The University of Edinburgh

School of Philosophy, Psychology and Language Sciences

The diachronic development of substitutive ᴅᴏ in Old to Middle French and Middle English

A comparative study using parsed corpora

Rebecka Elm

Master of Arts with Honours Linguistics and English Language

Supervisor: Robert Truswell

Date of Submission: 13 May 2019

Word count: 9,997

Acknowledgements

A heartfelt thank you to my supervisor Rob Truswell, for your endless patience, insight and commitment.

Thank you also to the anonymous native French speaker for helping me with grammaticality judgements.

Thank you to my ever-supportive Swedish and Scottish families, and to the Edinburgh swing dance community for keeping me grounded (mostly).

Thank you to Kim for the 'I told you so'.

The diachronic development of substitutive DO in Old to Middle French and Middle English:

A comparative study using parsed corpora

Abstract

This study examines the use of substitution with DO in French and English from the medieval period to the present day, starting from Miller's (1997) observations that Old French (OFr) and Middle French (MFr) had a form of substitution which closely resembles that found in Modern English (ModE). A definition of substitutive DO is developed based on Hankamer and Sag's (1976) distinction between deep and surface anaphora. Two corpus studies are presented, using data from the parsed MCVF (Martineau, Hirschbühler, Kroch & Morin, 2010) and PPCME2 (Kroch & Taylor, 2000) corpora to track the diachronic development of substitutive DO in OFr and MFr on the one hand and Middle English (ME) on the other hand. The results confirm Miller's findings for OFr and MFr, and show that the construction was well-attested throughout ME. Following a discussion on contact-induced and endogenous change, and an evaluation of potential contact scenarios, it is found that there is not enough evidence to reject the null hypothesis, namely that the diachronic development of substitutive DO in French and in English was independent of contact between the two languages.

The diachronic development of substitutive DO in Old to Middle French and Middle English:

A comparative study using parsed corpora

Table of contents

The diachronic development of substitutive DO in Old to Middle French and Middle English:

A comparative study using parsed corpora

1. Literature review

1.1 Introduction

In the literature on English diachronic syntax, there is a larger focus on periphrastic *do* in the literature than on substitutive *do*. However, the latter is a central feature of ModE[1] and an interesting phenomenon in its own right. Interesting observations by Miller (1997) reveal that a similar construction existed in OFr and MFr, but apparently disappeared in the 17th century.

This study aims to replicate Miller's findings regarding French, and add a comparison with ME, the corresponding period in English. This is done through a corpus investigation using the *Modéliser le Changement: Les Voies du Français* (MCVF) corpus of Historical French (Martineau, Hirschbühler, Kroch & Morin, 2010) with the *Penn supplement to the MCVF Corpus* (Kroch & Santorini, 2010), and the *The Penn-Helsinki Parsed Corpus of Middle English*, second edition (PPCME2) (Kroch & Taylor, 2000).

---

[1] (i) shows a list of the time periods associated with the different linguistic stages of English and French:

(i) Old English (OFr): 450-1150 AD

Middle English (ME): 1150-1500 AD

Early Modern English EModE): 1500-1700 AD

Modern English (ModE): 1700 AD- the present day   (Gelderen, 2014)

Old French (OFr): 842-1300 AD

Middle French (MFr): 1300-1500 AD

Renaissance French: 1500-1600 AD

Classical and Neo-Classical French: 1600-1789 AD

Modern French (ModFr): 1789 AD- the present day   (Ayres-Bennett, 1996)

The diachronic development of substitutive DO in Old to Middle French and Middle English:

A comparative study using parsed corpora

Section 1 will review the syntactic and semantic features of different anaphoric constructions, formulating a definition of substitutive DO informed by Hankamer and Sag's (1976) notions of deep and surface anaphora. Substitutive DO is defined as a surface anaphor whereby DO replaces a VP, or in certain subtypes a VP minus the NP complement. This definition includes VPE, PG, English *do so* and French *si/non faire*. Miller's (1997) observations of substitution in OFr and MFr will be compared to the situation in ModFr.

Section 2 will present the two corpus studies, looking firstly at OFr and MFr and secondly at ME. Some general methodological advantages and considerations associated with parsed digital corpora will be discussed, as well as the specific challenges of identifying substitutive DO in the two corpora. For each study, background information on the corpora will be introduced, followed by a description of the queries and a review of the results.

The discussion in section 3 will compare the results of the two studies, and evaluate the hypotheses that substitutive DO was influenced by French-English or Celtic-English contact. The hypothesis that English substitutive *do* is a precursor to periphrastic *do* will also be considered, leading into a general discussion on endogenous and contact-induced change.

Regarding the terminology used in this paper, a language-specific lexical item is referred to in italics whereas a non-language-specific construction is referred to in small capitals. Thus, substitutive *do* refers to the construction in English, and substitutive DO to the construction in English and French. In all examples, linguistic antecedents are underlined and anaphoric expressions, both deep and surface anaphora, are marked in boldface. This notation has been added to all examples cited from other sources, where it was not already present.

1.2 Anaphora with *do* in Modern English

Using Modern English examples, this section will introduce anaphoric constructions involving *do*. It will rely on definitions by Miller (1997). Then, following an overview of different syntactic analyses of anaphora and ellipsis, Hankamer and Sag's (1976) theory of deep and surface anaphora will be adopted to establish which constructions should be categorised as substitutive ᴅᴏ for the purposes of this study.

1.2.1 Verb Phrase Ellipsis (VPE)

According to Miller (1997), 'VPE is characterized by the ellipsis of a VP leaving behind one (or more) auxiliaries. If the understood VP has no auxiliary, then supportive *do* appears […]' (p. 120). Although VPE may involve different auxiliaries, the focus of this study will be VPE with *do*.

> (1) a. Abe <u>ate a starfruit </u>and Betty **did** too.
>
> c. I <u>feel nauseous</u> whenever you **do**.
>
> d. He <u>loves me</u> more than my pet lizard **does**.

1.2.2 Pseudogapping (PG)

> (2) a. Abe <u>ate </u>a starfruit and Betty **did** a pear.
>
> b. Surprisingly, the café <u>sold </u>more tea than it **did** coffee.
>
> c. He <u>loves</u> me more than I **do** my pet lizard.

In PG, an overt auxiliary substitutes the elided material, as in VPE. Here, however, the entire VP is not deleted; the NP complement of the main verb remains overtly present next to the ellipsis site.

1.2.3 Gapping

This latter aspect is shared with the elliptical construction referred to as Gapping, demonstrated in (3):

(3) Abe <u>ate a starfruit</u> and Betty __ a pear.

Unlike PG and VPE, however, Gapping leaves no overt auxiliary and only occurs in coordinate structures (Miller, 1997). Since there is no overt DO, Gapping will not be considered substitutive DO in this study. Based on the above criteria, Miller (1997) analyses VPE and PG as instances of the same construction, distinct from Gapping. The idea that Gapping is a different construction is well-established and widely debated in the literature (see for example Hankamer, 1971, 1973; Siegel, 1984, & Johnson, 2009). Central to Miller's (1997) analysis, however, is the concept of pro-predicates, meaning that VPE and PG can substitute non-agentive as well as agentive verbs. (1b-c) and (2c) above are examples of substitution with non-agentive antecedents.

PG is especially frequent in, but not restricted to, comparative structures such as (2c) (Miller 1997, p. 121). Miller (1997, p. 121) stresses that PG, although less common than VPE, is more frequent than it has traditionally been considered, well-attested in conversational as well as literary data.

1.2.4 Analyses of Anaphora: Deletion or pro-form

Simply put by Hestvik, Nordby, and Karlsen (2005), '[a]n anaphor is a linguistic expression that does not carry meaning by itself, but inherits its meaning from another expression' (p. 229). There is no consensus among syntacticians about whether anaphoric or elliptical constructions should be analysed as involving deletion or pro-forms. However, Haddican (2007) summarises the traditional view as follows:

A traditional distinction in the literature is between anaphora that have internal structure (typically, bare deletions […]) and those that do not have internal structure (which often take the form of some phonetically overt pro-form). In a derivational framework, this difference is often expressed by positing a derivational difference between these two kinds of anaphora such that true elisions have a full-fledged structure that undergoes deletion, while pro-forms are base-generated as anaphors. (p. 540)

According to Dagnac (2019), however, the term ellipsis comprises various anaphoric devices which may or may not involve deletion. As previously mentioned, Miller (1997) analyses VPE and PG as involving pro-predicates rather than deletion. Baltin (2012), on the other hand, argues that all anaphoric constructions involve deletion. He analyses *do* in VPE as a pro-form, in the sense of 'a functional head whose complement is missing' (p. 416). In his view, there is a false dichotomy between deletion and pro-forms, since pro-forms themselves involve deletion. However, the size of the deleted constituent differs between substitution with VPE, PG, and *do so* (Baltin, 2012).

In Hankamer and Sag's influential paper from 1976, the authors identify two distinct anaphoric processes: deep and surface anaphora. Surface anaphora undergoes syntactic control, referring to a linguistic antecedent in the discourse. It is derived through deletion under identity at a superficial level of the syntax. This type of anaphora only requires its antecedent to be syntactically coherent at surface level, not to represent a coherent semantic unit at an underlying level.

