THE UNIVERSITY
of EDINBURGH

Syntactic change during the anglicisation of Scots: Insights from the Parsed Corpus of Scottish Correspondence

Lisa Gotthard

Doctor of Philosophy

The University of Edinburgh

2022

For Erik

14.04.92–22.04.21
Simbelmynë

Abstract

Variation and change in syntax is particularly challenging to measure quantitatively, as such investigation requires syntactically annotated (parsed) corpora; a parsed digital corpus allows for retrieval of all instances of a construction or particular word order in a fraction of the time it would take to retrieve the same information by hand. Compared to English, research on syntactic change in the history of Scots has been limited, in part due to the lack of such a resource. In order to meet these demands, this thesis presents the new *Parsed Corpus of Scottish Correspondence* (PCSC), consisting of 270,000 words of parsed data from the *Helsinki Corpus of Scottish Correspondence 1540-1750* (Meurman-Solin and VARIENG 2017), and demonstrates the process in turning strings of words into searchable clause tokens by using a combination of automated and manual methods. The PCSC provides data from the 16th to 18th century, a previous blind spot within Scots syntax research despite being a highly interesting time period to investigate; these centuries saw a shift in the relationship between Scots and English, as English started to exert influence over Scots as a more socio-politically prestigious variety – consequently, salient Scots features were increasingly replaced by English ones in writing. Thus, the 16th-18th century marks a period of great change in Scots, as it went from being a more distinct variety on a standardisation trajectory, to the mixed variety we encounter in Scotland today.

Using the new parsed data from the PCSC, I present results from three case studies on syntactic change in 16th to 18th century Scots, thus beginning to fill the gaps in our knowledge of this period. The findings of the case studies reveal the transformative nature of Scots syntax in the 16th to 18th century, as the language undergoes dramatic changes in its subject-verb agreement system through the decline of the *Northern Subject Rule* and the rise of *do*-support, and further rearrangement in the verbal paradigm through the rise of verbal *-ing* in both participial and gerundive function. On assessing whether these changes can be attributed to influence from English, or whether they are simply parallel developments in closely related language varieties, it is found that the nature of

contact between Scots and English in the 16th-18th century, and the timing in which the changes take place, speaks in favour of these changes being contact-induced. However, further fine-grained investigation into the functions and distribution of the features involved, in Scots compared to English, will be needed before more firm conclusions can be drawn regarding the origin of the changes.

Lay Summary

Before the 16th century, a different language than English was utilised for all formal and informal purposes in Scotland: *Scots*. Scots and English share the same ancestor language, but started to develop differently from each other around the 11th century. Based on written evidence from the time, and commentary in such written evidence, it appears that speakers of Scots and English understood each other well, but still perceived these two language varieties as different – similar to how we today perceive other closely related languages, such as Danish and Norwegian, or Spanish and Catalan. However, towards the end of the 16th century, the status of Scots changed in Scotland, as significant socio-political events took place which caused English to become associated with influential institutions, such as the church (as English became the language associated with protestantism) and royal court (after the Union of Crowns 1603). By the time of the Union of Parliaments in 1707, Scots had been ousted from most formal writing, became associated with domestic and vernacular speech, and began to be perceived more like a dialect of English. Today, the language variety called Scots is made up of a range of dialects, and speakers often mix their speech with English to varying degrees depending on contexts. The process in which Scots went from its more prestigious position in the 16th century, to become gradually replaced by and mixed with English, is often called the *anglicisation* of Scots. Anglicisation effects have been observed in the the spelling system and lexicon of Scots, in that distinctively Scots words and spellings declined in favour of English ones.

Thus, the 16th-18th century is a transformational period for Scots, and marks an important shift in the development of the language. Despite this, there are still many stones unturned as regards linguistic changes in this period. Particularly, the effects of *anglicisation* processes on the deepest level of language structure, the *syntax*, has been largely unknown. This may be because the syntax, which concerns itself with structures of sentences, appears to be very similar between English and Scots at a first glance (but may not be so at closer investigation). Another reason

which has implicated research on 16th to 18th century Scots syntax is that the same type of resources which linguists utilise for such research, e.g. in English, have not existed for Scots. Thus, one part of the PhD project described in this thesis has involved creating such a resource; the *Parsed Corpus of Scottish Correspondence* (PCSC) consists of digitised personal letters written in the 16th-18th century, which have been enriched with annotation so that each word and sentence contains information of the function of that word or sentence in Scots syntax. With this resource, it is possible to measure proportions of one sentence structure compared to another over time, which is what has been done for the three case studies presented in this thesis. In this way, we can attempt to measure the influence from English on Scots, by looking at the extent to which an English structure is used compared to a Scots one; the structures investigated are typically assumed to be examples of a type of grammar, and the structures that replace them as a different type of grammar.

The findings of these case studies indeed give evidence of change in the syntax of Scots, which is visible in that each structure type investigated becomes gradually replaced by another throughout the course of the 16th-18th century: (i) the tendency in pre-anglicisation Scots to use *-s* inflection on verbs with subjects that are plural nouns, and pronouns in certain contexts, (for example: *the girls dances*, *I sing and dances*) declined in favour of a Standard English system (*the girls dance*, *I sing and dance*), (ii) the tendency in pre-anglicisation Scots to place the verb before negation in a sentence (*I eat not cake*) became replaced by a system in which the auxiliary *do* is inserted before the negative element (*I don't (do not) eat cake*; the same change took place in English, but ca. 200 years earlier), and (iii) nouns ending in *-ing* more frequently become used as verbs (*the closing of the door*, where 'closing' is a noun, becomes *closing the door*, where 'closing' is a verb), and pre-anglicisation Scots verbs ending in *-and* change to end in *-ing* instead (*closand the door* becomes *closing the door*).

The final contribution of this PhD thesis is to assess whether these changes came about as a result of influence from English, or whether what we see is parallel developments in closely related languages. The discussion reveals that the nature of the relationship between Scots and English

in the 16th-18th century, and the timing in which the changes take place, speaks in favour of that these structures arose from English influence on Scots. However, further fine-grained investigation into the functions of these features in both languages will be needed before more firm conclusions can be drawn, which I hope will be explored in future research.

Acknowledgements

I have had the luxury of being supported by a fantastic duo of supervisors throughout my time as a PhD student: Prof. Bettelou Los and Dr. Rob Truswell.
Bettelou has been supervising and training me since I was a masters student, and her teaching has shaped how I conduct research. It is truly a privilege to have access to Bettelou's deep well of knowledge, which contains insightful and intelligent answers to any question I may have (usually accompanied by a book to be borrowed from her office bookshelf!). Rob's endless patience with me as I have been navigating learning the methods required for this project has been admirable, and he has offered perspectives on the features investigated in this project which I never would have thought of myself. He has always put my best interest first, and seems to have a keen intuition for when I need to be challenged. They both have been a constant source of support during the ups and downs of my time as a PhD student. This thesis would not be anything to hang in a christmas tree (to use a Swedish expression) if it were not for Bettelou and Rob. (The shortcomings are, of course, all my own.)

At different stages of this project, I have also received supervision and training from Dr. Rhona Alcorn (Dictionaries of the Scots Language) and Dr. Beatrice Santorini (University of Pennsylvania). Without Rhona, this would be a very different project (if a project at all), as she gave me the idea to build a corpus for studying syntactic change in Scots. Rhona taught me all the fundamentals of the history of Scots, and provided me with exciting opportunities for outreach and engagement which taught me even more about this language I study. Beatrice has provided crucial training in PPCHE parsing methods, which continued long after my research visit to UPenn ended (after being interrupted by the outbreak of the Covid19 pandemic). She has been a joy to work with, a reliable source for clever solutions to difficult coding problems as well as many anecdotes. Thank you both!

I am part of a community of brilliant PhD students at the University of Edinburgh. In particular, having Alex Lorson, Mattia Zingaretti, and Sarah van Eyndhoven by my side during this time has kept me sane – our friendship is everything, and it has carried me over every hurdle. Many students in my cohort, and the ones before and after, have made PhD life beautiful, with a special shoutout to Tobias Ungerer and the girls in 1.15 – thanks for all the lunches and office shenanigans! Everyone pursuing a PhD journey should have the opportunity to be surrounded by such intelligent and funny peers, who listen to endless complaining about work and life, and provide much needed distractions as well as good advice when needed. I am also privileged to be part of the exceptionally supportive and collegial academic community at Linguistics and English Language in Edinburgh, where staff have shown admirable resilience in the face of multiple industrial actions and a global pandemic. I stand in solidarity with them, as the fight continues for the improvement of pay and conditions in UK Higher Education.

This thesis was made possible with essential financial support from the Arts and Humanities Research Council, through the Scottish Graduate School for Arts and Humanities. I am also indebted to Prof. George Walkden (Universität Konstanz); thanks to George, I had access to a brilliant student assistant, Sarah Einhaus, who helped with manual corrections of POS tags (danke, Sarah!), and by working part-time as a postdoc for George in the final semester of my PhD studies I was able to financially support myself during the last leg of this project. Thank you!

I am blessed with many different family constellations to care for me during this extended period of childhood that PhD studies sometimes are. Paul's family has been a rock during this project, always providing important advice on practices in academic work, and work/life balance, as well as offering their homes to me and Paul whenever we have needed it.
I want to thank my own family, for their endless love and support during this endeavour which sometimes has seemed incomprehensible, and despite the fact that it has taken me far away from them. I want to especially thank my brother Simon, for paving the way for me to pursue this in

so many different ways.

Finally, and most importantly, this thesis could not have been completed without Paul by my side. Indeed, I would not be complete without Paul. Tack för allt!

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Lisa Gotthard

List of tags and abbreviations

PPCHE part-of-speech tags and syntactic tags

This is a list of all the *Penn Parsed Corpora of Historical English* tags used in this corpus. The complete list of tags can be found in the PPCHE annotation guidelines.

– ADJ/ADJP

  Adjective/Adjective phrase

– ADV/ADVP

  Adverb/Adverb phrase

    -LOC – Locative

    -TMP – Temporal

– BED/BEP

  Past tense/Present tense (auxiliary and main verb) *be*

– CP

  Complement clause or question

    -REL – Relative

    -ADV – Adverbial

– IP

  Clause other than CP

    -MAT – Matrix clause

    -PPL – Participial clause

    -SUB – Sub-clause

– NEG/NEGP

  Negator/Negator phrase

– N/NS/N$

  Noun/plural noun/possessive noun

- NP

  Noun phrase

  - -COM – Complement
  - -OB1 – Direct object
  - -POS – Possessive
  - -SBJ – Subject

- NPR/NPR$

  Proper noun/Possessive propern noun

- P/PP

  Preposition/Preposition phrase

- PRO/PRO$

  Pronoun/Possessive pronoun

- PUNC

  Punctuation

- V – Verb

- VAG/VAN/VBN

  Present participle/Passive participle/Perfect participle

- VB

  Infinitival verb

- VBD/VBP

  Past/present tense finite verb

- Q/QP

  Quantifier/Quantifier phrase

Abbreviations

Abbreviations listed here are used in-text. Not included are abbreviations only used in captions or legendes of figures or tables.

Corpora

- EEBO

  Early English Books Online

- CEEC

  the Corpus of Early English Correspondence

- CSC

  the Helsinki Corpus of Scottish Correspondence

- DOST

  the Dictionary of the Older Scottish Tongue corpus

- HCOS

  the Helsinki Corpus of Older Scots

- LAOS

  a Linguistic Atlas of Older Scots

- PCEEC

  the Parsed Corpus of Early English Correspondence

- PCSC

  the Parsed Corpus of Scottish Correspondence

- PLAEME

  a Parsed Linguistic Atlas of Early Middle English

- PPCHE

  the Penn Parsed Corpora of Historical English

- PPCEME

  the Penn-Helsinki Parsed Corpus of Early Middle English

Other

– 1sg

  First person singular

– 2p

  Second person

– 3sg

  Third person singular

– AdvP

  Adverb Phrase

– CRH

  the Constant Rate Hypothesis

– CS

  CorpusSearch

– EModE

  Early Modern English

– ME

  Middle English

– NSR

  the Northern Subject Rule

– OE

  Old English

– ON

  Old Norse

– PDE

  Present-Day English

– pl/plNP

  Plural/non-pronominal plural subject

– POS

  Part-of-speech

– SSE

  Scottish Standard English

– StE

  Standard English

– S-V

  Subject-Verb

Contents

Figures and Tables

Figures

Tables

Chapter 1

Introduction

"the history of the development of Scots from the late 16th century to the 18th is largely still to be written"

– Caroline Macafee, in Macafee and Aitken (2002: 2.5.2).

The sixteenth to eighteenth century saw a dramatic decline of salient Scots features in writing in favour of English ones, a process often referred to as the *anglicisation* of Scots (e.g. Murison 1979, Macafee and Aitken 2002); in this period, the prestige of English in Scotland increased as a result of significant socio-political events which caused English to become associated with influential institutions, such as the church and royal court. Thus, the period of *anglicisation* was a transformational time for Scots, as the language went from being a distinct variety, apparently on a separate standardisation trajectory from Southern English and of equal prestige to English, to the highly variable, regionally and socially marked variety we encounter in Scotland today. To highlight the dynamic nature of this period, and noting the gravity of this period for the trajectory of Scots, Kopaczyk (2013) labels the sixteenth-eighteenth century the Transition period for Scots.

Despite the importance of this period in Scots history, there are still many stones unturned as regards linguistic change in this period, as implied by the quote from Macafee and Aitken (2002) given above. When written Scots changed to be more similar to English, it became increasingly seen as less autonomous, which has contributed to the sparsity of research on post-anglicisation Scots – the lack of available morpho-syntactically annotated Scots corpora covering the sixteenth-eighteenth century has turned it into a particular blind spot for syntactic study. This thesis begins to fill this gap in our knowledge, by providing a new resource for diachronic syntactic study of sixteenth-eighteenth century Scots, and by uncovering new insights into the trajectory of three

syntactic changes taking place during the Scots Transition period: the decline of the *Northern Subject Rule* pattern of Subject-Verb agreement, the rise and regulation of Scots *do*-support, and the rise of verbal gerunds and participles in *-ing*. The findings of these studies reveal that dramatic syntactic change took place in Scots during the anglicisation period, and that some of these changes could be outcomes of contact-induced change from anglicisation pressures. Hence, this thesis presents quantitative diachronic studies which demonstrate how the new *Parsed Corpus of Scottish Correspondence* (PCSC) can be used to investigate syntactic variation and change in Scots in a way which previously has not been possible. In this way, I hope to have contributed to writing the history of the development of Scots from the late sixteenth century to the eighteenth.

1.1 Summary and organisation of chapters

In Chapter 2, I give an overview of the history of Scots from the arrival of Germanic tribes to Britain until the 21st century, with a particular focus on the period from the emergence of Scots as a distinct variety until the decline of written Scots as a consequence of anglicisation pressures in the Transition period (Ch2, Section 2). This overview reveals that, in the context of Scotland, the contact between English and Scots in the sixteenth-eighteenth century has the hallmarks of a contact scenario in which structural borrowing is likely to take place (e.g. Thomason 2001; Thomason and Kaufman 1988); the languages are closely related and typologically similar, the contact is long and intense, and there is evidence of transfer in other areas of the grammar, such as in Scots orthography and lexis. The shift in prestige between English and Scots in the Transition period meant that the nature of contact shifted from involving two equally prestigious and linguistically similar languages, to one (English) exerting influence over the other (Scots) as a socio-politically dominant variety, which increases the likelihood for linguistic transfer (Thomason 2001: 66). In Section 3 of Chapter 2, I give a brief description of syntactic change in Early Modern English, and account for the known changes in Scots morpho-syntax during the same period, which reveals further how the effects of anglicisation on Scots syntax has been less researched.

Chapter 3 describes a new syntactically annotated (parsed) corpus of sixteenth-eighteenth century Scottish correspondence; the *Parsed Corpus of Scottish Correspondence* (PCSC) consists of 270,000

words of correspondence data from the *Helsinki Corpus of Scottish Correspondence 1540-1750* (Meurman-Solin & VARIENG 2017) which has been parsed according to the *Penn Parsed Corpora of Historical English format* (PPCHE; Kroch, Santorini, and Delfs 2004; Kroch, Santorini, and Diertani 2016; Kroch and Taylor 2000). This places the PCSC within a family of sister corpora which use the PPCHE annotation conventions, and makes the PCSC the first Scots corpus which can be easily used for comparative study within this family of corpora. Chapter 3 also accounts broadly for how I have used the PCSC to retrieve results for the case studies presented in this thesis (Ch3, Section 4), with more detailed methodological description provided for each case study in their respective chapters.

The subsequent chapters contain case studies which are designed with the aim to provide insights into the extent of influence from English on Scots syntax. For this reason, I am investigating (i) a distinctively Scots feature which is known to be operational in Scots pre-anglicisation (the NSR) (ii) a feature which is often assumed to have emerged in Scots from English influence (*do*-support), and (iii) a feature which is well-researched in English, but its trajectory in Scots is largely unknown (the rise of verbal -*ing*). For each case study, a secondary investigation is made into the effect of the writer gender on the trajectory of changes. The case studies can be summarised as follows:

– The first case study, in Chapter 4, investigates grammar competition between two subject-verb (S-V) agreement systems: the *Northern Subject Rule* (NSR), a distinctively Scots present tense agreement pattern, and Standard English (StE) present tense agreement. The NSR, which originated in Northumbrian Old English, has only been found to operate categorically in Older Scots (Montgomery 1994; Rodríguez Ledesma 2013, 2017). I explore how operational the NSR is in the PCSC data, and whether there is a change over time in the pattern, finding that the NSR system declines in favour of a StE system over time. This study also uncovers surprising variation in 3sg inflection, which is inconsistent with both an NSR and StE system, and reveals that female writers use more NSR-like inflection patterns overall.

– The second case study, on the rise and regulation of *do*-support in Scots, is presented in Chapter 5. This study has two aims: firstly, I present proportions of *do*-support structures over time in a way more comparable to quantitative studies on the rise of do-support in English (e.g., Ecay 2015; Ellegård 1953; Kroch 1989), and corroborate suggestions from previous research of a later emergence and slower rise of Scots *do*-support compared to English (Gotthard 2019; Meurman-Solin 1993a). Secondly, I investigate what constraints may have conditioned the development of Scots *do*-support, by investigating (i) whether Scots *do* enters a similar intermediate stage as is reported for English *do* by Ecay (2015), and (ii) whether Scots *do*-support enters grammar competition with the NSR, as implied by analysis of the phenomena by Bobaljik (2002) and de Haas (2011). The findings reveal similarities between early Scots *do*-support and intermediate *do* in English, and show that, while grammar competition between the two systems is not ruled out, the rise of *do*-support does not appear to be conditioned by NSR subject type constraints in the PCSC data. Finally, it is found that women are leaders at the late stage of the rise of Scots negative declarative *do*-support, which follows what is found by Nurmi (2011) for negative declarative *do* in English.

– The third case study, in Chapter 6, explores a widely researched feature in Early Modern English: the diachronic development from nominal to verbal gerunds (see e.g. Fonteyn 2019: 11ff, for a summary). The trajectory of the syntactic structure of gerunds in Scots has received little attention in previous research, but morphological differences between English and Scots participles have been documented: Older Scots retained *-and* forms for participles until the sixteenth century, when they became replaced by forms in *-ing*, ca. 400 years later than the same development in English (e.g. King 1997: 180, Görlach 2002: 96). In this study, I quantitatively explore the trajectory of change from nominal to verbal gerund structures in the PCSC data, and the decline of participial *-and* in favour of *-ing*. The findings indicate differences in the trajectory of verbal *-ing* compared to English. As regards gender differences, it is found that women do not lead the verbalisation of *-ing* in Scots, which differs from findings on Early Modern English data by Nevalainen and Raumolin-Brunberg (2003: 121-2), but that they appear to lead the change in the decline of participial *-and* in favour of *-ing*

Finally, the findings of these case studies are discussed in Chapter 7, where they are assessed against the likelihood that they are contact-induced changes, syntactic outcomes of influence from English during the anglicisation period. This discussion concludes that the social context and timing of the changes fit an analysis of contact-induced change, and that the decline of the NSR is most likely to be a contact-induced change, but that further investigation into the distribution and function of *do* and verbal *-ing* in Scots is needed in order to more confidently assess the origin of these features in Scots.

Chapter 2

Historical background: The divergence and convergence of Scots and English

2.1 Introduction

While the original research conducted for this thesis targets sixteenth to eighteenth century Scots, it is nonetheless important to understand the significance of this time period by also considering what preceded it and what emerged after it. Thus, this chapter outlines the entire history of Scots, from the linguistic context that preceded the emergence of Scots, until the present-day variety and its status in Scotland. The chapter is divided as follows: 2.2.1 outlines the linguistic context in Scotland in the time when Germanic varieties, ancestors to Scots, became established there, and 2.2.2 describes the emergence of Scots as a distinct variety from its southern sister variety, English. In 2.2.3, I describe the period under investigation for this thesis, commonly known as the *anglicisation* of Scots, when we see convergence between Scots and English again, and how that period catalyses the complex relationship between Scots and English in modern times. In summarising the linguistic history of Scotland in Section 2.2, I have drawn on detailed accounts by Murison (1979), Aitken (1985), and Macafee and Aitken (2002) as my main sources, unless otherwise specified. As will be seen, every stage of the history and trajectory of Scots has been shaped by contact with other languages, and most significantly by contact with English; Section 2.2 of this chapter will therefore pay particular attention to Scots as a contact variety, and describe its relationship with English through the lens of language contact in each of the time periods described here. Finally, in Section 2.3, I give a brief overview of syntactic features of English and Scots in the Early Modern period; while Early Modern English syntax (2.3.1) shows pattern of stabilisation

after the Middle English period of high variability, and settles into patterns we recognise in the modern grammar, Early Modern Scots syntax (2.3.2) is less documented and often described as identical to English despite distinct patterns being documented in both its modern and early history.

2.1.1 Brief note on the periodisation of Scots

The typically reproduced periodisation of Scots was first introduced by Aitken (1985), as shown in Figure 2.1, visualised by Kopaczyk (2013: 239). Aitken (1985: xiii) provides little motivation for this periodisation, mainly pointing out differences between Older and Modern Scots but not between the sub-periods he identifies. This, and in particular the placement of Middle Scots contemporary with Early Modern English, leads Kopaczyk (2013) to reconsider Aitken's established periodisation in light of findings from more recent research on Scots, as well as considering extra-linguistic criteria in Scotland independent of the English context. Kopaczyk (2013) finds that there is indeed motivation to assume an Early Modern Scots period contemporary with Early Modern English, with the end of that period, from 1560 until 1700, being more difficult to define due to the *anglicisation* of Scots (see 2.2.3 in this chapter) and therefore assigned the label *Late* or *Transition* Scots; her full proposal for a revised periodisation of Scots is seen in Figure 2.2, from Kopaczyk (2013: 253).

Figure 2.1: Aitken's (1985) periodisation of Scots, compared to English

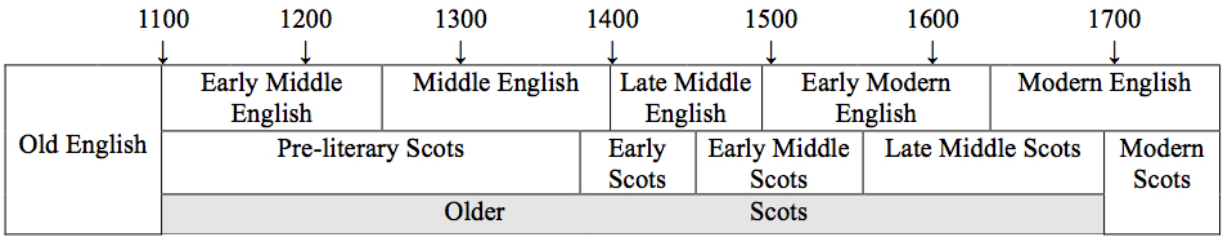
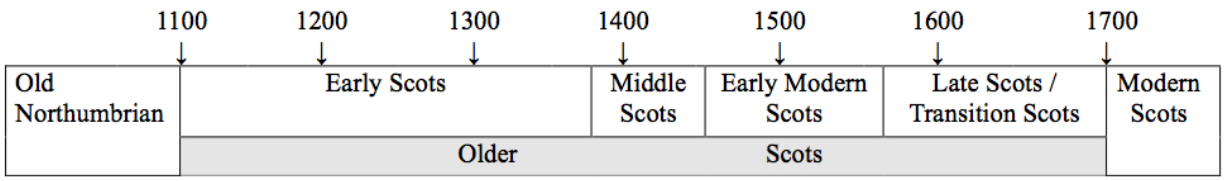


Figure 2.2: Kopaczyk's (2013) revised periodisation of Scots



In this thesis, when it is necessary to refer to distinct periods, the labels in Kopaczyk's (2013) periodisation will be used; Kopaczyk's "Transition" label will be used interchangeably with 'Early Modern' to describe the period from the sixteenth to mid-eighteenth century in Scots, and the term "anglicisation" will refer to the process in which salient Scots features declined in favour of English ones in this period.

2.2 The history of Scots and contact with English

2.2.1 The Scottish linguistic context pre-1100

Before any Germanic language became established in Scotland, the region was inhabited by speakers of Pictish and Gaelic. Starting in the fourth century, the Gaelic speakers, called *Scoti* or *Scots*, migrated from Ireland and brought their language with them. In the ninth century, the Scots and Picts had united under one kingdom, Alba, which is the first kingdom comparable to modern Scotland. While it is traditionally assumed that the Pictish rulers voluntarily entered such a union, it seems more likely that the kingdom of the Picts was taken over by the Gaels (Woolf 2007: 321). Subsequently, the Gaelic language of the Scoti, the ancestor of today's Scottish Gaelic, replaced Pictish as the dominant language in the region. However, simultaneously in the South-East, a Germanic language had spread and become established. I will now describe the historical context which led to this early Germanic variety, an ancestor of Scots and English, emerging and spreading in today's Scotland.

Evidence for the very earliest origins of Scots is sparse, but we can postulate a beginning of Scots (or, pre-Scots), to take place after the Anglo-Saxon settlement of Great Britain in the sixth century. The Angles established the kingdom of Bernicia, which incorporated the south-eastern part of what today is Scotland, with its northern border aligning with the Firth of Forth. In the seventh century, Bernicia and the kingdom of Deira, to the south of it, joined into *Northumbria* (see Figure 2.3), and later Northumbria seized control of parts of Cumbria to the west, specifically the kingdom of Rheged, which meant that most of today's Scotland south of the central belt would have been Northumbrian by the eighth century. Evidence from place names tells us that

the Germanic variety of the Angles which inhabited the region, *Anglian* or *Northumbrian*, was spoken throughout Northumbria in this period, and, thus, a Germanic language, a predecessor to Scots, was first spoken in the southern regions of Scotland as early as what has been established for the origins of English in England; of course, the England-Scotland border as we know it today did not exist at this time, so the history of Northumbria is the history of England and Scotland simultaneously.

Figure 2.3: Kingdoms in Southern Scotland and Northern England after the Anglo-Saxon settlement

(*Map from Lowe (1999), reproduced in Macafee and Aitken (2002: 2.1)*)

Northumbria becomes severely weakened during the Viking invasions, carried out by Vikings from today's Denmark, in the tenth and eleventh centuries, which leaves it vulnerable to lose its northern regions to the Scots. Continuous disputes between Scots and Danes over Northumbria and Cumbria lead to the final drawing of a border which effectively cut Cumbria and Bernicia in half; this is roughly what we recognise as today's Scotland-England border.

However, Anglian is not the only Germanic variety with a claim to Scots ancestry. In England, the Viking invasion and settlement led to the establishment of the Danelaw (see 2.4), where a mixed language variety subsequently emerged, often referred to as Anglo-Scandinavian (e.g., Graham-Campbell, Hall, Jesch, and Parsons 2001; Townend 2002). It is not clear how much political influence the Danes had over the southern part of Scotland during the era of the Danelaw, but it is likely that the Anglo-Scandinavian variety was also spoken north of the Scotland-England border at this time, and thus in contact with the already established Anglian variety spoken there. It should be noted that the situation was different in the Northern and Western Isles, where there had been settlement by Vikings from what today is Norway; in the Northern Isles, a Scandinavian variety, Norn, was spoken until the eighteenth century (e.g., McColl Millar 2020: 117-21). The Scandinavian varieties spoken on the Scottish Isles were not influential in shaping the emergence of Scots, but did influence the Scots and English dialects spoken there to this day.

Figure 2.4: Kingdoms in medieval England in the time of the Danelaw

(*Map from Dow (1926: 47)*)

It is not until after the Norman Conquest in the eleventh century, however, that we see more concrete influence from Anglo-Scandinavian in Scotland. Up until the conquest, the language of the Scottish court was (the ancestor of) Gaelic, but, after the Norman invasion, the English princess Margaret fled to Scotland and married the Scottish king Malcolm III. While the anglicisation of the Scottish court and church began already with the influence from Margaret, it was her youngest son, king David I of Scotland, who was most instrumental in the spread of Anglo-Scandinavian across Lowland Scotland: David was partly raised in the Anglo-Norman court, and after he became king, in the twelfth century, he granted Scottish land to Norman barons, in order to establish a feudal system in Scotland. These feudal tenants brought with them retainers from the north of England who spoke Anglo-Scandinavian. David also established monasteries with a likely Anglo-Scandinavian-speaking parochial administration. Furthermore, during and after the reign of David I, Norman-style burghs were established in Lowland Scotland, which became important hubs for trade and attracted immigration, often merchants and craftsmen, not only from England but also the Netherlands, Flanders, the Rhineland, and France. The burghs thus became linguistic melting pots, with Germanic language varieties being mostly represented, but also French and Gaelic dialects.

In the words of Macafee and Aitken: "The burghs, as foci of internal and external trade, were crucial in the spread of Lowland Scots, although the population even of the largest would have numbered hundreds rather than thousands" (2002: 2.3.2). The French spoken by the Norman barons never became as important in Scotland as it was in England; "[w]ith rare exceptions, it was Latin that was employed in administration [...] French was a familial language amongst the Normans, and was of use for wider communication (with England and France), but [pre-Scots] was the shared language of feudal overlords (secular and clerical), their vassals, and the freemen of the burghs" (Macafee & Aitken 2002: 2.3.2). Gaelic, on the other hand, became suppressed further into the Highlands, despite it so recently having been more widely spoken both geographically and numerically.

The marginal influence from Gaelic on Scots, which is mostly visible in lexis, with words of

particular cultural significance (e.g. *quaich*, a traditional two-handed drinking cup) or describing nature features (e.g. *loch*, *cairn*), suggests that Gaelic did not have a strong position in the Lowlands while pre-Scots established itself there, neither numerically nor politically (cf. Thomason and Kaufman 1988: 116) – indeed, as Macafee (1997: 199) notes, this also suggests the low status of Northumbrian compared to Anglo-Scandinavian, as we could postulate more mixing from the Gaelic-Northumbrian contact situation. The subtle influence from Gaelic is most visible in the North-East, where Scots and Gaelic would have been in contact the longest; e.g., North-Eastern Scots dialects will have /f/ in Wh-words instead of the more widely found /ʍ/, and word-initial /vr/ clusters instead of /wr/ in words like *wrong*, which may have been substituted by Gaelic speakers as Gaelic does not have /ʍ/ and /w/ (Macafee & Aitken 2002: 2.4). We can draw from this that, in line with Scots increasingly being seen as the language of business and administration, already at the earliest stages of the emergence of Scots, Gaelic was the less prestigious language, and only exerted a minor substratum influence on Scots. It is, however, difficult to gauge how much influence the Scots-Gaelic contact may have had on, e.g., the development of areal features, which are not as salient contact outcomes as structural or lexical transfers.

Thus, while the royal family was Gaelic speaking until 1100, and the feudal lords, who had a great impact on the Scottish infrastructure in the twelfth century, were Norman French-speaking, the people who came to have the most visible influence on the language of the people of Lowland Scotland spoke Anglo-Scandinavian. This may have been aided by the fact that a closely related language was already established among inhabitants of the South-East when Anglo-Scandinavian speakers arrived there. Consequently, a mixed variety – a combination of Anglo-Scandinavian, and the Northumbrian variety already spoken in this part of Scotland, with some, mainly lexical, influence from the other languages spoken in the burghs – spread rapidly across the Lowlands and became a *lingua franca* there, which we can now start referring to as *Scots*.

2.2.2 Scots-English divergence: The rise of Scots

From the twelfth century and in the 200 years that follow, the early Scots variety rises in status and continues to spread across Lowland Scotland and even to Caithness and the Northern Isles, while Gaelic continues to be spoken in the Highlands. In the fourteenth century, this variety, which was referred to as 'Inglis' in Scotland up until the turn of the fifteenth-sixteenth century, also started to be used for a wider range of functions, and even became the language of the higher social classes and the aristocracy. Before the late fourteenth century, there is very little evidence of what this early variety of Scots looked like, as writing was primarily in Latin and French at this time. When the first piece of written Scots emerges in 1360 – an interlinear gloss on a Latin text – Scots is showing some divergence from (more) southern contemporary English varieties, but the lack of sufficient material often leads to speculation based on sparse evidence; for example the use of *-and* endings on present participles is often highlighted as an example of this divergence (see e.g., Macafee and Aitken 2002: 2.3.3), but these endings were also used in Northern English dialects in the fourteenth century (Jack 1988, see also Chapter 6 of this thesis).

The first piece of literary writing in Scots, John Barbour's *The Brus*, appears in 1375 and, with this, our written record of Scots truly begins; it is this early written record that provides the best evidence of Scots having developed separately from English. Importantly, Scots as spoken and written north of the border during the fourteenth to fifteenth century does not appear to have been affected by written norms south of the border, where a form of standard language was spreading from London (not to be confused with the so-called "Chancery Standard", but rather a standard formed by dialect contact in London (e.g., Brinton and Bergs 2017: 149)). In this way, Scots showed signs of being autonomous with respect to English.

