Richard Charing
7	11	uall Rebellion therefore this deponent cannot gett satisfaction from them The totall of his losses amounts to sea	Thomas Browne
8	16	leman and therefore this deponent cannot gett any satisfaccion from any of them. The totall of his Losses amount	Willyam Haynes
9	1	this deponent knoweth not any way how to gett any satisfaccion from any of them the totall of her losses amounts	Margery Hazard
10	2	lbells, & therefore this deponent cannot gett any satisfaccion from any of them.	Thomas Whiteby
11	13	uall rebellion therefore this deponent cannot gett satisfaction from them.	William Harrish
12	14	his rebellion therefore this deponent cannot gett satisfaction from them	William Kingwell
13	3	erished by meanes of the rebellion he cannot gett satisfacc ion from any of them	John Harte
14	15	tants (by) meane of this rebellion he cannot gett satisfaction from any of (them)	William ward
15	12	turned papists therefore this deponent cannot gett satisfaction from him.	Thomas Willis
16	8	in the said Cittle) this examinant demanding then satisfaccion for the same sheepe the said Gerrald answered the	Peeter Mainsell

Figure 7: Concordance of ‘satisfaction’.

Another example of a high occurring parallel structure can be found from the concordance of ‘turned’. All uses of this verb are connected to religion ‘papists’ and rebellion ‘rebells’. Again, this structure appears to result from a hidden interrogative or other verbal prompts from the commissioner. Linell and Jönsson (1991) suggest that the institutional frameworks imposed by the legal team disadvantage lay peoples testimonies, as their narrative generally conflicts with the more ‘anonymizing case-type’ institutional perspective (in Johnston, 2010, p. 163). This suggests the deponents’ input to the deposition is highly controlled, and deposition commissions are liable to propose leading questions that predispose the addressee towards giving a specific answer.

N	Concordance	File
1	reputed protestants & now since this rebbellion turned Papists	John Bradish
2	a reputed protestant, & now since this rebellion turned papist, & [] & Elish Oge of the same spinster &	Richard Winchester
3	ofore a reputed protestant & since this rebellion turned papist. pist.	William Haynes
4	formerly reputed protestants & since this rebellion turned Papists. sts.	Thomas Whiteby
5	formerly reputed protestants & since this rebellion turned pap ist and Morr ogh Harbard of D rumin aforesaid	Edward Clare
6	formerly reputed protestants but since this rebellion turned papists	Thomas Browne
7	formerly a protestant but since this rebellion turned papist & rebell And although he accounted the sam	Peeter Peacocke
8	formerly reputed protestants but since this rebellion turned papists ??	Peeter Peacocke
9	merly reputed protestants but since this rebellion turned papists s	William Hodkins
10	an formerly a protestant but since this rebellion turned papists therefore this deponent canot gett satisf	Thomas Willis
11	ing formerly protestants are since this rebellion turned papists ts	James Lillies
12	erly a reputed protestant is since this rebellion turned papist & the wife of one Holmes likewise liueing	Joane Mlavan
13	ed a professed protestant is since this rebellion turned papist.	Augustine Kingesmill
14	fore this rebellion reputed protestants are since turned papists	Henry Briggs
15	nt formerly a protestant but since this rebellion turned papist & Rebell in the said County gentleman form	Peeter Peacocke
16	er being a protestant ere this rebellion is since turned papist t	Richard Winter
17	the said County labourer is since this rebellion turned rebell & one of the said John o Kenedyes souldier	john fox
18	Cardiffe Richardson that was a protestant, and turned papist now in actual Rebellion	Thomas Turner
19	souldiers Michell Hudson of the same is likewise turned rebell	john fox
20	ew & his wife & his wife father are al likewise turned papists	Joane Mlavan
21	rnd papists, & William Wigmore of the same Cowper turned papist & rebell. Valentine Palmer of the same sh	Thomas Whiteby
22	& rebell. Valentine Palmer of the same shoemaker turned papist & rebell. and also Telg o Grady of Kikel	Thomas Whiteby
23	Cullin aforesaid Merchant, & his wife & children turned papists, & William Wigmore of the same Cowper tur	Thomas Whiteby

Figure 8: Concordance of ‘turned’.

Finally, we will now consider the concordance of hostile (Figure 9). Interestingly this example differs from the previous concordances as the utterance only appears in four out of the forty depositions. Although it is low frequency, the concordance is striking because it contains the near-identical string ‘with collors flyeing in a hostill and rebellious manner’. Figure 10 demonstrates that all four of these depositions were collected in Limerick, and all four depositions mention a besieged Castle. Most similar is N1: Grady and N3: Browne, who spoke of the same castle and gave their testimonies on the same day, 19th November 1642.

Considering the other two depositions, they also provide a similar date (16th and 17th April 1642); however, these deponents gave their testimonies a year apart, still using identical strings in their descriptions. One explanation for this is ‘reports of reports’, i.e. accounts of what had been reported to them by other victims, frequently from places far removed from themselves (Canny, 2001 in Fennell-Clark, 2011, p. 28). Alternatively, the similarity could result from textual borrowing whereby the scribe copies the statement from one deposition to another to create a cohesive account that supports the overall narrative of the deposition commission (Grund & Walker, 2011, p. 50). Regardless of the cause, however, four identical strings would be highly improbably to occur without a degree of influence upon the deponent (Coulthard, 2004; Kredens, 2002).

N	Concordance	File
1	st a thousand armed men with collors flyeing in a hostill & rebellious maner came vpon this deponents said	Dermod Grady
2	lagh mc Mohowne his son with collors flyeing in a hostill maner came to and besieged the Castle or house o	Peeter Mainsell
3	tlemen in the said Country with collors flying in hostill & rebellious maner came to besiege this deponent	Thomas Browne
4	the said Edmond then and there in a rebellious & hostill maner with collors displayed, assaulted the Castl	John Ward

Figure 9: Concordance of ‘hostill’.