Deep anaphora may also be syntactically controlled, but it has the additional option of being pragmatically controlled. Hence, the anaphora need not refer to a linguistic antecedent in the discourse as it is resolved through pragmatic reasoning. Deep anaphora is not derived through

deletion but is already present at an underlying level. Unlike surface anaphora, then, deep anaphora relies on a deep semantically coherent, not necessarily linguistic, unit as its antecedent and does not require a superficial syntactically coherent unit. Hankamer and Sag (1976) use the following examples to illustrate their point:

> (4) a. [Sag produces a cleaver and prepares to hack off his left hand]
>
> Hankamer: #Don't be alarmed, ladies and gentlemen, we've rehearsed this act several times, and he never actually **does**.
>
> b. [Same context]
>
> Hankamer: ... He never actually **does it**. (p. 392)

Here, (4a) constitutes surface anaphora and (4b) deep anaphora. (4a) is ungrammatical because there is no linguistic antecedent, whereas (4b) is allowed because the anaphora can be resolved using pragmatic information. One can imagine an alternative example (5a), where a linguistic antecedent licenses the surface anaphora:

> (5) a. Sag: I will now hack off my left hand with this cleaver.
>
> Hankamer: Don't be alarmed, ladies and gentlemen, we've rehearsed this act several times, and he never actually **does**.
>
> b. Sag: I will now hack off my left hand with this cleaver.
>
> Hankamer: ... He never actually **does it**.

(5b) is also grammatical since deep anaphora may be syntactically controlled too.

We have seen from the above examples that *do* used in a VPE construction is a surface anaphor whereas *do it* is a case of deep anaphora. Furthermore, Hankamer and Sag (1976) show that *do it* resembles sentential *it*, illustrated below in (6) and (7), which also constitutes deep anaphora since it allows pragmatic control:

(6) Hankamer [observing Sag successfully ripping a phone book in half]:

I don't believe **it**.

(7) Sag [same circumstance]:

It's not easy. (p. 407)

1.2.5 Do so

Opinions differ in the literature regarding the status of anaphora involving *do so*. However, the following discussion will conclude that it should be considered a type of substitutive *do.* Hankamer and Sag (1976) identify three different kinds of *do so* anaphora, presented in (8-10), which will be referred to in this paper as Type A, B, and C.[2]

8) Type A: He wanted to <u>buy a new shirt</u>, and he **did so**.

9) Type B: He wanted to <u>buy a new shirt</u>, and **so** he **did**.

10) Type C: He <u>wanted to buy a new shirt</u>, and **so did** I.

Note that the antecedent constituent differs between (8-9) on the one hand and (10) on the other hand.

Type A is the construction that is generally used as an example when *do so* is mentioned in the literature (cf. Miller, 1997; Haddican, 2007; Baltin, 2012). (11) illustrates that Type A can occur in the same context as the bare *do* VPE in (1). However, as shown in (12) it cannot have an overt NP complement as in PG:

---

[2] Hankamer and Sag (1976, p. 415) also identify anaphora where *so* is combined with verbs like *believe* or *seem*.

This is not relevant for the present discussion, which concerns substitutive 'do' only.

> (11) Abe <u>ate a starfruit </u>and Betty **did (so)** too.

> (12) Abe <u>ate a starfruit </u>and Betty **did (*so)** a pear.

Furthermore, Haddican (2007, p. 544) and Miller (1997, p. 130) note that this type of *do so*, like *do it*, requires an agentive antecedent, whereas bare elliptical *do* does not:

> (13) Betty <u>burnt the letters</u>, and Abe **did (so)** too.

> (14) Abe <u>loves starfruit</u>, and Betty **does (*so)** too.

This is because in Type A, *do* is a dummy main verb rather than an auxiliary (Hankamer & Sag, p. 417). It can combine with auxiliaries, as in *Betty has **done so** too. Betty has **done** too* is a similar but separate construction referred to as British *do* (Baltin, 2012; Haddican, 20017). In fact, Haddican (2007, p. 539) argues that British *do* is a structurally deficient form of *do so.*

Like Type A, Type B cannot occur in a PG-like configuration:

> (15) *Abe wanted to <u>eat</u> something, and **so** he **did** a starfruit.

However, Type C allows an overt NP complement. Miller (1997) notes that these examples are rare in English, but attested. The overall acceptability of this construction is nevertheless debatable. The following example, for instance, sounds somewhat odd:

> (16) ?Abe <u>ate</u> a starfruit, and **so did** Betty a pear.

In Type B and C, *so* is fronted, and *do* is an auxiliary. They are compatible with non-agentive antecedents, as demonstrated below with the verbs *feel* and *want*:

> (17) Type B: I expected her to <u>feel betrayed</u>, and **so** she **did**.

> (18) Type C: He <u>wanted them to know the secret</u>, and **so did** I.

Type C differs from the other types in displaying subject-auxiliary inversion. It also differs semantically from A and B; (18) can be paraphrased using VPE and the adverbial *too*, as in (19) below:

(19) He wanted them to know the secret, and I did too.

Hankamer and Sag (1976, p. 416) suggest that the fronted *so* is a realisation of the adverbial element *too* or *also*, with which *so* is in complementary distribution, and that the rest behaves like VPE but with inverted word order. They point out that the *so*-fronting is contingent on the VPE. Furthermore, the fact that *so* is obligatory in Type C, whereas it is optional in Type A and B, may be due to its additional semantic value.

Crucially, Hankamer and Sag (1976, p. 417) point out that while the three *do so* constructions display subtle syntactic and semantic differences as exemplified above and are not necessarily related to one another, they are similar in a fundamental regard: like VPE and PG, they are all surface anaphora. The authors use examples (20 a-c) to illustrate that unlike *do it*, neither type of *do so* can undergo pragmatic control.

(20) a. [Hankamer again attempting to pass 12" ball through 6" hoop]

Sag: #I don't think you can **do so**.

b. [Sag plays William Tell Overture on recorder]

Hankamer: #And **so can** I.

c. [Hankamer plays William Tell Overture on recorder]

Sag: #And **so** he **did**. (p.418)

(20c) illustrates Type C with the auxiliary *can*, but the same ungrammaticality occurs with *do* in a sentence like *And so did I yesterday*. Analysing Type A *do so*, Baltin (2012, p. 421) concurs that it is surface anaphora.

The diachronic development of substitutive DO in Old to Middle French and Middle English:

A comparative study using parsed corpora

In the literature, discussions of *do so* anaphora generally refer only to Type A and fail to identify Type B and C (e.g. Miller, 1997; Haddican, 2007; Baltin, 2012; Garrett, 1998). Miller (1997, p. 119) distinguishes between substitution with main verb and auxiliary *do*. Based on the agentivity criterion, he groups *do so* with *do it* and *do that*, which all involve a main verb and are therefore distinct from substitutive *do* formed through VPE or PG. Miller (1997) does not explicitly acknowledge Type B or C of *do so*, which are in fact formed with auxiliary *do,* as similar to VPE or PG. However, in a different section he mentions a construction which is in fact Type C *do so.* He equates it to the *si faire 'de reprise'* construction which he groups with VPE and PG (see section 1.4 below for a description of *si faire*).

Interestingly, Miller (1997, pp. 121-122) invokes surface anaphora status in support of his unified analysis of VPE and PG. This is inconsistent with his decision to group surface anaphora *do so* with deep anaphora such as *do it*. In this study, surface anaphora status will be considered more important than auxiliary status for categorising substitutive DO.

Garrett (1998, pp. 300-301) addresses *do so* in a footnote. He rules it out as separate from VPE and PG based on the agentivity criterion and the fact that it does not allow an NP complement. However, it has been demonstrated above that these constraints only hold for Type A. The OE example he cites, quoted below in (21), has the word order SBJ + *so* + *do*. This differs from all three ModE *do so* constructions, shown in (8-10).

> (21) Se cing het hi <u>feohtan agien Pihtas</u> & hi **swa dydan**
>
> 'The king ordered them to fight against the Picts, and they did so' (ChronA 449.5-6, cited in Garrett, p. 301).

Although the OE example is translated as Type A in ModE, there is no syntactic evidence to indicate that this is the case. This highlights the difficulty of teasing apart the three types of *do*

so in historical data. It is beyond the scope and purpose of this study to account for the diachronic development of each type of *do so* separately. In fact, since the relevant criterion is surface anaphora status rather than auxiliary status, there is no need to separate them. All three constitute relevant examples of substitutive *do.*

1.3 Anaphora with *faire* in Modern French

Dagnac (2019) notes that unlike ModE, ModFr does not allow VPE or PG:

> Instead, French uses substitutive proforms […]: roughly, agentive vPs are anaphorized by *le faire* "do it" […]; adjectival and passive predicates by *le* "it" […]. Clausal complements can be represented by various pronouns (*le* "it", *en* "of it", *y* "to it", *ça* "that") according to the verb and the context . (p. 785)

This is demonstrated in (22):

(22) Agathe <u>mange une pomme</u> et Boris *(**le**) **fait** aussi. [3]

Agathe eats an apple and Boris *(it) does too.