Murison calls the fifteenth-sixteenth century the "heyday of the Scots tongue" (1979: 9), as this is the time when the language was most clearly developing separately from Southern English, and also seem to be used for the most communicative functions and genres. During the process of language shift from French and Latin to Scots in formal and influential domains, Scots undergoes extensive vocabulary elaboration to meet the new demands. Mostly, new lexis is transferred from

French; not Norman French, which was usually the source of French lexis in Scots pre-fourteenth century, but central, Parisian French. Loans from other languages, like Dutch, Old Norse, Old English, and Gaelic, are mostly found in domains which concern everyday life and nature. This elaboration is an important step towards what could have been a Scots standardisation process, which never came fully to completion. In Haugen's (1966) influential model for standardisation, a language undergoes processes of *selection* of standard forms, *elaboration* to serve a "maximal elaboration of function" and *codification* to "achieve minimal variation in form" (Haugen 1966: 932), and finally *acceptance* as standard by the community who uses it. King (1997: 157) argues that a standard written language was established for Scots in the 1450-1550 period, noting the functional *elaboration* taking place as Scots was used for a wide range of genres and styles, and that Scots had completed the stages of *selection* and *acceptance*, as demonstrated by Agutter (1988, 1990), who also notes that most official Scots writing from the period do not show regional features (which would be found in e.g. personal correspondence or local burgh records), suggesting the presence of a supra-local standard which writers could converge on (see also Kopaczyk (2013: 249-51)).

However, the language was not a homogenous standard variety; there was plenty of variation within Scots, and the distinction between Scots and the northern English dialects was also not clear-cut. Although Scots exhibits this great variation in form and spelling, with much overlap with Northern English forms, Meurman-Solin (1997) does indeed find that sixteenth century Scots is still quite distinct from contemporary English, and that it is the combination of a set of features, some shared with Northern English varieties and some uniquely Scots, which makes up the distinctiveness; this is also argued by Agutter (1988), who maintains that fifteenth-sixteenth century Scots, albeit distinctly identifiable through this set of features, is not an autonomous standard variety, but another standard English variety, similarly to Australian or American English. Nonetheless, it does appear that Scots was perceived as different from English by outsiders, evidenced by contemporary commentary, e.g, by the Spanish ambassador to Scotland regarding the language of king James IV being different from English, or by the English ambassador suggesting that James V was not literate in English, and by the fact that Margaret Tudor, the English-born wife of James IV,

switches between English and Scots depending on the purpose of her writing (see McColl Millar 2020: 79-80).

Despite these appearances that Scots was well underway of becoming a national language of Scotland in the beginning of the sixteenth century, it is not obvious that the Scots speakers themselves saw Scots as something distinct from English when this process was ongoing (see e.g. McClure 1994: 34); there was no pronounced political movement to preserve Scots from mixing with English, the same label, 'Inglis', was often used to refer to both languages, and the languages continued to be in close contact. Towards the end of this period, Scots writers show signs of "involuntary language shift" (in the sense of Joseph 1987), as the linguistic similarity between the languages, and their similar status, means that English is easily utilised and incorporated in Scots writing, resulting in plenty of mixing. As Görlach (2002: 16) notes, rather than cultivating a linguistic identity, Scottish people before the union distinguished themselves from the English through their religion and politics. It is not until much later, when the political and social relationship between Scotland and England has changed drastically, that writers express sentiments and opinions about a Scottish mother tongue.

Although they are two typologically similar and closely related languages, we can still arguably describe the contact scenario between Scots and English in the early sixteenth century as some type of convergence contact: the two varieties were in close contact during centuries without either group shifting language, and, despite the usual description of convergence contact, it is not necessary for languages to be typologically different from each other for convergence to take place; for example, the loss of past tense distinctions in French, Romantsch, Northern Italian dialects, and Southern German is likely to have originated from convergence contact between these languages in border regions (Hock & Joseph 1996: 397). The long duration of contact between Scots and English, the fact that there was no strong, established, Scots written norm, and that the varieties did not appear to compete politically, facilitates precisely the type of mixing we observe between written Scots

and English; Muysken (2012: 714) describes a type of code-switching as a typical outcome of a contact scenario with the traits listed here, *congruent lexicalisation*, which involves inserting words from either language into a shared structure in cases where languages are structurally similar – this aptly describes the Scots-English form alternation in Scottish texts.

2.2.3 Scots-English convergence: Anglicisation and present-day status

In the latter half of the sixteenth century, the trajectory of Scots started to change, and even more dramatically so in the seventeenth century. In this period, several socio-political events took place which effectively limited the use of Scots in every influential institution in Scotland, and instead made (Southern) English the language associated with high and formal registers. These events and their specific consequences for Scots are broken down in Table 2.1. In this period, we see the use of salient Scots features decline rapidly in writing, in favour of English ones; hence, the period from the late sixteenth century until the modern era is commonly referred to as the *anglicisation of Scots*. Meurman-Solin (1997: 8) reports that, while there was variation between Scottish and English forms in writing before 1600, the use of distinctively Scots variants in the *Helsinki Corpus of Older Scots* (HCOS; Meurman-Solin 1995) decrease significantly after 1650. The genres where Scots continues to be used in HCOS are ego-documents (e.g. correspondence and journals) and legal writings, the latter being a genre which favours the use of Scots long into the modern period.

Table 2.1: Socio-political events aiding the anglicisation of Scots

Events	*Consequences*
The Reformation, 1560	English was considered a more appropriate language for protestant writing, as it made use of a more simple and vernacular style than Scots in writing. Furthermore, the biblical translation favoured by the reformers was in English, and the invention of the printing press meant that the English translation of the bible could spread widely in Scotland. This had the consequence that English became the language associated with the church, one of the most influential institutions in Scotland.
The proliferation of printed works, late 1500s	The invention of the printing press also affected Scots in other written genres than religious texts; English printers were more productive, which meant that printed works in English spread quicker, but many Scotland-based printers also shifted to printing in English as it meant having a wider reader-base, and thus making more profit. This meant that English encroached on more and more written domains in Scotland until it became the language of the majority of written works; Bald (1926: 115) finds that, out of writing printed in Scotland between 1560-1600, the average proportion of writing in English compared to Scots is 22.5%, but between 1600-1625 the same proportion is 82.5%.
The Union of Crowns, 1603	After the death of Elizabeth I of England, James VI of Scotland inherited the English throne and became James I of England and Scotland. He moved to England and ruled the united monarchy from there, and subsequently used more and more English in his writing. This had the effect that English became the language associated with the royal court, another crucial institution in Scotland.
The Union of Parliaments, 1707	Referred to as "the last act in the story" by Murison (1979: 9), the Union of Parliaments meant the move of legislature and political administration to England, thus making English, not Scots, the language of government.

The association of English with influential institutions had the natural consequence that the prestige of English increased in relation to Scots, and Scots started to be seen as a more vulgar, provincial, and uneducated variety. The union of the two countries also meant that Scots became increasingly seen as heteronomous to English, a dialect of English, rather than being its own, autonomous, variety. After the restoration of 1660, the Scottish aristocracy began to shift language to English, intermarried with English aristocracy, sent their children for schooling in England, and overall spent more time in England than before. This contributed even further to the shift in prestige of English, and the idea of English as the formal variety trickled down and affected the lower classes as well – no doubt aided, in the eighteenth century, by the *age of politeness* which saw etiquette and elocution guides correcting the use of Scots language features to fit the English model. It is during this time we see the emergence of *Scottish Standard English* (SSE), an accent of English which is modelled on the speech of the English aristocracy of the time, which was, e.g., rhotic and had word-initial /hw/ for today's /w/, but with some Scots influences, predominantly in the vowel inventory. While spoken Scots survives, mainly in working-class urban and rural speech, SSE becomes the dominant speech variety in Scotland and remains so until the present day.

Scots was, however, by no means eliminated in writing. After the Union of Parliaments, in the latter end of the eighteenth century, authors and poets such as Allan Ramsay and Robert Burns spearheaded a revival of Scots in vernacular verse, poetry and popular literature, as a backlash to the view of Scots as inferior. This has been termed the *vernacular revival* (see e.g. Aitken 1979; McClure 1994). Furthermore, a recent study on the use of Scots in chapbooks – cheap, printed, pamphlets or story books distributed on the streets – show Scots indeed surviving in printed works, in genres such as verse and song, in line with the vernacular revival, but also in local news (van Eyndhoven, Gotthard, & Filgueira 2021). In the use of written Scots today, we see a continuation of this pattern, with Scots translations of popular literature, and original writing in Scots, but also non-literary use of Scots intended for local audiences, such as on Twitter, in tweets relating to the 2014 Independence Referendum (Shoemark, Kirby, & Goldwater 2017).

Defining the status of Scots today is no easy task. The 2011 Scottish census showed that app.

37% of the Scottish population had some skill in Scots, e.g. could understand, read, or speak it, but the pre-census question testing report also showed that many respondents did not grasp what 'Scots' referred to, whether it was a label for SSE, the name for a particular dialect, or even if it referred to something that was 'bad English' (Eunson, Murray, & Ipsos MORI Scotland 2009: 7). Stuart-Smith (2004), following the line of Aitken (1979, 1984), describes the present-day situation as a bipolar continuum, with Scots on one end and the range of English varieties spoken in Scotland, including SSE, on the other; rather than switching cleanly between two varieties, which would be the case in a typical diglossic situation (in the sense of Auer 2005), speakers are more likely to mix features of both varieties, using more features from one or the other end of the continuum depending on context. Scots never was and still is not a homogenous variety, and it never had an established written or spoken standard to converge on, which has meant that regional and social variation has flourished in Scots speech. However, this has also continued to make Scots vulnerable to dialect levelling towards English, or SSE, supra-local norms. The heavy intra-and-inter-speaker variability resulting from this simultaneous divergence and convergence contributes to the difficulty in defining Present-Day Scots as a language distinct from English.

The difficulty in determining the distinction between Scots and English historically and today also implicates the definition of the term *anglicisation* itself. The consensus in the literature is that there is a visible shift towards English forms at the expense of Scots in the written record of Scots, and this is usually discussed in terms of lexis and orthography (notably by Devitt (1989) and Meurman-Solin (1993b), with accounts of the extent to which these orthographic changes reflects change in pronunciation, e.g., Kniezsa (1997)). However, it is less clear whether this shift represents a change in Scots, or a change in the behaviour of writers; that is, did Scots itself transform to be more like English, or did writers simply switch to English? The fact that Scots remains a spoken (and, albeit limited, written) variety in present-day Scotland is evidence for that Scots as a variety did not change in the same dramatic way as the written record from the Transition period indicates, which suggests that the written record reflects more of a language *shift* than language *change* during the anglicisation period. We find further support for the language shift interpretation in that it affected different genres to different extents; for example, the correspondence genre, which is the

same genre which the data used for this project belongs to, is among the last to have participated in this shift, and genres associated with more formal registers (with the exception of legal texts) among the first. This genre sensitivity suggests an awareness in the writers of what variety was appropriate for a given purpose. Moreover, while the anglicisation process is typically understood as a gradual process, certain more saliently distinct features, such as the change from <quh-> to <wh-> spellings in Wh-words, saw a very rapid shift taking place within less than 100 years (as demonstrated by van Eyndhoven and Clark (2019)). Thus, what is often carelessly referred to as the "anglicisation of Scots" should perhaps not be understood as a transformation of the language, but the term *anglicisation*, as used in this thesis, should be understood as the increasing favouring of English over Scots in writing.

2.2.4 Conclusion

In this section, we have seen a development pattern in which Scots and English first diverge from each other as sister varieties from the same ancestor, only to later converge again through the influence of Southern English on Scots during the anglicisation period. There is no consensus on whether Scots could be considered a standard, homogenous, language before anglicisation started, but it is clear that it was markedly different from standard Southern English, and on a standardisation trajectory. Thus, the period between the sixteenth to eighteenth century appears to mark a distinct change in Scots, as it goes from being a standardising variety, to the mixed, complicated, variety we encounter in Scotland today, which is not quite Scots as we recognise it historically, but also not quite English. We have also seen that Scots and English were always in close contact, but, with the rise of English as the more prestigious variety in Scotland, the type of contact changes from involving two equally prestigious and linguistically similar languages, to one exerting influence over the other as a socio-politically dominant variety; if we tentatively described the contact situation in the fourteenth to sixteenth century as a form of convergence contact between Scots and, mainly Northern, English (see 2.2.2), this arguably shifted to a substratum-superstratum relationship between Scots and Southern Standard English beginning in the late sixteenth century.

Kopaczyk (2013: 253) suggests labelling the period from 1560 to 1700 "Transition Scots", to

emphasise the nature of this period for Scots independent of English, i.e., that this period marks a transition, a "diluting", of Older Scots into the form of Modern Scots dialects. Hofman (2019) also notes that this label brings attention to the transformational and dynamic nature of this period, as opposed to the sense of decline which arises from the "anglicisation" label, and highlights the potential for more insight into language variation and change in investigating the Transition Scots period, which he deems a "contact scenario involving two related but markedly different written varieties on the verge of standardisation" (Hofman 2019: 39). This thesis aims to shed more light on this, still largely obscure, Scots period, by investigating variation and change in syntactic patterns. By this, I also hope to find out more about the extent of English influence on Scots, and understand better whether anglicisation manifests similarly in Scots morphosyntax compared to what has been found in other areas of Scots grammar. Indeed, "the history of the development of Scots from the late sixteenth century to the eighteenth is largely still to be written" (Macafee & Aitken 2002: 2.5.22).

2.3 Syntactic change in Early Modern English and Scots

2.3.1 Early Modern English syntax

Early Modern English (EModE) syntax is characterised by processes of standardisation and regulation of features which came about through more dramatic changes, and great variation, in the Middle English period (see e.g. Rissanen 2008: 187). During the EModE period, the use of written English expanded so much in quantity and range that it is possible to investigate socially conditioned variation in this period, which catalysed the field of historical sociolinguistics (initiated by Romaine 1982, see also Nevalainen 2006a). During this functional elaboration of English, in a period when literacy and education increased and there was a desire to use English for all communicative purposes, we see variation and change being more conditioned by, e.g., style and genre, and influences from Latin structures, in a movement to make English more suitable for advanced communication (e.g., Görlach 1991; Rissanen 2008; Nevalainen 2006b). As Görlach (1991) notes, the changes which took place in the Early Modern period were fundamental, in the sense that this is when we largely see English taking the structural shape it has today. For example,

as a movement towards refining the complex clause, there was increasing usage of a wider range of subordinating conjunctions, and more allowance for subordinate clauses in subject and object positions (Görlach 1991: 121-122). Furthermore, the use of double negation became stigmatised, and mostly used by writers of lower social ranks, which gave rise to what became the standard pattern of using a negator combined with a negative polarity item (Nevalainen 2006b: 112-113; Rissanen 2008: 263).

One of the major changes which took place in the Middle English period was the decline of more free word order options; The loss of verb-second (V2) word order (and possibly the decline of inflectional endings) meant that a more strict subject-verb-object (SVO) order was employed. A V2 grammar is identifiable through its systematic subject-verb inversion, when the first element of the clause is something other than the subject, and this type of inversion still remained in EModE in specific contexts, mainly in clauses initiated by adverbs, but also with direct objects or subject complements in first position. However, the frequency of clauses with inversion declines in this period, and this feature, too, is influenced by text type and the author's idiolect (Rissanen 2008: 265). The increasingly fixed word order compromised the options available for the ordering of given information before new (see, e.g., Los 2009). This led to an increase in the use of passives and cleft constructions to highlight certain information in the clause.

The move towards SVO word order also facilitated another major structural change in the Early Modern period: the grammaticalisation of the auxiliary *do*, resulting in the mandatory employment of *do*-support in interrogatives and negative declaratives in modern English. This change was also happening as part of a general movement of modal verbs into their own auxiliary verb category, and the rise of periphrastic structures to express tense, mood, voice, and aspect – this is also when we see the rise of the progressive *be* + *-ing* construction to replace simple present tense (e.g. Görlach 1991: 113-114; Rissanen 2008: 210). The stabilisation of the subject-verb order and the grammaticalisation of *do* are frequently mentioned as the most important syntactic changes in EModE (see e.g. Rissanen 2008: 263).

2.3.2 Early Modern and Transition Scots syntax: A blind spot

Descriptions of Early Modern Scots syntax traditionally focus on areas in which it differs markedly from English – this typical element of any description or codification of Scots grammar is referred to by Kopaczyk as the "English backdrop" (2013: 240) to research on Scots. However, syntactic structures which look identical on the surface may not be so underlyingly, and by studying seemingly similar patterns we may uncover differences in their diachronic development, or reveal deeper structural differences that could further our understanding of the syntactic feature under investigation (cf. Harris 1984, Beal 1997: 335-6, Kayne 2000). The typical description of Older Scots syntax describes it as almost identical to English, and that differences between English and Scots is "negligible" (here in the words of Görlach (1991: 96), but representative of various descriptive works). However, more recent studies on modern Scots dialects show distinct differences from modern English in, e.g., negation (e.g. Smith 2000; Thoms 2017), auxiliary contractions (Thoms, Adger, Heycock, & Smith 2019) and the operation of *do*-support (Jamieson 2015; also Smith 2000). As it stands, we do not know to what extent these modern patterns are recent innovations, or whether they have roots in historical developments.

Despite the relatively (compared to English) sparse descriptions, there are well-documented features of Older Scots syntax (some of which overlap into Modern Scots of the eighteenth century) which set it apart from English, as listed in brief below (compiled from McClure 1994, Beal 1997, Moessner 1997, King 1997, Görlach 2002, and others mentioned where appropriate).

1. Older Scots made use of relative markers *that*, *at*, and ∅, instead of *wh*-forms. The *Wh*-markers emerged in Scots from English influence during anglicisation.
2. Differences in use of modals, e.g. the use of *shall* where English would have *will*, or *can* instead of English *may*.
3. The use of negative clitic *na*, and the post-verbal negator *no* (alternating with forms of *not*). Furthermore, pre-verbal negation, e.g. *ne*, was retained longer in Scots than in English (Macafee 1992/1993: 33).

4. *Be*-perfect was retained longer in Scots than in English, and *be*-progressive started being used earlier in Scots than in English.
5. *Do*-support emerged later in Scots, and verb-raising is retained longer (see also Jonas 2002; Meurman-Solin 1993b, and Chapter 5 of this thesis)
6. The employment of the *Northern Subject Rule* for subject-verb agreement, which often shows up as the use of *-s* inflection on the verb for all persons (see also Chapter 4 of this thesis). While Görlach notes that there is some "confusion" (1991: 121) in the use of -s inflections in EModE, the pattern of never using -s with a plural subject has stabilised by the end of the period, while the Northern Subject Rule pattern continues in Scots, in some version, into the present day.
7. The use of double modal constructions, e.g. *He'll can help us the morn* (Miller 1993: 119-120) is worth noting, even though this feature is not attested until the early nineteenth century.

Further to this list, Görlach (2002) also notes the many similarities between Early Modern Scots and English syntax when it comes to functional elaboration patterns, related to increasing written communication (as described in 2.3.1). While features like those in 1 and 2, and to some extent in 3, may be argued to be lexico-semantical differences in the use of functional items, i.e. variation in form rather than structure, some of the features listed here indicate potential structural differences which are yet to be investigated in detail, in particular features 5 and 6 which relate to word order and subject-verb agreement. Görlach (2002: 99) suggests that distinct structures are not attested in formal writing due to the move towards English in the anglicisation period, and "if Scots was to have developed independent structures this could have happened only in informal/spoken registers – but these are insufficiently researched (and documented!)". Indeed, the fact that differences in functional lexis, rather than e.g. word order, are more well-documented and studied is likely to do, at least in part, with the lack of large annotated corpora for Older Scots. This issue of data availability is also noted by Görlach (2002), who highlights the difficulty in determining the level of influence from English on Scots in this period, or whether some similar syntactic developments would have happened in Scots independently, pointing out that attempts, such as by Meurman-Solin (1993b), to investigate this question do not have enough data to reach reliable conclusions on this matter.

The lack of available digitised corpora of Early Modern and Transition period of Scots has turned it into a blind spot for syntactic investigation. Thanks to syntactic research on modern dialects, now facilitated by the ScoSyA project (Smith et al. 2019), we know more about the present-day synchronic variation, but not much about what diachronic processes led to it. What we know of Older Scots morphosyntax is thanks to studies making use of data from the earliest period of written Scots, 1380-1500, using the part-of-speech-tagged *Linguistic Atlas of Older Scots*, or which have used smaller, unannotated, corpora of specific texts, collections, or authors, and thus relied on more qualitative methods. In Chapter 3, I describe a new parsed corpus of sixteenth to eighteenth century Scots, which hopefully will be an aid in filling this data gap, and shed more light into the development of Scots syntax and outcomes of contact with English in the dynamic Transition period.

Chapter 3

Methodology: Building a parsed corpus of Older Scots correspondence

3.1 Introduction

This chapter describes a new parsed corpus of Older Scots, built by syntactically annotating data from the *Helsinki Corpus of Scottish Correspondence 1540-1750* (CSC; Meurman-Solin and VARIENG 2017) according to the *Penn Parsed Corpora of Historical English* (Kroch et al. 2004, 2016; Kroch & Taylor 2000) format and framework. This corpus is called the *Parsed Corpus of Scottish Correspondence* (PCSC). I first describe the background for the data selection (3.2.1), as well as the version of the CSC made available to us for this project (3.2.2) and the modifications made for our purposes (3.2.3). I then introduce the PPCHE format (3.3.1), and outline the process of enriching the data with part-of-speech (POS) tags (3.3.2) and subsequent parsing (3.3.3). Finally, in Section 3.4, I describe my methods when using the data for quantitative studies like those in Chapters 4, 5, and 6.

3.2 Selection of data

3.2.1 Background

As has been established in Chapter 1 and 2, the wider aim of this project is to investigate syntactic change in Scots in the Early Modern period, i.e. the period of *anglicisation*. For this, I needed a corpus which fit three criteria: (i) it should contain writing from Scotland or produced by Scottish writers, (ii) the data should be produced in the sixteenth to eighteenth centuries, and (iii) it should be a corpus suitable for syntactic study. Table 3.1 gives an overview of the available corpora for Older Scots, i.e. fulfilling criteria (i), at the time this project started.

Table 3.1: Corpora of Older Scots

Name	*Time period*	*Morpho-syntactic annotation*	*Size (m words)*
A Linguistic Atlas of Older Scots (LAOS)	1380-1500	Yes: POS-tagged and lemmatised	ca. 0.4
The Aberdeen Burgh Records	1398-1511	No	?
The Records of the Parliaments of Scotland	1424-1707	No	?
The Dictionary of the Older Scottish Tongue (DOST) Corpus	1450-1600	No	0.88
The Helsinki Corpus of Older Scots (HCOS)	1450-1700	No	0.83
The Breadalbane Collection	1548-1583	No	ca. 0.08
The Helsinki Corpus of Scottish Correspondence (CSC)	1540-1750	No[a]	0.42

a. The documentation (Meurman-Solin & VARIENG 2016) describes a POS-tagged, lemmatised, version, but this is not publicly available.

There were limited options for corpora which fit criteria (ii) and (iii); the only POS-tagged corpus available was LAOS (The University of Edinburgh 2008), which covers an earlier time period, but the corpora covering the relevant time period were not annotated with morpho-syntactic information. Thus, to fill this gap, I annotated an existing corpus which met criteria (ii). The new parsed corpus, then, facilitates investigation of change in Scots grammar over time as it provides morpho-syntactically annotated data produced shortly after the time period covered by LAOS.

It should however be noted that it is not an extension of LAOS as such, since the genres of these corpora are different; LAOS consists of mainly legal texts. In order to access a language as close to the spoken variety as possible for textual data, I selected the CSC as my main data source, as it consists solely of correspondence data. Personal letters, as well as other autobiographical writings, is part of the genre of *ego-documents*, which is considered a particularly suitable genre for historical linguistic study; ego-documents are both considered to represent a speech-like language, and also usually represent a less high and formal variety than other more widely available textual data, such as literary, legal, or political writing (see van der Wal and Rutten 2013; Dossena 2013). DOST (Aitken, Bratley, & Hamilton-Smith 1984) was therefore ruled out as it mainly consists of literary data, such as poetry and folk tales, and while HCOS (Meurman-Solin 1995) does contain some ego-documents (e.g. correspondence, autobiographies, diaries), I ruled it out on the basis that it also contains other genres so that the total number words of only ego-document data in HCOS (206,500, also including diaries and autobiographies) is only about half of the almost 420,000 words of correspondence in the CSC.

3.2.2 The Helsinki Corpus of Scottish Correspondence

The original corpus compilers of the CSC have digitised the texts with careful consideration of preserving original language and manuscript features, e.g. by not modernising spelling and punctuation, and by adding annotation in the form of {bracketed} comments which mark features like tears in the manuscript, ink stains, deletions or insertions, and comments by the editors. There are also html codes marking formatting features, e.g., line breaks, and some forms have been annotated with special symbols; e.g. * and % which mark the contracted part of a contracted form (see 4.3 in Meurman-Solin and VARIENG (2016) for more detailed information on comments and annotation). (1) gives an extract from the CSC with examples of the annotation before any modifications have been made. Further to this, in selecting data for the CSC, priority was given to autograph letters by a single writer, to the extent that it has been possible to identify the hand of the writer (there seems to be more variation in hand in the sixteenth century letters; Meurman-Solin and VARIENG 2016: §4.1).

(1)
writtin
at
Edinburgh
ye
xxiij
day
of
\
</line>
<line>
november
Be
{space}
{a wide space vertically}
\\
</line>
<line>
Zour
gracis
hvmill
s*er%uatris
\\
</line>
<line>
Katherine
Bellenden
{address>}

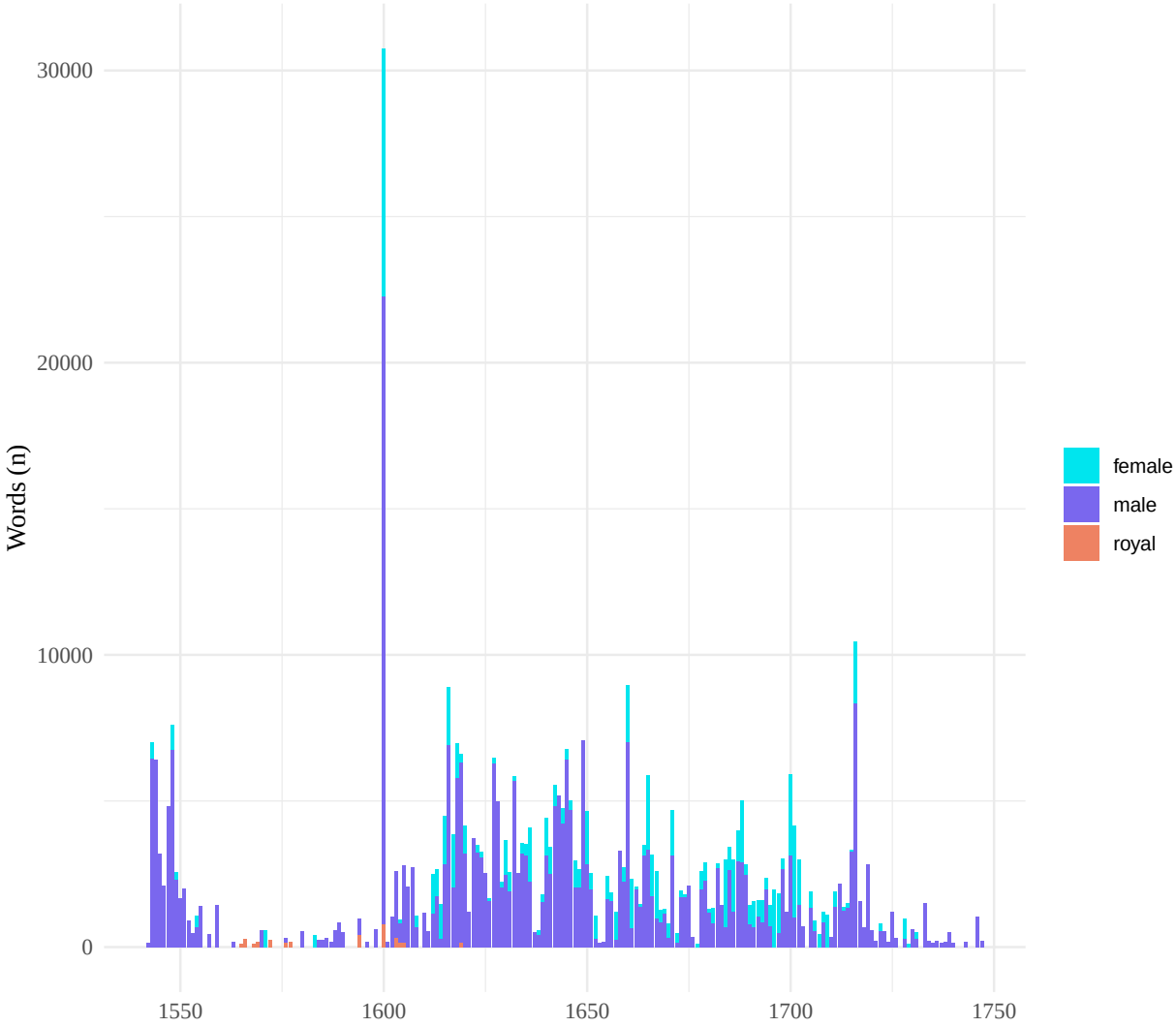
The corpus is grouped by writer gender and time period, but there is also a parameter value "royal", under the writer gender category, for the royal writers included in the corpus (Mary, Queen of Scots; James Stewart, 1st Earl of Moray; King James VI; and James Douglas, 4th Earl of Morton). The proportion of words, writers, and letters by gender is seen in Table 3.2.

Table 3.2: Gender statistics in the CSC

Gender	Words	Letters	Unique writers
Female	86,870	337	117
Male	327,612	1,008	320
Royal	3,227	17	6
Total	417,709	1,362	443

The time periods are specified as 1540–1599, 1600–1649, 1650–1699, and 1700-1749. The earliest and latest letters are produced at 1543 and 1747, respectively. Figure 3.1 gives the distribution of words across time, also showing proportions of writer gender.

Figure 3.1: Distribution of words in the CSC over time, by gender



Most of the letters are produced widely across Scotland, but some are localised to England or continental Europe. The provided metadata specifies the town/locality where the letter is produced (category: *lclet*), and the county (category: *lcinf*) and wider region (category: *largeregion*; 'Central', 'South-East', 'South-West', 'North', 'North-East') in which the writer is based, when possible; about 200 letters have "unspecified" for *lclet*, 52 have 'unlocalised' for *lcinf* and *largeregion* (with only one *lclet*: 'unspecified' letter overlapping). Furthermore, all letters with the gender variable

'royal' have their *lcinf* and *largeregion* specified as 'Court'. Similarly, writers who are lawyers, or members of the army or clergy, have their *lcinf* and *largeregion* specified as 'Professional' (Meurman-Solin & VARIENG 2016), which is likely due to the difficulty in pin-pointing these writers to a particular region due to their mobile lifestyles.

The metadata, which is given as a line of html code in the beginning of each letter file, also specifies various features of script, manuscript information, dates, word count, and the gender and addressee of each letter. (2) gives an example of a metadata line.

(2) <text datefrom="15431123" dateto="15431123" from="Katherine Bellenden" to="Mary of Lorraine, Queen Dowager" largeregion="South-East" year="1543" fraser="unspecified" lettertype="autograph" scripttype="secretary" lettertype2="information unavailable" scripttype2="information unavailable" id="52" bi="previously edited by Annie I. Cameron in the Correspondence of Mary of Lorraine, 36" ms="NRS SP2/1" fn="BellendenKatherine5431123" wgr="female" agr="royal: lcinf="Lothian" lclet="Edinburgh" wc="359" st="a copy in the CSC archive" date="1543 November 23" timefrom="000000" timeto="235959">

3.2.3 Data preparation and modifications

It was necessary to make modifications to the annotation of the data in order to make automatic POS-tagging more efficient. Firstly, the line of html code containing metadata was extracted for each file, and the metadata for all letters was instead organised into a spreadsheet, with each entry having a unique ID corresponding with the name of its letter file. The file names are formatted so that the first number is the number of the letter in the whole corpus, the second number is the number of the letter within its gender and time period group, and the final part has a *m/f/r + year* combination, referring to the gender classification of the writer and the start year of the time period group the letter belongs to; thus, 1_1_f1540 is the first letter of the corpus, and also the first letter in the group of female writers in the 1540–1599 time period.

All html codes which described formatting features were removed (e.g. </*line*>), as well as

backslashes indicating line breaks. Html codes representing special characters were converted into their corresponding characters; e.g., '&amp:' in the CSC is replaced by '&' in the PCSC. I was more conservative with the editors comments, in {brackets}, but still only kept comments which were essential for the interpretation of sentence structure, and not comments on manuscript features. Three categories of comments were kept: Comments indicating deletion of a form or extract, but where the deleted form(s) is still legible and thus transcribed (3-a), comments indicating damage in the manuscript which causes only fragments of a clause or word to be legible and thus transcribed (3-b), and editor comments providing possible interpretations of a form (3-c).

(3) a. sua sone as oportunitye ser{del} micht s*e%rue

b. we desy???{partly torn; possibly<desyre>;} you firmly

c. yaj think &{=if} yis realm' cu*m% till an' stabill way [that...]

The special symbols *, %, and =, representing contractions or superscripts, were also left in, but parentheses, '(' and ')', were replaced by '<lparen>' and '<rparen>', respectively, in order for automatic revision of POS-tags and parsing to work (see Section 3.3). After these modifications were carried out, the extract from (1) looks like (4).