N	File	Date being referenced	Date of deposition	Location	Castle
1	Dermod Grady	15th of January	19th Nov 1642	Lymerick	Castle of Cullen
2	Peeter Mainsell	sixteenth of Aprill	7th Feb 1642	Lymerick	Castle or house of Aghanish
3	Thomas Browne	Twentyeth of december	19th Nov 1642	Lymerick	Castle of Cullen
4	John Ward	17th day of Aprill	25th April 1643	Lymerick	Castle of Tromroe

Figure 10: Depositions containing utterance ‘with collors flyeing in a hostill & rebellious maner’.

5 Conclusion

To conclude, although the 1641 depositions are a valuable source of evidence of the Irish rebellion, this essay has shown that the testimonies recorded are the product of several co-authors who we cannot separate today. Section 4 showed an overwhelming tendency for high-frequency patterns in the texts that suggest the reports given by deponents consist of partly constrained speech. These findings have implications for the reliability of the depositions. Taken together, this analysis of explicit and implicit institutional influence has shown that we cannot assume the reliability of a witness deposition without considering the broader historical context and the institutional intentions of the authors. Future study would be beneficial to expand this research to multiple commissioners and develop an understanding of variations between commissioner idiolectal style and its impact on the depositions. Furthermore, the construction of two corresponding corpora in the present study has allowed for statistical analysis such as type token ratio comparisons in future research.

6 References

- Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.
- Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. London: Longman.
- Canny, N. P. (2003). *Making Ireland British, 1580-1650*. Oxford: Oxford University Press.
- Churches, C. (1996). ‘The most unconvincing testimony’: the genesis and historical usefulness of depositions in Chancery. *The Seventeenth Century*, 11(2), 209-227.
- Collins, D. E. (2001). *Reaminated voices: Speech reporting in a historical-pragmatic perspective*. Amsterdam: John Benjamins Publishing.
- Cope, J. (2001). Fashioning victims: Dr. Henry Jones and the plight of Irish Protestants, 1641. *Historical Research*, 74(186), 370–391.
- Cotterill, J. (2010). How to use corpus linguistics in forensic linguistics. In: M. McCarthy & A. O’Keeffe, (Eds.) *The Routledge Handbook of Corpus Linguistics*, (pp.578-590). London: Taylor & Francis Group.
- Coulthard, M. (2004). Author Identification, Idiolect and Linguistic Uniqueness. *Applied Linguistics*, 25(4), 431-447.
- Darcy, E. (2013). Manufacturing massacre: The 1641 depositions and the Wars of the Three Kingdoms. In: *The Irish Rebellion of 1641 and the Wars of the Three Kingdoms*, (pp.77-101). Online: Boydell & Brewer.
- Durant, A. & Leung, J. H. (2016). *Language and Law: A resource book for students*. London: Routledge.
- Fennell-Clark, B. (2011). ‘Dodgy dossiers? Hearsay and the 1641 depositions’. *History Ireland*, 9(3), 26-29.

- Grund, P. (2012a). “The Nature of Knowledge: Evidence and Evidentiality in the Witness Depositions from the Salem Witch Trials. *American Speech*, 87(1), 7–38.
- Grund, P. J. (2012b). Textual History as Language History? Text Categories, Corpora, Editions, and the Witness Depositions from the Salem Witch Trials. *Studia Neophilologica*, 84(1), 40–54.
- Grund, P. J. & Walker, T. (2011). Genre Characteristics. In: M. Kyrö, P. J. Grund & T. Walker (Eds.) *to Language and Life in Early Modern England*, (pp.15-56). Amsterdam: John Benjamins Publishing Company.
- Hiltunen, R. (1990). *Chapters on legal English: Aspects past and present of the language of the law*. Jyväskylä: Gummerus.
- Johnson, T. (2014). The Preconstruction of Witness Testimony: Law and Social Discourse in England before the Reformation. *Law and History Review*, 32(1), 127- 147.
- Johnston, A. (2010). *The Routledge Handbook of Forensic Linguistics*. London: Routledge.
- Kredens, K. (2002). Towards a corpus-based methodology of forensic authorship attribution: a comparative study of two idiolects. In B. Lewandowska-Tomazsczyk (Ed.), *Practical Applications in Language Corpora* (pp. 405-437). Frankfurt am Mein: Peter Lang.
- Kredens, K. & Coulthard, M. (2012). Corpus Linguistics In Authorship Identification. In P. T. Tiersma & L. M. Solan (Eds.) *The Oxford Handbook of Language and Law*, (pp.504-516). Oxford: Oxford University Press.
- Kytö, M., Walker, T. & Grund, P. (2007). English witness depositions 1560-1760: An electronic text edition. *ICAME Journal*, 31, 65-85.
- Langbein, J. H. (2003). *The Origins of Adversary Criminal Trial*. Oxford: Oxford University Press.
- Language and Linguistic Evidence in the Depositions (2010). *Reported Speech?*. Retrieved 18/12/2021 from <<https://1641dep.abdn.ac.uk/exhibits/show/reported-speech/reported>>
- Lass, R. (2004). Ut custodian litteras: Editions, corpora and witnesshood. In M. Dossena & R. Lass (Eds), *Methods and Data in English Historical Dialectology* (pp. 21-48). Bern: Peter Lang.
- Leht, A. (2017). Binomials and Multinomials in Early Modern English Parliamentary Acts. In: J. Kopaczyk & H. Sauer (Eds.), *Binomials in the History of English: Fixed and Flexible (Studies in English Language)*, (pp. 261-278). Cambridge: Cambridge University Press.
- MacLeod, N. (2012). ‘Rogues, Villaines and Base Trulls’: Constructing the ‘Other’ in the 1641 Depositions. In A. Margey, E. Darcy & E. Murphy (Eds.), *The 1641 Depositions and the Irish Rebellion*, (pp. 113-127). London: Routledge,.
- Ohlmeyer, J., 2009. Anatomy of plantation: the 1641 depositions. *History Ireland*, 17(6), 54-56.
- Palmer, F. R. (1986). *Mood and Modality*. Cambridge: Cambridge University Press.
- Scott, M. (2020). *WordSmith Tools version 8*. Stroud: Lexical Analysis Software. Section 114 (1) CJA, .. 2003. *Hearsay*, s.l.: UK Legislation.
- Stubbs, M. (1994). Grammar, Text, and Ideology: Computer-assisted Methods in the Linguistics of Representation. *Applied Linguistics*, 15(2), 201-223.
- Trinity College Dublin Library (2010). 1641 Depositions. Retrieved 10/01/2021 from <<https://1641.tcd.ie>>
- Turell, M. T. (2010). The use of textual, grammatical and sociolinguistic evidence in forensic text comparison. *The International Journal of Speech*, 17(2), 211-250.
- Walker, T. & Grund, P. (2017). “Speaking base approbious words”: Speech representation in Early Modern English witness depositions. *Journal of Historical Pragmatics*, 18, 1-29.