'Agathe eats an apple and Boris does too.'

(23a) and (23b) show that deep anaphora *le faire* may strand adjuncts, but not arguments as English PG may.

(23) a. Agathe <u>offrira une pomme à Boris</u> lundi et Céline **le fera** demain.

Agathe give.FUT an apple to Boris Monday and Céline it do.FUT tomorrow.[4]

---

[3] Grammaticality judgements for ModFr examples have been provided by a native French speaker.

[4] List of morphological abbreviations in alphabetical order:

1   first person

2   second person

'Agathe will give an apple to Boris on Monday, and Céline will tomorrow.'

b. Agathe <u>a offert</u> une pomme à Boris et Céline **l'a fait** *une orange/*à Darius.

Agathe has given an apple to Boris and Céline it-has done *an orange/*to Darius.

'Agathe gave an apple to Boris and Céline did *an orange/*to Darius.'

Hankamer and Sag's (1976) example (4) in ModFr translation below in (24) shows that *le faire* constitutes deep anaphora, just like *do it*.

(24) [Sag produces a cleaver and prepares to hack off his left hand]

Hankamer: Ne vous inquiétez pas, mesdames et messieurs, nous avons répété ce

numéro plusieurs fois et il ne **le fait** pas vraiment.

Of course, the anaphor in (24) may also be syntactically controlled by the linguistic

antecedent in (25) which corresponds to that in (5):

---

3   third person

COND   conditional

FUT   future

IMP   imperative

NEG   negation/negative element

PL   plural

RFL   reflexive pronoun/clitic

SBJV   subjunctive

SG   singular

Person and number distinctions are specified in cases where there is no overt subject, and to avoid ambiguity

where the English pronoun 'you' is used in the gloss. The French T-V politeness distinction is not specified:

formal singular *vous* is marked the same as plural *vous* (2PL).

(25) Sag: Je vais maintenant <u>trancher ma main gauche avec ce couperet</u>.

Hankamer: Ne vous inquiétez pas, mesdames et messieurs, nous avons répété ce numéro

plusieurs fois et il ne **le fait** pas vraiment.

1.4 Anaphora with *faire* in Old and Middle French

Original observations by Miller (1997) reveal that substitutive DO existed in Old and Middle

French, along with substitutive uses of *estre* 'be' and *avoir* 'have'. He cites the following

examples (26) and (27) which involve *faire* in VPE and PG respectively:

(26) Si <u>souffrira</u>. – C'est bien dit, voirement **fera**.

So he suffer-FUT. – Well said, he does-FUT indeed.

'He will suffer indeed.' (Miller, 1997, p.122[5])

(27) L'<u>ont</u> miex <u>ben'ie et sacree</u>

Que il n'**ont** une autre contree

'They have blessed and sanctified it better than they have another land.'

(Miller, 1997, p.123)

Another substitutive use of *faire* identified by Miller (1997) is *faire* preceded by *si* or *non*, as

in (28) and (29).

(28) <u>Del dol s'asist</u> la medre <u>jus a terre</u>,

Si **fist** la 'spose dan Alexis

'The mother sat on the ground because of her pain,

So did the wife of Sire Alexis.' (Miller, 1997, p. 125)

---

[5] See Miller (1997) for the original sources of cited examples.

(29) 'Car <u>retornons en France</u>..!' –'Non **ferons**' […]

' "Let us return to France!" – "We shall not" […] '   (Miller, 1997, p. 126)

Miller (1997, p. 126) states that the *si faire* '*de reprise*' found in (28) has a 'very similar discourse function to the *so* + AUX + SBJ construction in ModE' (p.126), referred to here as Type C. Although not explicitly stated, this may indicate that Miller (1997) considers the English Type C construction a case of substitutive *do,* as opposed to Type A which he excludes on the basis of the agentivity criterion. Although there is no mention of Type B, it would presumably be grouped with Type C based on the same criterion.

In any case, based on the framework of the present study, the similarity between *si faire* and *do so* noted here provides further support to treating *si faire* – and the similar construction *non faire* – as cases of substitutive DO.

Based on examples from secondary sources, Miller (1997, p. 127) suggests that all substitutive uses of auxiliary *faire*, as well as *avoir* and *estre*, disappeared together during the sixteenth and seventeenth centuries. However, he notes that 'serious corpus-based research would be required to establish the details of these developments' (p. 127). Indeed, this is what section 2.1 will be dedicated to.

1.5 A note on the interpretation of historical data

The definition of substitutive DO outlined above is based on the distinction between deep and surface anaphora. Naturally, it is very difficult given the written medium of the source material to confirm whether anaphora in historical French or English texts can undergo pragmatic control or not. However, the aim is not to provide a detailed analysis of French and English diachronic syntax, but rather to use this as a framework to identify a group of related

constructions, and guide the search for substitutive DO. In practice, this involves including *do so* whereas certain studies (e.g. Miller, 1997; Garrett, 1998) do not.

2. Corpus studies

Parsed digital corpora are an especially useful tool in historical linguistics, which, unlike synchronic linguistics, has to rely entirely on text as a data source. Compared to the traditional method of manually reading manuscripts and counting occurrences of a given phenomenon, this process is much less time-consuming and less prone to errors by omission or by inconsistent analysis of ambiguous cases.

As in all diachronic research, however, access to complete and representative data is more limited than in synchronic studies. The textual record inevitably overrepresents certain dialects and sociolects, for instance as literacy rates differ between social classes, and the variety of a single productive scribe may not reflect the general language use. Apart from reported speech and meta-descriptions of spoken language use, which are typically scarce, the textual record offers little information about spoken language. Furthermore, there is generally a time gap between when a feature is introduced in spoken language and when it appears in writing (Ingham, 2014, p. 29). The length of this lag may be affected by sociolinguistic factors such as dialect attitudes, degree of standardisation, and writing norms within a given language. However, if considered together with other factors such as sociolinguistic situation and linguistic structure, diachronic data offers vital insight into linguistic change. Below, I will present two corpus studies examining the diachronic development of substitutive DO, looking first at Old and Middle French, and then Middle English.

2.1 Study 1: Old and Middle French

2.1.1 Corpus: MCVF

The MCVF (Martineau, Hirschbühler, Kroch & Morin, 2010) is a syntactically annotated digital corpus containing just over 1 million words, and additionally nearly 225,000 words in the Penn supplement (Kroch & Santorini, 2010). The corpus comprises manuscripts dated between 842 and 1738, covering the Old, Middle, Renaissance, and Classical French periods. The manuscripts dated before the 17th century are assigned to four large dialect areas: Anglo-Norman, Picardian, Champenois/Parisian, and Central-Western, and data from the 17th and 18th centuries also includes Canadian texts (Martineau, 2008, p. 139). Furthermore, standardised annotation enables more reliable and efficient cross-linguistic comparisons, and the MCVF and the PPCME2 (described below in Section 2.2.1) use largely the same annotation system.

2.1.2 Queries

2.1.2.1 Substitutive DO coding query

The data was extracted using the software CorpusSearch 2 (Randall, 2005). A coding query (Fr1.c, reproduced in the Appendix[6]) was formulated to classify different syntactic constructions involving the verb *faire*. In a second version of the coding query (Fr2.c), each category was assigned the value 'Keep' or 'Exclude', based on whether the results generated by each column constituted cases of substitutive DO or not. An evenly distributed sample of the results was examined manually to verify the accuracy of the query. Based on this, the query was revised repeatedly to generate more precise results.

---

[6] All queries are reproduced in the Appendix.

Because *faire* does not have its own tag that sets it apart from other verbs in the MCVF corpus, like *do* does in the PPCME2, the first step was to compile a list of all instances of *faire* that occur in the corpus, to refer to the coding query. This was achieved through a query which listed all verb forms that began with the letter 'f'. These were subsequently organised alphabetically, and duplicates were eliminated automatically. After that, items that were not forms of *faire* were identified and deleted manually. The website DicFro was used as a reference for various historical spellings of *faire* (http://micmap.org/dicfro/search/tableaux-de-conjugaison/faire). This left 276 tokens of *faire*, with various spelling variations and inflections.

Subsequently, the coding query searches each IP node of the corpus. The first column of Fr1.c picks out finite forms of *faire*[7]. The second column classifies different syntactic constructions involving *faire*, by referring to the list of *faire* forms.

There are obvious difficulties in identifying substitutive ᴅᴏ in the corpus data, since elided VPs are not represented in the MCVF or the PPCME2, and there is no tag that specifies whether ᴅᴏ is a main verb in its own right or whether it is used anaphorically to replace a deleted VP. Hence, all instances of ᴅᴏ must be examined, and the relevant constructions are identified mainly through a process of elimination. The first 17 rows in column 2 rule out

---

[7] Most cases of substitutive *do* involve an auxiliary and will thus occur in finite form only. However, non-finite as well as finite forms occur in *do so* (type A) constructions which involve a dummy main verb rather than an auxiliary:

(ii) I have <u>gone swimming twice today</u>, and Betty **has done so**, too.