(4)
writtin
at
Edinburgh
ye
xxiij
day
of
november
Be
Zour
gracis
hvmill
s*er%uatris
Katherine
Bellenden

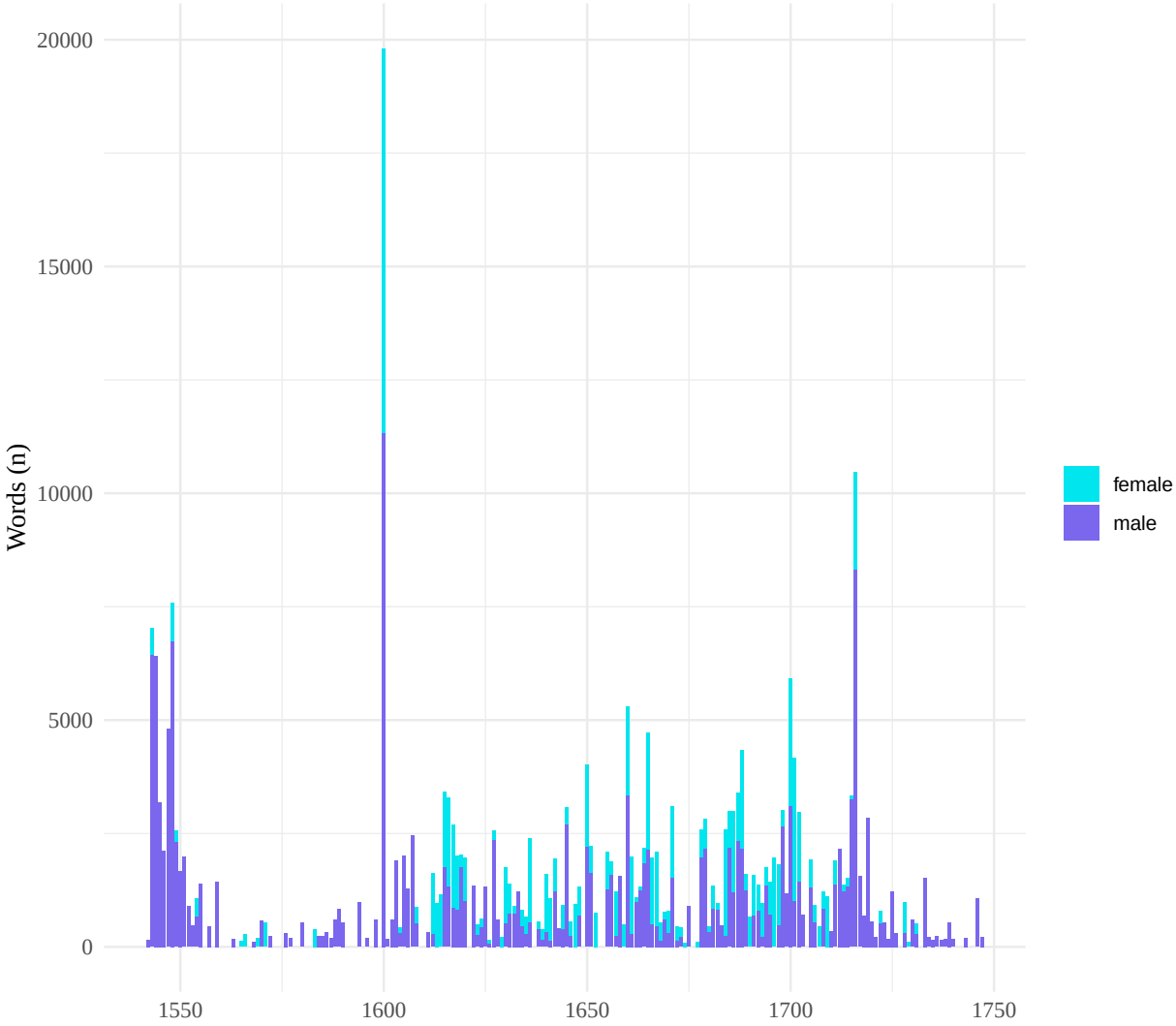
No further modifications to the language data was made, in order to keep the text as close to the manuscript forms as intended by the original corpus compilers. Thus, the PCSC can also be used to investigate orthographic variation at the lexical level.

After estimating the time commitment for this project, I decided to automatically tag and parse the whole corpus, but only manually correct a subset of it. This means that the size of the completed PCSC is 270,553 words, and the rest of the data is prepared for manual correction at a later time. The selection was made by randomly picking out 150 letters each from the male 1600-1649 group and male 1650-1699 group, which also made these time periods more gender balanced. Furthermore, I sorted the writers who had their *gender* parameter value set as 'royal' into 'male' or 'female', as their royal status is obvious from their IDs and other metadata values. Thus, the final distribution of words, letters and writers by time period and gender in the PCSC can be seen in Table 3.3, and the distribution of words over time by gender can be seen in Figure 3.2.

Table 3.3: Words, letters, and writers by gender and time period in the PCSC

Time period	*Gender*	*Words*	*Letters*	*Unique writers*
1540-1599	Female	3,603	14	9
	Male	48,048	160	70
	Tot:	51,651	174	79
1600-1649	Female	32,113	121	44
	Male	46,190	158	35
	Tot:	78,303	279	79
1650-1699	Female	37,041	152	49
	Male	47,385	150	56
	Tot:	84,426	302	105
1700-1750	Female	14,699	53	19
	Male	41,474	124	46
	Tot:	56,173	177	65
Total:		270,553	932	328

Figure 3.2: Distribution of words in the PCSC over time, by gender



Compared to the distribution shown in Figure 3.1, we now see a more balanced word distribution by gender in the seventeenth century, while there is still significant imbalance before 1600 and after 1700. There is also a marked gap in the data from the 1560s until 1600 and after the early eighteenth century – in a future project, the pre-1600 gap could ideally be filled with a correspondence corpus like the *Breadalbane collection* (Dawson 1997; listed in Table 3.1), which has the bulk of its data produced between 1565 and 1583, or by adding correspondence data from HCOS.

I modified the metadata so that the parameter values 'Court' and 'Professional' were no longer variables under the *lcinf* and *largeregion* categories but instead under the new category *class*, and

I set the value 'unspecified' for the rest of the writers. This resulted in 2% of authors classified as 'Court' (5,212 words, 27 letters), 3.6% classified as 'Professional' (16,937 words, 36 letters), and the remaining 291 authors classified as 'unspecified'. I then replaced the 'Court' and 'Professional' values with 'unspecified' under *lcinf* and *largeregion*.[1]

3.3 Annotation

3.3.1 Framework

The parsed CSC is annotated according to the format and principles of the *Penn Parsed Corpora of Historical English* (PPCHE), which include the *Penn-Helsinki Parsed Corpus of Middle English* (Kroch & Taylor 2000), the *Penn-Helsinki Parsed Corpus of Early Modern English* (PPCEME; Kroch et al. 2004), and the *Penn Parsed Corpus of Modern British English* (Kroch et al. 2016). This places the PCSC within a family of sister corpora which use the PPCHE annotation conventions, together with those listed in Table 3.4, and makes the PCSC the first Scots corpus which can be easily used for comparative study within this family of corpora.

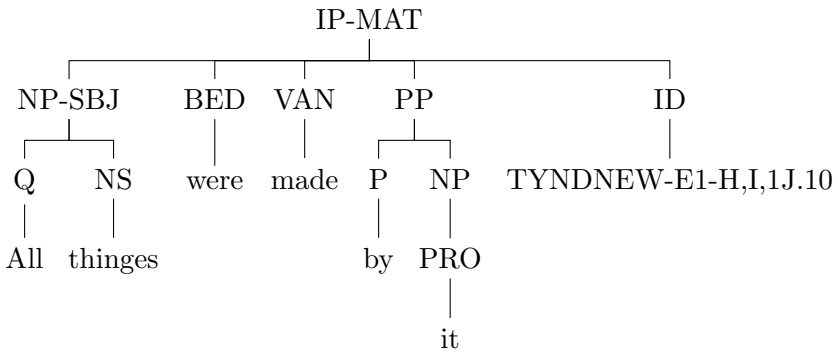
Table 3.4: Selection of corpora parsed according to the PPCHE annotation schema

the *York-Toronto-Helsinki Corpus of Old English Prose*	(Taylor, Warner, Pintzuk, & Beths 2003)
the *Parsed Corpus of Early English Correspondence* (PCEEC)	(Taylor, Nurmi, Warner, Pintzuk, & Nevalainen 2006)
the *Parsed Corpus for Early New High German*	(Light 2011)
the *Icelandic Parsed Historical Corpus*	(Wallenberg, Ingason, Sigurdsson, & Rögnvaldsson 2011)
the *Parsed Corpus of Middle English Poetry*	(Zimmerman 2014)
the *HeliPaD* for Old Saxon	(Walkden 2016)
the *Tycho Brahe Parsed Corpus of Historical Portuguese*	(Galves, de Andrade, & Faria 2017)
the *Audio-Aligned and Parsed Corpus of Appalachian English*	(Tortora, Santorini, Blanchette, & Diertani 2017)
a Parsed Linguistic Atlas of Early Middle English	(Truswell, Alcorn, Donaldson, & Wallenberg 2017)
the *Penn-BFM Parsed Corpus of Historical French*	(Kroch & Santorini 2021)
the *MCVF Corpus* for historical French	(Martineau, Hirschbühler, Kroch, & Morin 2021)
the *Penn Parsed Corpus of Historical Yiddish*	(Santorini 2021)

1. The complete modifiied metadata for the PCSC is in Appendix 1.

The PPCHE framework aims to facilitate automated searches, and, in doing so, offers a flatter parsing system which generalises across different generative analyses and can be applied to a range of languages. Most strikingly, the PPCHE parsing system does not commit to an analysis of VP structure, and the verb is instead an immediate daughter of the root IP and sister to all its complements and adjuncts, which renders a flat, multi-branched structure like that illustrated by Truswell, Alcorn, Donaldson, and Wallenberg (2019) in Figure 3.3. The framework aims to be as consistent, non-subjective, and non-controversial as possible, leaving little room for intuition in the parsing process – this again is to facilitate accurate automated searches within and across corpora. In order to further streamline work on and with PPCHE-type corpora, two tools have been developed which can be used for automatic revision and querying, and for hand-correction of parsed data: CorpusSearch (CS; Randall 2000/2013) and Annotald (Beck, Ecay, and Ingason 2011; see Sections 3.3.2, 3.3.3).

Figure 3.3: PPCHE tree structure



(Parsed token from the PPCEME, visualised in Truswell et al. (2019: 21))

Thus, the PPCHE family of corpora is an excellent resource for studying diachronic change and synchronic variation, as they allow us to extract structural patterns in a fraction of the time it would take to retrieve the same results from an un-tagged corpus. They are, however, limited in their size, range of dialectal variation, and lack of more detailed lexical annotation and lemmatisation. The PCSC fills some of the gap in size and dialectal variation by providing more data and widening the

range of dialects available for syntactic study; the PCSC is particularly suitable for comparative study with the PCEEC. In Section 3.4, we will see that, by using the CorpusSearch tool, it is indeed possible to search for lexically constrained variation on unlemmatised PPCHE-style corpora such as the PCSC.

3.3.2 POS-tagging

I first aimed to tag the corpus as accurately as possible using automatic methods. I created a Python script which takes a token from the CSC and matches it with a token from a dictionary of part-of-speech (POS) tagged forms, and then assigns the dictionary token's key (the POS tag) to the CSC token (see (5)). That is, the script reads forms in both the tag dictionary and corpus in lower case and without special characters, and, if a match is found, the output is a form which is POS-tagged according to the PPCHE format.

(5)

CSC form	*Dictionary entry*	*Output*
ye	'ye' : 'PRO'	(PRO ye)

Two tag dictionaries were used; the script attempts to first find a match in a tag dictionary of the *Linguistic Atlas of Older Scots* (LAOS; The University of Edinburgh 2008), and then in the Penn Treebank (Taylor, Mitchel, & Santorini 2003) tag dictionary. If a match is not found, the form was assigned an ERROR tag. I used one Scots and one English dictionary due to the high variability between Scots and English forms in this corpus, particularly in the later time periods, and as the LAOS dictionary uses older Scots forms and spellings than what is used in the CSC data (which is representative of a period when Scots spellings begin decline in favour of English ones).

LAOS is POS-tagged according to the *Linguistic Atlas of Early Middle English* format, so I used a Python script developed for the PLAEME project (Truswell et al. 2017), to convert the tags in the LAOS tag dictionary into the PPCHE format. As there were some differences between

the tag labels in the Penn Treebank and PPCHE, I also wrote a script which converted the tag labels in the Penn Treebank into corresponding PPCHE labels. These modifications gave us two python-readable dictionaries with forms in Scots and English with assigned PPCHE tags. (6) gives an excerpt from the LAOS dictionary.

(6) 'brugh' : 'N'
'bruik' : 'VB'
'bruk' : 'VB'
'bruk' : 'VBP'
'bruke' : 'VB'

Initially, I manually added frequently occurring ERROR-tagged forms to the LAOS dictionary, with their correct tags, and re-ran the dictionary-matching script until the remaining ERROR tags were found only with particularly idiosyncratic forms, such as proper names or individual spelling patterns. I also changed the dictionary keys where I noticed that the output was consistently incorrect, e.g. when there were multiple entries for the same form in the dictionary and the less frequent option was "chosen" for the match; in (6) we can see two entries for the same form 'bruk', in which case one would have been removed from the dictionary (typically the finite form, as verbs were tagged as infinitival by default). I then used various methodological options for correcting the remaning ERROR tags: (i) manually replacing them by their appropriate tag, (ii) using Python and RegEx to capture general orthographic patterns and assigning the most likely form based on that pattern; e.g., a form ending with some version of *-ous* would get the tag ADJ(ective), and (iii) using CorpusSearch revision queries to change ERROR tags based on their structural position; e.g. in a D(eterminer) ERROR N(oun) pattern, the ERROR was changed to ADJ.

When no ERROR tags remained, I used CS revision queries to automatically correct the rest of the tags based on their structural position, in cases where errors where consistent enough to be corrected in this manner; e.g., if a 'to' form tagged as P(reposition) occurred immediately

before an infinitival verb (VB), the tag is changed to TO (the tag label used for the infinitival marker). Manual correction of POS-tags was partly carried out after this stage[2], or at the stage of hand-correcting the parsed data. When POS-tagging was completed, the extract in (4) looks like (7)

(7) (VAN writtin)
(P at)
(NPR Edinburgh)
(D ye)
(NUM xxiij)
(N day)
(P of)
(NPR november)
(P Be)
(PRO$ Zour)
(N$ gracis)
(ADJ hvmill)
(N s*er%uatris)
(NPR Katherine)
(NPR Bellenden)

3.3.3 Syntactic annotation

The initial, automated, part of the syntactic annotation was done by using CS revision queries. Starting at the word level, each tagged form was wrapped in its corresponding phrase, as seen in (8-a). The next step expanded phrases to include complements and adjuncts, as can be seen in (8-b) where the PP extends over its NP complement. Finally, the whole clause is wrapped in the highest phrase level (8-c). Due to the non-modernised punctuation in the CSC, I could not use

2. Many thanks to Sarah Einhaus, then a student at the University of Konstanz, for helping to manually correct a portion of the automatically tagged data!

punctuation as an indication of root IP/sentence boundary. Therefore, I applied the same method as for other phrases when determining the root clause: I formed IP clauses by starting at the verb level and expanding it across potential complements and adjuncts to the verb. Note that, in the PPCHE system, CP is only used for subclauses and not for the root clause level.

(8) a. (IP-PPL (VAN writtin))
(PP (P at))
(NP (NPR Edinburgh))

b. (IP-PPL (VAN writtin))
(PP (P at) (NP (NPR Edinburgh)))

c. (IP-PPL (VAN writtin) (PP (P at) (NP (NPR Edinburgh))))

At this automatic stage, some functional elements could be added; e.g., for any Wh-phrase (WNP), a corresponding category with a movement trace (NP *T*) could be added to the subclause it is connected to, as well as an empty complementizer (C 0) as the Wh-phrase's immediate sister. I could also add functional tag extensions to some constituents, where this could be done with some accuracy; e.g. adding -SBJ to noun phrases which precede the verb (and are thus likely to be the subject of the sentence). Finite clauses were all assigned the IP-SUB phrase level, and CPs were not assigned any type extensions (such as -REL, -ADV). At the end of this automated stage, I had a "parse skeleton" which provided a starting point for manual correction – this parse skeleton forms the annotation for the automatically parsed CSC data not included in the PCSC.

The remaining errors, such as wrong labels or misplaced constituents, were then hand-corrected using Annotald (Beck et al. 2011) – a purpose-built tool for correcting PPCHE-style corpora with a user-friendly interface and using quick keyboard commands. The final parse of a simple sentence can be seen in (9).

(9) *Sentence*: "s=r George Lackent{<?} is not Come heer yet,"
Parse:
((IP-MAT
(NP-SBJ (N s=r=) (NPR George) (NPR Lackent{<?}))
(BEP is)
(NEG not)
(VBN Come)
(ADVP-LOC (ADV heer))
(ADVP-TMP (ADV yet))
(PUNC ,))
(ID 290_151_F1650,.8))

3.4 Retrieving results

3.4.1 Querying and further annotation

By using the CorpusSearch tool (Randall 2000/2013), we can query the PCSC and extract syntactic structures which occurrences in the corpus can then be measured and compared against alternative patterns or variables from the metadata. CS is a java program which has been developed specifically to query PPCHE-style corpora, and makes use of its own search-function calls as well as boolean logic operators in its query language. Thus, in using CS, we can write a query to search for a particular construction, such as (10-a), run that query on our corpus file, and the output contains all instances of this construction, e.g. (10-b), repeated from (9) but with the relevant construction in bold font.

(10) a. *query*: (BEP iPrecedes NEG) AND (NEG iPrecedes VBN)
b. *Output:*
((IP-MAT
(NP-SBJ (N s=r=) (NPR George) (NPR Lackent{<?}))
(BEP is)
(NEG not)

(VBN Come)
(ADVP-LOC (ADV heer))
(ADVP-TMP (ADV yet))
(PUNC ,))
(ID 290_151_F1650,.8))

In CS, definition (.def) files are used to assign multiple forms or tags to one label, which can then be used for more efficient query writing. For example, if we want to write a query which finds all instances of finite verbs preceding its subject, we can define a label *finV* which finds all instances of finite verbs and auxiliaries. This means that instead of writing "VB[PD]|HV[PD]|BE[PD]|DO[PD] precedes NP-SBJ", we can simply write "finV precedes NP-SBJ". This CS feature becomes useful when analysing potential lexical conditions on syntactic variation: say we wanted to investigate whether the type of pronominal subject has an effect on word order (cf. Chapter 4). This would require us to measure word orders with different types of pronoun subjects, but with the base annotation we can only get instances of NP-SBJ dominating a pronoun, and would then need to sort between the pronoun types in the results manually. However, by extracting all NP-SBJ dominating a pronoun, and manually assigning labels to the different pronoun forms in a .def file, we could then write revision queries which added temporary extensions to the existing tags; e.g., we first extract "(NP-SBJ (PRO" using the Unix "grep" command, then we define a subject type "2PRO" which covers all forms of the 2nd person pronoun found in the extracted list (11-a), then we write a CS revision query that adds the extension -YOU to NP-SBJ if the NP-SBJ dominates a 2PRO pronoun (11-b). Finally, we can formulate a CS query using our .def labels, e.g. 'finV precedes NP-SBJ-YOU'. In this way, CS allows us to look for variation conditioned by different types of lexical items within a syntactic category, even within a non-lemmatised corpus like the PCSC.

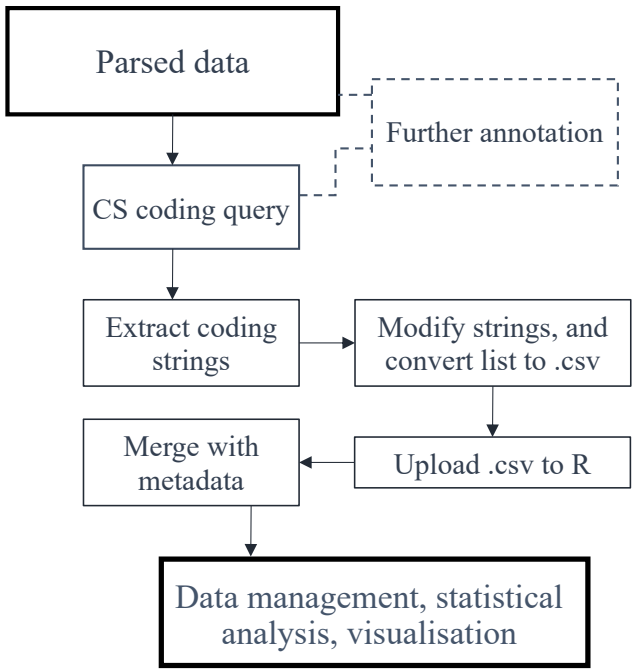
(11) a. 2PRO: yo=w=|yo=u=|you|you?|you_|youe|yow|yow_|_yow|yowe|yw|z=w=|ze|zee|zei|zo?w?|zou|zow|zow?|zowe|ziow|zw|Yee|You|You_|Yow|Ze|z*e%|zhe|zie|Zou|zou|zoue|Zow|zo=w=|ye|ye'|yhe|yee

b. query: (NP-SBJ iDoms [1]PRO) AND ({1}[1]PRO iDoms 2PRO)
append_label{1}: -YOU

3.4.2 Workflow

The process of querying the parsed data, and turning the results into a dataframe for statistical analysis, is visualised in the flowchart in Figure 3.4 and further detailed in this section.

Figure 3.4: Workflow



In the PPCHE format, each clause is given an ID node, which becomes part of the structure itself and functions as a branch of the structure which can be included in a query (see Figure 3.3). Each letter in the PCSC is assigned an ID which matches the file name (as described in 3.2.3), and which is an exact match of the IDs in the metadata. When individual sentence structures are extracted from a CS query, each clause will have this ID and a unique clause number added to it (cf. the ID

node in (9)).

CS also allows you to arrange queries into columns, generating a coding string output which can be used for multivariate analyses. For example, continuing our mock study from 3.4.1, we want to investigate word order variation by exploring both *pronoun subject type* and *tense* as potential factors conditioning variation. We then create coding query columns, and sort all queries related to pronouns in one column, all related to tense in the next, and those related to word order in a third – we can add more columns for each clause feature we want to investigate as a potential conditioning variable. When we run this query on the corpus, each clause in the results has a coding string associated with it, in which each part of the string is a parameter value associated with a coding column: say we only have three columns in our coding query: column 1 groups queries extracting subject types where the value for 2nd person pronoun subjects is labelled '2PRO', column 2 groups queries which extract tense where the value for present tense clauses is labelled 'present', and column 3 groups queries which extract word order where the value for subject-verb inversion is labelled 'verb-sbj'. Then, for a structure in the output which looks like 'yesterday came you here' (fabricated for the purpose of demonstration), the associated coding string would be: '2PRO:past:verb-sbj'.

From the output of a coding query, we can extract the coding strings using the CS *print_only* command. This renders a list of coding strings from the coding query output with their associated clause IDs, looking like (12); the coding string is followed by an @ sign and the clause ID.

(12) 2PRO:present:verb-sbj@1491_1_r1540,.2
N:present:sbj-verb@100_82_f1600,.5
2PRO:past:sbj-verb@1106_1_m1650,.4

With some string modifications, this list can be converted into a CSV file where each row is a clause and each coding parameter value and the IDs are arranged in separate columns. The CSV file can then be uploaded as a dataset to a statistical computing tool; for this project, I have used R (R Core Team 2021) for data management, statistical analysis, and visualisation. Once we upload our coding string dataset, we assign informative labels to the columns and merge the data with the metadata spreadsheet we created from the extracted metadata information (see 3.2.3), by the ID/filename column. The resulting dataset now has one row for each clause in the coding query output, which contains not only information about the clausal features we queried for, but also all the metadata information available for the associated ID, e.g. looking like 3.5.

Table 3.5: Example of data entry

filename	clause ID	sbj_type	tense	word_order	author	year	large_region	writer_gender
1491_1_r1540	2	2PRO	present	verb-sbj	James VI	1572	Unspecified	male

In this way, we can easily explore the results of our CS queries on the PCSC quantitatively, and investigate various factors affecting variation and change. This is demonstrated in more detail in the case studies presented in Chapters 4-6.

Chapter 4

The Northern Subject Rule

4.1 Introduction

This chapter explores a present tense agreement systems in Scots, the *Northern Subject Rule* (NSR), and how it manifests itself in the PCSC compared to Standard English present tense agreement. The NSR is a subject-verb (s-v) agreement system which is historically a distinctive feature of Scots and Northern English dialects. Versions of the NSR remain operational in dialects of Present-Day Scots, while Scottish Standard English makes use of Standard (Southern) English s-v agreement in the present tense. The two systems are similar – they only differ under specific conditions, and the differences are apparently structurally conditioned; as demonstrated in Table 4.1, in the typical NSR pattern, the options for present tense verbal inflection are -*(i)s*[1] or -∅, and the latter applies only variably with 1st person singular (1sg) and all plural (pl) pronouns, conditioned by their adjacency to the verb. Put differently, the present tense verb is inflected with -*(i)s* unless it is immediately adjacent to a 1sg or (any person) plural pronoun subject (e.g., Montgomery 1994, King 1997: 175-7, Pietsch 2005b).

1. The inflectional ending -*is* is typical of Older Scots specifically, and the vowel is dependant on the verb stem ending (as witnessed in Table 1).

Table 4.1: the Northern Subject Rule

Subject type		*Agreeing verb*
He/she/it/the girl		sing-is
I/we/they/you		sing-∅
The girls		sing-is
He/she/it/the girl,	while dancing furiously,	sing-is
He/she/it/the girl	sing-is and	dance-s
I/we/they/you,	while dancing furiously,	sing-is
I/we/they/you	sing-∅ and	dance-s
The girls,	while dancing furiously,	sing-is
The girls	sing-is and	dance-s

This makes the NSR a particularly interesting object of study for a syntactic description of Scots, because it brings us away from the typical description of differences between Scots and English in functional lexis as described in Chapter 2.3.2. Investigating the NSR requires retrieving clauses with particular types of pronoun subjects and with different types of verbal inflection, which are not specified features in the default PPCHE annotation for Noun Phrases (NPs), pronouns, or verbs. To tackle this, further annotation, specific to the NSR, was added to the PCSC, thus demonstrating how a non-lemmatised corpus can be queried to extract clauses based on lexical as well as structural features (as described further in Chapter 3.4.1).

The Chapter is organised as follows: Section 2 gives a brief account of the origin of the NSR in Northern Britain, and how it has developed and spread across the English- and Scots-speaking world (4.2.1), as well as describes the operation of the rule and how it differs from a Standard English system (4.2.2). The research question and hypotheses are stated in Section 4.3, including a brief background to why potential differences in gender are also investigated. In section 4.4, the method of annotating and extracting clauses coded for NSR constraints is described (4.4.1), as well as how the results were analysed and tested for significance (4.4.2). Finally, the results are presented in section 4.5; there is a decline of verbal inflection with subject types which would occur with inflected verbs in an NSR system but not a Standard English agreement system. These

results, as is concluded in section 4.6, indicate a move from a Scots system to a Southern English system, and are thus consistent with a contact-induced change with respect to this syntactic feature. Furthermore, female writers in the PCSC tend to use the NSR system of s-v agreement more than male writers (4.5.3); this suggests that Scottish women are more conservative in their use of Scots syntax, similarly to what has been found for other Scots features. Unexpected patterns of variation with 3rd person singular inflections are also found, for which no more definitive explanation is given than that the nature of this variation is possibly predicted by analyses of the NSR which identify the -∅ inflections as true subject-verb agreement.

4.2 Background

4.2.1 Origin and spread

The *Northern Subject Rule* pattern of s-v agreement was first named so by Ihalainen (1994: 221), but it is sometimes referred to as the 'Northern Present Tense Rule', 'Northern Personal Pronoun Rule', or 'Northern Verbal Concord'. These names all hint at the well-established theory that the rule's origins lie in Northumbrian Old English (e.g., King 1997: 175-7, Pietsch 2005b, de Haas 2011). It is generally argued that the subsequent NSR paradigm developed from the reduction of the verbal inflectional affixes -*eð*/-*að*/-*iað*/-*is* to neutralized -*es* (cf. forms in Table 4.2), which was a change already in progress in mid-10th century Northern texts (Pietsch 2005a: 173). By the time of the first Northern Middle English (ME) texts, the NSR pattern is already established (de Haas & van Kemenade 2015; de Haas 2011; McIntosh 1989). Table 4.2 gives a simple representation of present tense inflection suffixes on verbs with pronoun subjects in Older Scots compared with contemporary Northern and Southern English; the -*(i)s* inflection in Scots is assumed to have its origin in the Northern English paradigm, where we find -*as* and/or -*es* throughout the entire paradigm except for in 1sg contexts, whereas -*(i)s* inflection in 1sg was either extended by analogy with the other forms or transferred from contact with Scandinavian varieties spoken in the North of England (Rodríguez Ledesma 2013: 152). The alternation seen in Older Scots and Northern ME, with ∅-inflection in 1sg and plural subject contexts, is dependent on S-V adjacency (cf. Table 4.1, this alternation is further detailed in 4.2.2). This pattern may have emerged when Southern

English inverted V-SBJ constructions without verbal inflection (e.g., 'sing we') were introduced to the already reduced verbal inflection system in Northern ME; West Saxon OE tended to delete plural suffixes before the plural pronouns *we* and *ge* (=you(pl.)). The introduction of this system to the Northern paradigm, Pietsch (2005b: 173-8) suggests, would have led to enough variation in form that a new pattern emerged which considered subject type and adjacency rather than person and number.

Table 4.2: Subject-verb agreement with pronouns

	OSc	Southern ME	Northern ME
1sg	-(i)s / -∅	-∅	-o,-e / -∅
2sg	-(i)s	-(e)st	-as
3sg	-(i)s	-(e)th	-as,-es
Pl (all)	-(i)s / -∅	-∅,-(e)n	-as,-es / -∅

OSc = Older Scots, ME = Middle English
(Table adapted from King (1997: 175-6) and Pietsch (2005b: 175))

However, it is likely that the pattern arose from a combination of internal and external factors. The similarity of the NSR system to Welsh agreement has led to suggestions of substrate effects from Brythonic-English bilingualism in Northumbria as a possible origin of the rule (Klemola 2000); with 3rd person subjects, Modern Welsh uses the singular verb form with a nominal subject regardless of number, but the plural verb form when the subject is an adjacent plural pronoun, which is likely to be an inherited pattern from the Brythonic ancestor of Welsh (Klemola 2000: 337-8). As is persuasively argued by de Haas (2011), it is plausible that language contact, both with Brythonic Celtic and Old Norse, did play a role in the establishment of the pattern; language contact would have introduced more variation to Northumbrian, leading to the "confusion" between forms which gave rise to re-analysis, and substratum influence from Brythonic is a likely influence causing the pattern to eventually settle on subject type/adjacency conditions, rather than a number/person distinction (as in Southern English).

The NSR pattern appears to have been part of Scots grammar from the time of the divergence of Scots from English, when Scots developed from Northern English varieties and continued to form a dialect continuum with Northern English dialects. By the time of the earliest written record of running prose in Scots, the pattern is already near-categorical (Montgomery 1994). From its place of origin in Northern Britain, the NSR pattern spread both domestically and globally. Versions of the NSR have been recorded as a minority pattern in the south of England historically and today, and the pattern is robustly attested in Irish English and particularly in Ulster (in all probability due to the high number of settlers from Northern Britain there); see e.g., Pietsch (2005a, 2005b); Henry (1995); Corrigan (1997) for modern Irish English NSR studies. The pattern was brought with Irish and British emigrants to North America where it is now associated with, e.g., Appalachian, Ocracoke, and African-American English (see Montgomery 1997, McCafferty 2003, and de Haas 2011: 108-16, for an overview). In present-day Scots, a variable NSR pattern is recorded for some present-day dialects spoken in Scotland, e.g., in the South-East (Pietsch 2005a, 2005b), Buckie in the North-East (Adger & Smith 2010), and Hawick in the Scottish Borders (Buchstaller, Corrigan, Holmberg, Honeybone, & Maguire 2013).

4.2.2 Operation of the rule

The paradigm given in the beginning of this chapter (table 4.1) demonstrates what Pietsch (2005b: 5-6) describes as the "idealized" version of the NSR, formulated as follows (13):

(13) *The Northern Subject Rule (A)*:
Every agreement verb takes the *-s* form, except when it is directly adjacent to one of the personal pronouns *I*, *we*, *you* or *they* as its subject.

(Pietsch 2005b: 5)

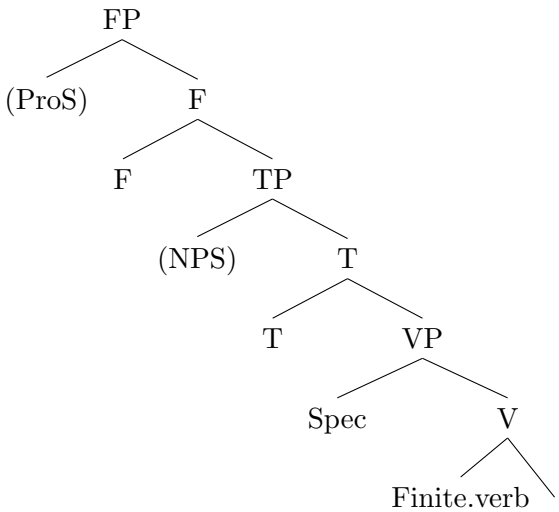
We can also phrase (13) as: default agreement in an ideal NSR system is some type of *-s* inflection; only if the subject is of a particular type (*I*, *we*, *you* or *they*) and is immediately adjacent to the agreeing verb, verbal inflection is ∅. This description assumes that the default, unmarked, inflection is *-s* and the exception, the marked option, is -∅ (cf. Pietsch 2005a: 179-81).[2] Theoretical accounts of the NSR differ in their analyses regarding which of the -s or -∅ inflection is the true agreement form. Corrigan (1997) analyses -∅ inflection as "anti-agreement", in that it appears with pronouns to avoid double-marking of features, as pronouns have distinct case and number forms. This is developed further by Börjars and Chapman (1998), who propose two lexical sets of pronouns – *he, she, it, I_1, you_1, we_1, $they_1$,* and *I_2, you_2, we_2, $they_2$* – of which set 2 pronouns are part of the verb morphology (and thus show up with -∅ inflection), while set 1 pronouns are part of the syntax (thus triggering true s-v agreement in the form of an -s suffix).