- Wilson, J. & Walker, H. (2015). Pragmatic markers as implicit emotive anchoring: Modality as evidence of trauma in the 1641 depositions. In C. P. Amador-Moreno, K. McCafferty & E. Vaughan (Eds.), *Pragmatic Markers in Irish English*, (pp. 248-269). Online: John Benjamins.
- Wright, D. (2013). Stylistic variation within genre conventions in the Enron email corpus: Developing a text-sensitive methodology for authorship research. *International Journal of Speech Language and the Law*, 20(1), 45-75.

7 Appendix

List of Depositions Used for Section 4 Analysis

Full depositions available at <https://1641.tcd.ie>. (Accessed: 10 January 2021).

N	Deponent	Topic	Commissioners	Year
1	Ambrose Martin	Apostasy	Henry Rugg and Philip Bisse	1642
2	Ann Graham	Apostasy	Henry Rugg and Philip Bisse	1642
3	Augustine Kingesmill	Apostasy	Henry Rugg and Philip Bisse	1642
4	Danyell Spratt	Apostasy	Henry Rugg and Philip Bisse	1642
5	Dermod Grady	Apostasy	Henry Rugg and Philip Bisse	1642
6	Edward Clare	Apostasy	Henry Rugg and Philip Bisse	1642
7	Elizabeth Cooke	Apostasy	Henry Rugg and Philip Bisse	1642
8	Henry Briggs	Apostasy	Henry Rugg and Philip Bisse	1642
9	Isaak Graneere	Apostasy	Henry Rugg and Philip Bisse	1642
10	James Lillies	Apostasy	Henry Rugg and Philip Bisse	1642
11	Joane Fflavan	Apostasy	Henry Rugg and Philip Bisse	1642
12	John Bradish	Apostasy	Henry Rugg and Philip Bisse	1642
13	John Fox	Apostasy	Henry Rugg and Philip Bisse	1642
14	John Harte	Apostasy	Henry Rugg and Philip Bisse	1642
15	John Symson	Apostasy	Henry Rugg and Philip Bisse	1642
16	John Ward	Apostasy	Henry Rugg and Philip Bisse	1643
17	Leonard Webber	Apostasy	Henry Rugg and Philip Bisse	1643
18	Margery Hazard	Apostasy	Henry Rugg and Philip Bisse	1642
19	Nicholas Roberts	Apostasy	Henry Rugg and Philip Bisse	1643
20	Peeter Mainsell	Apostasy	Henry Rugg and Philip Bisse	1642
21	Peeter Peacocke	Apostasy	Henry Rugg and Philip Bisse	1642
22	Richard Canes	Apostasy	Henry Rugg and Philip Bisse	1642
23	Richard Chaning	Apostasy	Henry Rugg and Philip Bisse	1642
24	Richard Smith	Apostasy	Henry Rugg and Philip Bisse	1642
25	Richard White	Apostasy	Henry Rugg and Philip Bisse	1642
26	Richard Winchester	Apostasy	Henry Rugg and Philip Bisse	1642
27	Richard Winter	Apostasy	Henry Rugg and Philip Bisse	1642
28	Robert Coyne	Apostasy	Henry Rugg and Philip Bisse	1642
29	Thomas Browne	Apostasy	Henry Rugg and Philip Bisse	1642
30	Thomas Canes	Apostasy	Henry Rugg and Philip Bisse	1642
31	Thomas Jones	Apostasy	Henry Rugg and Philip Bisse	1642
32	Thomas Turner	Apostasy	Henry Rugg and Philip Bisse	1642
33	Thomas Whiteby	Apostasy	Henry Rugg and Philip Bisse	1642
34	Thomas Wills	Apostasy	Henry Rugg and Philip Bisse	1642
35	William Harrish	Apostasy	Henry Rugg and Philip Bisse	1642
36	William Hodkins	Apostasy	Henry Rugg and Philip Bisse	1642
37	William Kingwell	Apostasy	Henry Rugg and Philip Bisse	1642
38	William Tomlinson	Apostasy	Henry Rugg and Philip Bisse	1642
39	William Ward	Apostasy	Henry Rugg and Philip Bisse	1642
40	Willyam Haynes	Apostasy	Henry Rugg and Philip Bisse	1643

Acknowledgements

Thank you to my supervisors, Dr Nicci Macleod and Dr Billy Clark, for providing guidance and invaluable feedback throughout this project. I would also like to express my appreciation to the English Language department at Northumbria University for their continuous assistance during my degree. Finally, I would like to thank Charlie A. Brown for all the support and encouragement throughout my studies.