For practical reasons, however, in the interest of eliminating noise, corpus searches were limited to finite uses of ᴅᴏ.

irrelevant constructions, of which the most frequent ones will be reviewed here. Firstly, *faire* is most commonly used for reported speech, as in (30):

(30) - Por quoi? fet Galaad.

- Why? did Galaad.

'- Why? said Galaad.' (1225-QUESTE-P,60.1649)

Moreover, cases where *faire* has an IP complement are ruled out. This captures the *faire* + infinitive construction meaning 'cause X to Y' exemplified below, which still exists in ModFr:

(31) E! reis celeste, tu nus i fai venir!

And! king celestial, you.2SG us here make come!

'And! celestial king, you make us come here!' (10XX-ALEXIS-V-PENN,67.612)

Furthermore, there are numerous cases where non-substitutive main verb *faire*, meaning something like 'make', takes an NP complement as a direct object.

(32) signes fazen per podestad.

miracles make.3PL by power.

'they do miracles by their power.' (1000-PASSION-P-BFM,124.368)

In the same way, deep anaphor *le faire* was eliminated:

(33) **fist lo** mul ben.

did.3SG it very well

'He did it very well.' (0980-LEGER-V-PENN,XIV.96)

However, excluding cases of non-substitutive transitive DO also automatically excludes PG. For this reason, PG will be treated in a separate query (see next section). Another non-

substitutive use is main verb *faire* with a 'that'-clause complement, which means roughly 'bring about consequence X'.

> (34) et si le trouvez, faictes qu' il soit ramené mort ou vif.
>
> and if him find.2PL, do.2PL.IMP that-he be.SBJV brought dead or alive
>
> 'and if you find him, make sure that he is brought back dead or alive.'
>
> (1527-BAYART-P-PENN,68.1360)

Exemplified in (35) and (36) respectively, reflexive uses of *faire* are excluded, as is the *faire + à +* infinitive construction which is translated by Spearing (1993, p.104) using a passive construction in English:

> (35) et si telz achaptz se font, nous voullons qu' ilz soient et demourent en nostre main.
>
> and if such purchases RFL.3PL make.3PL, we want that-they be.SBJV and remain.SBJV in our hand.
>
> 'and if such purchases are made, we want them to remain in our hands.'
>
> (1309-JOINVILLE-P,349.5205)

> (36) Ne feseit mie a refuser!
>
> NEG did.3SG NEG to refuse!
>
> 'It was not to be disdained!'   (116X-MARIE-DE-FRANCE-R,77.1567)

Once the above constructions have been excluded, the next eight rows in the query define instances of substitutive DO. This includes *si faire* and *non faire*:

> (37) La dame portoit son filz en son bras, et pensa qu' ele se reposeroit a la fontainne, et elle **si fist**,

The lady held her son in her arm, and thought that-she RFL rest.COND by the

fountain, and she so did,

'The lady held her son in her arms, and thought that she would rest by the fountain,

and so she did,'   (127X-CASSIDORUS-P-PENN,335.2601-2603)

(38) «Sire», dist le conte, «ne vous <u>doubtez de moy</u>.»

«**Non fay** je, sire», dist l' abbé,

«Sir», said the count, «not RFL.2PL doubt.2PL.IMP of me.»

«Not do I, Sir», said the-abbot,

'"Sir", said the count, "do not doubt me."

"I do not, Sir", said the abbot,' (133X-PERCEFOREST-P-PENN,123.1105-1106)

Furthermore, there are comparative structures involving substitutive DO:

(39) Mielz <u>valt</u> mesure que ne **fait** estultie.

More is-worth moderation than NEG does pride.

'Moderation is worth more than pride.'   (1100-ROLAND-V,131.1722)

(40) qu' il <u>gisoient</u> el mileu del feu come il **feissent** en un lit de roses

that-they lay in-the middle of-the fire as they do.SBJV in a bed of roses

'that they lay in the midst of the fire as they would in a bed of roses'

(125X-EUSTACE-MURRAY-P-PENN,43.597)

Substitutive DO was also identified in subordinate adverbial clauses:

(41) S' ensi est por kai <u>aparurent</u> il dons anzois as pastors des berbiz k'il ne **fesissent**

as rois de la terre nen as prestes del temple?

If-so is.3SG why appeared they then so to-the herders of-the sheep than-they NEG

do.SBJV to-the kings of the earth nor to-the priests of-the temple?

'If so, why did they appear to the shepherds rather than to the kings of the earth and the priests of the temple?' (1190-SBERNAN-P-BFM,71.2390)

Categories ordered below this section of the coding query were naturally excluded.

2.1.2.2 PG queries

Due to the impossibility of automatically separating PG from non-substitutive transitive faire in Fr1.c, two queries were formulated to investigate PG separately. Firstly, a coding query (FrPGcmp.c) looked for PG in comparative clauses, where the construction is most frequent (Miller, 1997, p. 121). The first column identifies the relevant clause type, namely an IP which dominates a finite verb which in turn has a comparative clause as a sister. The second column picks out cases where the finite verb is something other than *faire*, since the aim is to find cases where *faire* substitutes for another verb. Finally, the third column ensures that the comparative clause dominates a subordinate IP containing a finite verb *faire* which has a direct object NP as a sister.

(42) car il <u>congnoissent</u> otant bien toute la marce ou il estoient, aloient et venoient, que il **font** lor pais d' Escoce.

for they knew as well whole the area where they were, came and went, than they did their land of-Scotland.

'for they knew just as well the whole area where they were, came and went, as they did their Scottish land' (1369-FROISSART-1-P,137.2158)

Secondly, another query (FrPGadv.q) finds PG examples in in subordinate adverbial clauses, as in (43):

(43) Car premierement il leur <u>moustre</u> les vins et les viandes qui sont beles et delicieuses (comme il **fist** a Eve la pomme),

For first he them shows the wines and the meats that are beautiful and delicious (as he

did to Eve the apple),

'For first he shows them the beautiful and delicious wines and meats (as he did the

apple to Eve),'   (1279-SOMME-ROYAL-P-PENN,1,51.1364)

2.1.3 Results

2.1.3.1 Substitutive DO results

The coding query identified 643 cases classed as substitutive DO in the corpus, out of 403,039

IPs in total. 15,952 IPs contained a finite form of *faire*. Averaged across time periods,

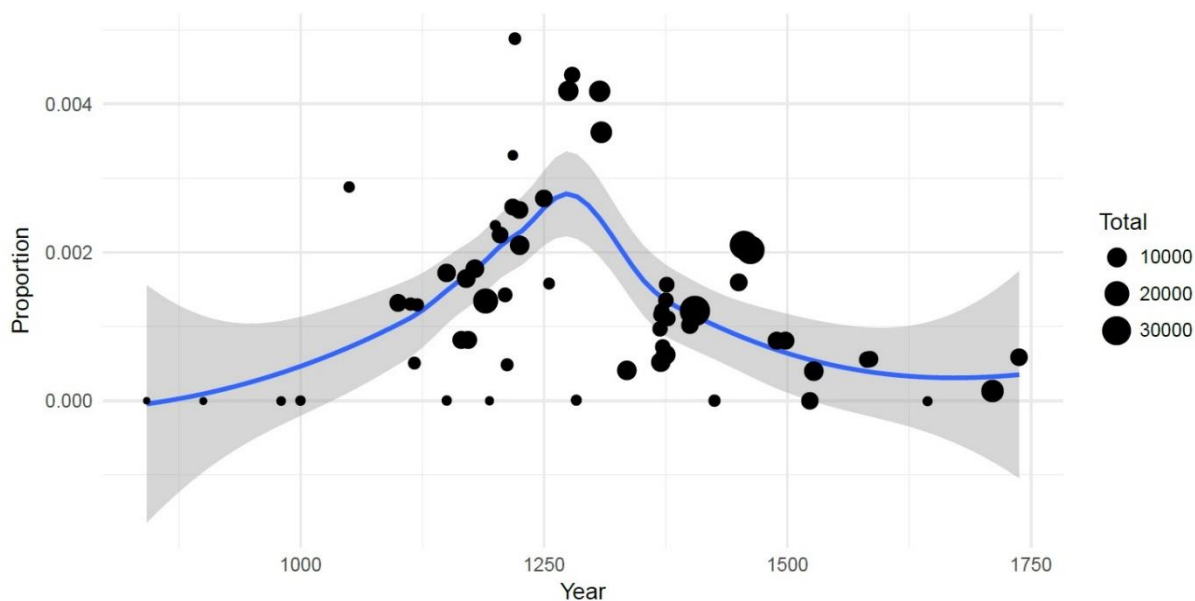substitutive DO occurs in 0.16% of all IPs and constitutes 4.0% of the occurrences of finite

faire.

Figure 1: Substitutive DO in OFr and MFr

The results were visualised in RStudio (RStudio Team, 2015). Figure 1 shows the diachronic

frequency of substitutive DO. The X axis shows which year each text was written, and the Y

axis represents the proportion of IPs containing substitutive DO out of the total number of IPs

in the text. Each dot represents one text, and the size of the dot illustrates the total number of IPs in that text.