An apparently opposite categorisation of the inflectional endings is given by Henry (1995) who, in formulating an analysis of the version of the NSR found in Belfast, analyses -∅ as real agreement, a result from checking of features between the subject and verb which have raised to the specifier and head positions of an AGR(eement)P(hrase) projection; it is only if subjects are morphologically marked for nominative case that this move takes place, and, otherwise, the subject and verb move to T(ense)P(hrase) and the verb receives a default tense marking (i.e., -s inflection). De Haas (2011) builds on this theory, and develops an account in which -s inflection is a default agreement suffix which applies when subject-verb agreement fails, and that agreement is also dependent on s-v adjacency at the spell-out stage (in line with Bobaljik (2002)) – this theory will be revisited in more detail in the analysis of Scots *do*-support in Chapter 5. These analyses by Henry (1995) and de Haas (2011) assume that pronoun subjects merge in a different subject position from nominal subjects, comparable to the high subject position of Old English subject pronouns (as proposed in seminal work by Haeberli (2002)), which allows for subject-verb agreement if it is projected (cf. Figure 4.1, which has de Haas's (2011) label 'F(unctional)P' for this higher subject position).

---

2. As is noted by Pietsch (2005a: 179-81), whether a suffix is considered 'marked' in an analysis of the NSR is based on the interpretation of the term 'marked' – this will not be discussed further here.

Figure 4.1: NSR tree structure, adapted from de Haas (2011: 117)



ProS = Pronoun Subject, NPS = Nominal Subject

By comparing the NSR to Standard English (StE) S-V agreement, as seen in Table 4.3, we find that (i) in S-V adjacent clauses, only clauses with plural nominal subjects behave differently between the two systems, and (ii) in non-adjacent clauses, only clauses with 3rd person singular (3sg) subjects categorically behave the same between the two systems.

Table 4.3: the Northern Subject Rule compared to Standard English agreement

Subject type		*Agreeing verb*	*Grammatical in StE?*
	S-V adjacent clauses		
He/she/it/the girl		sing-is	✓
I/we/they/you		sing-∅	✓
The girls		sing-is	✗
	Non-adjacent clauses		
He/she/it/the girl,	while dancing furiously,	sing-s	✓
He/she/it/the girl	sing-is and	dance-s	✓
I/we/they/you,	while dancing furiously,	sing-s	✗
I/we/they/you	sing-∅ and	dance-s	✗
The girls,	while dancing furiously,	sing-s	✗
The girls	sing-is and	dance-s	✗

The examples in (14) and (15), from the PCSC, demonstrate an ideal NSR pattern operating – subjects are in bold font, and the agreeing verbs are underlined. First, in (14), with plural nominal NPs: in both S-V adjacent (14-a) and non-adjacent clauses (14-b), the agreeing verb will take *-(i)s* inflection if the subject is nominal and plural, unlike a StE system.

(14) a. for **all his freyndis** thynkis it suld be sa

Hew Campbell of Loudoun (sheriff of Ayr), 1548

b. **zour frendis** yat ar heyr' consellis yow all to hald furht zour purposs and' keip zour daye affixit to zow

Richard Kincaid, 1543

(15) a. **I** being young and hauing little experience knows little how to doe in so criticall a time

John Murray (Duke of Atholl), 1700

b. **I** se na help bot be zowr grace & hoppis na vdyr [..]

Alexander Gordon (Postulate of Caithness), 1549

The examples in (15) show different types of non-adjacency with a 1sg subject, as also seen in the paradigms in Tables 4.1 and 4.3; In (15-a) we see S-V adjacency interrupted by a parenthetical clause and thus *-s* inflection on the agreeing verb. In (15-b) we see adjacency between the 1sg subject and first verb, and thus *-∅* inflection, whereas the conjoined verb (which I assume to have an unpronounced subject, anaphorically dependent on the subject of the first verb) takes *-is* inflection because it is not immediately adjacent to the overt subject.

NSR varieties usually do not exhibit this ideal NSR system consistently, and there is often variation between the NSR and StE agreement in the environments where the two systems differ. This leads Pietsch (2005b: 6) to formulate a definition of the rule which distinguishes between its two constraints, as seen in (16). Henceforth, the Type-of-Subject and Position-of-Subject constraints will be referred to as *subject type* and *adjacency*, respectively. The type of construction in (15-b),

with one overt subject and two conjoined verbs, the first being adjacent and un-inflected and the second non-adjacent and inflected, is considered the most typical example of the NSR constraints at work (Pietsch 2005b: 7-8).

(16) *The Northern Subject Rule (B):*

a. All third singular subjects (and, where preserved, the old second singular *thou*) always take verbal -*s*.

b. Type-of-Subject Constraint: All other subjects except the personal pronouns *I*, *we*, *you*, *they* (and, where it exists, *youse*) take verbal -*s* variably

c. Position-of-Subject Constraint: Non-adjacency of subject and verb favours verbal -*s*.

(Pietsch 2005b: 6)

There are two types of intervening elements which seem to be invisible for the *adjacency* constraint: simple Adverb Phrases (AdvP) and stranded quantifiers (e.g. 'all' in *they all sing*). These types of intervenors are frequently found to not invoke the -s inflection on the verb when the subject is *I*, *we*, *you*, *they*, despite seemingly interrupting adjacency between the subject head and the agreeing verb (Pietsch 2005b: 9-10, Henry 1995: 25-6). While they are strictly not the same type of clausal element, they can both be analysed as "weak" elements which allow transfer of features between the subject and verb, e.g. in terms of their position as specifiers rather than heads of their own projection (cf. Kroch 1989: 236 and van Kemenade 2000, who make similar analyses regarding the negator *not* in the history of English), or in terms of their similar loose connection to the clause which allows them to be pronounced in different positions (cf. Bobaljik 2002: 212-20).

The *adjacency* constraint is frequently found to operate more variably than *subject type*, particularly in the modern varieties but also in Northern Middle English (e.g., Fernández Cuesta 2011), leading de Haas (2011: 107) to suggest that the *subject type* constraint is overall more stable, and more essential to the NSR, than *adjacency*. This difference in strength of constraints is probably caused by grammar competition between the ideal NSR and StE agreement, as those dialects

which historically use an NSR pattern with both constraints operational are the same dialects that have experienced less exposure to standardising pressures from Southern English (e.g., Pietsch 2005b: 11). In fact, the NSR has only been recorded as operating near-categorically in Older Scots; Rodríguez Ledesma (2013) finds adjacency to be as "strong" a condition as subject type in 14th-15th century Scots data from the *Linguistic Atlas of Older Scots* (LAOS; 1380-1500), looking only at 1sg contexts. In another study of a selection of 100 letters from the *Breadalbane Collection* (1548-1583), Rodríguez Ledesma (2017) finds the *adjacency* condition operating near-categorically with all relevant subject types. Montgomery (1994) finds the constraints operating with equal strength in sixteenth-17th century Scots data from a range of genres; he includes data from the *Scottish Correspondence of Mary of Lorraine* (1543-1546) and the *Memorials of the Montgomeries* (1611-1634), which are also part the *Helsinki Corpus of Scottish Correspondence* (CSC) [3], and thus included in the PCSC. However, Montgomery (1994: 86) only includes the first 100 instances of potential NSR clauses for each of these collections (the total number of contexts investigated by Montgomery is 1,534). These findings on earlier Scots data suggest that the NSR was a productive rule in the beginning of the period covered by the PCSC. Furthermore, Montgomery (2013) finds that both NSR constraints operate in eighteenth century correspondence of emigrants from South Argyll in North Carolina, which Rodríguez Ledesma (2017) argues follows from the strength of the constraints in sixteenth century writing from Argyll in the *Breadalbane Collection*. Given that versions of the pattern – that is, NSR patterns which exhibit *subject type sensitivity* but a weakened *adjacency* effect – also survive in more traditional present-day dialects, such as Buckie (Adger & Smith 2010) and Hawick Scots (Buchstaller et al. 2013), it seems that a decline in the pattern in the Transition period would not solely be due to language-internal processes.

Analysing 1sg pronouns as participating in the NSR pattern has been contested: King (1997:175-6) suggests the -∅ inflection in the 1sg is actually a result of the loss of weak 1sg endings in the Old Northumbrian present tense paradigm (cf. Table 4.2), and de Haas (2011) excludes 1sg pronouns from her analysis of the NSR. However, Rodríguez Ledesma (2013) finds robust evidence for

3. Regarding the letters to Mary of Lorraine, the CSC manual states: "Only the letters identified as autograph in the earlier edition of these letters by Annie I. Cameron have been included" (Meurman-Solin & VARIENG 2016: §4.1).

1sg pronouns participating in the rule through their near-categorical sensitivity to the *adjacency* constraint – whether they were part of the rule originally or not, it seems, then, that 1sg pronouns are analysed as participating at the time of the LAOS data. The 2nd person (2p) pronouns are also dealt with differently across previous literature, due to the fact that they only seem to adhere to the NSR if there is a salient number distinction in the 2p form (*thou/you* in Older Scots, and *you/youse* in modern varieties; cf. (16)), and that, otherwise, the form *you* behaves as in plural or 1sg contexts regardless of whether it has a singular or plural referent, in that it occurs with ∅-inflected verbs in S-V adjacent clauses and *-(i)s* elsewhere (see also Smith 2000: 252). This has parallels in the loss of the 2p number distinction in English, as the 2p singular *-st* disappeared when the plural form *you* was used for both numbers. While the *thou/you* distinction was retained longer in Scots than in English, with *thou* surviving in religious contexts into the nineteenth century (King 1997: 170), there are no examples of *thou* forms in the PCSC and thus the 2p number distinction is not expected to be relevant to the current study; in Rodríguez Ledesma's (2017) study on sixteenth century Scots data, 2p subject are found to align with plural pronouns and 1sg subjects in their agreement patterning, and the same is found by Montgomery (1994).

4.3 Research Question and predictions

The main research question driving this case study is as follows:

RQ: Is there competition between Standard English subject-verb agreement and the Northern Subject Rule during the Scots Transition period?

This question feeds into the wider aim of this thesis in that it investigates whether anglicisation processes in the Scots Transition period had an effect on the syntax of Scots, which will contribute to our understanding of the likely outcomes of the type of contact scenario we see between Scots and English in Scotland in this period, and particularly the shift in prestige between the languages. In light of this, the hypothesis in A. predicts that the Northern Subject Rule system declines in favour of a Standard English S-V system.

A. English influence in the Transition period led to a syntactic change in Scots.

There is reason to hypothesise that, if there is change over time in this feature, the female writers will be more conservative: It is frequently found to be the case that women use more Scots features in seventeenth-eighteenth century data (e.g., Meurman-Solin 1993b: 160-2), which may be related to that women had less access to education, and thus Standard English, while Scots remained the language of the household (see also McColl Millar 2012: 28). This contradicts the usual principles of the role of women in language change, as formulated by Labov (2001: 261-93), which state that women are leaders of language change both in cases of change from above (adapting to prestigious norms) and below (being innovators of new linguistic norms); the paradox can be summarised as "Women conform more closely than men to sociolinguistic norms that are overtly prescribed, but conform less than men when they are not" (Labov 2001: 293). Therefore I also investigate whether there is a difference between male and female writers with respect to the NSR. If the relative linguistic conservativeness of Scottish women applies to syntax, the prediction in B should hold:

B. Female writers use higher frequencies of NSR-pattern constructions, and the change from NSR to Standard English agreement is led by the male writers.

In addition to the more clearly stated predictions A and B, this chapter has the more exploratory aims of investigating whether there is a difference in strength of NSR constraints in this data, and to what extent a parsed corpus of correspondence is suitable for investigating this type of variation.

4.4 Method

4.4.1 Annotation and retrieval

In order to retrieve frequencies of verbal inflection by type of subject, it was necessary to add more detail to the PCSC annotation. For this, I used the method described in Chapter 3.4.1 to extract subject pronouns and present tense verbs from the PCSC, and added the annotation in (17) using CorpusSearch (CS; Randall 2000/2013) revision queries. Henceforth, the group of subjects

which variably occur with ∅ inflection – that is, plural and 1sg pronouns – is referred to as "NSR subjects". The 2p pronouns were not separated by number, but analysed as a single subject type (motivated by the lack of salient singular forms of the 2p pronoun in the corpus, and that no number distinction was found by Rodríguez Ledesma (2017) and Montgomery (1994))

(17) All pronoun subjects received a tag extension specifying their relevant person and number features:

- 1sg subjects are tagged as PRO-1sg
- 2p subjects are tagged as PRO-2p
- 3sg pronoun subjects are tagged as PRO-3sg
- Plural pronoun subjects are tagged as PRO-pl

Present tense verbs (VBP, DOP, HVP, BEP) with a sibilant inflectional ending received the tag extension -s, resulting in the following tags:

- VBP-s, DOP-s, HVP-s, BEP-s

The verbal inflections are a mixture of typically Scots and Southern English inflectional endings (e.g.: *-s*, *-is*, *-es*, *-eth*, *-th*), testifying to the high variability between the forms in the data. The verbs tagged with the inflection extension -s were extracted from the corpus and checked manually to ensure that only truly inflected verbs were tagged as such. This revealed a number of ambiguous cases, arising from verb stems ending in <s> (or <th>, but the only example of this is the verb BREATHE). The ambiguous cases, which amounted to 242 tokens (out of which forms of PLEASE constituted ca. 60% (n = 148))[4], could be grouped into three categories:

A. Likely uninflected: Forms ending with a single <s>, e.g.: <ples>, <pleis>, <pleas>

B. Likely inflected: Forms ending with <*s%>, e.g.: <pleas*s%>, <pless*s%>

C. Truly ambiguous: Forms ending with <ss>, e.g: <pleass>

4. Besides BREATHE (n = 6) and PLEASE (148) , the other verbs on the list are: ADDRESS (1), ADVERTISE (2), BANISH (1), BLESS (24), CAUSE (7), CEASE (1), CHOOSE (1), CONFESS (10), DRESS (1), EXPRESS (1), INCREASE (4), MENACE (1), PASS (7), POSSESS (2), PRESS (2), PROMISE (4), PURPOSE (3), REFUSE (4), SUPPOSE (8), USE (1), WIT (2; spelled <vis>, <viss>), WITNESS (1).

Items which fit into group A (n = 182) were re-coded as uninflected. Group B items (n=32) were determined to be cases of inflected verbs because the contraction markers <*%> are generally used for inflectional endings in the corpus (cf. <intend*is%>, <occurr*is%>), and, thus, they were counted as cases of verbal inflection. The truly ambiguous cases, group C (n = 28), were not included as tokens.

The next challenge was to determine the subject type of omitted subjects; in the PPCHE framework, an elided subject under conjunction gets assigned the form *con*, and does not have a POS-tag but is immediately dominated by the NP projection. In order to find and categorise these subjects, I first added further annotation by percolating features of the subject onto the NP-SBJ projection and then to the IP clause:

– NP-SBJ dominating an NSR subject (*I*, *we*, *you*, *they*) received an -NSR tag extension, NP-SBJ dominating a *con* received -CON, non-pronominal plural NP-SBJ received -PL, all other NP-SBJ are tagged as NP-SBJ-NO
– IP-SUB and IP-MAT with present tense verbs received a tag extension based on the subject tag extension:
  - The extension on IP-SUB/MAT is -NSR if the subject of the clause is NP-SBJ-NSR
  - The extension on IP-SUB/MAT is -NO if the subject of the clause is NP-SBJ-NO
  - The extension on IP-SUB/MAT is -CON if the subject of the clause is NP-SBJ-CON
  - The extension on IP-SUB/MAT is -PL if the subject of the clause is NP-SBJ-PL

Then, the -CON subjects were categorised through the following steps:

1. IP-SUB-CON was changed to IP-SUB-NO, IP-SUB-NSR or IP-SBJ-PL if its conjunct was IP-SUB-NO, IP-SUB-NSR or IP-SUB-PL, respectively
2. The extension on IP-SUB was percolated down to the NP-SBJ level, so that NP-SBJ-CON was changed to NP-SBJ-NSR/NO/PL accordingly
3. Since it is not possible to find IP-MAT conjuncts with CorpusSearch, the IP-MAT-CON clauses were corrected manually; for each IP-MAT-CON, its corresponding conjunct with an -NSR, -NO or -PL tag was identified and the -CON tag extension changed accordingly
4. IP-MAT-NSR/NO/PL extensions were percolated down to the NP-SBJ level, so that NP-SBJ-CON was changed to NP-SBJ-NSR/NO/PL accordingly

When the annotation was added, a CS coding query waas written to find instances of present tense clauses coded for *subject type*, *finite verb type* (i.e., whether it is *be*, *have*, *do*, or a lexical verb), *S-V adjacency*, and *inflection*.[5] The definitions for different types of adjacency in the coding query are as follows:

– A clause is S-V adjacent if the head of the subject noun phrase immediately precedes the finite verb, and that head is not an empty element (e.g., *con*).
– A clause is weak-adjacent if there is an element occurring between the subject and verb, and that element is an Adverb Phrase (AdvP) or stranded quantifier (i.e., a simple Q(uantifier)P occurring in final position of the NP-SBJ)
– A clause is non-adjacent if there is an element other than an AdvP or QP occurring between the subject and verb, or if the head of the subject noun phrase is an empty element (e.g., *con*).

The output of the coding query was then uploaded to R and merged with the PCSC metadata, so that interactions with, e.g., *year* and *gender* could be tested.

4.4.2 Analysis

Three *subject type* groups were investigated (note that the plNP group would be expected to align with non-NSR subjects in a categorical NSR grammar):

NSR subjects:
PRO-1sg, PRO-2p, PRO-pl, and NP-SBJ-NSR dominating the empty element *con*
non-NSR subjects:
PRO-3g, any singular non-pronominal NP-SBJ, and NP-SBJ-NO dominating the empty element *con*
plNP subjects:
plural non-pronominal NP-SBJ, and NP-SBJ-PL dominating the empty element *con*

Any *con* subject counts as a non-adjacent subject, in terms of the *adjacency* constraint. That is, if an NP-SBJ dominating a *con* with an NSR subject referent immediately precedes the finite verb, this clause is counted as a non-adjacent clause. I initially separated 2p pronouns from the main subject type categories, to analyse the pattern of this subject type independently, but noted

5. The coding query, definition file, and query output are in Appendix 2.

that they aligned with NSR subjects in terms of their S-V agreement patterning and thus included them in the NSR subject group for the main study.

Table 4.4 shows the absolute frequencies of each subject type in the data. While 1sg is clearly the most frequent subject type, which could be expected from a correspondence corpus, there are still plenty of occurences of each type of subject. Thus, we can assume that the results for each subject type group represents a relatively balanced sample, and not just the behaviour of one dominating type.

Table 4.4: Absolute frequencies of different subject types included in the study

Subject type	*n*
1sg.	5,083
2p.	979
pl. pronoun	678
**con* w. NSR-subject referent	346
Tot. NSR subjects:	7,086
3sg. pronoun	1,587
NP (sg.)	3,292
plNP (incl. *con* with plNP referent)	488
**con* with non-NSR subject referent	256
Total non-NSR subjects:	5,629

The dependent variable is the presence of verbal inflection on the finite verb in a present tense clause (i.e., the presence of VBP-s, DOP-s, HVP-s, or BEP-s). The frequencies of inflection were calculated in S-V *adjacent* and *non-adjacent* clauses separately, and by *subject type*. Three different methods were used to analyse the proportions of verbal inflection over time. First, the average frequencies were taken per 20-year time intervals, calculating the 95% confidence interval for each sample using the binom package in R (Dorai-Raj 2014) with the Wilson method, which has been shown to perform better than the normal approximation (see e.g. Wallis 2013). This ensures that we get a sense of the data without any parametric assumptions, but binning the data over intervals does not make full use of the variation in the data. Next, to reduce some of the noise and thereby

give a complementary visual representation, I also fit the data to a LOESS curve, with *year* as a continuous rather than binned variable. The LOESS method fits a smooth but flexible curve to the data, and may help to visualise general trends, although it adds a level of abstraction from the raw data.

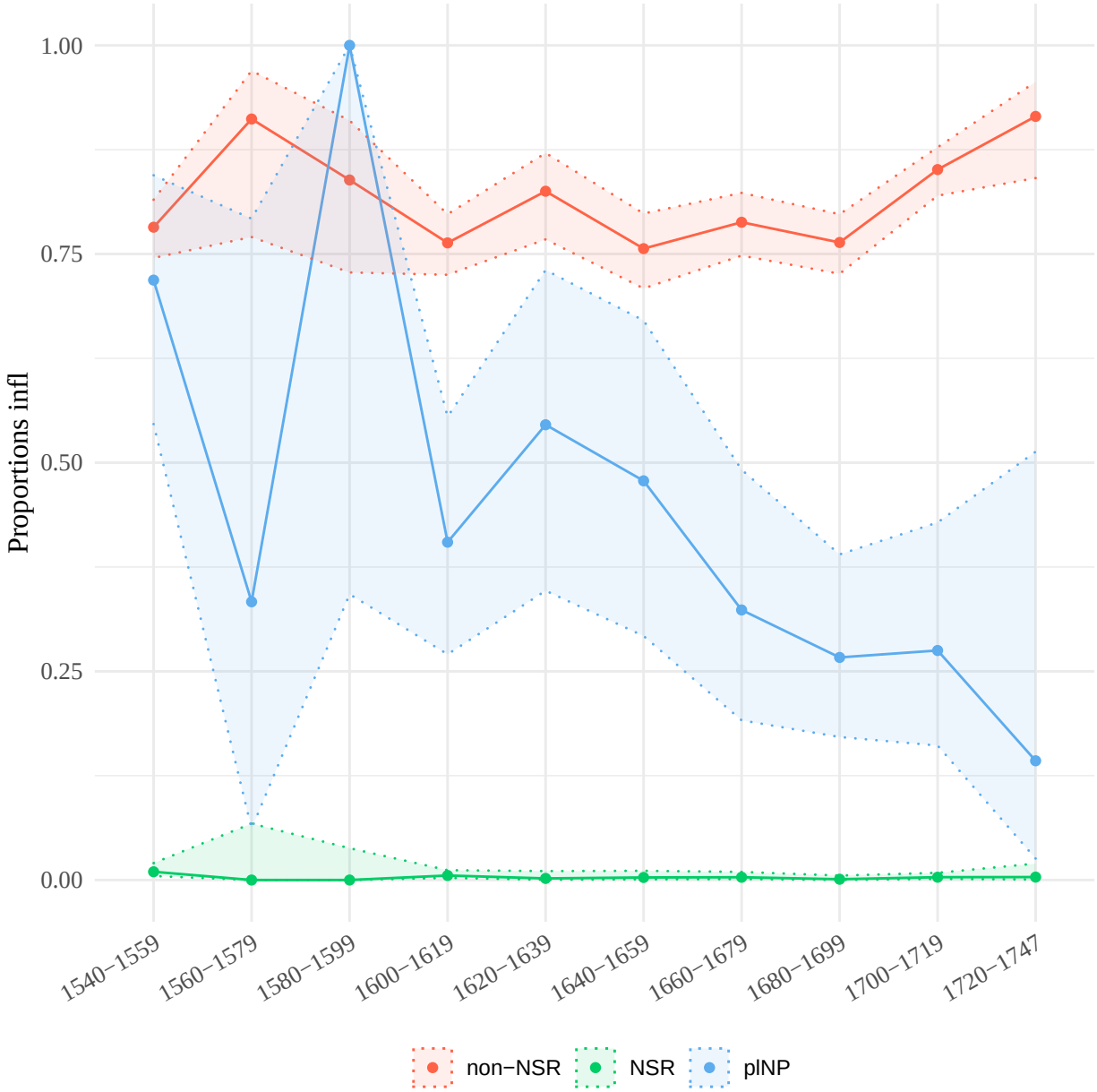
A downside of both the method of binning the data and fitting it to a LOESS curve is that a non-parametric approach does not allow for significance testing of a general time trend. One way to allow for this, is to instead fit a parametric model. It is generally believed that language transition follows an S-shaped curve (e.g. Kroch 1989). Thus, a popular class of model used for analysing language change over time is a logistic regression model, where time is a continuous variable. Fitting a logistic model also allows us to control for additional factors influencing the change over time; this will reduce bias in case the proportions of the additional factor variables change in the data over time. Therefore, to allow for significance testing of the cases where the LOESS and binned plots suggest a change over time, I fitted a mixed effects logistic regression model to the data, with *inflection* as dependent variable, *year*, *subject type*, and *gender* as fixed effects, and *filename* as random effect. I also included an interaction of *subject type* and *year*, but did not include interactions of *subject type* and *year* with *gender*; I omitted the interaction with *gender* as the model would not converge under this specification – when I constructed models which tested the interaction of *year* and *gender* separately for each subject type, there was no significant difference in time trends between genders for any of the subject types. Hence, I inferred that I could remove the year*gender interaction from the baseline model specification. The model was also compared with a simpler model excluding *gender* as a predictive variable, using the ANOVA() function in R, and a significant preference for the more complex model was found.

4.5 Results

4.5.1 Subject-verb adjacent clauses

In S-V adjacent clauses, we would expect an ideal NSR system to have categorical verbal inflection with non-NSR and plNP subjects, and to never have verbal inflection with NSR subjects. Conversely, in a Standard English system, plNP would align with NSR subjects in S-V adjacent clauses, in showing -∅ inflection. Figure 4.2 shows the results where the data is sorted into bins based on 20-year intervals. We indeed find, as expected from both types of s-v agreement systems, very low frequencies of inflection with NSR-subjects (1sg and all plural pronouns). The non-NSR subjects appear with inflected verbs at high and fairly stable frequencies, but not as categorically as both an NSR and StE system would predict; the proportions range between 75-91%.

Figure 4.2: Frequency of inflection in S-V adjacent clauses (e.g. 'the girls dance(s)'), by subject type, where non-NSR subjects exclude plNP.



There are, naturally, fewer observations of plNP subjects than the other groups; 265 observations, out of which 105 show up with verbal inflection (cf. 2,684/3,375 for non-NSR subjects and 10/6,587 for NSR subjects). Furthermore, the sample sizes in the 1560-1579 and 1580-1599 bins only consist of 3 and 2 observations respectively, so should not be treated as reliable[6]. In (18-a), repeated from (14-a), we see a clause compatible with the NSR pattern, and, in (18-b), a clause compatible with StE agreement.

(18) a. for **all his freynd*is%** thynk*is% it suld be sa

(PCSC ID: 408_64_f1540; Hew Campbell, 1548)

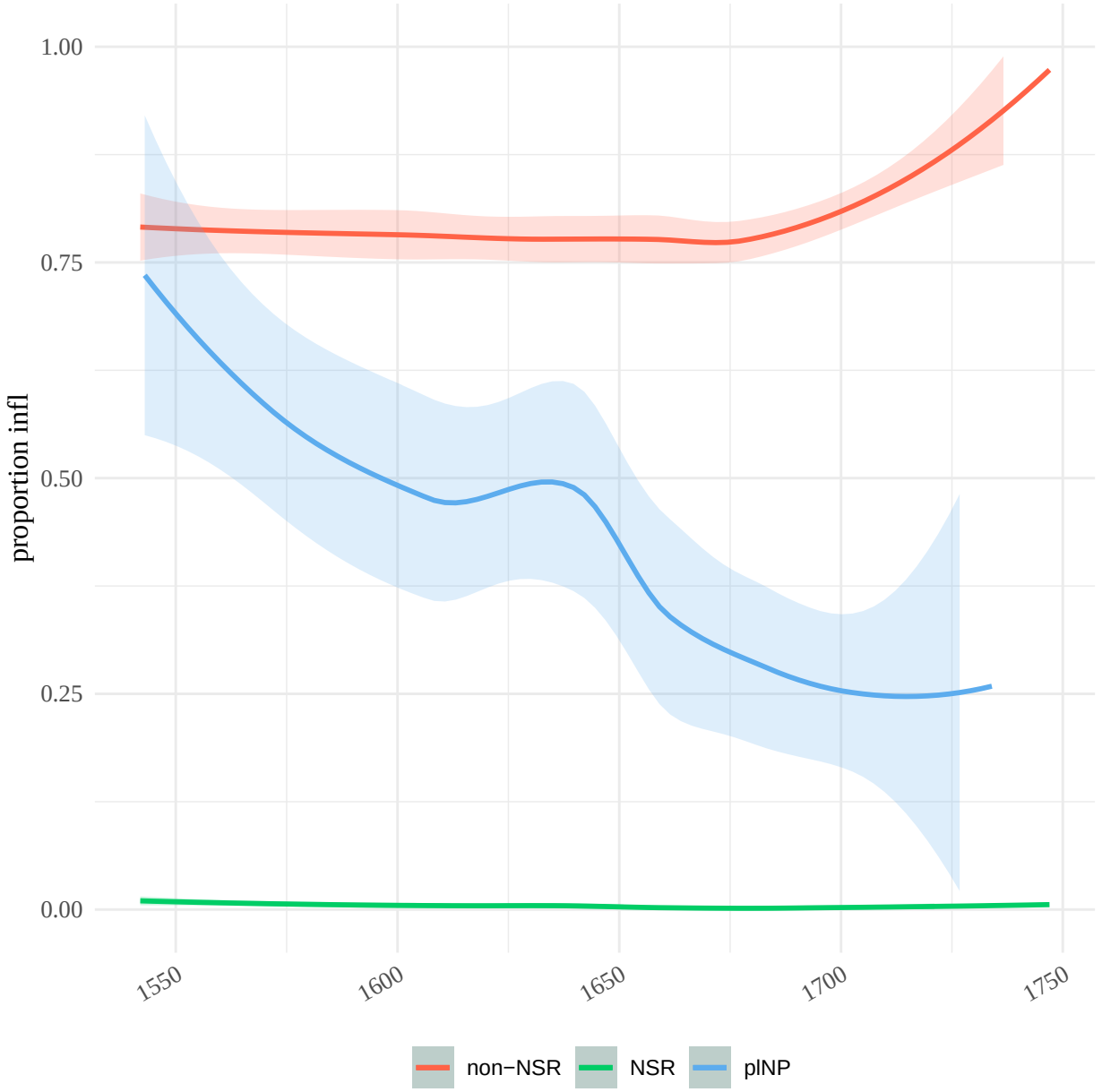
b. I am content yat **freind*is%** hear me and him Both

(PCSC ID: 277_138_f1650; Margaret Sinclair, 1652)

However, despite the noise caused by the small sample sizes before 1600, we can still see a trend of decline of plNP subjects appearing with inflection, suggesting a move from an NSR system to a StE system. This trend is clearer when visualised with a LOESS curve, as in Figure 4.3. Estimates based on the mixed effects model show that the decline of plNP subjects with verbal inflection is significant ($p < 0.001$).

6. This gap coincides with the one identified in the PCSC word distribution over time in Chapter 3.2.2.

Figure 4.3: Frequency of inflection in S-V adjacent clauses (e.g. 'the girls dance(s)'), by subject type, where non-NSR subjects exclude plNP (LOESS curve).

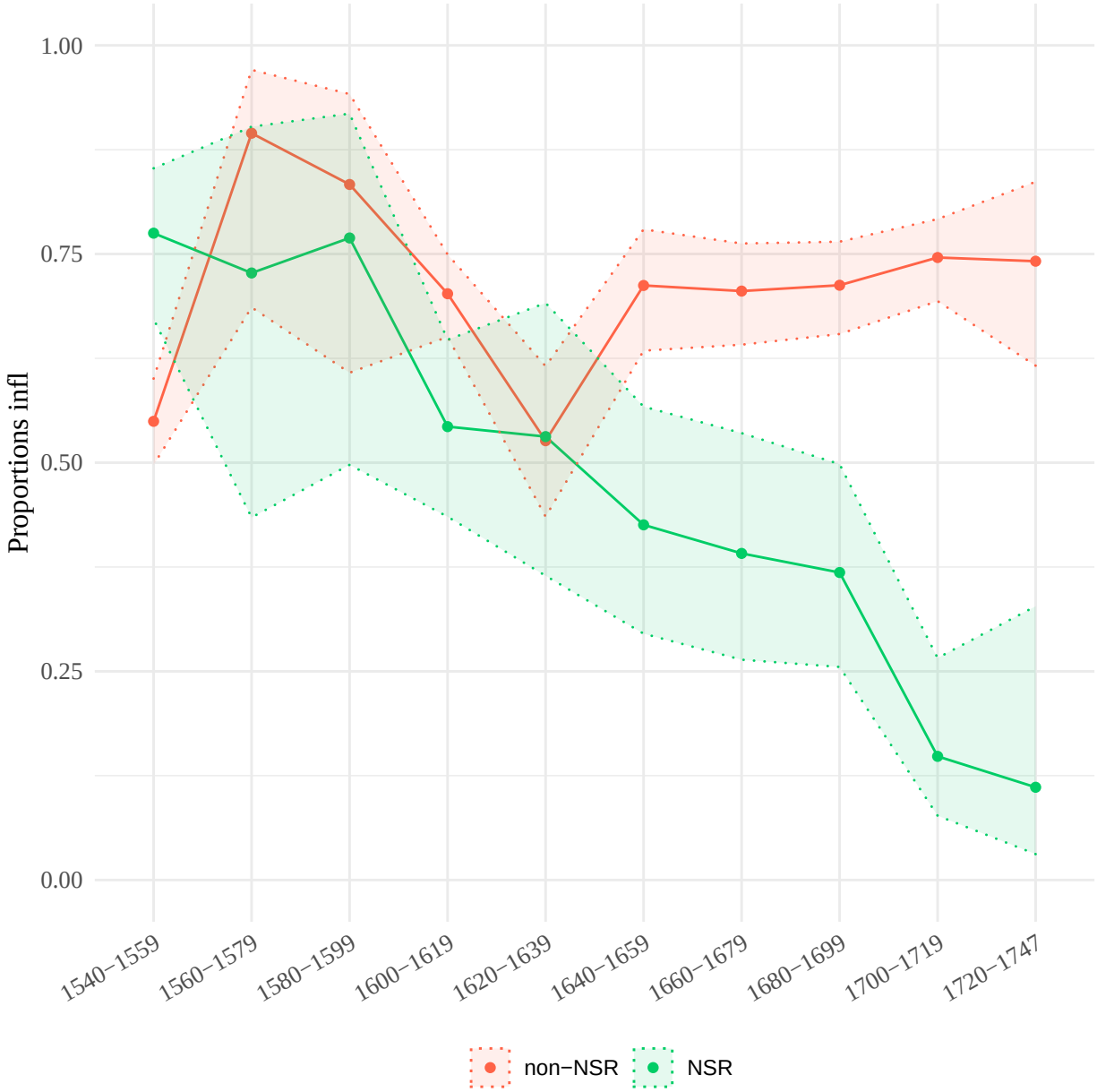


The LOESS curve also gives a clearer image of the trend with non-NSR subjects, where the frequency of inflection is stable for most of the period and then increases after 1675. The results from the mixed effects model show that there is not a significant 'S-shaped' increase in inflection with 3sg subjects when testing over the whole period ($p = 0.0513$), but in a subset of the data from 1650 to 1750 the test confirms that the increase is significant at the end of the period ($p < 0.001$).