P-to-Q Entailment in Slovakian

Andrej Gregus

University of Edinburgh

Abstract. I present new data from Slovakian to test the P-to-Q Entailment hypothesis recently proposed by Roelofsen & Uegaki (2020). P-to-Q Entailment builds on work by Spector & Egré (2015), Theiler et al. (2018), and Steinert-Threlkeld (2020) to formulate a novel semantic universal in the domain of “responsive predicates”: predicates which can embed both declarative and interrogative complements (Lahiri 2002). P-to-Q Entailment asserts that for a responsive predicate V and agent x , “ x Vs that P ” entails “ x Vs that Q ”. For example, “It matters to me that like it” (P-sentence) entails “It matters to me whether you like it” (Q-sentence). Based on a list of common predicates compiled by Roelofsen & Uegaki, I survey 48 predicates in Slovakian to see whether they are P-to-Q entailing. Specifically, I judge whether example P-sentences entail example Q-sentences for each predicate. I conclude that while the majority of the predicates are indeed P-to-Q entailing, a handful are not: namely, *dozvedieť sa* (to learn), *myslieť* (to think), *mať obavy* (to be worried) and *rozhodnúť sa* (to decide). I claim that these could (but need not necessarily) constitute counterexamples to P-to-Q Entailment—further study of the semantics of these predicates is needed to better understand their relationship to P-to-Q Entailment. Particularly, I note that special focus should be given to the subjunctive/future-oriented flavour of some of the predicates and how this might affect the inference pattern from P-sentences to Q-sentences.

Keywords: P-to-Q entailment; semantic universal; Slovakian; responsive predicates

1 Introduction

So-called “responsive” predicates are those predicates that can embed both declarative and interrogative complements (Lahiri, 2002). A responsive predicate generally features in one of the following two kinds of syntactic structure, where the responsive predicate V takes two arguments: an entity X and a complement — either a declarative complement P (as in 1a), or an interrogative complement Q (as in 1b):

(1a) declarative-embedding structure: X Vs P

(1b) interrogative-embedding structure: X Vs Q

The question arises how to account for the fact that responsive predicates can embed both types of complements. There are two main approaches. The first approach keeps declaratives and interrogatives as different kinds of objects, and so requires two lexical entries for each verb: one which selects for declaratives, and one which selects for interrogatives. The second approach treats declaratives and interrogatives as essentially the same object, and so has to reduce one to the other. For instance, one can treat propositions as primary, and reduce questions to propositions (this has been dubbed “Q-to-P reduction” by Uegaki 2019, or just simply “Reduction” by Theiler et al 2018). Alternatively, one can treat questions as primary, and reduce propositions to questions (dubbed “P-to-Q reduction” by Uegaki, 2019; or “Inverse Reduction” by Theiler et al.).

The question of how to account for the existence of responsive predicates naturally feeds into a second — and for this paper more relevant — question of what universal constraints can be placed on responsive predicates, i.e. the relationship between P and Q. Much more about this will be discussed below, but broadly speaking, all of the proposed universals attempt to capture different truth-relations that are supposed to hold between P and Q.

The universal that will be of central concern for this paper is "P-to-Q Entailment", proposed by Roelofsen and Uegaki (2020). Specifically, I look at whether this universal holds of a set of responsive predicates in Slovak. I begin, in Section 2, by sketching out three important theoretical precursors to Roelofsen and Uegaki (2020): Spector and Egré (2015), Theiler et al. (2018), and Steinert-Threlkeld (2020). I then, in Section 3, provide a more detailed characterization of P-to-Q Entailment. Lastly, in Section 4, I present and discuss my preliminary findings about P-to-Q Entailment in Slovakian. I consider several problematic cases, but ultimately conclude that they can either be reasoned away as unproblematic, or else require further study into their individual semantics to render a conclusive result. As such, I find no straightforward counterexamples to P-to-Q Entailment in Slovakian.

2 Theoretical precursors

2.1 Spector and Egré (2015)

Spector and Egré (S&E) (2015) develop a framework that posits two lexical entries for each responsive verb, and on this basis formulate a universal constraint on the relationship between responsive verbs and their complements. Interesting though their two-lexical-entries framework is, I will, given the limited real-estate I have, focus not on *how* their universal follows from their framework, but rather on *what exactly* their universal says.

As such, let us turn to the following universal proposed by S&E (p. 1732):

Veridicality Universal (VU)

V is veridical w.r.t. Q iff V is veridical w.r.t. P

There is some nuance to the term "veridical" (see p.1737 footnote 7)⁶, but I'll bracket it here and focus mainly on the following fact: in order for **VU** to be true, the antecedent and the consequent of the biconditional either have to be both true, or both be false. Counter examples thus have to show that V is either veridical w.r.t to Q but not w.r.t. P, or vice versa.

Among the potential counter-examples S&E consider are communication verbs, such as *tell*, which are ambiguous between two readings. Consider:

(2a) Max told us who he saw steal his tobacco.

(2b) Iona stole Max's tobacco.

⁶ I will, to some extent, conflate the notions of 'veridicality' and 'factivity' in my exposition of the universal, because maintaining this distinction, as S&E do, would take me too far afield. For the same reason I also do not expand in any length on S&E's account of presuppositions. I do my best, however, to be underlyingly faithful to the original spirit of their discussion.

(2c) Max told us Iona stole his tobacco.

Reading 1 of (2a): Max is telling us the truth (namely, 2b), so we can infer (2c).

Reading 2 of (2a): Max could be mistaken (he could've seen someone other than the actual culprit steal his tobacco), so inferring (2c) would be erroneous.

Given that there are two possible readings of (3a), the biconditional in **VU** holds in virtue of us being able to pick whichever reading would make the biconditional true.⁷ In the absence of further counterexamples, S&E thus conclude that **VU** holds universally.