The first occurrence of substitutive DO is recorded in the middle of the 11th century. The construction is well attested c. 1100-1475. It is particularly well-attested around 1250, during a short but apparently robust period of common use. Then, from 1500 onwards, substitutive DO is sparsely attested but does not completely disappear. These findings support Miller's (1997) observations about when substitutive DO starts to decline, however attestations as late as the mid-18th century indicate that the construction is not completely gone by 1700 as Miller (1997) hypothesises.

2.1.3.2 PG results

Table 1 below shows the number of French PG cases generated by the two queries together:

Time period	1100-1200	1200-1300	1300-1400	1400-1500	Total
Number of PG cases	6	7	9	1	23

Table 1: PG in OFr and MFr

Across the corpus, PG (at least as detected in these two contexts) occurs in 0.006% of the total number of IPs, and in 0.15% of the IPs containing a finite form of *do*. PG is attested between 1100 and 1500. There is not enough data to draw any conclusions about the development of this feature. The number of cases may be limited because the search is only conducted on two clause types, as opposed to the other cases of substitutive DO, however it is clear from manual inspection of the corpus that PG is a less frequent construction.

The first query, which targeted comparative clauses, generated 18 cases whereas the adverbial clause query added 5 cases. It was more difficult to eliminate noise while searching for PG in adverbial clauses, however as the query returned only thirteen results, the relevant cases were

picked out manually. The other eight results from that query were discarded, either because they were duplicates or otherwise irrelevant. For instance, four involved the set phrase *faire lou samblant*, equivalent to ModFr *faire semblant.*

2.1.3.3 Substitutive DO in Anglo-Norman and Continental French

One can imagine a methodological problem with using a mix of Continental French (CF) and Anglo-Norman (AN) data, as in the MCVF corpus, to represent features of AN. However, examining a wide range of grammatical features, Ingham (2012) finds that Anglo-Norman follows the syntactic development of Continental French in the 11th to 14th century, the period in which AN is used the most, rather than assimilating to the syntactic features of English. He states the following regarding CF influence on AN:

> […] [C]ore syntactic distinctions continued to be reliably transmitted even without equivalents in Middle English. It was only in the 14th century, notably with noun gender assignment, that the native-like transmission of syntactic distinctions showed signs of breaking down. (p. 162)

To test whether the structural similarity holds between CF and AN regarding substitutive DO, another subset of the MCVF data was analysed. The corpus contains five AN texts, all from the 12th century, which were compared to six 12th century CF texts. The titles and dates of the manuscripts are listed in the Table 2 below. 'X' indicates an unknown value.

AN text	Year	CF text	Year
La vie de St Alexis	10XX	*Lapidal*	1117

La chanson de Roland	1100	*Les lais de Marie de France*	116X
Le voyage de St Brendan	1120	*Yvain*	1179
Li quatre livre des reis	1150	*Le roman de l'estoire dou Graal*	1190
Leis Willelme	1150	*Bernan*	1190
		Chièvres	1194

Table 2: AN and CF manuscripts

In AN, the Fr2.c coding query identified 28 cases of substitutive ᴅᴏ found in a total of 18,066 IPs, of which 609 IP contained finite *faire.* Thus, substitutive ᴅᴏ occurred in 0.15% of all IPs, and in 4.6% of all finite uses of *faire*. By comparison, the CF texts contained 69 cases of substitutive ᴅᴏ out of 53,584 IPs, of which 1,899 involved finite *faire*. The frequency of the construction was 0.13% of all IPS and 3.63% of finite uses of *faire*. Substitutive ᴅᴏ is slightly less frequent in 12th century CF texts than in AN, however the difference is not large enough to conclude that AN diverges from CF. In addition to the cases of substitutive ᴅᴏ reported above, the PG queries identified only two cases of PG in CF and none in AN.

2.2 Study 2: Middle English

2.2.1 Corpus: PPCME2

The PPCME2 comprises roughly 1.2 million words from manuscripts dated between 1150 and 1500 (Kroch & Taylor, 2000). The dialect areas represented in the corpus are Kentish, Southern, West Midlands, East Midlands, and Northern (Kroch & Taylor, 2016, https://www.ling.upenn.edu/hist-corpora/PPCME2-RELEASE-4/index.html).

2.2.2 Queries

2.2.2.1 Substitutive DO coding query

Unlike the MCVF (with *faire*), the PPCME2 corpus provides a separate tag for *do*, which makes the identification process easier. However, constructions involving *do* still had to be categorised as substitutive or non-substitutive. The coding queries (ME1.c, ME2.c) and the process closely resembled those adopted for the MCVF data. Again, the search focused on finite forms of *do.* As with the French data, numerous examples of *do* with an IP complement, as in (44) or direct object NP complement (45) were excluded.

> (44) Huanne he ous dede come to þe cristenedome we were poure and naked / and child of yre / and of helle.
>
> When he us did come to the christendom we were poor and naked and child of ire and of hell.
>
> 'When he brought us into Christendom, we were poor and naked, and children of ire and of hell.'   (CMAYENBI-M2,101.1979)

> (45) We rede also of othir maydens, whan thei were put in the fyre to be brent, the fire vanysshed and did hem no harme.
>
> 'We also read of other maidens, that when they were put in the fire to be burnt, the fire vanished and did them no harm.'   (CMAELR4-M4,11.284)

Subsequently, substitutive *do* was identified. *Do so*, and *do thus* which was analysed as an identical construction, were included in the results as substitutive DO. These are illustrated in (46-48):

> (46) This is to seyn that when eny sterre fix is passid the lyne meridional, than begynneth it to descende; and **so doth** the sonne.

'This is to say that when any fixed star has passed the meridional line, it begins to descend; and so does the sun.   (CMASTRO-M3,673.C2.385)

(47) Lotryn had sent to Camber, his Brother, þat he schuld <u>come also to hym with all þe power that he myght</u>, <u>hym for-to helpe</u>; & **so** he **dede**, with good will.

Lotryn had sent to Camber, his brother, that he should come also to him with all the power that he might, him for to help; and so he did, with good will.

'Lotryn had sent for Camber, his brother, to come to him too with all the power that he could muster, to help him; and so he did, happily.'   (CMBRUT3-M3,12.341-342)

(48) and sythen, þay <u>ledde Hym dreryly to þe dede</u> and ȝitt neuer He sayde till thaym anes why þay **swa dyde**.

and then, they led Him drearily to the death and yet never He asked to them once why they so did.

'and then, they led Him drearily to his death, and yet he never asked them why they did so.'   (CMEDTHOR-M34,43.618-619)

Moreover, the coding query identifies substitutive *do* in comparative structures as in (49-50) below:

(49) for ye <u>trespassen</u> so ofte tyme as **dooth** the hound that retourneth to eten his spewyng.

for you trespass so often time as does the hound that returns to eat his vomit.

'for you trespass as often as the dog that returns to eat his own vomit.'

                                        (CMCTPARS-M3,290.C1.78)

(50) for sche <u>weryd white clothyng</u> mor þan oþer **dedyn** whech wer holyar & bettyr

þan euyr was sche as hym thowt.

'for she wore white clothing more than others did, who were holier and better than she

ever was, as he thought.' (CMKEMPE-M4,84.1889)

Substitutive *do* is also found in subordinate adverbial clauses (51), and in constructions of the

type 'if SBJ *do*' (52):

(51) <u>Wrastle irnestly wit þy God</u> as Iacob **dede**,

'Grapple earnestly with your God as Jacob did,' (CMAELR3-M23,43.494)

(52) Sain Benet spekis of þa þat ere sent owte ani erand and sal cum o-gain þat ilke

day, þat tay ne be noht sua hardy at ete owte, þoz man pray þaim, Bot yef þabbes <u>giue</u>

<u>þaim leue</u>; yef þai **do**, þai sal be cursid.

Saint Benet says of those that are sent out any errand and shall come again that same

day, that they NEG be not so hardy to eat out, even if someone asks them, But if the

[abbotts] give them leave; if they do, they shall be cursed.

'Saint Benet says of those that are sent out on any errand and shall come back that

same day, that they should not be so hardy as to eat out[side the abbey], even if

someone asks them to, unless the abbots give them leave. If they do [eat out], they

shall be cursed.' (CMBENRUL-M3,34.1111-1112)

2.2.2.2 PG queries

As in study 1, two queries (MEPGadv.q and MEPGcmp.q) were used to identify PG in ME

adverbial clauses and comparative clauses, exemplified in (53) and (54) respectively:

(53) and seyde: "Certes, sire, sauf youre grace, I love youre honour and youre profit as

I do myn owene,

and said: "Certainly, sire, save your grace, I love your honour and your profit as I do

my own,

'and said "Certainly, Sir, with all due respect, I love your honour and care about your

health as much as I do my own,'   (CMCTMELI-M3,235.C2.716)

(54) He dredyth þe mor þan thow dost hym.

'He dreads you more than you do him.'   (CMKEMPE-M4,51.1143)

2.2.3 Results

2.2.3.1 Substitutive *do* results

The coding query identified 740 cases classed as substitutive ᴅᴏ, out of 181,423 IPs in total.