4.5.2 Non-adjacent clauses

In an NSR system, all verbs in clauses with "heavy" elements (not AdvP or QP) between the subject and the verb, and all verbs in clauses with elided subjects, should be inflected. In a StE system, verbs are only inflected with 3sg subjects (= *non-NSR* subjects) regardless of adjacency. Thus, both NSR-subject clauses and plNP-subject clauses are of interest here; if they occur with verbal inflection, it is indicative of an NSR system, and if they do not, it is indicative of a Standard English system. The results (figure 4.4) show a clear decline of NSR subjects appearing with inflected verbs in non-adjacent contexts, while non-NSR subject clauses remain at high, but non-categorical, frequencies, just as in S-V adjacent contexts but with apparently more variation and a dip in frequency in the early seventeenth century.

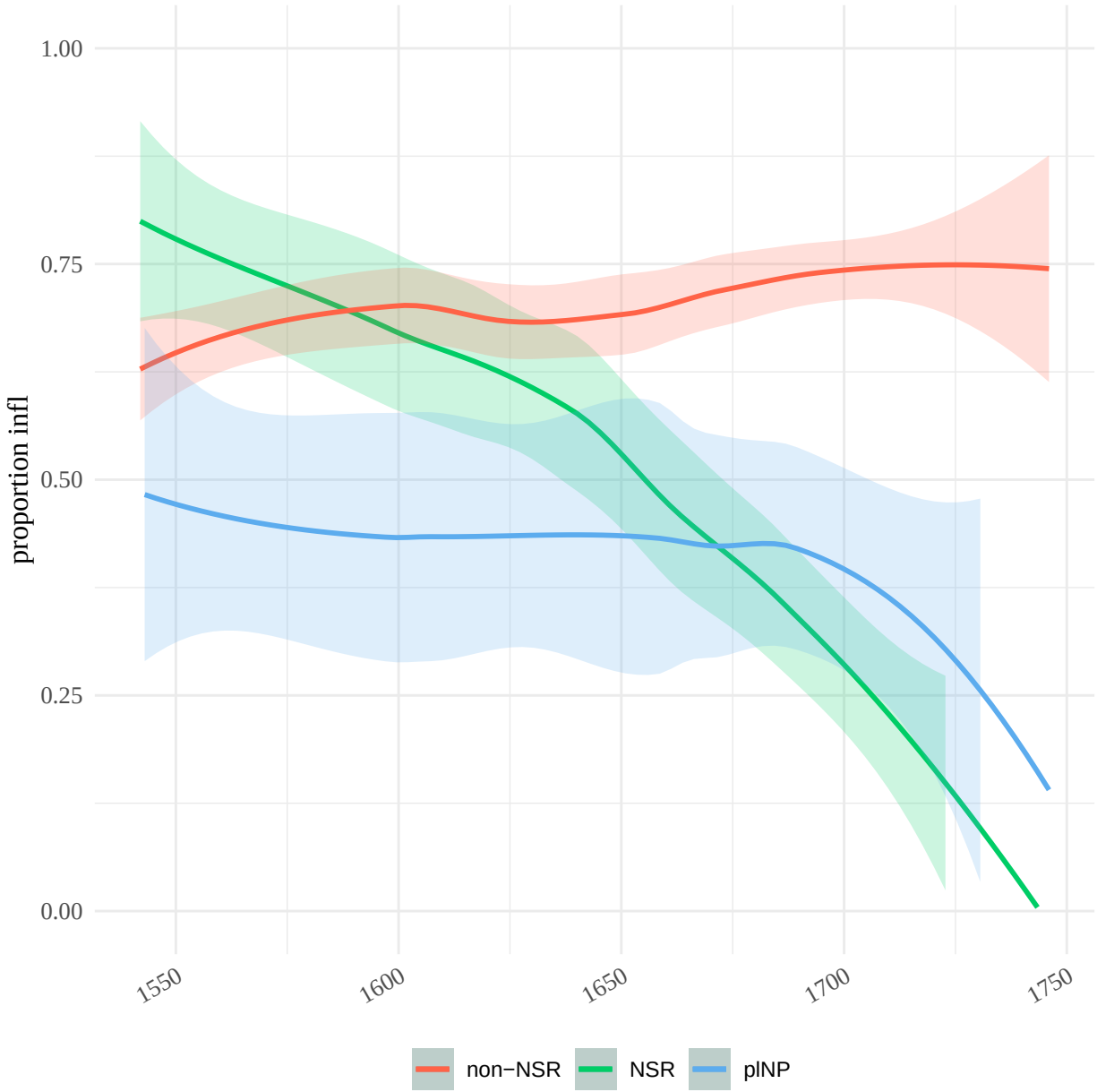
Figure 4.4: Frequency of inflection in non-adjacent clauses (e.g. 'the girls, while dancing furiously, sing(is)'), by subject type, where non-NSR subjects exclude plNP.



I excluded plNP from the graph in Figure 4.4, as the sample sizes per time period were too small to analyse the pattern over time accurately; aggregating the data over the entire time period, the overall proportions of plNP subjects appearing with verbal inflection in non-adjacent clauses were lower than expected for an NSR grammar, only at 42% (76/181), and significantly lower ($p < 0.001$ in a two-sample t-test) than the same proportion for other non-NSR subjects (1217/1801 =

68%). The fact that they appear with inflection on the agreeing verb at all is unexpected for a StE grammar. When visualising the same data in a LOESS curve, the variation in plNP-subject clauses seems more stable in non-adjacent clauses than what was observed for S-V adjacent clauses, but the mixed effects model shows that the decline of plNP subjects occurring with inflected verbs over time is significant when estimating the model across the whole time period ($p < 0.05$).

Figure 4.5: Frequency of inflection in non-adjacent clauses (e.g. 'the girls dance(s)'), by subject type, where non-NSR subjects exclude plNP (LOESS curve).



In (19), repeated from (15), we see examples of clauses with the expected NSR pattern, and, in (20), examples of clauses not adhering to the NSR *adjacency* constraint (subjects marked in bold font, agreeing non-adjacent verb underlined).

(19) a. **I** being young and hauing little experience knows little how to doe in so criticall a time

(PCSC ID: 1384_18_M1700; John Murray, 1700)

b. **I** se na help bot be zowr grace & hoppis na vdyr [..]

(PCSC ID: 436_92_M1540; Alexander Gordon, 1549)

(20) a. **they** that hes no vit nor gret moyen get pensions

(PCSC ID: 78_60_F1600; Agnes Keit, 1600)

b. **I** follow my trade as closs as I can here, & hope ere Long, to come home & put you to your pro\mise

(PCSC ID: 1155_50_M1650; Patrick Hume, 1687)

In (20-a) we see an example of the NSR being violated with the matrix verb, but note that the finite verb in the relative clause, *have*, has the s-inflected form. In an NSR system, relative clauses typically favour -s inflection with all subjects, regardless of whether the relative pronoun is overt or omitted (Pietsch 2005b: 8). Relative clauses where the subject is gapped were not included in the study, as the only subject-less clauses tested were conjoined clauses.

The mixed effects model also shows that the decline of NSR subjects occurring with verbal inflection is highly significant ($p < 0.001$). The LOESS curve (figure 4.5) gives a different picture of the variation in non-NSR-subject clauses than what was suggested in Figure 4.4; the dip in the early seventeenth century is barely present, and the use of non-NSR subjects occurring with verbal inflection seems to increase over time. This increase is, in fact, also significant according to the

mixed effects model ($p < 0.001$). The overall proportion of inflected verbs with non-NSR subjects in non-adjacent clauses (68%) is significantly lower ($p < 0.001$) than the same proportion in S-V adjacent clauses (79.5%), suggesting more variation in the agreement pattern of non-adjacent clauses.

Weak-adjacent clauses

Finally, I also tested clauses with "weak" S-V adjacency (i.e. adjacency is only interrupted by an AdvP or QP). Examples of this clause type are given in (21), with a simple adverb (21-a) and a stranded quantifier (21-b) interrupting adjacency.

(21) a. **they** probably exceed the power of common imagination:

(PCSC ID: 1464_98_M1700; James Maule, 1733)

b. **ve** al luke for sum ass\surance from zou

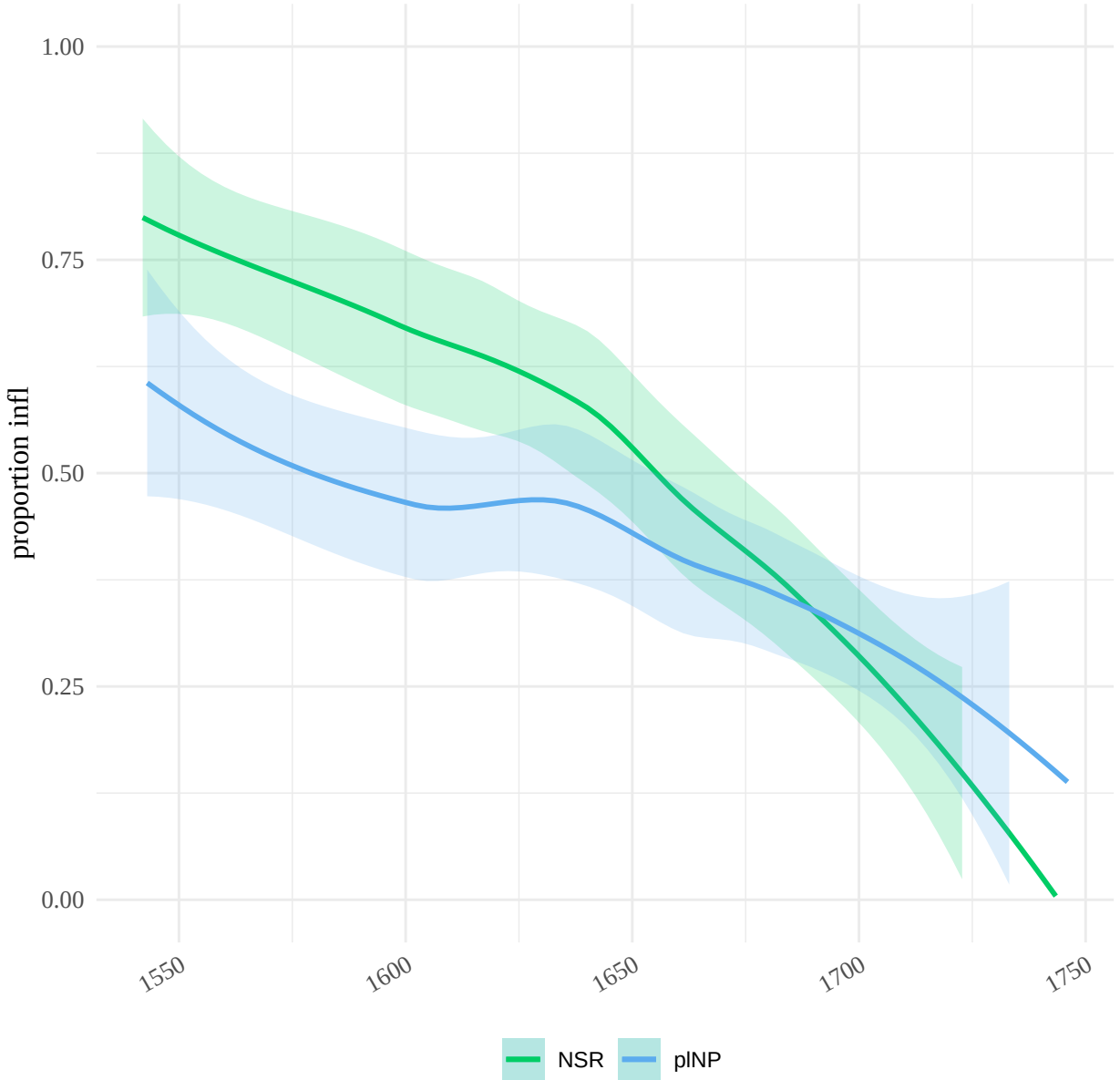
(PCSC ID: 75_57_F1600; Agnes Keith, 1600)

Due to a small sample size, a fine-grained comparison over time was not possible for this clause type. Comparing the average use of inflection in the earlier half of the time period to the later half also did not yield a significant result. Aggregating the data over the entire time period, we see non-NSR subjects appearing with inflected finite verbs 76 times out of 106 occurrences (71.7%), while the proportion for NSR subjects is 2/105 (2%), and plNP subjects occurred 16 out of 38 times (42%) with an inflected verb. Thus, the pattern in weak-adjacent clauses shows similar overall proportions of inflection as in S-V adjacent and non-adjacent clauses for each subject type, and we can draw no conclusion regarding the "weak" status of these adverbs and stranded quantifiers compared to other types of intervening elements.

Interim summary

Figure 4.6 shows the decline of verbal inflection in favour of -∅ inflection with plNP subjects and non-adjacent NSR subjects – the critical environments where the presence of inflection indicates an NSR grammar. Thus, it can be seen that the NSR pattern declines during the Scots Transition period, and it becomes replaced by a grammar more similar to StE present tense agreement.

Figure 4.6: Frequency of inflection with NSR subjects in non-adjacent clauses (e.g. 'we sing and dance(s)'), and plNP subjects in both clause types. (LOESS curve)

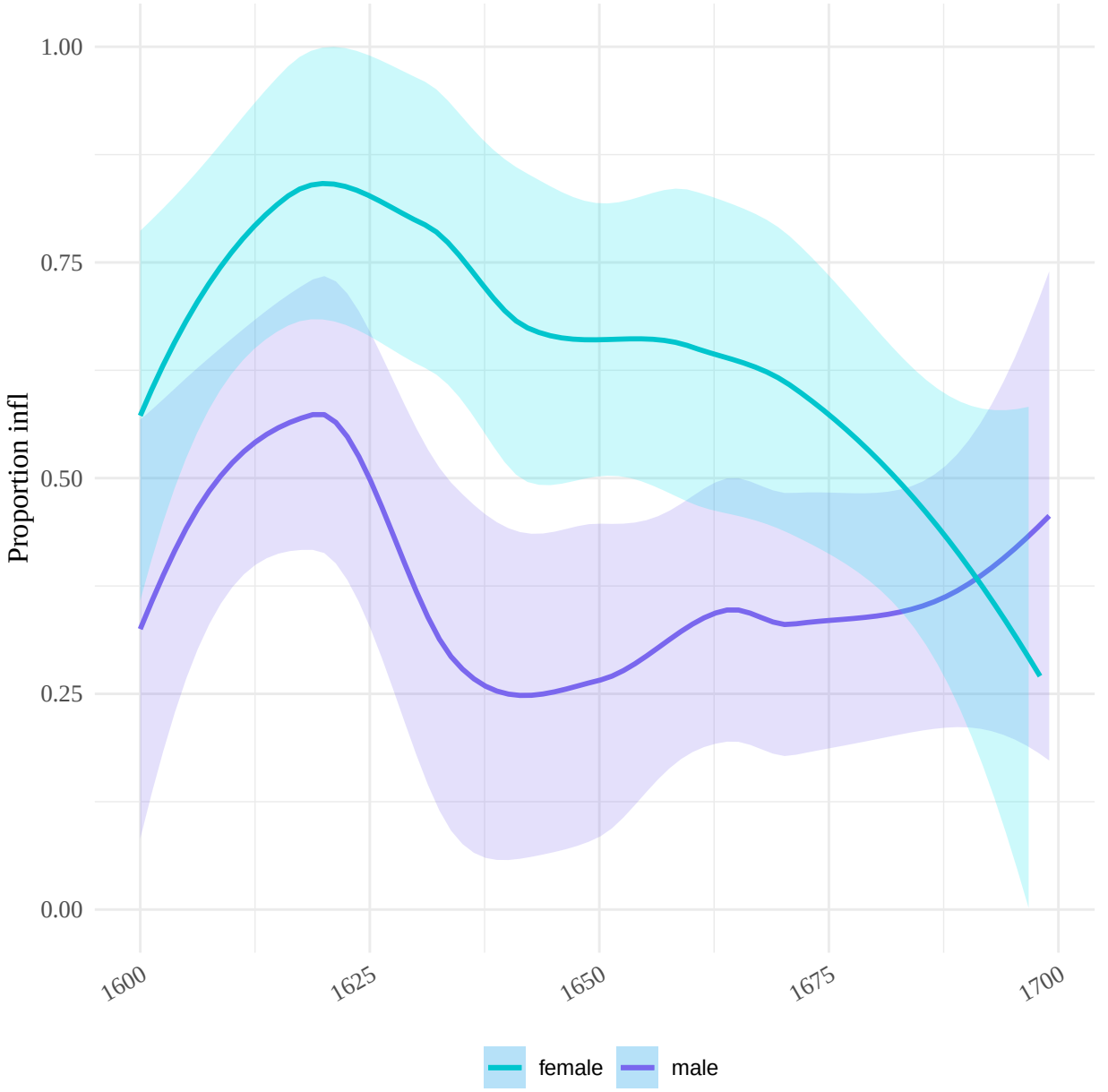


As regards the strengths of the *adjacency* and *subject* type constraints, there is no indication that the *adjacency* effect is weakening more rapidly or earlier than the *subject type* effect; plNP subjects, which are not concerned with S-V adjacency in their inflectional patterning, favour -∅ inflection at similar rates to non-adjacent NSR subjects.

4.5.3 Gender

The interaction *year*gender* was not statistically significant (see 4.4.2), and, thus, the effect of *gender* appears to be stable over time. However, there is a significant main effect of gender ($p < 0.01$) in the overall use of inflection in non-adjacent clauses between male and female writers, whereby women use inflection more than men on average. To analyse this difference, and test prediction B (section 4.3) regarding women's preference for Scots features, I investigated data from the seventeenth century (as that is when the corpus is most balanced for gender (see Chapter 3.2.3)) and tested the difference in frequency of inflection with non-adjacent NSR subjects. More use of inflection with non-adjacent NSR subjects would indicate a more NSR-like grammar. The results are in Figure 4.7.

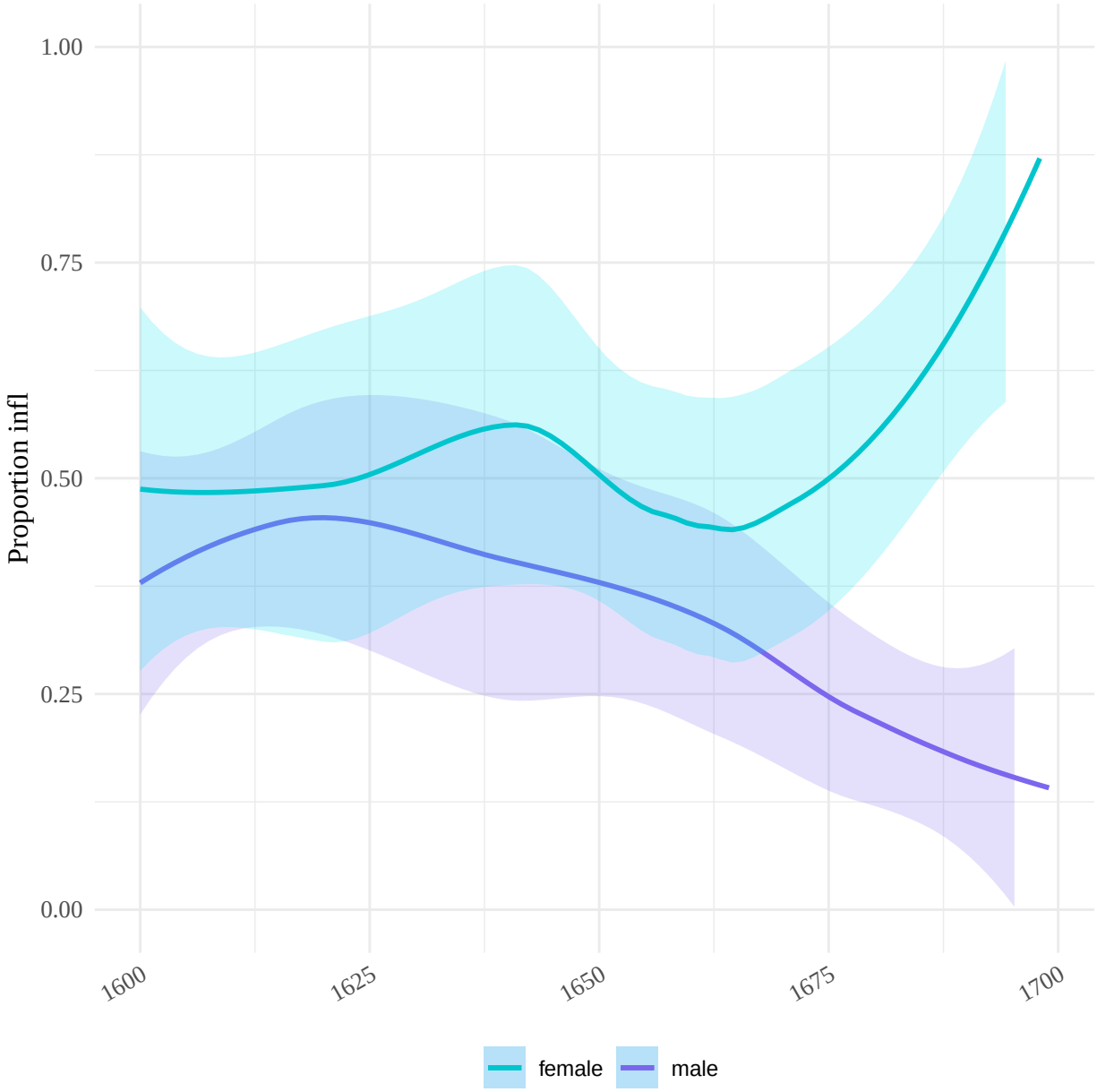
Figure 4.7: Frequency of inflection with NSR subjects in non-adjacent clauses (e.g. 'we sing and dance(s)'), by writer gender. (LOESS curve)



For most of the period, female writers appear to be more conservative in their use of verbal inflection with NSR subjects, but they converge with the male writers towards the end of the period (the fact that the line representing the male writer data turns upwards in the end is slightly misleading, as there is a trend of decline in the use of inflection with NSR subjects for both writer groups after 1700). Finally, I also tested the frequency of inflection with plNP subjects, regardless of adjacency.

With these subjects, a similar pattern would be expected; female writers using more inflection overall, but converging with male writers at the end of the period. However, the results in Figure 4.8 show a different trend.

Figure 4.8: Frequency of inflection with plNP subjects by writer gender. (LOESS curve)



Interestingly, the frequency of inflection with plNP subjects increases significantly with female writers at the end of the period, whereas it steadily declines with male writers. Here, we see a move from fairly similar usage in the beginning of the period, towards divergence at the end, with women favouring an NSR pattern and men moving towards a StE pattern. Thus, surprisingly, plNP and non-adjacent NSR subjects do not show the same pattern of development over time between female and male writers in the PCSC.

4.6 Discussion and conclusion

The picture emerging from the findings presented in this chapter is one where the NSR pattern appears to be in competition with StE agreement throughout the period under investigation, and that StE eventually ousts the NSR. Thus, an anglicisation effect is suggested here, and prediction A holds: English influence in the Transition period led to a syntactic change in Scots. The gender differences between the non-adjacent NSR contexts and the plNP contexts remain mysterious. However, both results suggest, in different ways, that women indeed tend to use more Scots features in their writing, thus (tentatively) supporting prediction B: Female writers use higher frequencies of NSR-pattern constructions, and the change from NSR to Standard English agreement is led by the male writers. Further investigation should seek to uncover what may have caused the change in the linguistic behaviour of women during the seventeenth century, e.g., whether there might be a change in education level which would have given women access to a new linguistic repertoire.

An unexpected finding was that inflection with 3sg subjects was not categorical. It may be that these numbers are affected by the subjunctive verbs in the corpus, which do not take inflection with any subject as they are not distinguished from regular present tense verbs in the part-of-speech tag system (in accordance with PPCHE conventions). We could also perhaps understand this as expected variation resulting from an unstable, changing, grammar; during times of change, sociolinguistic factors (such as stigmatisation of variables, or hypercorrection) may influence the trajectory of a variable, as is, e.g., demonstrated by Warner (2005) for the sudden drop in the rise of negative declarative *do*-support in the sixteenth century. However, this does not account for why S-V adjacent and non-adjacent clauses should behave differently in 3sg contexts, nor why there was

a significant increase in the use of verbal inflection with 3sg subjects in non-adjacent clauses, as well as at the end of the period for S-V adjacent clauses. Another theory which could be explored further to understand this variation, but should for the moment be taken as speculative, is that this is a result of a time lag between weakening of NSR constraints and full adoption of StE agreement, under the assumption that *-s* inflection is a sign of failed agreement (in line with Henry (1995) and de Haas (2011)); if S-V adjacency is no longer a requirement for agreement, then we could expect a period of increase of ∅-inflections in non-adjacent clauses to mark successful agreement. As regards the *subject type* condition, it may be that the weakening of the *adjacency* constraint coincides with a loss of the separate subject position system, which took place earlier in Southern English (in the ME period, e.g., Biberauer and van Kemenade 2011), possibly leading to confusion in the paradigm which would also affect 3sg subjects. This confusion would make the Scots s-v agreement vulnerable for influence from an incoming robust system, such as StE agreement – and, indeed, we do see a move towards StE agreement in later part of the period, as inflection increases over time with 3sg subjects. Another possibility for what accounts for this, what has hitherto been referred to as, confusion in the paradigm, is that the possibility for 3sg subject to occur with ∅-inflection may indicate the presence of another type of grammar in competition, alongside the NSR and StE. That is, a S-V agreement system similar to that of, e.g., Swedish, without person and number distinctions, or subject type/adjacency constraints, wherein the default verbal marking is ∅. For now, these speculations are left to be investigated in future research.

Chapter 5

The rise of Scots *do*-support

5.1 Introduction

This chapter explores the rise of *do*-support in Scots, and investigates the nature of Scots *do*-support during its emergence; whether the feature showed similar functions to English 'intermediate *do*' (as analysed by Ecay (2015)) before grammaticalising into its current function, or whether the feature interacts with the *Northern Subject Rule* (NSR), which is suggested in an analysis by de Haas (2011). *Do*-support is the mandatory insertion of the auxiliary *do*, which has historically been bleached of its semantic meaning and today functions as a 'dummy' element with a strictly morpho-syntactic function: carrying tense and agreement features. The environments where *do*-support occurs in Present-Day English (PDE) and Scots are clauses with NICE properties: Negation (22), Inversion (23), Code (that is, in constructions which signify, or 'code', another verb phrase, as in (24)), and Emphasis (25).

(22) a. *Negative declarative*: I **do**n't eat cake
b. *Negative imperative*: **Do**n't eat cake!

(23) a. *Affirmative/negative polar question*: **Do**/**do**n't you eat cake?
b. *Wh-question*: What **do** you eat?

(24) a. *Ellipsis*: I eat cake, and Alex **does** too
b. *Tag question*: You eat cake, **do**n't you?

(25) *Emphasis*: I **do** eat cake

English *do*-support is extensively researched, with quantitative accounts mapping the emergence and development of the feature as early as that of Ellegård (1953). Since then, various theories have emerged which aim to account for the emergence of the dummy auxiliary in English, the quantitative and qualitative process of grammaticalisation of the auxiliary, and why it became mandatory in this way in English alone of the Germanic languages (Denison 1985; Ecay 2015; Garrett 1998; Kroch 1989; Poussa 1990; Tieken-Boon van Ostade 1990; van der Auwera and Genee 2002; Warner 2002, to name a selection of notable contributions). The feature in Scots has received relatively less attention; Scots *do*-support emerged during the Transition period, and, while the feature is mostly grammaticalised, it has remained variable in some Scottish dialects until the present day (Gotthard 2019; Jamieson 2015; Meurman-Solin 1993a; Smith 2000).

This chapter is organised as follows: First, I establish my theoretical assumptions for the development and operation of *do*-support in English (5.2.1), and then summarise the main scholarly theories for its origin and subsequent development in English (5.2.2), as well as what is known regarding the feature's emergence in Scots (5.2.3). The background to the hypothesis that *do*-support in Scots may be conditioned by NSR constraints is given in Section 5.5.3. The research questions are given in Section 5.3, where predictions are made regarding the role of women in the regulation of *do*. The methodology, detailing the process in extracting relevant clauses conditioned by NSR and intermediate *do* constraints, is in Section 5.4. Four sets of results are presented in Section 5.5: first, the overall proportions of *do*-support over time (5.5.1); then results relating to intermediate *do* (5.5.2) and NSR constraints on *do* (5.5.3) are presented and summarised; finally, results for the rise of *do* stratified by gender are presented in Section 5.5.4. The findings, as discussed in Section 5.6, indicate that (i)*do* in the PCSC exhibit similar behaviour to intermediate (pre-1575) *do* in English, (ii) NSR constraints do not seem to have an effect on the environments where *do* emerges, and, when it appears, the auxiliary *do* slots into the NSR pattern similar to lexical verbs, and (ii) female writers are leaders in the rise of *do* after 1675.

5.2 Background

5.2.1 *Do*-support in English: Theoretical assumptions

Do-support is the name of a syntactic phenomenon involving insertion of the auxiliary *do* in a clause in order to pronounce tense and agreement features when the transfer of these features between the subject and main verb is interrupted and no other auxiliary is present. In the generative framework, *do* merges in a position typically analysed as the head of a functional projection located below C but higher than VP in the syntactic structure, e.g. T(ense) or AGR(eement), called I(nflection) in the days before Pollock's (1989) Split IP-Hypothesis. *Do*-support is often described as a *last-resort operation* to establish subject-verb agreement, which became required after the loss of *verb-raising*, i.e., main verb movement from VP to such a higher position. As much of the work on the rise of *do*-support in English was carried out in the 1980s, before Pollock's innovation, this movement is traditionally referred to as V-to-I movement. The rationale for V-to-I movement in syntactic theorising is usually presented as a need to express the feature-content of the syntactic head I to agree with the subject in Spec,IP. After the loss of verb-raising, this operation became restricted to auxiliaries only; *do* is in complementary distribution with other auxiliaries (except for in imperative clauses, where *do* can co-occur with progressive and passive *be*, and perfective *have* (see, e.g., Ecay 2015: 5)).

Thus, after the loss of verb-raising, lexical verbs could only achieve subject-verb agreement if there was no intervening material between I and V, which necessitated the insertion of *do* in those clauses. Examples a. in (26) and (27) demonstrate what V-to-I would surface as in modern English (which is now, of course, ungrammatical), the failed agreement resulting from loss of verb-raising is in b.; in questions like (26), I with its features has moved above the subject, to the head of C, which means that the subject itself becomes an intervening element for the transfer of agreement features. In a negative declarative clause like (27) the negator *not* is the intervening element. In c., we see the grammatical clause with *do*-insertion.

(26) a. *what$_k$ eats$_i$-*3sg*$_j$ Alex t$_j$ t$_i$ t$_k$?

b. *what$_k$ Ø-*3sg*$_j$ Alex t$_j$ eat t$_k$?

c. what$_k$ does-*3sg*$_j$ Alex t$_j$ eat t$_k$?

(27) a. *Alex eats$_i$-*3sg* not t$_i$ cake

b. *Alex Ø-*3sg* not eat cake

c. Alex does-*3sg* not eat cake

In PDE and Scots, the auxiliary itself has no meaning, and no other purpose than to spell out features in the manner described – for this reason, the auxiliary is sometimes referred to as a 'dummy' auxiliary.

5.2.2 *Do*-support in English: its origin and development

The origin of *do*

Before it grammaticalised into the function we recognise in PDE, the auxiliary *do* had other functions, as illustrated in the Old English (OE) examples in (28); causative *do* (28-a), anticipative *do* (28-b), substitute *do* (28-c).