2.2 Theiler et al. (2018)

However, Theiler et al. contend that there is in fact another class of counterexamples which does successfully undermine **VU**: namely, predicates of relevance (pp. 446-453). As I mentioned, the antecedent and the consequent of **VU** have to both be true or both be false, otherwise **VU** fails to hold. Theiler et al. restate this idea in a slightly different way:

For **VU** to be true, both of the following have to either be true or false:

(**VU**₁) If V is veridical w.r.t. P, then V is veridical w.r.t. Q

(**VU**₂) If V is veridical w.r.t. Q, then V is veridical w.r.t. P

However, Theiler et al. submit that **VU**₁ is not true of predicates of relevance like *matters*.

Consider:

Veridicality w.r.t. P

(3a) It matters to Max that Iona stole his tobacco.

(3b) Iona stole Max's tobacco.

Veridicality w.r.t. Q

(4a) It matters to Max who stole his tobacco.

(4b) Iona stole Max's tobacco.

(4c) It matters to Max that Iona stole his tobacco.

While the inference from (3a) to (3b) is valid, the inference from (4a) to (4c) is not, because Max might fail to know that it was indeed Iona who stole his tobacco. Thus, *matters* is veridical w.r.t. P (the antecedent of **VU**₁ is true), but it is not veridical w.r.t. Q (the consequent of **VU**₁ is false). Thus, **VU**₁ fails to hold, and **VU** as a whole is undermined.

In addition to discounting **VU**, Theiler et al. formulate another universal:

Clausal Distributivity (CD)⁸

⁷ This framing of their conclusion is my own, but I believe it puts their point more transparently.

⁸ Their formulation of the universal (found on p. 448) is a bit involved, so I have opted instead to borrow Roelofsen and Uegaki (2020)'s formulation.

$x \text{ Vs } Q$ iff there is an answer p to Q such that $x \text{ Vs } p$

However, as pointed out by Roelofsen and Uegaki (R&U) (2020), predicates of relevance are equally a counterexample to **CD**. To show this, we can re-run the same example which I used above to discount **VU**₁, but the essential common point is this: biconditional formulations of a universal on responsive predicates (which both **VU** and **CD** are) fail because the Q-to-P entailing direction fails⁹. Thus, to foreshadow Section 3 to come, in order to arrive at a universal which can overcome these counterexamples, we have to discount the problematic Q-to-P direction and restrict ourselves instead to the P-to-Q direction.

2.3 Steinert-Threlkeld (2020)

Before we turn to P-to-Q entailment, though, I would like to very briefly consider an interesting contribution by Steinert-Threlkeld in support of the stronger, bidirectional **VU**. He devised a computational experiment which aimed to show that verbs which satisfy **VU** ‘are easier to learn than those that do not’ (Steinert-Threlkeld, 2020, p. 139). Unlike the previous two proposals, which have focused on formulating a rule (entailment pattern) that is meant to hold universally and examining *whether* it holds, Steinert-Threlkeld offers an explanation of *why* such a pattern should hold. Thus, **VU** and **CD**, though liable to counterexamples, have the empirical backing of a plausible justification for their existence.

3 P-to-Q Entailment

Picking up on the concluding idea of **Section 2.2**, in order to arrive at a more feasible universal constraint on clause embedding we have to get rid of the problematic Q-to-P direction of entailment and limit ourselves to the P-to-Q direction. As a result, Roelofsen and Uegaki (2020) propose P-to-Q Entailment as the best candidate universal on responsive predicates, given that it is able to overcome all of the counterexamples that undermined its predecessors. I will consider here the two counterexamples mentioned in the previous Section of this paper (for a more detailed list see R&U pp. 9-15) by assessing whether the following schema holds:

Schema for P-to-Q Entailment (PQE)

$x \text{ Vs that } P \Rightarrow x \text{ Vs that } Q$ (provided Q is exhaustivity-neutral¹⁰)

Starting with predicates of relevance, the inference from (5a) to (5b) seems valid:

(5a) It matters to Max that Iona stole his tobacco.

(5b) It matters to Max whether Iona stole his tobacco.

Likewise with the inference from (6a) to (6b) in regard to communication verbs:

(6a) Max told us that Iona stole his tobacco.

⁹ This is slightly simplified. The refutation of **VU** in Roelofsen & Uegaki (2020), following Theiler et al. (2018), construes **VU** as pertaining to *exhaustivity-neutral* interrogative complements (i.e. complements like polar questions and *whether*-interrogatives, whose answers are taken from two-membered sets), whereas my example does not.

¹⁰ Seeing as this is an explicit condition on **PQE**, I will adjust my examples to be exhaustivity-neutral.

(6b) Max told us whether Iona stole his tobacco.

Although R&U hold **PQE** to be robust, they concede that there are at least a handful of *prima facie* counterexamples that seem to violate **PQE**: namely, Buryat ‘hanaxa’, Turkish ‘bil’, and Tagalog ‘magtaka’. However, they conclude that further study of these is required to provide a more wholistic picture. One result of such study might be that the semantics of these predicates can be re-conceptualized in a way that does conform to P-to-Q Entailment. Alternatively, P-to-Q Entailment could be modified such that it is able to successfully account for these counterexamples. Lastly, to reinforce the empirical robustness of P-to-Q Entailment, it would be useful to research into *why* **PQE** holds, in the same way that Steinert-Threlkeld (2020) argued **VU** holds due to ease of learnability.

4 P-to-Q Entailment in Slovakian

Having established the theoretical background to **PQE**, the following Section will present my preliminary findings from Slovakian and argue that there are no straightforward counterexamples to **PQE** in Slovakian. I will begin by outlining the methodology (Section 4.1) and then move on to discuss the results of my research (Section 4.2). I present several problematic predicates that resist a straightforward **PQE** judgement, either because of translation issues, the examples they appear in, or the complex presuppositional semantics of the declarative that seem not to transfer over into the interrogative. However, I conclude that none of these cases are straightforward counterexamples to **PQE**, and that further study is needed to resolve the cases which encode presuppositions.