2,812 IPs contained a finite form of *do*. Averaged across time periods, substitutive ᴅᴏ occurs

in 0.41% of all IPs, and constitutes 26.32% of the occurrences of finite *do*. This last

percentage is noticeably higher than the corresponding percentage in French (4.0%).

However, finite *do*, occurring in 1.5% of all IPs is overall less frequent than finite *faire* is in

French, at 4.0% of all IPs. This may be because a larger proportion of finite *faire* than finite

do is used in non-substitutive transitive contexts, since English also has *make* for that purpose.
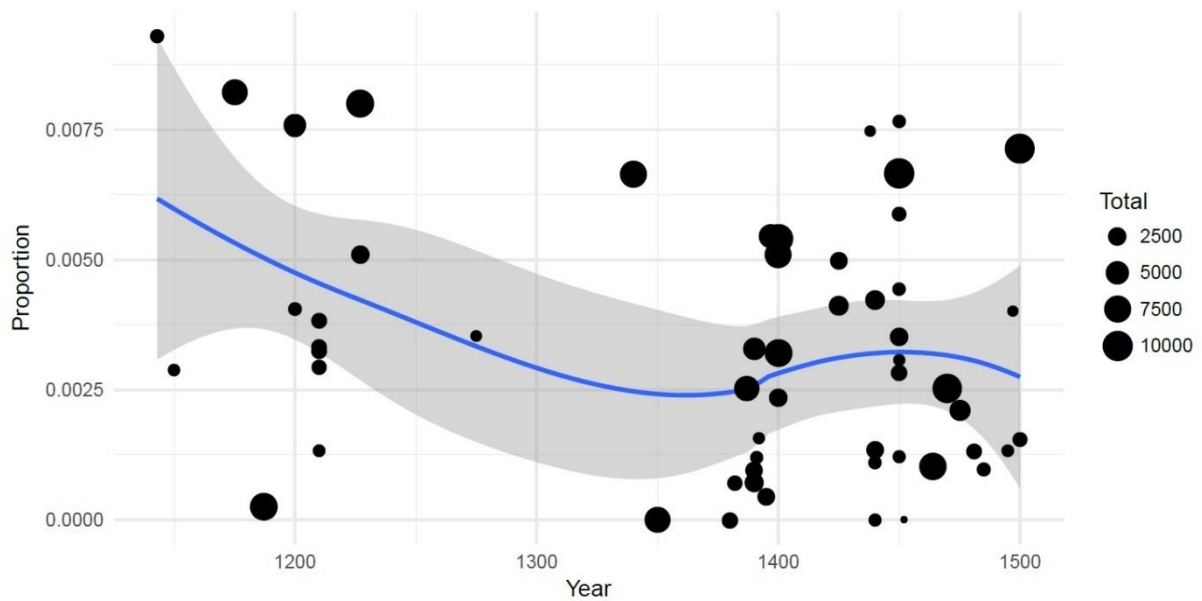
Figure 2: Substitutive do *in ME*

Substitutive ᴅᴏ is present throughout the ME period (see Fig. 2). There is a gradual decrease from approximately 0.6% out of all IPs in the 12th century to around 0.3% in the 15th century.
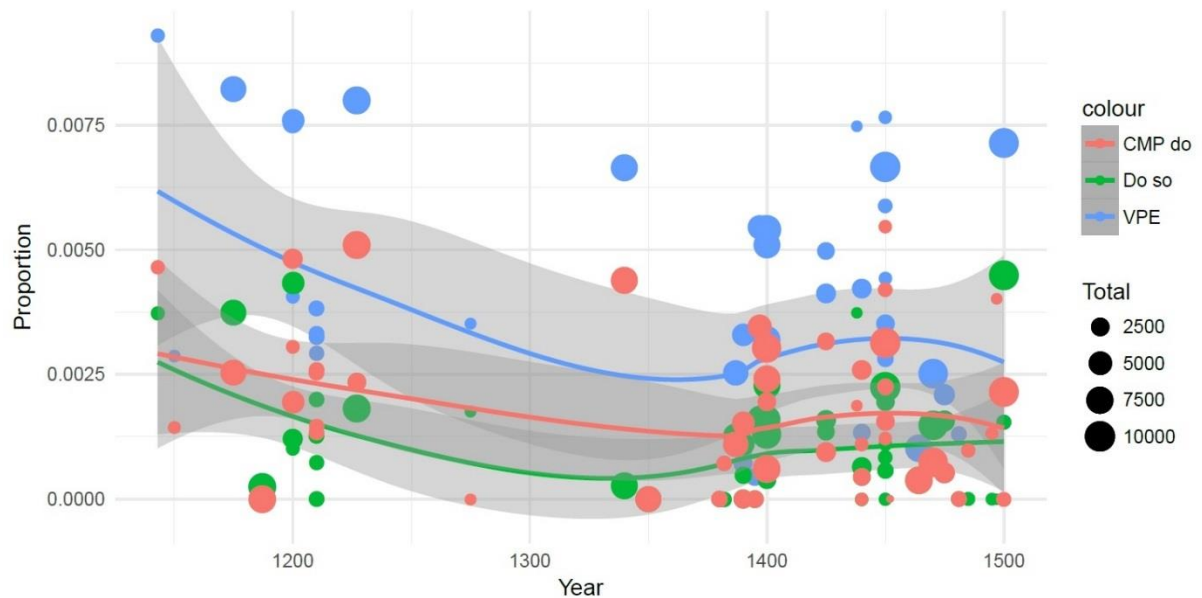

Figure 3: Subsets of substitutive do *in ME*

Figure 3 tracks the diachronic development two large subsets of the instances of substitutive *do*, namely *do so* (labelled 'Do so') and substitutive *do* in comparatives (labelled 'CMP do').

Warner (1993, pp.117-118) notes that these constructions were common in OE, and this observation is evidently true of ME as well. These subsets follow essentially the same pattern as each other and as substitutive *do* in general, further supporting the analysis of *do so* as a type of substitutive *do*.

2.2.3.2 PG results

Across the corpus, PG occurs 0.018% of the total number of IPs, and in 1.14% of the IPs containing a finite form of *do*. Again, these results are limited to comparatives and adverbial clauses.

Time period	*M1 (1150-1250)*	*M2 (1250-1350)*	*M3 (1350-1420)*	*M4 (1420-1500)*	*Total*
Number of PG cases	29	7	20	18	32

Table 3: PG in ME

The time period reflects the date of the manuscript (Kroch & Taylor, 2016

https://www.ling.upenn.edu/hist-corpora/PPCME2-RELEASE-4/index.html). In the present study, manuscript date is prioritised over composition date in cases where these differ.

3. Discussion: Evaluating contact hypotheses

We have seen that substitutive DO is well-attested in French around 1100-1475, with a peak in usage around 1250, and sparsely attested from 1500 onwards. It is robustly attested in CF and AN in the 12th century, with a marginally higher frequency in AN. Substitutive DO is overall more frequent in ME than in French in the corresponding period. The usage is relatively stable, but decreases slightly and gradually throughout ME. PG is attested in both languages, slightly more frequently in ME. However, as the PG results are limited to two specific

syntactic contexts, they cannot reliably indicate the overall distribution and diachrony of this feature.

It is a reasonable hypothesis to pursue that the linguistic parallelism of substitutive DO between French and English in the medieval period is the result of contact influence. In order to establish whether a change is the result of language contact, one must address linguistic, sociolinguistic, and diachronic evidence.

Thomason and Kaufman (1988, p. 37) define two different patterns of contact-induced change: borrowing and shift. Borrowing is characterised by 'strong long-term cultural pressure from source-language speakers on the borrowing-language speaker group' (Thomason & Kaufman, 1988, p. 37). Lexical items are transferred first, before phonology, syntax, and inflectional morphology.

A shift, on the other hand, involves interference through imperfect learning. Speakers shift to a dominant target language (TL), bringing features from their original language which are subsequently adopted as part of the TL. This is also known as substratum interference. Unlike borrowing, this change begins with syntax, phonology, and occasionally morphology. Lexical influence is typically very limited (Thomason & Kaufman, 1988, p. 39). Structural change through a shift can happen quickly, even in the space of a generation, whereas structural change through borrowing typically requires centuries of intense contact (Thomason & Kaufman, 1988, p. 41).

3.1 French influence on English

As established above, substitution in OFr and MFr is linguistically very similar to that of ModE. The traditional view is that there was extensive lexical borrowing from AN into ME in the centuries following the Norman Conquest, however only minimal syntactic borrowing

(e.g. Thomason & Kaufman, 1988). Thomason and Kaufman (1988, p. 275) argue that the number of French speakers in England was always low, and that they started to abandon the language around 1235. They estimate that from 1265 onwards, use of French in England was artificial and largely ungrammatical, serving mostly as a class marker. This position is challenged by Haeberli (2010), Trotter (2003), and Ingham (2010, 2012, 2014), who claim that the possibility of structural borrowing has been underestimated.