(28) a. ...and **de**þ hi sittan, and he gæþ sylf and hym þenað
...and does them sit, and he goes self and them serves
'. . . and makes them sit down, and goes himself and serves them'

b. ...ac utan **don** swa us þearf is, gelæstan hit georne
...but let-us do as us need is, perform it carefully
'. . . but let us do as we should, i.e. carry it out with care'

c. and hit þær forbærnð þæt mancyn, swa hit her ær **dyde**.
and it there burns-to-death that people, as it here before did
'and it will burn those people to death, as it has done here before'

(*Sleg.* (Ld) 4.118; from Fischer and van der Wurff 2006: 154)

The latter two, anticipative *do* and (28-b), substitute *do* (28-c), are still functions of *do* in PDE (*code*, under the NICE properties). Causative *do*, as in (28-a), is used in other modern West-Germanic varieties, e.g. Dutch (van der Horst 1998: 57), but the causative function of *do* is lost in PDE – it is precisely this early function of the *do* auxiliary which commonly has been identified

as the most likely source construction for the grammaticalisation of *do*, initially by Ellegård (1953) and by many scholars since. While the presence of the more semantically vacuous *do* auxiliaries with *code* properties may have also influenced the development of a dummy *do* auxiliary, the causative construction has the advantage of combining a *do* auxiliary with an infinitival verb, which is the same syntactic pattern of PDE clauses with *do* (except for in *code* contexts; e.g., Fischer and van der Wurff 2006: 154-155; Garrett 1998: 287-291).

In Ellegård's (1953) early proposal for the grammaticalisation of causative *do*, he suggests that the meaning of *do* was vulnerable for re-analysis in clauses where the agent is ambiguous, as in the Middle English (ME) clause in (29). Here, the main verb has its own causative meaning, which could lead to an interpretation like (29-b), rendering the *do* auxiliary redundant. Eventually, *do* auxiliaries with unambiguously non-causative meaning started to emerge (as in (30)), with the earliest examples from the western Midlands in the ME period.

(29) Whan he was at London, a haule he did vp wright.

(?a1400 (1338) Mannyng Chron Pt2 (Petyt) 88/17)

a. *Indirect agent (causative* do): When he was in London, he had a hall built.
b. *Direct agent (re-analysis)*: When he was in London, he built a hall.

(Adapted from Garrett 1998: 287)

(30) Hir self for sorow dide scho slo
'In sorrow she slew herself

(?a1400 (a1338) Mannyng *Chron. Pt.1* (Petyt) 2544; from Garrett 1998: 288)

The grammaticalisation of *do* is not an isolated syntactic change; as also mentioned in Chapter 3.3.1, the reanalysis of *do* and its subsequent spread through various syntactic contexts in the Early Modern period was part of the wider shift of modal verbs into their own auxiliary category, facilitated by the move towards a more fixed SVO word order (e.g. Görlach 1991: 113-114; Fischer and van der Wurff 2006: 156-8; Rissanen 2008: 210). Neither is it likely that this is solely a language-internal change, as both dialect- and language contact may have played a role in the emergence of dummy *do*. As causative *do* was more common in the east, and semantically vacuous *do* emerges in the west where causative *make* was more prevalent, it has been suggested that the re-analysis of causative *do* happened when Western ME speakers re-interpreted the use of causative *do* by Eastern ME speakers (Ecay 2015: 74). The rise of the feature in the West Midlands has also given rise to theories of Celtic substratum effects from contact with a Brythonic ancestor of Welsh in the OE period, most notably argued by Poussa (1990), which is evidenced by the use of periphrastic *do* constructions in Celtic languages, as seen in Middle Welsh (31).

(31) Mynet a **oruc** Padric y Iwerdon
go(VN) PTCL **do**.PRET.3SG Patrick to Ireland
'Patrick went to Ireland'

(adapted from Garrett 1998: 285)

Some scepticism to theories of OE-Celtic contact origins of *do* arise from that they depend on the assumption that some form of *do*-support was in use in the West Midlands but unattested before the ME period, which is difficult to prove and causes the theory to become less reliable (e.g. Garrett 1998: 286). However, as is noted by Hickey (2012: 501-3), we cannot assume that this is a one-directional transfer, and it is possible that the *do* constructions in early Celtic and English dialects arose from mutual influence, thus being an areal feature, particularly since there are *do* constructions in other Germanic languages which were not in contact with Celtic. Indeed, the strength of Poussa's (1990) account lies in what she calls a "universalist dummy aux innovation theory" (Poussa 1990: 415). The theory states that situations of language contact would lead to more frequent use of dummy auxiliaries, such as *do*, witnessed in data from English pidgins, creoles

and L2 varieties where *do*-support occurs more frequently than in standard varieties, and supported by the fact that *do*-support occurs more frequently in code-mixing, as well as in children's speech (see also Tieken-Boon van Ostade 1990: 19-24; van der Auwera and Genee 2002: 287; Hickey 2012: 502; this is often the function of *do* periphrasis in other Germanic languages). This theory aligns with suggestions that *do* was initially more frequent with verbs borrowed from French, to aid speakers in not having to fit new verbs into their native inflectional system by pronouncing inflectional feature on the *do* auxiliary instead (e.g. Fischer and van der Wurff 2006: 155), and finds further support in a more recent study by Shaw and de Smet (2022), who find a strong preference for non-finiteness in the use of borrowed verbs from French in ME.

The question of the origin of the dummy auxiliary *do* is still unsolved. There do seem to be parallels between the origin of *do* and the rise of the NSR pattern (see 4.2.1), in that the origin may not be solely do to language-internal or external circumstances. A *do* auxiliary was already present in OE, and its presence in other Germanic languages suggests a common Germanic origin (c.f. Tieken-Boon van Ostade 1990). This auxiliary spread to more environments and became bleached of its semantic meaning in varieties spoken in Britain only, a process which apparently began in an area with a history of contact with Celtic languages which had verbal periphrasis with *do*, i.e. the West and South-West of England. Without accepting an origin of auxiliary *do* structures as direct transfers or calques into English, it is still feasible that this unique contact situation could accelerate or catalyse a different development of *do* than what we see in other Germanic languages. This does not rule out that the change could also be facilitated by influence from French constructions with *faire*, as suggested by Ellegård (1953) and Denison (1985, 1993). Likewise, the most significant change which made the grammaticalisation of *do* possible, the loss of verb-raising, is argued to have been brought on by contact with Old Norse (e.g., Kroch, Taylor, and Ringe 2000). It seems counter-intuitive to assume that these Britain-specific contact situations would have had no effect on this emerging structure, even if the origin of the pattern is language-internal.

The development of *do*-support in English

Thanks to the extensive data collection carried out on *do*-support in English, it has been possible to reliably map its spread through the grammar quantitatively. Figure 5.1 illustrates the spread of *do*-support in different English clause types, starting from when the non-causative *do* construction starts to appear, using data from Ellegård (1953). As seen in Figure 5.1, it was for a period possible to apply *do*-support in affirmative declarative clauses, which is analysed a case of non-emphatic, semantically vacuous *do* and not the emphatic function of *do* recognised from PDE. Despite being the most common type of *do* in absolute frequencies in the early days of *do*-support, this affirmative declarative *do*-support began to decline while *do*-support increased in other contexts; the rise and fall of affirmative declarative *do* has been deemed a "failed change" which fuelled the increase of *do*-support in other constructions (Postma 2010).

Figure 5.1: The spread of English *do*-support

AffDecl=Affirmative Declarative; NegDecl = Negative declarative; AffQ = Affirmative Question; NegQ = Negative Question; NegImp = Negative Imperative

(Visualisation adapted from Los (2015: 111))

Kroch (1989: 219) finds that *do*-support was introduced into all appropriate contexts in English at the same time, but in different frequencies, i.e., a *simultaneous unequal actuation*, but that the rate of change from the actuation is the same across all contexts. In this way, the rise of English *do*-support supports Kroch's (1989) Constant Rate Hypothesis (CRH): "[...] when one grammatical option replaces another with which it is in competition across a set of linguistic contexts, the rate of replacement, properly measured, is the same in all of them. The contexts generally differ from one another at each period in the degree to which they favor the spreading form, but they do not differ in the rate at which the form spreads" (Kroch 1989: 200). Differences in frequencies at a given time during the change may be affected by stylistic or functional factors, but, as predicted by the CRH, the rate of change typically follows an S-curve.

During the period of variability in its development, writers converged on various extra-syntactic functions for the use of *do*, i.e., as peak marker (Stein 1985: 292-6), highlighting that the peak of a narrative, its central reportable event, is being reached, or to facilitate rhyme and metre in poetry (e.g., Tieken-Boon van Ostade 1990: 26-27). Ultimately, of course, *do*-support was integrated into Standard English grammar as a syntactic rule, but its regulation continued to be conditioned by social factors throughout the period of variability. Warner (2002: 237-9) notes that writer age is a significant variable in the early use of *do*, although he finds no evidence of generational change, but of age-grading, with writers reverting to not using *do* as much when they get older – I will return to these results. In a study on *do*-support in the *Corpus of Early English Correspondence* (CEEC; Nevalainen et al. 1998), Nurmi (2011) finds that social factors influenced the rise of *do*; measuring social effects on negative declarative *do*-support from 1580 to 1681, she finds that, after 1600, the change is led by female writers, followed by the group of male writers identified as "social aspirers" (that is, men who are born into a lower social class and move upwards on the social scale during their lifetime; Nurmi 2011: 356). At the point when increased use of *do* by women and social aspirers starts, educated men use the feature more frequently than uneducated men, which suggests that the rapid adoption of the feature by women and social aspirers is catalysed by it becoming more prestigious at that time. This rise of *do* in negative declaratives also coincides with women using affirmative declarative *do*-support significantly more than men, before *do*-support declines overall

for this clause type (Nevalainen and Raumolin-Brunberg 2003: 125-6, reproducing results from Nurmi 1999). Thus, these findings regarding gender differences in the use of English *do*-support give some evidence of women leading the change overall, and apparently doing so after the feature becomes prestigious, i.e. a change-from-above scenario in terms of Labov's (2001) principles of the role of women in language change (cf. Chapter 4.3).

The syntactic function of *do* also varied before it became settled in its present-day usage. Denison (1985) analyses the ambiguous *do* contexts, which gave rise to the reinterpretation of the causative meaning of *do*, as cases of *perfective do*; that is, the focus of the clause is on the action being completed, rather than who is performing the action. This perfective *do* was not an auxiliary, but (in simple terms) took a clause with a PRO subject as its complement. This yielded surface structures like *subject-do-verb(inf.)*, causing *do* to become reanalysed as an auxiliary by analogy with other modals. Garrett (1998) proposes a *habitual do* as the origin structure for present-day *do*-support, which involved a re-analysis of instances of lexical *do* with a nominal object having the same surface form as an infinitival verb (e.g., *I do work*).

The analyses by Denison (1985) and Garrett (1998) aim to capture the variation in *do* functions at an intermediate stage between pre-*do*-support-*do* and present-day dummy *do*. A more recent account by Ecay (2015) provides more evidence for this intermediate stage by comprehensive quantitative study of various syntactic contexts affecting the rise of *do*, primarily replicating results by Ellegård (1953) and Warner (2005) using Middle and Early Modern English data from the *Penn Parsed Corpora of Historical English* (PPCHE; Kroch et al. 2004; Kroch and Taylor 2000). He then validates the findings of these replication experiments by testing the hypotheses again on a much larger dataset: the *Early English Books Online* corpus. Ecay's (2015) findings give more clarity to the different nature of *do* before and after the "dip" in frequencies of *do*-support in the period roughly placed around 1575 in Ellegård's (1953) data, as can be observed in Figure 5.1 – this is around the same time as Nurmi (2011) observes a change in behaviour of male and female writers with respect to *do* (in the CEEC data, this dip is slightly later than in Ellegård's (1953) corpus, around 1600). Warner (2005) suggests that this dip marks a stylistic reanalysis of negative *do*, as

the contraction of *do* and *not* became stigmatised for a time and then de-stigmatised again. This is evidenced by *do* being more frequent in higher-complexity texts pre-1575, and that the opposite correlation is found after 1575. Similarly, the age-grading effect, mentioned above, only applies after 1575. In replicating Warner's (2005) results, Ecay (2015: 66-73) does however find that there is an effect of complexity and age on affirmative *do* as well, which suggests that it is *do*-support itself that undergoes a re-evaluation, and not the *do* + *n't* contraction.

Ecay's (2015) analysis identifies an *intermediate do* auxiliary pre-1575, which exhibits a different syntactic behaviour than the version of *do* which increases in usage after 1575. Specifically, he proposes that intermediate *do* is an agentive marker which merges in a lower syntactic position than post-1575 *do*. The analysis of the syntactic and argument-selecting features of intermediate *do* is based on observations already made by Ellegård (1953); *do*-support is initially more frequent in post-adverbial position and with pronoun subjects, and seems to be resisted by a certain class of verbs. The timing of this proposed intermediate stage of English *do* is interesting for the purpose of this study; as will be seen in the next section (5.2.3), *do*-support emerges in Scots (at relevant frequencies) *after* the re-analysis in English, i.e. after the intermediate-*do* stage. For this reason, it is necessary to look at some of the details of Ecay's (2015) findings here.

Ecay (2015: 76) highlights the examples in (32) as evidence for the hypothesis that *do* initially held a low position in the clause.

(32) a. He hes **done** petuously devour
the noble Chaucer of makaris flour
'[Death] has piteously devoured the noble Chaucer, flower of makars [=bards]'

(Wm. Dunbar *Lament for the Makars*, c. 1505)

b. consequently it wyll **do** make goode drynke
'Consequently [barley] will make good drink'

(A. Boorde *Introduction of Knowledge*, a. 1542)

c. Fro the stok ryell rysing fresche and ying But ony spot or macull **doing** spring
'From the royal stock rising fresh and young / without any spot or blemish springing'
(Wm. Dunbar *The Thrissill and the Rois*, 1503)

Ecay (2015) does not address the fact that examples (32-a) and (32-c) are produced by a Scottish writer, William Dunbar. In fact, (32-a) is highlighted by Görlach (2002: 108) as an example of a form of *do* that was used in Older Scots but not in English. In Gotthard (2019: 6), I discard these types of constructions as a type of *do*-support, on the basis that the *do* is non-finite, and that they seem to mainly have the function of marking tense in verse (Macafee & Aitken 2002: 7.8.15). The other example from Dunbar, (32-c) above, seems to fall within this same category. The example from (the Englishman) Boorde in (32-b), on the other hand, may be more reliable as evidence of a low *do* analysis.

Ecay (2015: 77-78) goes on to test the hypothesis of intermediate *do*'s lower position by comparing the placement of *do* relative to clause-medial adverbs, in comparison to the position of modal auxiliaries and the auxiliary *have*. He finds that the other auxiliaries occur pre-adverbially already from the start of the fifteenth century, and, when *do*-support starts to become more established in the early sixteenth century, *do* initially appears post-adverbially (i.e., *he often did see it*) but increasingly move into pre-adverbial position (i.e., *he did often see it*) to converge with the behaviour of other modals towards the end of the seventeenth century.

Furthermore, Ellegård (1953) identifies a group of verbs which resist *do*-support: *know*, *doubt*, *care*, *fear*, *boot*, *list*, *mistake*, *skill*, *trow* (henceforth referred to as Ellegård's (1953) KNOW class of verbs). This leads Ecay (2015: 78-82) to investigate the behaviour of *do* across different lexical classes of verbs, in order to test argument structure effects on *do*: he extracts 12 high-frequency verbs (including their various spellings) from the PPCHE, sorted into 2 broad semantic classes, *unaccusative* and *experiencer-subject*, with 6 verbs in each class. Ecay's (2015) *unaccusative* class contains ARISE, COME, DIE, GO, RISE, STAND, and his *experiencer-subject* class, which corresponds to Ellegård's (1953) KNOW class, contains the verbs CARE, DOUBT, DREAD, FEAR, KNOW, LIKE;

Ecay (2015: 78) makes the generalisation that the KNOW class of verbs takes experiencer subjects, admitting that this might not be the most appropriate generalisation. Indeed, these types of verbs are typically used as hedging devices and in parenthetical commentary clauses (cf. Bolinger 1977: 127, Chan and Tan 2009: 102), and, for this reason, they can be expected to appear more frequently with first person subjects in their active form. Hence, I will continue to classify these types of verbs as KNOW-class verbs.

He then categorises the rest of the verbs as *unergative* (verbs without a direct object) and *transitive* (verbs with a direct object). His results are consistent with the hypothesis of a reanalysis of *do* after 1575: in negative declarative clauses, there is a clear preference for *do* with verbs taking agentive subjects before 1575, and the other classes of verbs start to ”catch up” in the beginning of the seventeenth century. The picture is slightly different in affirmative declaratives, where the *unaccusative* class of verbs is the only outlier, occurring less frequently with *do* support than any other verb class investigated. From this, he draws the interim conclusion that intermediate *do* functioned as an agentive marker. When testing this on the EEBO data, he finds that the semantic classification effect is not as clear-cut, and that some lexically-specific effects are at work, especially with verbs like REGARD and KNOW – from this, he concludes more tentatively that “(certain) unaccusatives, as a group, are delayed in their progress along this trajectory [and] other specific lexical items may also show oddities”.

The final piece of evidence from Ellegård (1953) investigated by Ecay (2015), is the preference for *do*-support with non-pronominal subjects. He finds that there is a robust subject type effect throughout the period when affirmative declarative *do*-support is used, whereby *do*-support appears with pronominal subjects at significantly lower frequencies than other subjects, and that negative declarative and affirmative question clauses lose this effect after the dip in 1575 (there was not enough data to confirm this effect in negative questions and imperatives). As pronoun subjects have a different information-structural status to nominal subjects, and they have historically inhabited a different structural position than nominal subjects in the clause (e.g., Haeberli 2002, Biberauer and van Kemenade 2011), differences between pronominal and nominal subjects in the Early Modern

data is not surprising in itself. The crucial finding here is that there is a similar behaviour of affirmative declarative *do* and pre-1575 *do* in other contexts, but that this constraint on *do*-support is lost post-1575. This suggests, in Ecay's (2015) analysis, that there are two underlying *do*-support grammars in Early Modern English, and that one changed in favour of the other around the reanalysis of 1575-1620, at the time when verbs stopped raising above adverbs completely. We could also argue that the the type of agentive *do* which Ecay (2015) refers to as "intermediate" does not in fact represent an intermediate stage from pre-*do*-support to PDE *do*-support, but that this "intermediate *do*" represents a different grammar altogether. That is, there are three types of auxiliary *do* grammars present in the history of English. Accounting for the mechanics of these grammars would go beyond the scope of the present study, but the syntactic behaviour of English *do*-support pre- and post-1575 is, as mentioned, relevant for this investigation into whether *do*-support is transferred from English to Scots in the late sixteenth century.

5.2.3 The development of Scots *do*-support

Scots *do*-support is far less researched than English *do*, and, to date, I am only aware of two studies which aim to capture the historical rise of Scots *do* quantitatively: Meurman-Solin (1993a) and Gotthard (2019). Further to this, Jonas (2002) provides a more qualitative account of verb-raising in Older Scots, with a particular focus on Shetland. Besides these studies, Older Scots *do*-support is mostly accounted for briefly in more descriptive works (such as Aitken 1979; Beal 1997; Görlach 2002), or not accounted for at all, which could be because the focus of description is on pre-anglicisation Scots features, or because the feature is assumed to operate identically to English. Evidence from modern Scots dialects suggests that *do*-support does not, in fact, operate identically to English, as evidenced in studies on negative declaratives in Buckie Scots (33), spoken in the North-East of Scotland, and questions and negation in Shetland Scots.

(33) I na mine fa come in
I NEG remember who came in

(Buckie; Smith 2000: 232)

(34) I sall du dat, as lang as it maks no da rates hicher
I will do that, as long as it makes NEG the rates higher

(Shetland; Jonas 2002: 252)

(35) Comes du hame a lot?
Comes you home a lot

(Shetland; Jamieson 2015: 52)

Jamieson's (2015) findings indicate that the *do*-less interrogative is an outgoing structure, with low acceptability ratings for verb-raised questions among younger speakers in Shetland. In line with Jonas (2002), they suggest that V-to-C raising in questions was retained in Shetland Scots, due to influence from Norn, the North-Germanic variety spoken in Shetland from ca. the ninth until eighteenth/nineteenth century. The accounts by Jamieson (2015) and Jonas (2002) assign an important role to language contact for these outcomes in Shetland Scots, both in the retention of V-to-C movement because of contact between Norn and Scots, and in later contact with anglicised Scots, from the sixteenth century onwards, which would have introduced the option of interrogative *do*-support (Jonas (2002: 270) suggests that the loss of V-to-I in Scots is an anglicisation outcome in mainland Scots).

Hence, the variation in *do*-support in Shetland Scots may have come about through the unique contact situations there, but the Buckie structure seems to be a different phenomenon; the Scots Syntax Atlas (ScoSyA; Smith et al. 2019) shows high acceptability ratings for 'I na want to be late' among both young and old speakers in the North-East of Scotland. Smith (2000: 251) suggests an unpronounced *do*-element for these structures, and I will present an alternative analysis in 5.5.3. Both the North-East and Shetland dialects are typically used as examples of "broad Scots", a label which implies less apparent anglicisation than, e.g., the speech of central Scotland, which would imply that the variable *do*-support is evidence of residual Older Scots grammar – i.e., that *do*-support is not a language-internal development, but an anglicisation feature.

However, previous studies on the rise of *do* in Early Modern/Transition Scots have not been able

to make conclusions regarding the origin of Scots *do*-support. The descriptive literature presents different speculative options, but with little discussion. The main argument for Scots having developed *do*-support independently of English is that Scots also had a causative *do* auxiliary (an inherited feature from Northern Old English), as well as the special *do* used in verse (as seen in (32-a)), which could have undergone the same grammaticalisation process as the English auxiliary. The fact that Scots and English had such similar grammatical systems could support such hypotheses. This theory does not, however, give a satisfactory explanation for why the development of *do*-support, or the loss of verb-raising which facilitated the regulation of *do*, takes place nearly 200 years later in Scots than in English.

The emergence of *do*-support in Scots has been dated to the mid-sixteenth century (Gotthard 2019; Meurman-Solin 1993a). Early use of *do* possibly had some stylistic function, as it gradually replaced the auxiliaries *gan/gouth*, *can*, and *couth*, which were all past tense auxiliaries used in narrative verse to indicate events, actions or change in behaviour (Macafee and Aitken 2002: 7.8.15; Gardela 2017; this function is also observed for *gan* in Southern Middle English (Brinton 1996)). In Meurman-Solin's (1993a) study on *do*-support in the *Helsinki Corpus of Older Scots* (HCOS), she indeed finds higher frequencies of *do*-support in texts where this kind of narrative function of *do* could be expected – trial transcriptions, pamphlets and diaries – and that *do*-support favours past tense, which is also expected if *do* has a narrative function (cf. Brinton 1996). From this, she notes that "emphatically interactive" and "overtly persuasive" text types favour *do* (Meurman-Solin 1993a: 265). Thus, this early usage of Scots *do* as a tense marker in verse, or as a narrative device, shows similarities to the early usage of *do* in English.

To estimate proportions of *do*-support in HCOS, Meurman-Solin (1993a) calculates the occurrences of auxiliary *do* over 10,000 words, within the different HCOS genres (in addition to the genres mentioned above, this also includes private and official correspondence, science texts, travel journals, religious texts, etc.). Hence, she does not measure proportions of clauses with *do* in a way comparable to results by, e.g., Ellegård (1953)[1], Kroch (1989), and Ecay (2015). Her findings

1. Ellegård (1953) estimates proportions of affirmative declarative *do* based on a sample of the data.

indicate that affirmative declarative *do*, which is introduced around the time when the usage peaks in English, continues to increase in Scots throughout the seventeenth century, and that affirmative declarative *do* overall occurs with higher frequencies than interrogative and negative declarative *do* – this is perhaps not surprising, given that the affirmative declarative clauses are more frequent as clause types. She does not give a more detailed analysis of *do* in negative declarative than to provide frequencies of occurrences. Finally, Meurman-Solin (1993a) finds that *do* is used more frequently in texts with more anglicisation features, which supports her conclusion that *do*-support is a transferred structure into Scots from Southern English.

In Gotthard (2019), I estimated proportions of negative declarative *do*-support by counting *do*-supported and non-*do*-supported negative clauses in samples of 50 clauses per 50-year time period, extracted from the *Helsinki Corpus of Scottish Correspondence* (CSC; Meurman-Solin and VARIENG 2017), in its original, untagged and unparsed, form (the clause tokens were extracted by searching for the lexical items *do* and *not*). From this, I found only 1 example of negative *do*-support pre-1600 (36), from 1548, and that proportions of negative *do* average at 14% during the seventeenth century (I only included data from 1540-1699, and no data from the *royal* group of writers). The first example of affirmative declarative *do* is seen in (37), and there are only 2 examples of *do*-support in this context before 1600. Thus, negative and affirmative *do* starts to appear at roughly the same time in Scots, but both occur at very low frequencies.

(36) for I **dow** not get yam to hym'
for I DO NEG get them to him

(PCSC-ID: 408_64_m1540, Hew Campbell, 1548)

(37) ye protecto*ur% of yngland **did** send for me
the protector of England DO.pst send for me

(PCSC-ID: 477_133_M1540, Adam Otterburn, 1547)

I could only give absolute frequencies of affirmative declarative *do*, and I speculate that these must represent a low proportion of affirmative declarative clauses (as the total number is 447, which I suspected was low for a corpus of over 400,000 words). I was not able to find examples of interrogative *do*. The trouble in extracting the relevant clause types from the CSC stemmed from the format the corpus was accessible in at the time, and particularly that the corpus could only be searched through a specific web-based corpus search software ('Korp', Borin, Forsberg, and Roxendal 2012)[2] which did not lend itself for the kind of analysis I was attempting to do (this is further detailed in Gotthard (2019: 10-14).

Thus, neither the study by Meurman-Solin (1993a) or my own 2019 study have been able to measure proportions of *do*-support in Scots over time, in the different contexts where it occurs in present-day Scots and English. In Gotthard (2019) I concluded that a morpho-syntactically annotated corpus would be essential to investigate *do*-support in Scots in a way comparable to quantitative studies on the rise of *do*-support in English, thus laying the foundation for the study in this chapter. Nevertheless, the studies presented here have been able to pinpoint a rough starting point for *do*-support in Scots, the mid-late sixteenth century, and the frequency data indicates that negative and declarative *do* remains at low frequencies throughout the seventeenth century. This suggests a slower rise of Scots *do* than in English, which is further supported by that Scots retains V-to-I raising as late as the eighteenth century (and even later in Shetland), and that *do*-support remains variable in some present-day dialects of Scots.

5.2.4 Potential interaction with the *Northern Subject Rule*

As suggested in the analysis by de Haas (2011), introduced in Chapter 4.2, the NSR can be described as a subject-verb (S-V) agreement system on a par with *do*-support, in that both features are last-resort operations when conditions for S-V agreement are not met (de Haas 2011: 163). This analysis finds support in the Buckie pattern of negation (recall (33)); Smith (2000: 245) finds that this variation is only present with subject types which historically occur with ∅-inflection in the

2. This is no longer the case, as the CSC text files have since then been made available for download: https://www.kielipankki.fi/corpora/scotscorr/

NSR paradigm; plural (pl) and 1st person singular (1sg) pronouns (henceforth: NSR subjects). That is, when the subject is a 3rd person singular (3sg) pronoun or a nominal NP (henceforth: non-NSR subjects), *do*-support is categorically present, whereas it is variable with NSR subjects (seen with *they* in (38))

(38) a. They ∅ na seem [to bide in the Beacons lang]
They ∅ NEG seem [to stay in the Beacons a long time]

b. They dinna ken [they're gan to wear a kilt]
They DO+NEG know [they're going to wear a kilt]

(Smith 2000: 232)

de Haas's (2011) analysis of the NSR provides a potential explanation for the variation in Buckie Scots: If we assume that *-(i)s* inflection is a last-resort operation similar to *do*-support, then the NSR would exhibit a similar kind of PF (spell-out) adjacency effect to Swedish object shift or standard English *do*-support, as analysed in Bobaljik (2002). Thus, both *do*-support and the NSR are solutions to the same problem, as illustrated in (39). When the conditions for S-V agreement are not met, e.g. due to interruption of adjacency or that the necessary subject position is not projected, the solution is a last-resort operation which requires the addition of an element (the auxiliary *do* in a *do*-support system, or *-(i)s* inflection in an NSR system). The absence of this element indicates successful s-v agreement.

(39)

Do-support

Derivation:	They [TP ∅.pst [NEGP not [VP sing]]]
PF merger:	Interrupted by *not*.
Solution:	Replace ∅ with *do*
Output:	They did not sing

NSR

Derivation:	They [TP ∅.pres [VP sing and dance]]
PF merger:	Merger is allowed for *sing*, interrupted for *dance*.
	Subject is of correct type
Solution:	Merge with first verb, apply default inflection for second verb
Output:	They sing-∅ and dance-s

Thus, this analysis would explain the Buckie pattern as such: Clauses with successful NSR S-V agreement, i.e. clauses which would have -∅ inflection on the verb, would also not require mandatory *do*-support. However, the negative element *na* appears to intervene between the subject and verb (recall (33) and (38)) – in a Standard English negative declarative clause, this would trigger *do*-support. One explanation for this difference is that the clitic negator *na* is a weak element, loosely connected to the clause and thus not an intervening element, similar to adverbs in Bobaljik's (2002) analysis (also mentioned in the environment of (16) in Chapter 4.2) – in English, PF merger works fine in *The cat quietly sneaks outside*). This analysis aligns with accounts by van Kemenade (2000), van Gelderen (2008), and Kroch (1989) regarding the re-analysis of negators from weak to strong elements in the history of English and cross-linguistically; Kroch (1989: 236) notes that word orders like "we not fear" did occur during the rise of *do*-support in English, when *not* was re-analysed as a weak element (see also Gotthard 2019: 17-8). This word order seems to also have been possible in the history of Scots, as seen in (40), where the relevant part of the clause is in bold font.

(40) Meckle easier e sall expone quhat it nocht beiris, than quhat it beiris
much easier you shall expone what it not bears, than what it bears

(1596, DOSTIV: 561; from Jonas 2002: 264)

Furthermore, Meurman-Solin (1993a) finds that *do*-support favours clauses with intervening elements between the subject and verb in HCOS, as witnessed in (41), from Meurman-Solin (1993a: 269), where the subject and *do*-auxiliary is underlined.

(41) a. **zoung[young] boys**, vith artificiall winges, at her entrey, **did** flee touards her, and presented her tuo[two] siluer keyes of ye[the]city.

(Birrel, Diary: 25)

b. **I** kept at Forres betwixt the children of Thornhil, and **did** see how carnal! affections and self lov[love] blinds and divids the most neer relations

(A. Brodie, Diary:311)

These examples would be consistent with an analysis of *do*-support favouring the same environments where we would expect to see a default *-(i)s* inflection in an ideal NSR grammar.

5.3 Research questions and predictions

The study in this chapter has two aims: Firstly, I explore the rise of *do*-support in the PCSC, extracting proportions of *do*-support over time in a way that was not previously possible for Scots data due to the lack of a parsed corpus. Thus, the first research question asks:

RQ1: What are the proportions of *do*-support over time during the Transition Scots period?

Findings by Meurman-Solin (1993), Jonas (2002), and Gotthard (2019) predict that *do*-support appears later and at lower frequencies than what is found in English.

Secondly, in light of the analyses outlined in Sections 5.2.2, on intermediate *do* in English, and 5.2.4, concerning the last resort status of *do* and *-(i)s* in processes conditioned by PF adjacency, I ask:

RQ2:

A. Does Scots *do*-support exhibit similar features to English "intermediate *do*"?

B. Does Scots *do*-support exhibit sensitivity to NSR constraints during its emergence and rise?

The findings of RQ2A will provide more clarity on the origin of *do*-support in Scots: If there are no intermediate *do* features in the early examples of *do* in Scots, it would support an analysis of *do*-support being an adopted structure, with post-1575 English *do* qualities. However, if early Scots *do*-support exhibits the same qualities as intermediate *do*, the origin of Scots *do* is more obscure. Such an outcome could either be seen as (i) evidence of an "intermediate *do* grammar" having spread from English to Scots, and the slightly later emergence in Scots is due to a time lag between the English and Scots data, or (ii) evidence of English and Scots *do*-support being parallel developments, following the same pattern of grammaticalisation, which we may interpret as *do*-support not being a borrowed structure – this does not rule out that the regulation of *do* it is not a contact-induced change in Scots; rather, the contact with English may have triggered such a development in Scots.

Likewise, the findings of RQ2B also provides more insight into grammar competition between similar systems: if an NSR condition is present in the rise of Scots *do*-support, in the way suggested by de Haas's (2011) analysis, it would predict that *do*-support will appear first and more frequently with 3rd person singular subjects and plural nominal subjects, i.e. in clauses that would normally resort to a default *-(i)s* inflection under the NSR, which is suggested by the pattern found in Buckie by Smith (2000). This prediction assumes, then, that *do*-support would slot into the existing NSR system, and not replace it in all contexts at once (at least not in the early stage of the change represented by the PCSC data).

Finally, I also investigate gender effects on the rise of *do*-support in Scots, making two opposite predictions. First, the findings by Nurmi (2011) regarding women driving the change in the rise of *do*-support in English, give rise to prediction I.

I. Female writers are leaders in the rise of *do*-support in Scots

However, if we assume that *do*-support is a structural transfer from English, this would give rise to the opposite prediction, in II, namely that *do*-support is dis-favoured by Scottish women – this prediction is now further supported by the results on gender differences in the use of the NSR pattern in Chapter 4.