4.1 Methodology

I have taken the list of paradigmatic responsive predicates graciously provided to me by Uegaki (2020, personal correspondence) as my departure point and translated them into Slovakian¹¹. This generates a table with the following structure:

Table 1: *Layout of responsive predicate table.*

Category of predicate	Predicate V in English	Predicate V in Slovak	x Vs P	x Vs Q	Does "x Vs P" entail "x Vs Q"?
-----------------------	------------------------	-----------------------	--------	--------	--------------------------------

I have chosen ‘že’ as the closest Slovak equivalent of ‘that’, and ‘i’ as the closest Slovak equivalent of *whether*. Admittedly, some of the translations are a bit strained, but I can conceive of a scenario, however niche, when they could be plausibly uttered¹². The example embedded clause I have used for P is ‘že prišiel domov’ [that he came home], and the one I have used for Q is ‘či prišiel domov’ [whether he came home]. In some cases, for the sake of readability (and to eliminate ambiguity when making judgements regarding entailment patters) I have made the matrix subject and the embedded subject differ in gender, and all the

¹¹ Translations are my own, though some have been corroborated by the go-to dictionary for Slovak-English translation found here: <https://slovníky.lingea.sk/anglicko-slovensky>.

¹² All grammaticality/well-formedness judgements are mine.

matrix verbs (the responsive predicates V) be in present tense and all the embedded verbs be in the past tense.

Here is a sample entry in the table:

Table 2: *Example entry in the responsive predicate table.*

assessment	<i>accept</i>	prijat'	Prijímajú, že prišiel domov.	Prijímajú, či prišiel domov.	YES
------------	---------------	---------	------------------------------	------------------------------	-----

And here is the translation of each of the two sentences from columns 4 and 5:

Table 3: *Sample translations of P- and Q-sentences.*

Declarative

Prijímajú	že	prišiel	domov
accept-3rd-PL	that	come-PAST-3rd-SG-MASC	home
<i>They accept that he came home.</i>			

Interrogative

Prijímajú	či	prišiel	domov
accept-3rd-PL	whether	come-PAST-3rd-SG-MASC	home
<i>They accept whether he came home.</i>			

4.2 Results & Discussion

Below is the complete, alphabetically ordered table of my findings:

Table 4: *Complete table of all responsive predicates tested in Slovakian.*

Category of predicate	Predicate V in English	Predicate V in Slovakian	x Vs P	x Vs Q	Does "x Vs P" entail "x Vs Q"?	Notes/reason for judgement in previous column
assessment	accept	<i>prijat'</i>	Prijímajú, že prišiel domov.	Prijímajú, či prišiel domov.	1	
doxastic	agree with X [stative:	<i>súhlasit'</i>	Súhlasia, že prišiel domov.	Súhlasia, či prišiel domov.	1	

	to be in agreement]					
communication	announce	<i>vyhlásiť</i>	Vyhlasuje, že prišiel domov.	Vyhlasuje, či prišiel domov.	1	
communication	argue	<i>tvrdiť (?)</i>	Tvrdí, že prišiel domov.	*Tvrdí, či prišiel domov.	NA	V is not responsive
inquisitive	ask	<i>pýtať sa</i>	*Pýta sa, že prišiel domov.	Pýta sa, či prišiel domov.	NA	V is not responsive
communication	assert	<i>vyhlásiť</i>	Vyhlasuje, že prišiel domov.	Vyhlasuje, či prišiel domov.	1	
doxastic	assume	<i>predpokladať</i>	Predpokladá, že prišiel domov.	*Predpokladá, či prišiel domov.	NA	V is not responsive
doxastic	be certain	<i>byť si istý</i>	Je si istá, že prišiel domov.	Je si istá, či prišiel domov.	1	
doxastic	be convinced	<i>byť presvedčený</i>	Je presvedčená, že prišiel domov.	*Je presvedčená, či prišiel domov.	NA	V is not responsive
inquisitive	be curious	<i>byť zvedavý</i>	*Je zvedavá, že prišiel domov.	Je zvedavá, či prišiel domov.	NA	V is not responsive
emotive factive	be happy	<i>byť šťastný</i>	Je šťastná, že prišiel domov.	*Je šťastná, či prišiel domov.	NA	V is not responsive
doxastic evaluative	be right	<i>mať pravdu</i>	Má pravdu, že prišiel domov.	*Má pravdu, či prišiel domov.	NA	V is not responsive
emotive factive	be surprised	<i>byť prekvapený</i>	Je prekvapená, že prišiel domov.	*Je prekvapená, či prišiel domov.	NA	V is not responsive

doxastic	be unaware	<i>nebyť si vedomý</i>	Nie je si vedomá, že prišiel domov.	Nie je si vedomá, či prišiel domov.	1	
bouletic: negative be worried		<i>mať obavy</i>	Má obavy, že prišiel domov.	Má obavy, či prišiel domov.	???	
doxastic evaluative	be wrong	<i>mýliť sa</i>	Mýli sa, že prišiel domov.	*Mýli sa, či prišiel domov.	NA	V is not responsive
doxastic	believe	<i>veriť</i>	Verí, že prišiel domov.	*Verí, či prišiel domov.	NA	V is not responsive
predicates of relevance	care	<i>záležať (niekomu) (na niečom)</i>	Záleží mi (na tom), že prišiel domov.	Záleží mi (na tom), či prišiel domov.	1	
communicat ion	claim	<i>vraviť</i>	Vraví, že prišiel domov.	Vraví, či prišiel domov.	1	
communicat ion	complain	<i>sťažovať sa</i>	Sťažuje sa, že prišiel domov.	*Sťažuje sa, či prišiel domov.	NA	V is not responsive
decision	decide	<i>rozhodnúť sa</i>	Rozhodla sa, že prišiel domov.	Rozhodla sa, či prišiel domov.	???	
directive	demand	<i>domáhať sa</i>	*Domáha sa, že prišiel domov.	Domáha sa, či prišiel domov.	NA	V is not responsive
assessment	deny	<i>poprieť</i>	Popiera, že prišiel domov.	Popiera, či prišiel domov.	1	
dubitative	doubt	<i>pochybovať</i>	Pochybuje, že prišiel domov.	Pochybuje, či prišiel domov.	1	
doxastic	expect	<i>očakávať</i>	Očakáva, že prišiel domov.	*Očakáva, či prišiel domov.	NA	V is not responsive