Ingham (2010, p. 1) stresses that French-English bilingualism was more extensive and long-lived than standard accounts allow. While English was the first language and the spoken vernacular of the majority of the population, AN had the status of a prestige written vernacular, having spread to virtually all written registers by the 14th century. In addition, AN was transmitted as a spoken language in French so-called 'song schools', where children were taught literacy by members of the clergy. This transmission is characterised by Ingham (2010) as 'childhood second language acquisition in a quasi-naturalistic context' (p. 462). Thus, AN constituted 'a spoken natural language among educated professionals', until its transmission was disrupted by the Black Death in the late 14th century (Ingham, 2012, pp. 161-162). AN then disappears from the written record in England in the early fifteenth century (Ingham, 2012, p. 162). Notably, Ingham (2010, p. 2) argues that syntactic features may have passed from French to English though the 'code-crossing' of scribes in the late fourteenth to early fifteenth century, who frequently calqued English sentences on French.

Ingham (2014) argues that elliptical tag questions and answers, [8] which constitute a subset of substitutive DO, are a structural borrowing from AN into English. A premise for Ingham's

---

[8] (iii) and (iv) exemplify elliptical tag questions and elliptical answers respectively:

       (iii) I have heard Peter, haven't I [VP heard Peter]?

conclusion is that elliptical tag questions and answers were absent in Old English, 'an expectation which seems correct' (p.35). Although elliptical tag questions and answers were not a specific focus of the corpus searches outlined above in section 2, *do* did not appear to occur frequently in these constructions. Substitutive *do* as a whole, however, is attested in OE, as seen in (21). Warner (1993, pp.117-118) confirms that substitutive *do* is attested in OE and ME in both VPE and PG contexts. He notes that substitutive *do* in OE mostly occurs in comparatives or with *swa*, which corresponds to ModE *so*.

Similarly, Haeberli (2010) investigates four aspects relating to the development of subject-verb inversion in Middle English, and proposes French contact influence as a plausible explanation for three out of the four phenomena. However, he concedes that while French is a plausible candidate, there is not enough evidence to confirm that this was the source of the change nor to rule out other potential sources (Haeberli, 2010, p.161).

However, the diachronic evidence on substitutive DO presented above argues against structural borrowing from French. If contact influence were the case, first appearance or an increase in usage of substitutive DO in ME would be expected some time after 1066, allowing for a delay before the feature appears in writing. However, given its early and stable incidence in Middle English, substitutive DO does not fit this description.

(iv) – Haven't you heard Peter?

– Yes, I have [VP heard Peter]   (Ingham, 2014, p. 26)

3.2 English influence on French

Considering the diachronic data in isolation, it is in fact more likely that the feature passed from English to French after 1066, causing a peak in French around 1300. It then declines and finally disappears in French, but remains in English until the present day.

However, borrowing from ME to AN is highly implausible given the sociolinguistic situation and linguistic evidence. AN was the prestige variety, and there is no evidence of heavy lexical borrowing from ME to AN, which would be a prerequisite for structural borrowing to occur. Structural influence without lexical transfer could however be the result of substratum interference. Since AN primarily existed in Britain as a second language, it is conceivable that ME syntactic features were incorporated through imperfect learning. Nevertheless, the situation is not typical for a shift since English was not rapidly abandoned in favour of AN, but remained as the first language throughout the three centuries of contact with AN.

Furthermore, as previously mentioned, AN consistently developed in line with CF during this period as part of the medieval French dialect continuum, rather than adopting ME syntactic features (Ingham, 2012; Trotter, 2003). The corpus data presented in 2.1.3.3 confirms this picture, showing that substitutive ᴅᴏ is attested in both CF and AN, at similar rates. The slightly higher frequency in AN is not large enough to suggest a contact influence from English. In a substratum interference scenario, the change would have had to travel from ME to AN, and from there to CF. This would contradict the directionality of linguistic influence described above, which is supported by lexical, syntactic, and sociolinguistic data.

3.3 Celtic contact hypotheses and periphrastic ᴅᴏ

Certain scholars propose that English periphrastic *do* is the result of Celtic contact influence (e.g. Preusler, 1938; Poussa, 1990; Tristram, 1997). While structural borrowing from Celtic to

English is unlikely given the scarcity of Celtic loanwords in English (van der Auwera & Genee, 2002, p. 284), substratum interference is more plausible.

By periphrastic DO, van der Auwera and Genee (2002, p. 286) intend the constructions also known as DO-support, where auxiliary DO is combined with a main verb to express negation, interrogation, and emphasis. Somewhat confusingly, they also use periphrastic DO as a wider term, including aspectual/habitual DO in nonemphatic positive assertions, causative DO, DO combined with a verbal noun, pragmatic DO, and, notably, substitutive DO. In this discussion, periphrastic DO will refer to DO-support uses only, as in Garrett (1998).

Van der Auwera & Genee (2002) argue that there are considerable linguistic similarities between use of periphrastic DO in English and Celtic languages. They express tentative support for a Celtic hypothesis, stating that there is good circumstantial evidence indicating a direct Brythonic influence on English periphrastic *do*, but that there is not enough direct evidence to prove or disprove this hypothesis. However, there is little mention of substitutive DO specifically, beyond the fact that it exists in Celtic languages (van der Auwera & Genee, 2002, p. 291).

According to Garrett (1998, p. 286), Celtic hypotheses lack sociolinguistic evidence to support the posited contact situations. Another important criticism raised by Garrett (1998) concerns the considerable time gap between the onset of Anglo-Celtic contact in early OE (van der Auwera & Genee, 2002, p. 296) and the first unambiguous attestations of periphrastic *do* in 1300 (Garrett, 1998, p. 284).

Van der Auwera and Genee (2002) fail to acknowledge, however, that substitutive *do* existed in OE. Thus, a Celtic origin appears more likely for substitutive *do* than for periphrastic *do.* In

order to test this hypothesis, further corpus research on the diachronic and diatopic development of substitutive *do* in OE would be required.

3.4 Endogenous change and multiple causation

A further point of interest for diachronic research is whether substitutive *do* is a precursor to periphrastic *do*, a hypothesis supported by Warner (1992). He states that the fact that substitutive *do* 'is formally closely related with the auxiliary group in Old English implies that we may need to rethink its significance' (p. 204). It seems plausible that the later auxiliary uses involved in periphrastic *do* developed as an extension of the grammatical functions of substitutive *do*, rather than completely independently.

Garrett (1998) proposes that this extension was driven by habitual uses of *do*, which originated in the 12th or 13th century from reanalysis of the lexical *do* + verbal action noun construction. Habitual *do* merged with substitutive *do* through reanalysis, bringing with it the ability to take an infinitive complement. This merger resulted in the emergence of periphrastic *do.* Garrett's (1998) theory has the merit of accounting for the otherwise puzzling time gap between the onset of substitutive *do*, at some point before ME, and periphrastic *do* in the 13th century.

As such, there is a plausible endogenous change scenario for the appearance of periphrastic *do*. In my view, the null hypothesis is that syntactic change is endogenous rather than contact-induced. As stated by Thomason and Kaufman (1988), structural borrowing requires intense and prolonged contact, since syntactic features are not as easily transferred as lexical items. Contact-induced syntactic change occurs more easily and quickly through a shift. However, this requires a rather dramatic sociolinguistic situation, where speakers rapidly abandon their native variety in favour of a target variety. Crucially, the change happens at the expense of the

shifting speakers' variety, which declines considerably or disappears. My view is that this situation cannot be posited as the norm for syntactic change, without sociolinguistic evidence to corroborate the occurrence of a shift. Nor can structural borrowing be posited without sociolinguistic evidence of extensive and prolonged contact, and linguistic evidence of heavy lexical borrowing.

 As a default scenario, it is more likely that syntax evolves through transmission over time, as specific features are reanalysed, discarded, and adopted in different contexts. However, as always when endogenous change is considered, the question of actuation must be addressed. It is also worth noting, as highlighted by Thomason & Kaufman (1988, p. 57) and van der Auwera and Genee, (2002, p. 302), that contact-induced and endogenous change are not mutually exclusive but may interact in rather complex ways. For example, it seems plausible that a feature that develops through endogenous change may be reinforced through contact with a variety which shares that feature. For instance, Tristram (1995, cited in van der Auwera & Genee, 2002) has argued for a Celtic-English mutual reinforcement, or Sprachbund, hypothesis relating to the progressive verb form. In addition, Thomason and Kaufman (1988, pp. 95-97) review the Balkan, Kupwar, and Pacific Northwest multilateral Sprachbund situations.

To summarise, there is no compelling evidence that substitutive DO was influenced by, or created through, language contact. Regarding French and English, it is more likely given the diachronic and sociolinguistic evidence that substitutive DO evolved independently in the two languages, at different rates and with different start and end points, existing in parallel but with no structural transfer during the centuries in which there was AN-ME contact. Future diachronic and diatopic research would be required on substitutive *do* in OE to establish

whether there was Celtic contact influence. Further studies could also examine which factors led to the decline and subsequent disappearance of French substitutive DO.

4. Conclusion

This study has examined the use of substitution with DO in French and English from the medieval period to the present day, starting from Miller's (1997) observations that OFr and MFr had a form of substitution which closely resembles that found in ModE.