II. Female writers use less *do*-support

5.4 Method

To extract frequencies of *do*-support in the PCSC in order to answer RQ1, no further corpus annotation was needed; all relevant clause types, tenses, and moods are already part of the standard annotation. For RQ2B, the same additional annotation as for the study on the *Northern Subject Rule* (described in Chapter 4.1) was used, and the same CorpusSearch (CS; Randall 2000/2013) coding query – this query already had a column labelled *verb type* under which the level *do* finds clauses with the auxiliary *do* (clauses with a DOP or DOD with an infinitival verb sister). A column specifying *mood* was added, which contained the levels *imperative*, and *declarative*. The *type* of clause in terms of its polarity was also added as a separate column, with the levels *negative* and *affirmative*.[3] A separate coding query was written to retrieve results for matrix questions, which contained the same columns as the original query but targeting questions as the root node.[4] Hence, I could retrieve results for clauses with various combinations of polarity, mood, tense, and subject type. To investigate NSR subject effects on *do*-support, I divided up the subjects into the same three groups as in the analysis of NSR in Chapter 4; NSR subjects (plural and 1st person singular (1sg.) prononus), non-NSR subjects (3rd person singular (3sg.) pronouns and non-pronouns) and

3. The coding query, definition file, and query output are in Appendix 2.
4. The coding query is in Appendix 3.

plNP (plural non-pronominal subjects)

In investigating RQ2A, i.e., whether the type of *do*-support emerging in Scots in the late sixteenth century shows similar behaviour to Ecay's (2015) intermediate *do*, or whether it has post-1575 English *do* behaviour, I wrote another coding query with columns based on Ecay's (2015) findings. The 'fundamental' columns remained the same, i.e. those specifying verb type (*do*, *have*, *be*, *lexical verb*), polarity, mood, and tense. The subject type column was altered to only specifying whether the subject is pronominal, nominal or null (including all forms of null subjects, e.g. subject traces, subjects omitted under conjunction, PRO subjects, etc.). I added a column for adverb position, which finds adverbs occurring between the subject and finite verb, or after the finite verb, in order to measure instances of pre- and post-adverbial *do* compared to other auxiliaries; I looked for adverbs immediately preceding or following the finite verb using the CS `iprecedes` search function, as I noticed a high error rate in the results when using the less strict `precedes` option – this may have caused some examples to be missed, but since I am only measuring proportions within the subset of clauses with clause-medial adverbs occurring immediately before or after the verb, I believe this will not cause a crucial skewing of the result.[5]

Finally, I used a similar method to (Ecay 2015: 78-9) in retrieving representative examples of lexical classes of verbs, using the following steps:

1: I extracted all finite verbs from the corpus, and all infinitival verbs that are sisters to finite *do* (i.e. main verbs in *do*-support clauses). This was done by running two simple CS search queries, one finding finite verbs and one finding *do*-supported infinitival verbs, and only printing the relevant nodes in the output. Then I ran a `make_lexicon` query on the output files which extracted all relevant verbs into a list, including the frequency of occurrences for each verb (which is a feature of the `make_lexicon` function)

5. The coding query and query output are in Appendix 3.

2: I sorted the list of verbs according to their frequency, manually searched through the list of high-frequency verbs, and sorted them into 3 semantic classes, with 7 representative high-frequency verbs per class. The 3 classes are: *agentive*, *unaccusative*, and *k-class* (from KNOW class). This initial selection, then, essentially determined 7 semantic lemmas per class.

3: After deciding on the 7 lemmas, I added them as variables in a CS .def file, and manually looked through the verb lists to add all forms and spellings belonging to the lemma to the relevant .def file variable.

The resulting lemmas can be seen in Table 5.1, along with the different lexical items they denote when more than one, and their absolute frequencies (including both finite verbs and infinitival verbs in *do*-supported clauses)

Table 5.1: Semantic classes and their representative lemmas, with absolute frequencies

Agentive		Unaccusative		K-class	
lemma	*tot. n*	*lemma*	*tot. n*	*lemma*	*tot. n*
DO	108	GO	163	KNOW (incl. *ken, wit*)	446
SEND	166	COME	375	DOUBT	110
SAY (incl. *tell*)	292	FALL	23	TROW (*find*)	130
GIVE	163	STAND	39	FEAR	58
WRITE	255	LIE	48	DESIRE (incl. *want, wish*)	444
TAKE	124	STAY (incl. *remain, rest*)	223	THINK	531
MAKE	87	LIVE	29	HOPE	379
tot. n:	1,195	*tot. n*:	900	*tot. n*:	2,097

Once these lemmas were specified in the .def file, I could add two more columns to the coding query to find the particular verbs and to categorise them according to their semantic class. This made it possible to investigate the behaviour of these semantic classes of verbs and the individual lemmas with respect to their proportions with *do*-support.

For this study, I only binned the data for the analysis of overall proportions of *do*, using the

same method as for the NSR data in Chapter 4; first taking the average frequencies within 20-year bins, calculating the 95% confidence interval for each sample using the binom package (Dorai-Raj 2014) with the Wilson method. I also fit the data to a LOESS curve with year as a continuous variable, for a complementary analysis of the overall data. For the analyses which consider the effect of *gender* (5.5.4), and other conditioning factors relating to intermediate *do* (5.5.2) and NSR constraints (5.5.3), I fitted the data to a LOESS curve.

5.5 Results

5.5.1 The rise of *do* in the PCSC

Interrogative and Imperative

Unfortunately, the results for interrogative and imperative clauses did not reveal any reliable findings. Out of a total of 25 matrix questions (21 affirmative, 4 negative), only 3 exhibited the relevant environment for *do*-support (i.e., Wh-object or polar question, and no other auxiliary present), and 2 of those had *do*-support, as seen in (42), the relevant part of the clause is underlined and *do* is in bold font.

(42) a. or proceed they from a free will guided by unstable Passions, & various interests,

(PCSC ID: 302_11_F1700; Margaret Hamilton, 1702)

b. bot **do** ze thenk yes xx days past hee heth let me kno hou my bessenes is theruoth:

(PCSC ID: 83_65_F1600; Elizabeth Ker, 1641)

c. **Do** you never rime now nor ver= versifie[sic] when you have so much time on y=r=[your] hands?

(PCSC ID: 1421_55_M1700; John Erskine, 1717)

Negative imperative *do* occurs first in 1660, and in total only 9/49 times in the corpus, whereof 9/22 times after 1660 (incl.) – thus, the difference between pre-1660 and and post-1660 is coarse-grained but significant with respect to imperative *do* ($p < 0.001$)). In (43) are examples of negative imperatives with and without *do*, where *do* is in bold font and the relevant part of the clause is underlined.

(43) a. bot **doe** not tye your selff to a falzie w=t=out yow sell w=t=in bounds,

(PCSC ID: 1115_10_M1650; John Gordon, 1660)

b. so **do** not feall to Come

(PCSC ID: 293_2_F1700; Christian Cameron, 1722)

c. and als*s%[also] fayll no=t= tyll[to] caus*s% murray [to] haist hym' heir

(PCSC ID: 406_62_M1540; John Cambell, 1543)

d. and let not the berer staye long

(PCSC ID: 73_55_F1600; Agnes Keith, 1600)

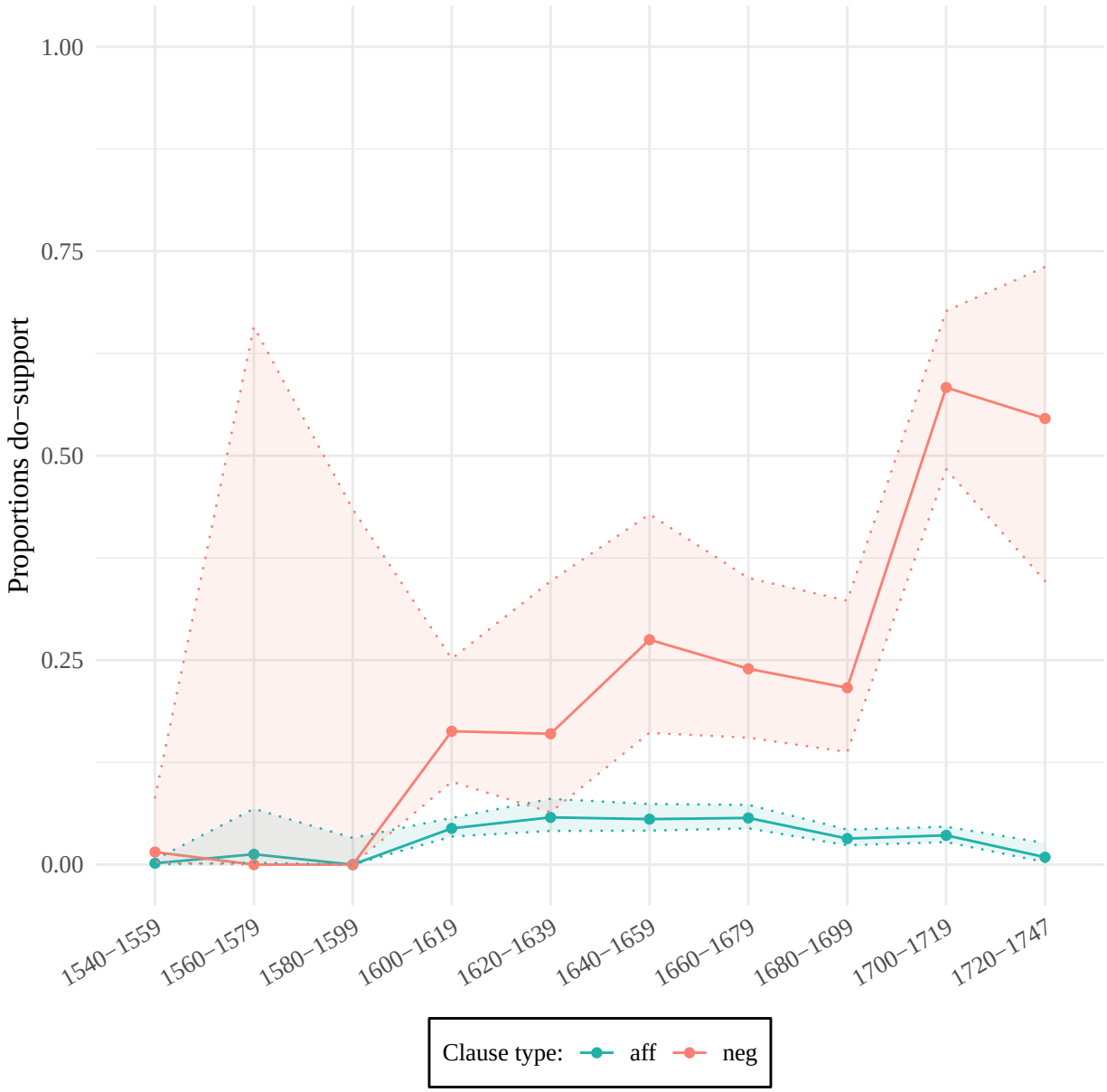
One reason why these constructions are rare in the corpus may be due to the level of formality of these letters; findings from recent and not-yet published work on eighteenth century CSC data by Christine Elsweiler[6] show that writers make use of a range of discourse strategies in order to make requests, which are often not formed by matrix questions. It is feasible to assume that this would also mean that imperatives, which express commands, are avoided for the sake of politeness.

6. Findings from this work in progress have been presented at several conferences in 2021-22, including the Virtual HiSoN 2021 Conference, the 21st International Conference on English Historical Linguistics, the 13th international conference of the Forum for Research on the Languages of Scotland and Ulster, and the HiSoN 2022 Conference.

Negative and Affirmative Declarative

Testing the proportions of *do*-support with negative and affirmative declarative clauses gives more robust results than for the other *do*-support environments. As seen in Figure 5.2, affirmative declarative *do* is very rare in this corpus (3.6% overall (292/16,092), and peaking at a 5.8% average in 1620-1639). In total, negative declarative *do* occurs 140 times out of 522 possible occurrences (26.8%).

Figure 5.2: Frequency of negative and affirmative declarative do-support



Negative declarative do-support appears to remain at stable levels throughout the seventeenth century (overall proportions are 20%), peaking at 24% in average proportion in the 1660-1679 bin, and then starts to increase more dramatically after 1700, reaching a 55% average in the 1720-1747 period, based on point estimates. I am also able to confirm that there is only one case of negative declarative *do* before 1600, the same example found in Gotthard (2019; seen in (36)). There are 3 examples of affirmative declarative *do* before 1600 – in my 2019 study, I missed one occurrence as it was produced by a writer in the *royal* group, which I excluded from the study. In (44) are the 2 other examples, in addition to the first example of affirmative *do* in (37)

(44) a. for tha do lyk noble me*n

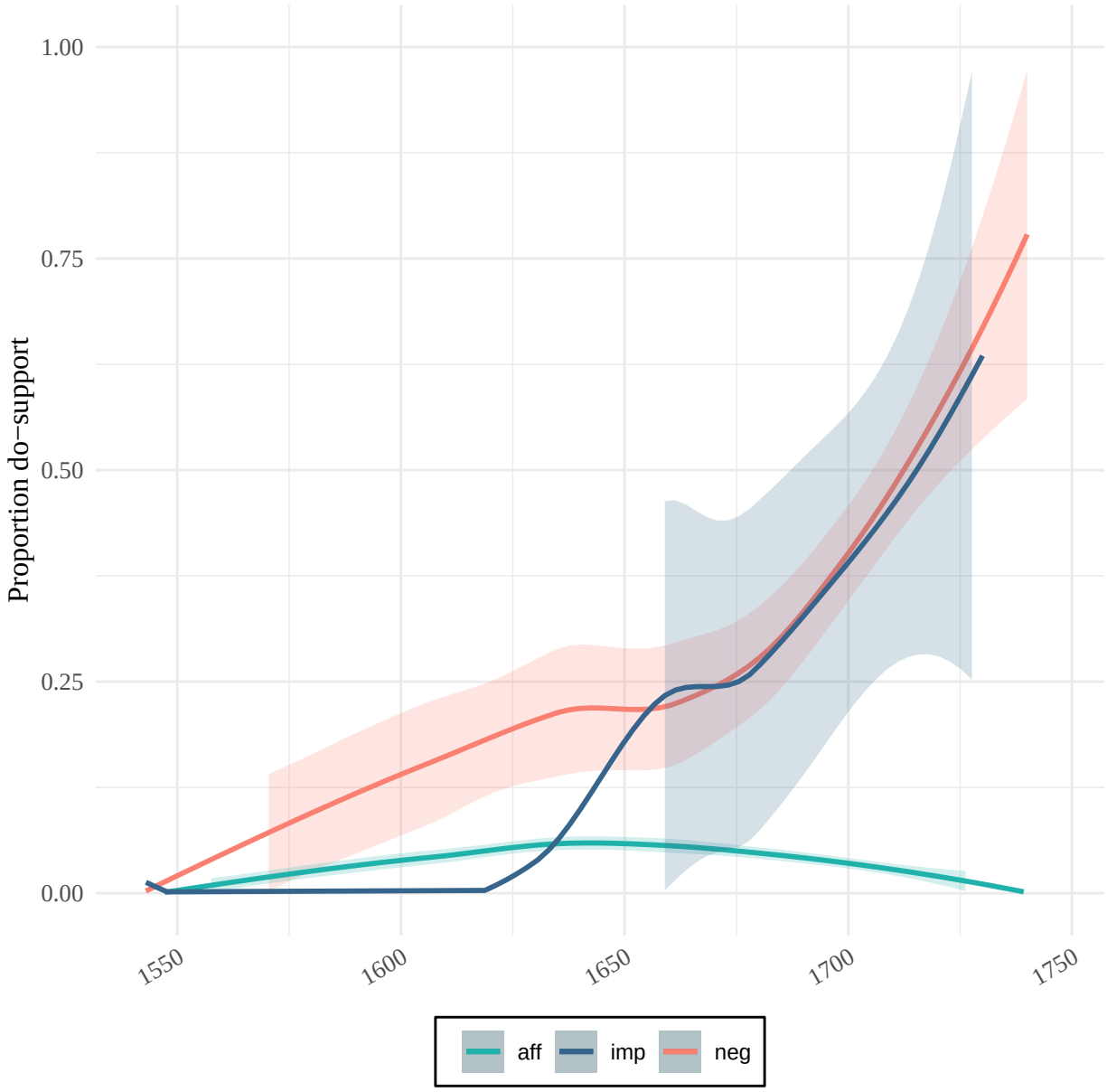
(PCSC-ID: 7_7_F1540; Marion Haliburton, 1549)

b. And wheare you*r% L[lordship] Dothe So frendly offer to advaunce the howse agayne to the auncyent estate

(PCSC-ID: 1492_2_R1540; Margaret Douglas, 1571)

Figure 5.3 shows the same data visualised in LOESS curves, and here I have also plotted the data of negative imperative clauses for comparative purposes – despite the low frequencies, the LOESS curve fits a trend of incline on the imperative data which matches the curve for negative declarative *do*.

Figure 5.3: Frequency of declarative do-support (LOESS curves)

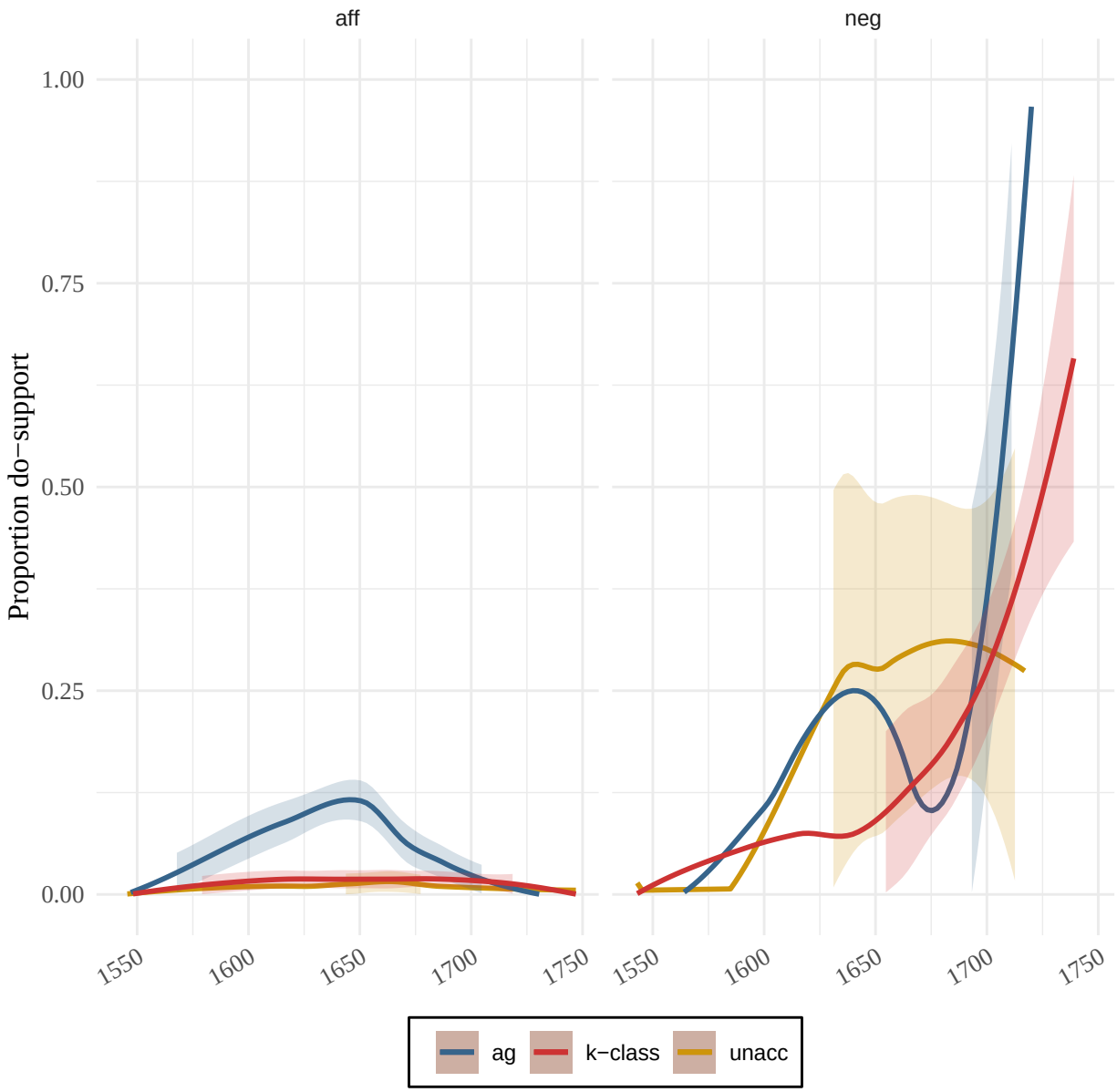


In both visualisations, we see a hint of the characteristic rise and fall of affirmative declaratives with *do*-support in English (Kroch 1989; Postma 2010; cf. Figure 5.1). The proportions of negative declarative *do* are similar to my 2019 estimations, and the results corroborate a later emergence of Scots *do*-support compared to English. As regards the predicted slower increase, this prediction holds until the late seventeenth century when the proportions of *do* rapidly increase.

5.5.2 Intermediate do

The first results of the investigation of whether early Scots *do*-support behave like intermediate *do* in English relate to the argument-selectional features of *do*. That is, if early Scots *do* is like intermediate *do*, we would expect it to occur more frequently with agent-selecting verbs. The results in Figure 5.4 confirm this prediction for affirmative declarative clauses, where *do*-support is clearly favoured with the agentive verb type compared to the other two verb type classes. There are 53 examples of *do*-support with this group of verbs, out of 1,110 possible occurrences (this constitutes ca. 18% of the overall frequency of affirmative declarative *do*-support). Unaccusative verbs only appear with affirmative declarative *do*-support 9/848 times in total, and the KNOW-class ('k-class' in the graph) 29/1,856 times.

Figure 5.4: Frequency of do-support by verb class (LOESS curves)

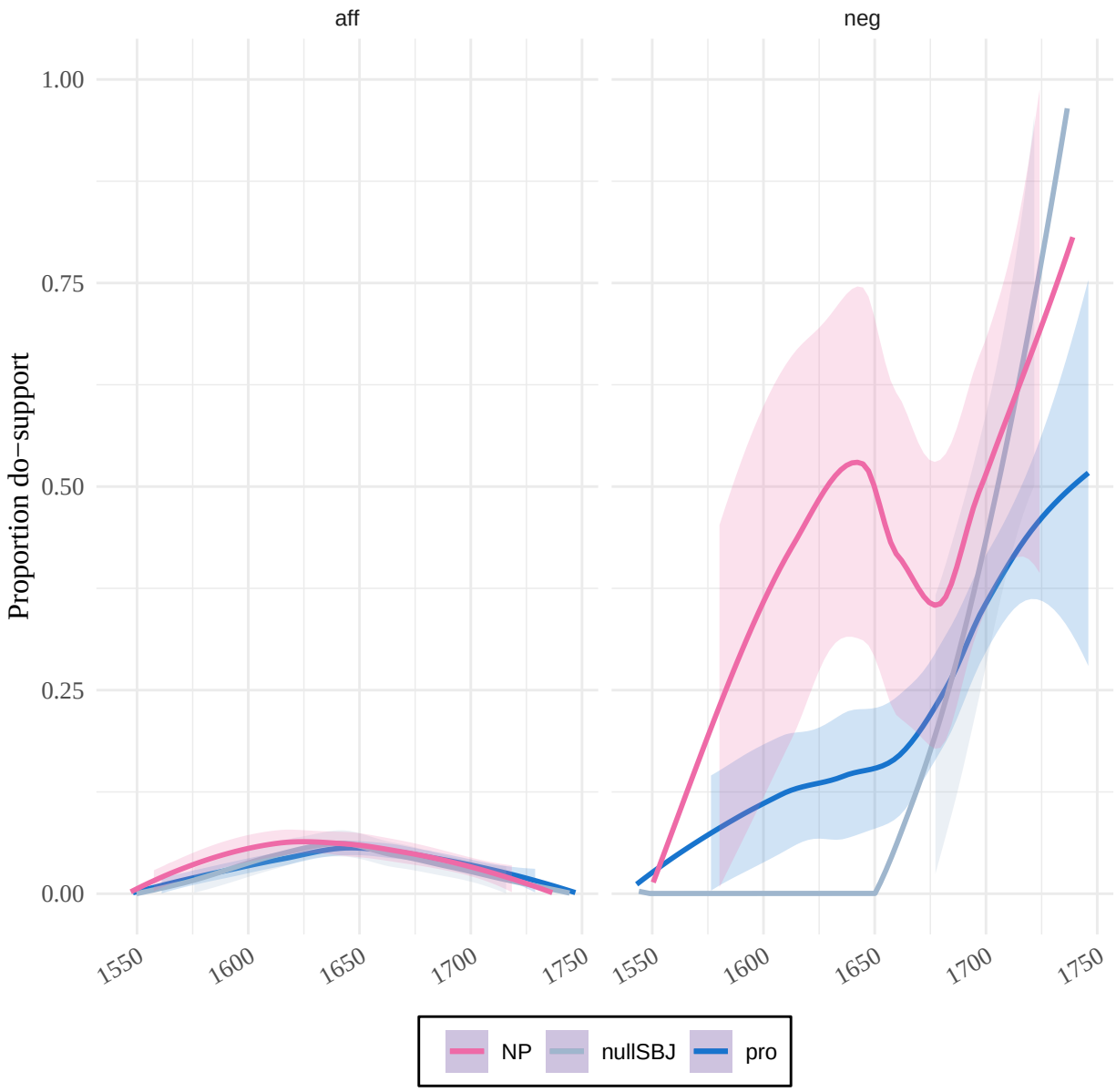


The findings for negative declarative *do* are not as easily interpretable. The results for the agentive and KNOW-class verbs are more robust from the later half of the seventeenth century, where both verb types sharply increase with respect to their *do*-support frequencies and *do*-support with agentive verbs has an apparent advantage. Before this point, there is not enough reliable data to make conclusions about the difference between the verb types. The curve for unaccusative verbs is fitted over 45 clause tokens, and only 9 of these have *do*-support; COME (2), GO (2), STAY(2),

LIVE(1), STAND (1). The unaccusative verbs appearing with *do*-support in the affirmative declarative clauses are COME (3), STAY (3), GO(1), LIE(1). In terms of Sorace's (2000) Unaccusativity Hierarchy, these verbs are mid-high on this hierarchy, and thus not expected to show agentive qualities. However, as Ecay (2015: 80) notes, the function of *do* could be to coerce agentivity on non-agentive verbs in these cases, i.e. *do* is used to indicate an external argument.

Next, I investigate whether there is a difference in frequency of *do*-support between pronominal and non-pronominal subjects, where *do* favours non-pronominal subjects. While there is no apparent subject type preference in affirmative declarative clauses, the difference is striking in the negative declaratives, where non-pronominal subjects give rise to higher rates of *do* than other subjects throughout the period. There is some noise in the data due to the relatively small number of negative declarative clauses with NP subjects (75, whereof 30 has *do*-support) which means that the dip in frequencies in the late sixteenth century should be observed with some caution.

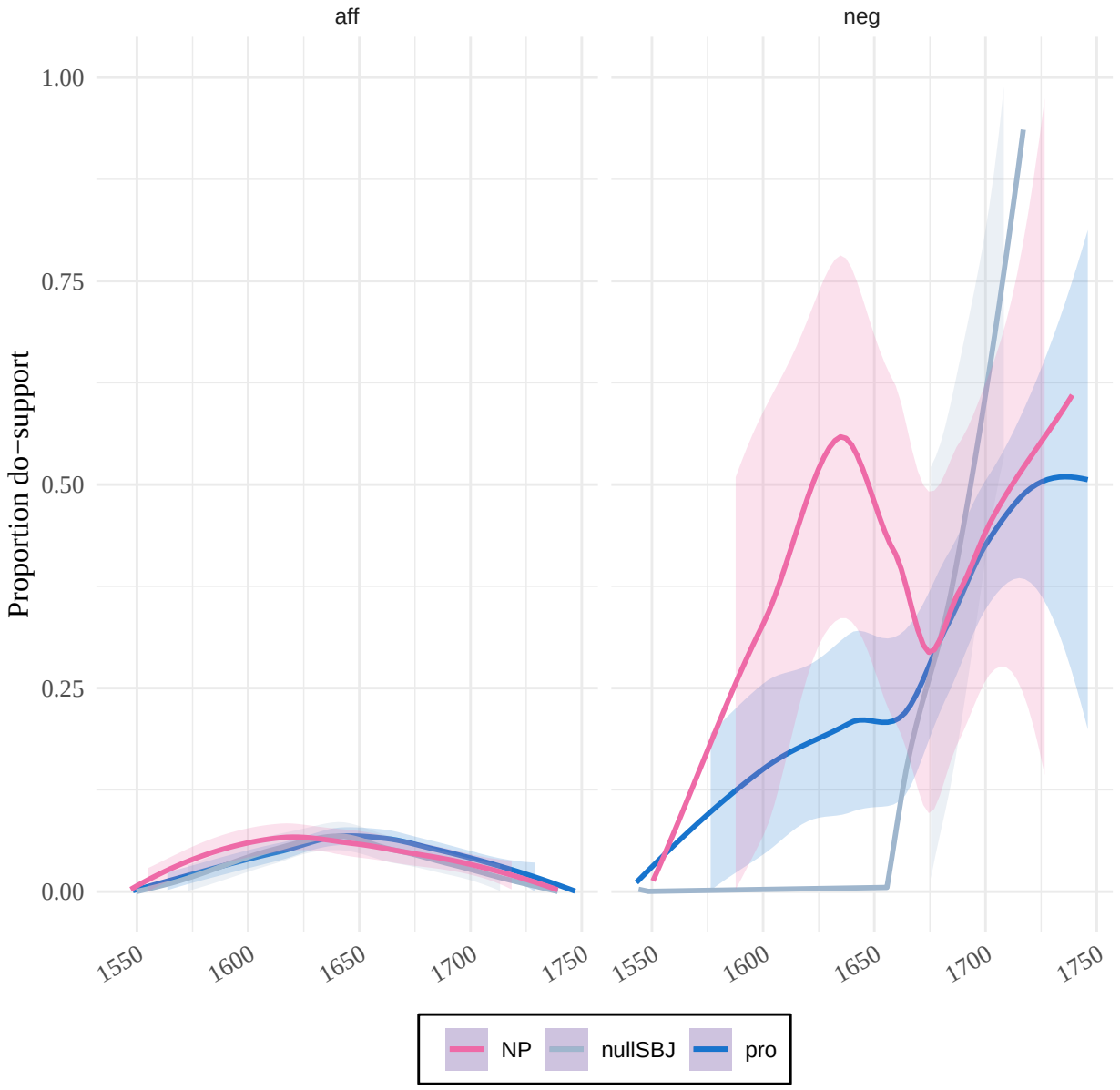
Figure 5.5: Frequency of declarative do-support by pronominal or nominal subject (LOESS curves)



Due to the special status of the KNOW class verbs as hedging devices, we can expect them to appear more with first person subjects, and therefore they might skew the results. For this reason, I tested the subject type effect again without the KNOW-class. There were 39 examples of pronoun subjects with KNOW-class verbs in total, which constitutes 39.4% of the total number of negative declaratives

with pronoun subject (n=99). The trend for subject type sensitivity without the KNOW-class is seen in Figure 5.6. The majority of pronominal subjects with KNOW-class verbs are 1sg. (1337/1689 (79%)), which is the most likely subject to be used in a hedging function.

Figure 5.6: Frequency of declarative do-support by pronominal or nominal subject, excluding know-class verbs (LOESS curves)

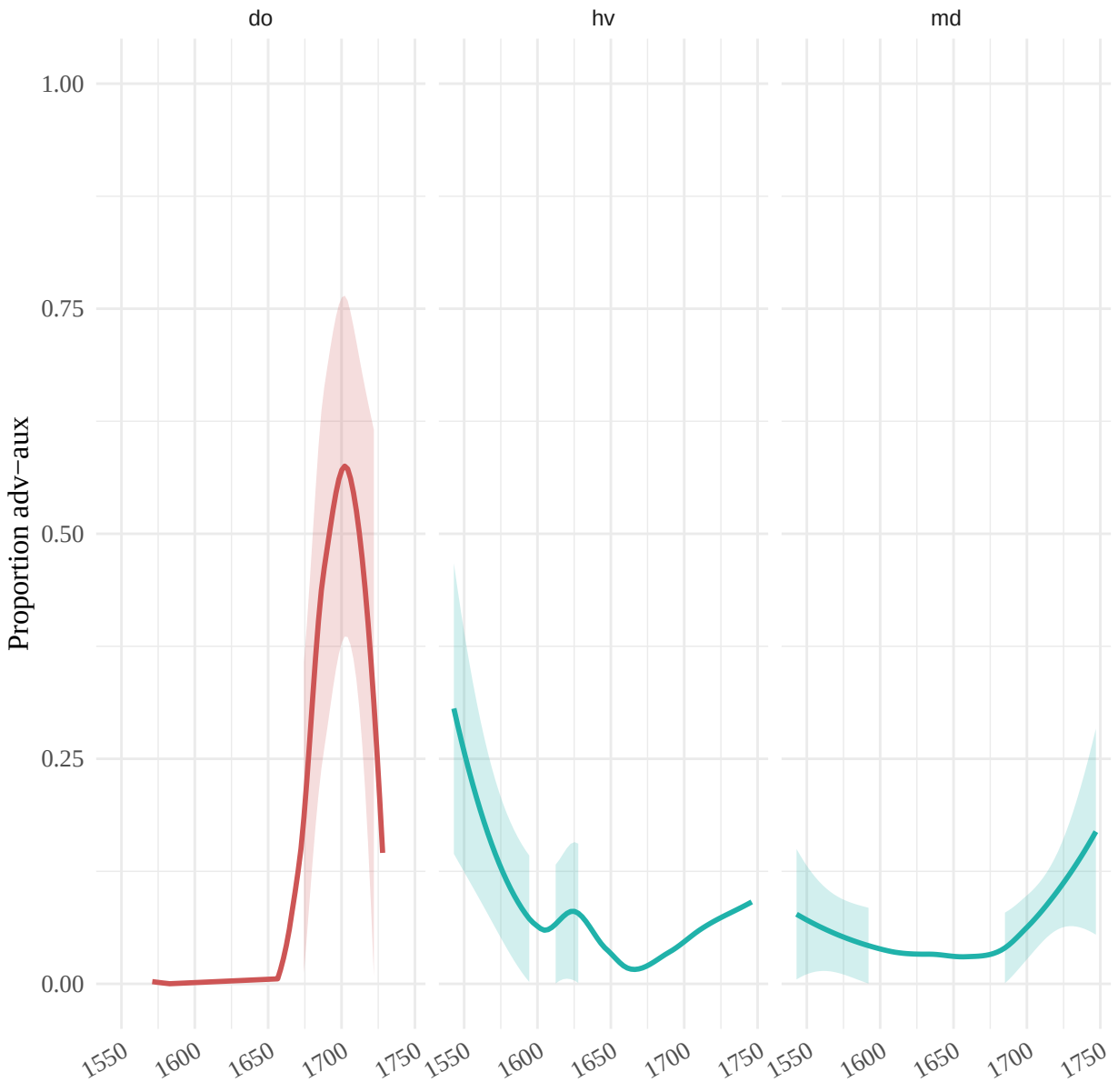


(*'NP' = non-pronominal subject, 'pro' = pronominal subject, 'nullSBJ' = null subject)*

Removing the KNOW-class verbs did indeed yield higher proportions of negative declarative *do* with pronoun subjects, which is most noticeable from the late sixteenth century. The results here appear more similar to Ecay's (2015) findings with respect to subject type; the effect is there, but after a critical point in time the effect is not significant any more. This critical point appears to be around 1675 in the PCSC data.