communicat ion	explain	<i>vysvetliť</i>	Vysvetluje, že prišiel domov.	Vysvetluje, či prišiel domov.	1	
bouletic: negative fear		same as "be worried"	doxastic: change of state	forget	<i>zabudnúť</i>	Zabudla, že prišiel domov.
Zabudla, či prišiel domov.	1					
bouletic	hope	<i>dúfať</i>	Dúfa, že prišiel domov.	Dúfa, či prišiel domov.	1	
communicat ion	inform	<i>oznámiť</i>	Oznámila, že prišiel domov.	Oznámila, či prišiel domov.	1	
inquisitive	inquire	same as "ask"	inquisitive	investigate	<i>skúmať</i>	*Skúma, že prišiel domov.
Skúma, či prišiel domov.	NA	V is not responsive				
doxastic	know	<i>vedieť</i>	Vie, že prišiel domov.	Vie, či prišiel domov.	1	
doxastic: change of state	learn	<i>dozvedieť sa</i>	Dozvedela sa, že prišiel domov.	Dozvedela sa, či prišiel domov.	???	
directive	order	<i>prikázať</i>	*Prikazuje, že prišiel domov.	*Prikazuje, či prišiel domov.	NA	V is not responsive.
bouletic	pray	<i>modliť sa</i>	*Modlí sa, že prišiel domov.	Modlí sa, či prišiel domov.	???	
bouletic	prefer	<i>preferovať</i>	Preferuje, že prišiel domov.	*Preferuje, či prišiel domov.	NA	V is not responsive.
directive	propose	<i>navrhnúť</i>	*Navrhuje, že	*Navrhuje, či	???	

			prišiel domov.	prišiel domov.		
ratification	prove	<i>dokázat'</i>	Dokázala, že prišiel domov.	Dokázala, či prišiel domov.	1	
emotive factive	regret	<i>mrziť</i>	Mrzí ju, že prišiel domov.	*Mrzí ju, či prišiel domov.	NA	V is not responsive
perception	see	<i>vidieť</i>	Vidí, že prišiel domov.	Vidí, či prišiel domov.	1	
doxastic	suspect	<i>tušiť</i>	Tuší, že prišiel domov.	Tuší, či prišiel domov.	1	
communication	tell (someone)	<i>povedať (niekomu)</i>	Povedala (mi), že prišiel domov.	Povedala (mi), či prišiel domov.	1	
doxastic	think	<i>myslieť</i>	Myslí, že prišiel domov.	*Myslí, či prišiel domov.	???	
bouletic	want	<i>chcieť</i>	*Chce, že prišiel domov.	*Chce, či prišiel domov.	NA	V is not responsive.
communication: manner of saying	whisper	<i>pošepkať</i>	Pošepkala, že prišiel domov.	Pošepkala, či prišiel domov.	1	
inquisitive	wonder	<i>zaujímať (niekoho)</i>	Zaujíma ma, že prišiel domov.	Zaujíma ma, či prišiel domov.	1	
communication: manner of saying	write	<i>písať</i>	Píše, že prišiel domov.	Píše, či prišiel domov.	1	

Predicates that are unambiguously responsive abide by **PQE** and receive a 1 in the judgement column. However, predicates which are not straightforwardly responsive (i.e., they are either rogative or antirogative) receive a judgement of NA (not applicable), because **PQE** does not apply to them.

Let us now turn to a handful of problematic verbs, highlighted in grey in the table. These are, in the order that they appear in the table:

Table 5: Potentially problematic predicates.

bouletic: negative	be worried	<i>mať obavy</i>	Má obavy, že prišiel domov.	Má obavy, či prišiel domov.	???
decision	decide	<i>rozhodnúť sa</i>	Rozhodla sa, že prišiel domov.	Rozhodla sa, či prišiel domov.	???
doxastic: change of state	learn	<i>dozvedieť sa</i>	Dozvedela sa, že prišiel domov.	Dozvedela sa, či prišiel domov.	???
bouletic	pray	<i>modliť sa</i>	*Modlí sa, že prišiel domov.	Modlí sa, či prišiel domov.	???
directive	propose	<i>navrhnúť</i>	*Navrhuje, že prišiel domov.	*Navrhuje, či prišiel domov.	???
doxastic	think	<i>myslieť</i>	Myslí, že prišiel domov.	*Myslí, či prišiel domov.	???

4.2.1 *Modliť sa* (to pray) and *navrhnúť* (to propose)

The Slovakian counterparts to pray and propose — ‘*modliť sa*’ and ‘*navrhnúť*’ — are ill-formed with the complementizers ‘*že*’ or ‘*či*’. Instead, these verbs require the subjunctive complementizer ‘*aby*’, which (as far as I am aware) does not have a direct counterpart in English. The closest approximation would be “so that” or “in order that”. The predicate ‘*navrhnúť*’ (propose) only works with ‘*aby*’, where the resulting complement clause is declarative. Since this is the only construction it can figure in, the predicate is not responsive in Slovakian, and thus is not subject to **PQE** judgements (since these apply exclusively to responsive verbs).