Applying Hankamer and Sag's (1976) distinction between deep and surface anaphora, substitutive DO was defined as a surface anaphor involving DO, whose antecedent is a VP, or in certain subtypes a VP minus the NP complement. Following Hankamer and Sag (1976), three types of ModE anaphoric *do so* constructions were identified, of which only one is generally acknowledged in the literature. The surface anaphora criterion justified the inclusion of all three types as forms of substitutive DO in this study.

Data from two large digital parsed corpora (MCVF and PPCME2) has allowed new insight into the diachronic development of substitutive DO in French and English. Two parallel corpus studies were conducted on OFr and MFr on the one hand and ME on the other, each study involving one main query which captured cases of substitutive DO, as well as two supplementary queries which identified PG in comparative and adverbial clauses.

The results show that substitutive DO was well-attested in French around 1100-1475, with a peak in usage around 1250, and sparsely attested from 1500 onwards. It was attested at similar rates in CF and AN in the 12th century, supporting Ingham's (2012) and Trotter's (2003) view that AN did not diverge from CF syntax. Substitutive DO was overall more frequent in ME than in French in the corresponding period. The usage was robust throughout ME, undergoing only a slight and gradual decrease from the start to the end of the period. PG is attested in

both languages, slightly more frequently in ME than in OFr/MFr. However, due to methodological limitations, the PG results only reflect two specific syntactic contexts and thus cannot give a reliable indication of the overall distribution of this feature.

The next section examined the hypothesis that substitutive DO emerged or increased in usage due to contact influence. Three potential scenarios for contact-induced syntactic change were evaluated based on linguistic, sociolinguistic, and diachronic information: firstly, structural borrowing from AN to ME; secondly, substratum interference from ME on AN, and thereafter from AN on CF; finally, substratum interference from Celtic on OE. The first scenario was contradicted by the diachronic data, and the second scenario by sociolinguistic and linguistic evidence. In order to evaluate the third scenario, further diachronic and diatopic research on OE is needed.

Finally, a brief review followed of a plausible endogenous change proposed by Garrett (1998), whereby substitutive *do* developed into periphrastic *do* through merger with habitual *do.* Based on Thomason and Kaufman's (1988) account of different types of language contact, I have argued that overall, endogenous rather than contact-induced change should be considered the null hypothesis when it comes to syntax. In the case of substitutive DO, there is not enough evidence to reject the null hypothesis, namely that its diachronic development in French and in English was independent of AN-ME contact.

Appendix

https://drive.google.com/drive/folders/1miRWTv07BzPbgqDOfxNdAZqz5NA61li6?usp=sharing

References

Computer Programs

Randall, Beth. 2005. CorpusSearch 2. http://corpussearch.sourceforge.net.

RStudio Team. 2015. RStudio: Integrated Development for R. RStudio, Inc. http://www.rstudio.com/.

Corpora

Kroch, Anthony & Beatrice Santorini. 2010. Penn supplement to the MCVF Corpus. Department of Linguistics, University of Pennsylvania.

Kroch, Anthony & Ann Taylor. 2000. The Penn-Helsinki Parsed Corpus of Middle English (PPCME2). Department of Linguistics, University of Pennsylvania. CD-ROM, 2nd edn, release 4, http://www.ling.upenn.edu/ppche/ppche-release-2016/PPCME2-RELEASE-4.

Martineau, France, Paul Hirschbühler, Anthony Kroch & Yves Charles Morin. 2010. Corpus MCVF (parsed corpus), Modéliser le changement: Les voies du français. Département de français, University of Ottawa. CD-ROM, 1st edn, http://www.arts.uottawa.ca/voies/voies_fr.html.

Print publications

Auwera, Johan van der & Inge Genee. 2002. English do: On the convergence of languages and linguists. *English Language* 6(2). 283-307.

Ayres-Bennett, Wendy. 1996. *A history of the French language through texts*. London: Routledge.

Baltin, Mark. 2012. Deletion versus pro-forms: An overly simple dichotomy? *Natural Language & Linguistic Theory* 30(2). 381-423.

Dagnac, Anne. 2019. Typological case studies: French. In Craenenbroech, Jeroen van & Tanja Timmermann (eds.), *Oxford Handbook of Ellipsis*, 785-814. Oxford: Oxford University Press.

Garrett, Andrew. 1998. On the origin of auxiliary *do*. *English Language and Linguistics* 2(2). 283-330.

Gelderen, Elly van. 2014. *A history of the English language*, revised edn. Amsterdam: John Benjamins Publishing Company.

Haddican, Bill. 2007. The structural deficiency of verbal pro-forms. *Linguistic Inquiry* 38(3). 539-547.

Haeberli, Eric. 2010. Investigationg Anglo-Norman influence on late Middle English syntax. In Richard Ingham (ed.), *The Anglo-Norman language and its contexts*, 143-163. Woodbridge, England: York Medieval Press, in association with The Boydell Press, an imprint of Boydell & Brewer Inc.

Hankamer, Jorge E. 1971. *Constraints on deletion in syntax*. New Haven, CT: Yale University P.h.D thesis. Available from ProQuest Dissertations & Theses Global, https://search.proquest.com/docview/288039125?accountid=10673 (last retrieved 08.05.2019).

Hankamer, Jorge E. 1973. Unacceptable ambiguity. *Linguistic Inquiry* 4(1). 17-68.

Hankamer, Jorge E & Ivan Sag. 1976. Deep and surface anaphora. *Linguistic Inquiry* 7(3). 391-428.

Hestvik, Arild, Helge Nordby & Geir Karlsen. 2005. Antecedent reactivation by surface and deep anaphora in Norwegian. *Scandinavian Journal of Psychology* 46(3). 229-238.

Ingham, Richard (ed.). 2010. *The Anglo-Norman language and its contexts*. Woodbridge, England: York Medieval Press, in association with The Boydell Press, an imprint of Boydell & Brewer Inc.

Ingham, Richard. 2012. *The transmission of Anglo-Norman: Language history and language acquisition*. Amsterdam: John Benjamins Publishing Company.

Ingham, Richard. 2014. Contact with French as a factor in the rise of English elliptical discourse constructions. In Tyne, Henry, Virginie André, Christophe Benzitoun, Alex Boulton & Yan Greub (eds.), *French through corpora: Ecological and data-driven perspectives in French language*, 26-45. Newcastle-upon-Tyne: Cambridge Scholars Publishing.

Johnson, Kyle. 2009. Gapping is not (VP-)ellipsis. *Linguistic Inquiry*. 40(2). 289-328.

Martineau, France. 2008. Un corpus pour l'analyse de la variation et du changement linguistique. *Corpus (Laboratoire Language)* 7. 135-155.

Miller, Philip Harold. 1997. Auxiliary verbs in Old and Middle French: A diachronic study of substitutive faire and a comparison with the Modern English auxiliaries. In Kemenade, Ans van & Nigel Vincent (eds.), *Parameters of Morphosyntactic Change*, 119-133. Cambridge: Cambridge University Press.

Poussa, Patricia. 1990. A contact-universals origin for periphrastic do, with special consideration of OE-Celtic contact. In Adamson, Sylvia M, Vivien A Law, Nigel Vincent &

Susan Wright (eds.), *Papers from the 5th international conference on English historical linguistics*, 407-434. Amsterdam: John Benjamins Publishing Company.

Preusler, Walter. 1938. Keltischer Einfluss im Englischen. *Indogermanischen Forschungen 56*. 178-191.

Siegel, Muffy E A. 1984. Gapping and interpretation. *Linguistic Inquiry* 15(3), 523-530.

Spearing, Anthony Colin. 1993. *The medieval poet as voyeur: Looking and listening in medieval love-narratives*. Cambridge: Cambridge University Press.

Thomason, Sarah & Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley, CA: University of California Press.

Tristram, Hildegard L C. 1997. Do-periphrasis in contact? In Ramisch, Heinrich & Kenneth Wynne (eds.), *Language in time and space. Studies in honour of Wolfgang Viereck on the occasion of his 60th birthday*, 401-417. Stuttgart: Franz Steiner.

Trotter, David. 2003. Not as eccentric as it looks: Anglo-French and French French. *Forum for Modern Language Studies* 39(4). 427-438.

Warner, Anthony R. 1992. Elliptical and impersonal constructions: Evidence for auxiliaries in Old English? In Fran Colman (ed.), *Evidence for Old English: Material and theoretical bases for reconstruction*, 178-210. Edinburgh: John Donald.

Warner, Anthony R. 1993. *English auxiliaries: structure and history*. Cambridge: Cambridge University Press.

Websites

Faire. n.d. *DicFro*. http://micmap.org/dicfro/search/tableaux-de-conjugaison/faire (retrieved 11.05.2019).

Kroch, Anthony & Ann Taylor. 2016. Penn-Helsinki Parsed Corpus of Middle English. https://www.ling.upenn.edu/hist-corpora/PPCME2-RELEASE-4/index.html (retrieved 11.05.2019).

Santorini, B. n.d. Corpus Construction. https://www.ling.upenn.edu/~beatrice/corpus-ling/ (retrieved 11.05.2019).