The predicate ‘*modliť sa*’ [pray], however, is not as easy to reason away. The declarative version requires ‘*aby*’ (rather than ‘*že*’) to be ill-formed, but, unlike ‘*navrhnúť*’ [propose], it is also well-formed in the interrogative version with ‘*či*’. So the question becomes: Does ‘*Modlí sa, aby prišiel domov*’ entail ‘*Modlí sa, či prišiel domov*’?

The issue is that the declarative ‘*aby*’ version has a future-oriented, wishful reading (“She prays that he would, in the future, come home”), whereas the interrogative ‘*či*’ version has a past tense, evaluative reading (“She prays about whether he had come home by now”). Thus, there is a mismatch in the tense that is projected by the predicate in the two sentences, which makes the judgement of whether P entails Q unstraightforward at best, and impossible at worst (how can a future event imply a past tense event?)¹³.

4.2.2 *Dozvedieť sa* (to discover)

¹³ I am assuming that, in order for us to assess whether **PQE** holds, the sentences have to be consistent in their tense-readings. If I am mistaken in making this assumption, then I believe *modliť sa* is by far the best candidate out of all the ones I discuss to mount a compelling challenge to **PQE**.

The next predicate, ‘dozvediet’ sa’, I have flagged simply because it is not an exact translation of the English "learn". ‘Dozvediet’ sa’ has the much wider reading of "learn new things", "discover", rather than the more restricted "absorb new information" reading that is more salient to the English version. If we grant my translation, however, **PQE** holds.

4.2.3 *Mysliet’ (to think)*

My example interrogative-embedding sentence with ‘mysliet’ [think] is ill-formed, but both the declarative and the interrogative example sentences can be modified by adding ‘už’ (which, roughly, means “already”) to make the latter grammatical:

Table 6: Comparison of P- vs. Q-sentence embedding *mysliet’* (‘think’) predicate.

	Declarative	Interrogative
Old version	<i>Myslí, že prišiel domov.</i>	<i>*Myslí, či prišiel domov.</i>
New version with "už"	<i>Myslí, že už prišiel domov.</i> (She thinks that he already came home.)	<i>Myslí, či už prišiel domov.</i> (She thinks whether he already came home.)

Adding ‘už’ makes the interrogative read like indirect speech, thus making it sound grammatical. To sum up: on the old version, ‘mysliet’ was not responsive so **PQE** did not factor in, and on the new version, **PQE** holds. Neither version thus poses a problem to **PQE**.

4.2.4 *Mat’ obavy (be worried)*

This predicate is problematic because the declarative version has the presupposition that the matrix speaker does not *wish* for the embedded subject to have come home, whereas the interrogative version has no such presupposition (it connotes straightforward worry about whether he did or did not come home). As such, since the wishing-for-a-negative-outcome presupposition is, at least in my judgement, absent from the interrogative, more research needs to be done into this predicate to unpack the semantics of its presupposition structure. As it stands, however, **PQE** is inconclusive on this case.

4.2.5 *Rozhodnúť sa (to decide)*

On its declarative version, this verb has performative force — as though the matrix speaker is deciding to undertake the action to come home. The interrogative version, on the other hand, creates the sense that the speaker is looking back and reflecting whether or not the embedded speaker has or has not come home. As such, this verb, just as ‘mat’ obavy’ in the previous Section, requires further study to be able to conclusively affirm or deny (or deem as inapplicable) **PQE**¹⁴.

¹⁴ Again, as in the case *modlit’ sa*, I am assuming that these issues can be remedied once the semantics of ‘mat’ obavy’ and ‘rozhodnúť sa’ are better understood. I am choosing to say, ‘this needs further investigation’ rather than saying, ‘this is a

5 Conclusion

This paper has hoped to serve a two-fold purpose: firstly, to present the P-to-Q Entailment universal and overview the literature that led up to it (Sections 2 and 3); and, secondly, examine whether P-to-Q Entailment holds in Slovakian (Section 4). The preliminary findings show that, within the verbs that I have tested for, there are no straightforward counter examples to P-to-Q Entailment.

I did, however, flag a handful of verbs which I believed deserve more detailed commentary. The first category includes verbs that require a subjunctive complementizer, which either disqualifies them from being responsive verbs ('navrhnúť') or else gives them differently tensed readings which are hard to evaluate whether or not they are P-to-Q Entailing ('modlit' sa'). The second category consists of verbs whose translations are either tentative ('dozvedieť sa'), and so an affirmative P-to-Q Entailment judgement is made on shaky ground; or which require a slight modification to the example they feature in to make them responsive, though they ultimately nevertheless come out as P-to-Q entailing ('myslieť'). The last category are verbs whose declarative clauses have presuppositions ('mať obavy') or performative force ('rozhodnúť sa') which does not carry over into their interrogative counterparts, and as such require further study to unpack their semantics. The upshot of this paper, then, is that I have found no straightforward counterexamples to P-to-Q Entailment; I have only found cases which require further study and thus yield an inconclusive P-to-Q Entailment judgement.

6 References

- Lahiri, U. (2002). *Questions and Answers in Embedded Contexts*. Oxford University Press.
- Spector, B., & Egré, P. (2015). A uniform semantics for embedded interrogatives: An answer, not necessarily the answer. *Synthese*, 192(6), 1729–1784.
- Steinert-Threlkeld, S. (2020). An explanation of the veridical uniformity universal. *Journal of Semantics* 37, 129–144.
- Theiler, N., Roelofsen, F., & Aloni, M. (2018). A uniform semantics for declarative and interrogative complements. *Journal of Semantics*, 35(3), 409–466.
- Roelofsen, F. & Uegaki, W. To appear. "Searching for a universal constraint on the possible denotations of clause-embedding predicates." *Proceedings of SALT 30*.
- Uegaki, W. (2019). "The semantics of question-embedding predicates." *Language and Linguistics Compass*, 13(1).

counterexample to **PQE**, because it is quite possible that there is something more complex going on in these verbs that, once better understood, might be elucidated in a way that is compatible with **PQE**.