Finally, the position of *do* in relation to mid-clausal adverbs was investigated. The results are in Figure 5.7; I compared the position of *do* with the auxiliary *have* ('hv' in the graph) and modals ('md' in the graph; e.g. *will*, *shall*), and calculated the proportions of adverbs occurring between the subject and finite verb (or modal), compared to when the adverb occurs after the verbal item. Thus, high proportions indicates the verbal item typically having a low position in the clause.

Figure 5.7: Proportions of post-adverbial (vs. pre-adverbial) verbs (LOESS curves)



As can be seen, the results are quite inconclusive for all the modal and auxiliary verbs tested; on average, modals appear in post-adverbial position 54/544 times, the auxiliary *have* 19/179 times and *do* 7/48 times. It seems that the auxiliary *have* still appears in a low position in the beginning of the period, but increasingly settles into a higher position in the clause, but there are too few data points after 1600 to make assumptions regarding its position then. Similarly, the seventeenth

century seems to be a bit of a blind spot regarding the position of modals, but they occur in post-adverbial position at quite low frequencies in the beginning and end of the period. Despite displaying some noise in the data, the curve for *do* indicates that *do* takes a low position more frequently than the other modals and the *have* auxiliary.

Summary: Intermediate *do*

In testing RQ2A, whether *do*-support in the PCSC is showing sensitivity to intermediate *do* constraints, the following pattern was found:

1. *Do*-support is favoured by verbs taking agentive subjects more than it is by k-class or unaccusative verbs. This difference is most apparent in affirmative declarative clauses, but can be seen after 1675 in negative declarative clauses.
2. *Do*-support occurs more frequently with non-pronominal subjects in negative declarative clauses. When KNOW-class verbs are excluded, this effect is lost after 1675.
3. There is no subject type effect on affirmative declarative *do*-support.
4. In the late seventeenth century, it appears that *do*-support inhabits a lower position in the clause than other auxiliaries and modals.

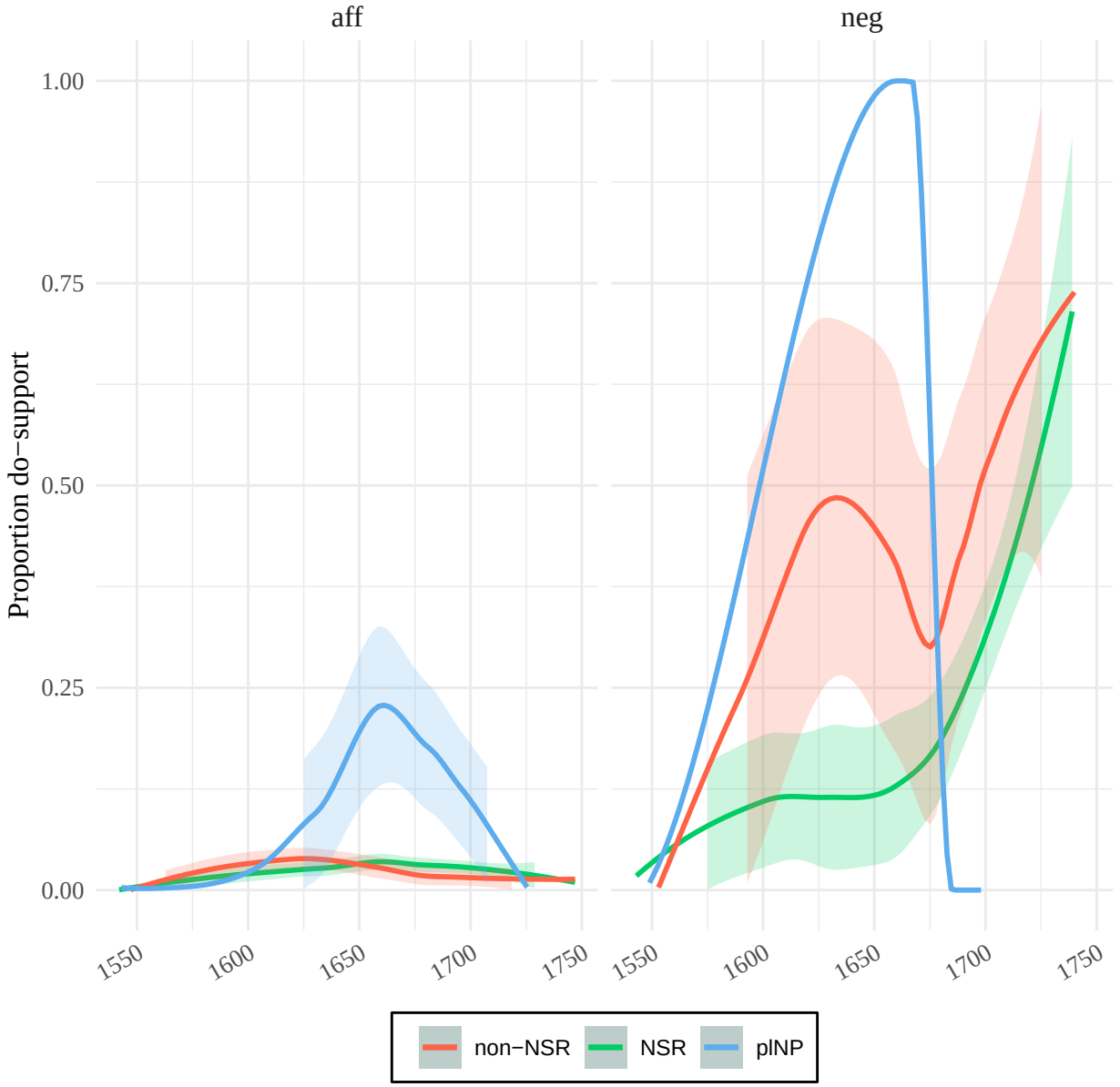
The findings in 1, 2, and 4 are consistent with what Ecay (2015) observed for intermediate *do* in English, whereas 3 is not. Similar to English *do*-support, some form of reanalysis seems to take place in negative declarative *do* around the time when affirmative declarative *do*-support declines, i.e. around 1675. This is visible in the findings for subject type effect and clausal position (findings 2 and 4), but not for the argument-selectional features of *do* (finding 1).

5.5.3 NSR subject type

In the previous section, we saw a subject type effect on *do*-support where, before ca. 1675, *do* occurred with pronominal subjects at lower frequencies than with non-pronominal subjects. The NSR subject group consists of only pronominal subjects (plural and 1sg.), but the non-NSR group consists of both nominal and pronominal 3sg subjects. For there to be an NSR subject type effect on

do-support, higher frequencies of *do*-support with non-NSR and plNP subjects would be expected. The results are in Figure 5.8. For this analysis, only present tense clauses were tested, as the NSR applies in the present tense. Hence, the total average proportion of affirmative declarative *do*-support is ca. 0.2% (136/5679), and for negative declarative *do* the proportion is 24.1% (90/374).

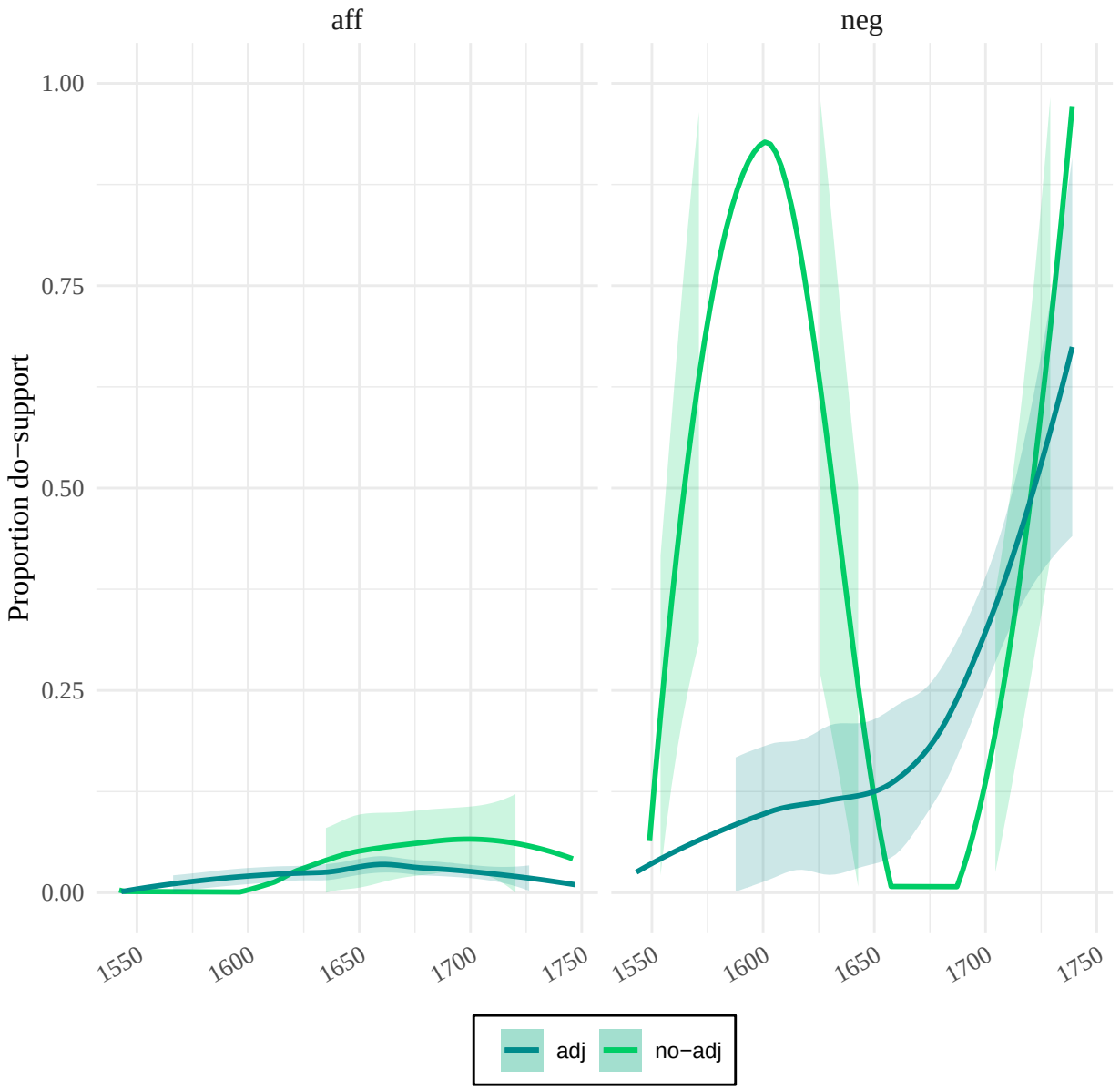
Figure 5.8: Frequency of affirmative and negative declarative do-support by subject type, where non-NSR subjects exclude plNP (LOESS curves).



The results for non-NSR and NSR subjects look nearly identical to the subject type effect in Figure 5.5. This is probably due to 3sg. pronoun subjects not being very frequent in the data (30 tokens, out of which 16 have *do*-support). There is also an indication of the same bump and dip in the mid-late seventeenth century as was observed for agentive and non-pronominal subjects. There are only 5 examples of plNP subjects in present tense negative clauses overall in the data, of which only 2 occurrences have *do*-support (cf. 27/63 for non-NSR subjects, and 71/306 for NSR subjects). While there are more examples of plNP subjects in present tense affirmative clauses (146), only 12 of these clauses have *do*-support (cf. 33/1,719 for non-NSR subjects, and 91/3,792 for NSR subjects). However, there still appears to be a preference for *do*-support with plNP subjects compared to the other subject types in the affirmative declarative data; aggregating the affirmative declarative data over the whole time period, plNP subject appear with *do*-support significantly more than the other subject types combined ($p<0.0001$), and there is no significant difference between the proportions of *do*-support with NSR and non-NSR subjects ($p = 0.187$).

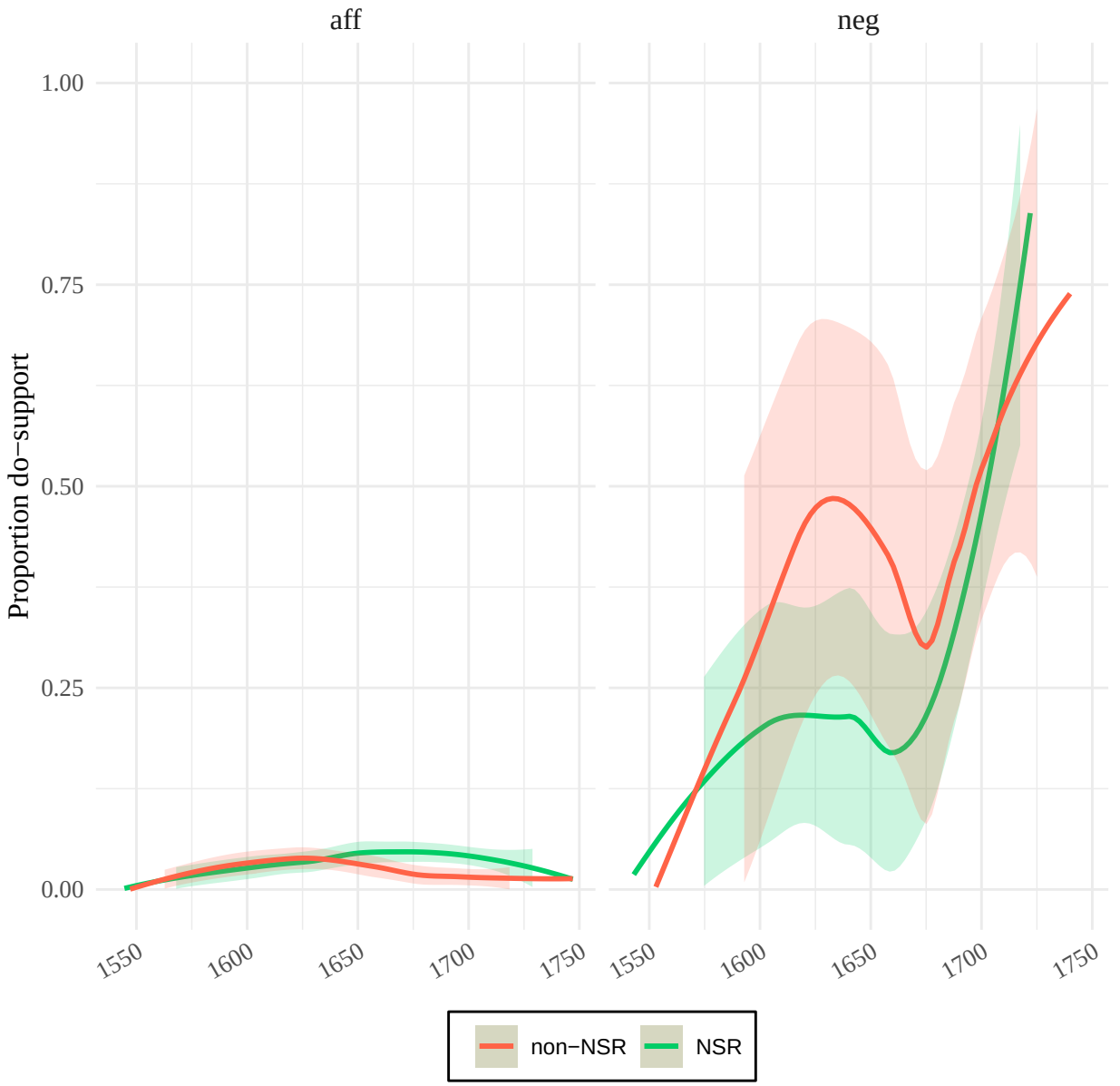
In Figure 5.8, I included both adjacent and non-adjacent clauses. The only subject type which would be expected to behave differently in non-adjacent clauses is NSR subjects, so I compared the proportions of *do*-support with NSR subjects only, as seen in Figure 5.9. As can be seen, the data is very noisy for non-adjacent negative declarative clauses (there are only 8 instances of this clause type with *do*-support). There appears to be an indication of an adjacency effect in the affirmative declarative data, despite the adjacency constraint already being weakened for NSR subjects in this time period (cf. Figure 5 in Chapter 4), but it is not statistically significant ($p = 0.52$); aggregated over the whole period, *do*-support with NSR subjects occurs 83/3,470 times in adjacent clauses, and 7/228 in non-adjacent clauses.

Figure 5.9: Frequency of affirmative and negative declarative do-support with NSR subjects, by S-V adjacency (LOESS curves)



Finally, I also tested whether the results changed if the KNOW-class of verbs were excluded, which can be seen in Figure 5.10. This time, I excluded the noisy plNP data from the graph.

Figure 5.10: Frequency of affirmative and negative declarative do-support by subject type, where non-NSR subjects exclude plNP (LOESS curves).



Just as seen in Figure 5.5, excluding the KNOW-class makes a difference for the behaviour of pronominal subjects, in this case NSR subjects, after 1675 – there is no apparent subject type effect for negative declaratives after the shift in the frequencies marked by the bump in the curve. Surprisingly, the opposite subject type effect than what was predicted from the analysis of the interaction of *do* and the NSR can be seen in the later part of the period for affirmative declarative clauses, wherein NSR subjects occur more frequently with *do*-support than non-NSR subjects.

Inflected *do*

Finally, I also tested whether *do*-support participates in the NSR pattern when it does occur, i.e., whether there is a preference for inflection on the auxiliary *do* with non-NSR subjects, or whether *do*-support is adopted with a StE agreement pattern. The results are in Table 5.2.

Table 5.2: Do-support with inflection on DO, by subject type

Affirmative declarative		**Negative declarative**	
	S-V adjacent		
Subject type	*Proportion inflection*	*Subject type*	*Proportion inflection*
non-NSR	21/25	non-NSR	13/22
NSR	0/83	NSR	0/58
plNP	0/2	plNP	0/0
	Non-adjacent		
non-NSR	7/7	non-NSR	3/4
NSR	5/7	NSR	0/3
plNP	8/10	plNP	0/2

While there are very small samples for some subject types, and particularly in non-adjacent contexts, a tendency of a pattern can be noted, in which *do* follows the overall S-V agreement pattern described in Chapter 4; *do* variably takes *-s* inflection when the subject is non-NSR and plNP (45), and with non-adjacent NSR subjects (46).

(45) a. **hee** do-es not allow yow to give him the sirup of peall Rosses

(PCSC ID: 1121_16_M1650; John Gordon, 1671)

b. that **she** doe not talk or pratll fals

(PCSC ID: 331_40_F1700; Isabel Mackenzie, 1711)

c. only as conserning y=r=[your] brother it is trew that **he & his lady** do-th not now liue together,

(PCSC ID: 125_107_F1600; Marie Stewart, 1600)

d. so I found by it, that **both your leaters and main** do-eth almost al miskearie,

(PCSC ID:1215_110_M1650; Alexander Montgomery, 1678)

(46) a. only **I** wish for weell do-es

(PCSC ID:1143_38_M1650; George Mackenzie, 1699)

b. and **I** pout my wholl troust under god on you and do-es estime you the chiefest anchour of my hopes

(PCSC ID:1185_80_M1650; James Ogilvy, 1658)

c. if **ye** that ar my frends doe nott urge the Erle of Mar

(PCSC ID: 125_107_F1600; Marie Stewart, 1600)

There are only 2 examples of weak-adjacent *do* in total, one with a non-NSR subject (47-a) and the other with an NSR subject (47-b), and there is only one example of the "Buckie pattern" of having pre-verbal negation and no *do*-support (47-c).

(47) a. cationry is what no body willingly **do-es** subject to [if] it can be avoided

(PCSC ID: 1466_100_M1700; Patrick MacDowall, 1711)

b. I but looked vpon it as I still **doe**

(PCSC ID:1375_9_M1700; John Gordon, 1705)

c. I not fayll [to] be at zour l[ordship]

(PCSC ID: 410_66_M1540; James Cockburn, 1546)

Summary: *Do* and the NSR

In testing RQ2B, whether *do*-support in the PCSC is showing sensitivity to NSR constraints, the following pattern was found:

1. Affirmative declarative *do*-support occurs more frequently with plNP subjects than with other subject types.

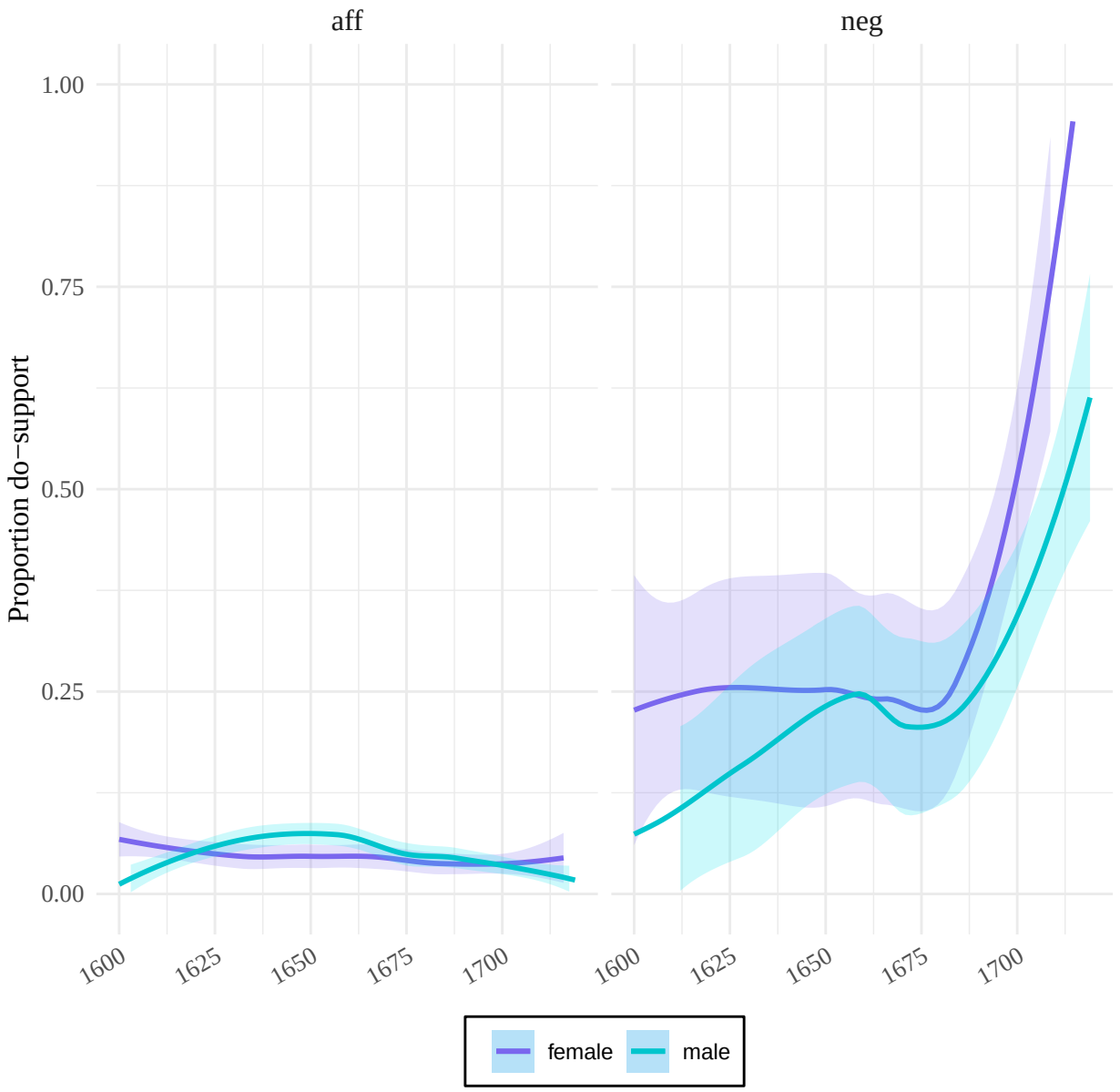
2. Affirmative declarative *do*-support does not occur more frequently with non-NSR subjects compared to other subject types.
3. Negative declarative *do*-support occurs more frequently with non-NSR subjects compared to other subject types.
4. There is no significant difference in *adjacency* effect on affirmative declarative *do*-support with NSR subjects, and there are too few examples of non-adjacent negative declarative *do* with NSR subjects to test the adjacency effect in those contexts.
5. When *do*-support appears, it follows the NSR pattern similarly to other verbs.

Only the finding in 3 is consistent with an NSR condition being present on *do*-support, in the way predicted in Section 5.3. Given the small number of 3sg pronoun subjects, it is not obvious which subject type effect applies here; whether the effect is connected to NSR *subject type* constraints, or the same subject type effect observed for intermediate *do*, as reported in 5.5.2. However, as the findings in 1, 2 and 4 are inconsistent with an NSR constraint being present on *do*, such an hypothesis does not find much support in these results. On the other hand, the finding in 5 indicates that *do* slots into the NSR system similar to other verbs, and does therefore not appear to compete with the default *-(i)s* inflection. The same bump in the late seventeenth century as observed in the intermediate *do* data is present for negative declarative *do*-support with non-NSR subjects, and with NSR subjects when KNOW-class verbs are excluded, and it is contemporary with this bump that we see the higher proportions of affirmative declarative *do* with plNP subjects.

5.5.4 Gender

The last potential factor affecting the rise of *do*-support in Scots investigated for this study is the effect of writer gender. Figure 5.11 shows the results for overall proportions of *do*-supported affirmative and negative declarative clauses from 1600-1720, stratified by gender.

Figure 5.11: Frequency of affirmative and negative declarative do-support by writer gender (LOESS curves)



In the negative declarative clauses, it appears that women are leading the change after 1675 – this result matches what Nurmi (2011) found for English *do*. An indication of a bump in the proportions of use of negative declarative *do* by male writers can again be seen in the late seventeenth century.

5.6 Discussion and conclusion

The findings presented in this chapter indicate that seventeenth century Scots *do*-support, as exhibited in the PCSC, shares many similarities with intermediate *do* in English, as reported by Ecay (2015). Primarily, these similarities lie in *do* occurring more frequently with verbs selecting agentive subjects, compared to KNOW-class and unaccusative verbs, and in that *do* occurs more frequently with non-pronominal subjects than with pronoun subjects. There was no convincing evidence for that a similar NSR subject type constraint as observed in Buckie Scots (Smith 2000) was present during the rise and regulation of *do* in the PCSC. In fact, *do*-support was found to be incorporated in the (declining) NSR pattern similarly to lexical verbs. One surprising result was the preference for affirmative declarative *do* with plural NP subjects; a speculative theory for why this may be, is that *do* could have briefly functioned as a facilitator auxiliary for speakers, in a period where *-(i)s* inflection, as conditioned by the NSR, was variable and declining with plNP subjects. This, however, does not explain why the same effect is not visible with non-adjacent NSR subjects, which are equally affected by the decline of the NSR.

At the end of the seventeenth century, indications of a similar bump in the trajectory of *do* as noted 100 years earlier for English can be seen in the PCSC data. This is only present in the data for contexts where *do*-support is favoured; it is, for example, only present for NSR subjects and pronoun subjects when the KNOW-class, which disfavours *do*-support with these subjects, is excluded. After the 1675 bump, *do*-support begins to increase rapidly in all contexts. The shift in 1675 coincides with the regulation of 3sg *-(i)s* inflection, as observed in Chapter 4 – more in-depth analysis of these S-V agreement mechanisms would be needed to determine whether this is a related change. Another future avenue of research which has not been addressed in this study, is to investigate the function of *do* as a facilitator verb, e.g. by testing whether *do* has a preference for appearing with English or Scots verbs.

Chapter 6

The rise of verbal -*ing*

6.1 Introduction

The development of verbal gerunds, that is verb forms ending in -*ing*, from nouns with the de-verbal suffix -*ing* is well-documented in English, with an extensive body of literature investigating the distribution of the gerund and theories of how this change came to be (Dal 1952, Tajima 1985, Jack 1988, Houston 1989, Fanego 1996, 2004, de Smet 2008, 2013, Fonteyn 2019, to name but a selection). The form in question and its syntactic environments in Present-Day English (PDE) are illustrated in (48) (adapted from Fischer and van der Wurff (2006: 178), with my own additions) where we see phrasal (or, nominal) and clausal (or, verbal) gerunds appearing as clausal complements or subjects to verbs (48-a), complements of P (48-b), and a free adjunct (48-c). The gerunds are in bold font, and the clause or phrase containing them is highlighted in square brackets.

(48) a. *Phrasal*:

[The **finding** of the body] was a crucial breakthrough.

Alex documented [the **finding** of the body] in the field journal.

Clausal:

Alex regretted [**having** to leave so early]

[**Leaving** the party early] was necessary to catch the train

b. *Phrasal*:

Alex was interrupted by [the **passing** of a train]

Clausal:

Alex was offended by [the train **passing** them]

c. [**Having** left the party early], Alex caught the train on time.

In English, the development of verbal gerunds is typically described as involving two different changes, the extension of *-ing* to replace the Old English (OE) participle suffix *-Vnd(e)*, and the verbalisation of nouns ending in *-ing*. This led to the current situation in which verbal stems with the *-ing* suffix are used to express a wide range of functions. While the first of these changes, from *-Vnd(e)* to to *-ing*, took place already in the early thirteenth century (e.g. Lass 1992), the verbalisation of nominal gerunds happened in stages, with the forms acquiring more verbal characteristics at each stage, starting in the Middle English (ME) period and developing more rapidly during the fifteenth-eighteenth centuries, i.e. the Early Modern period of English (e.g., Fanego 1996, Fischer and van der Wurff 2006: 178-9, Nevalainen and Raumolin-Brunberg 2003, Fonteyn 2019).

While there is no consensus in the literature on the causality of this change, i.e. whether the extension of *-ing* to present participles caused the verbalisation of nominal gerunds, evidence from Scots and Northern English suggests that these changes are more independent from each other than has traditionally been assumed (e.g. Jack 1988, Zehentner 2014): the change from *-Vnd(e)* to *-ing* took place about 400 years later in Scots than in English, in the sixteenth century, and at this point the verbalisation of nominal gerunds was already underway. The expansion of *-ing* to replace *-Vnd(e)* in Scots has been suggested to be induced or influenced by anglicisation pressures, but no such claims have been made for the verbalisation of the gerund. In fact, there is very little research done on this change in Scots, and remarkably few quantitative studies (Zehentner 2014 being a noteworthy exception).

The case study presented in this chapter, a quantitative investigation into the trajectory of the verbalisation of gerunds in the PCSC, fills this gap in our knowledge. The chapter is organised as follows: an overview of the of the existing scholarship about English and Scots gerunds is given in 6.2. Section 6.4 gives details of the method of retrieving the relevant phrases and clauses from

the corpus is detailed. The results, in Section 6.5, indicate that the trajectory of *-ing* is different in Scots from English, in that it does not appear to verbalise first in prepositional complement positions, as indicated by that the decline of nominal *-ing* in these typical gerund positions take place 100 years after nominal *-ing* declines overall. It is also found that women are favouring *-ing* forms over *-and* as participial *-and* is declining.

6.2 Background

The origin of the verbal gerund lies in the Old English suffix *-ing/ung* which initially derived abstract action nouns from verbs (e.g., *hunt* > *huntung*, Fanego 1996: 98). That is, the *-ing* forms, while having verbal stems, behaved syntactically as nouns in the Old English (OE) period (e.g., Jack 1988: 16). Eventually, the *-ing* suffix became productive enough that not only abstract nouns were derived, and any verb could be de-verbalised, which is still a function of the derivational suffix *-ing* in Present-Day English (PDE) and Scots. The verbal gerund emerged in the Middle English period, when the *-ing* forms increasingly showed more clausal qualities (e.g. Jack 1988: 17, Fanego 1996: 98); in a gradual process starting around 1200, gerunds began to appear with adverbs (49-a), followed by direct objects not governed by an *of*-phrase around 1300 (as in (49-b), where the second gerund, 'casting' takes a nominal direct object while the first, 'ordaining', takes an *of*-phrase), and finally, in the Early-Late Modern period, the verbalisation of the gerund has become more regularised so that the gerund forms are able to express tense and aspect (such as in (49-c); Fanego 2004).

(49) a. ... The quickly doing of it is the grace

(Jonson, Alchemist (Everym.) IV, ii, p.62; Fischer and van der Wurff 2006: 179)

b. Sain Jon was ... bisi In **ordaining of priestes, and clerkes**, And in **casting kirc werkes**
'Saint John was ... busy ordaining priests and clerics, and in committing himself to church works'

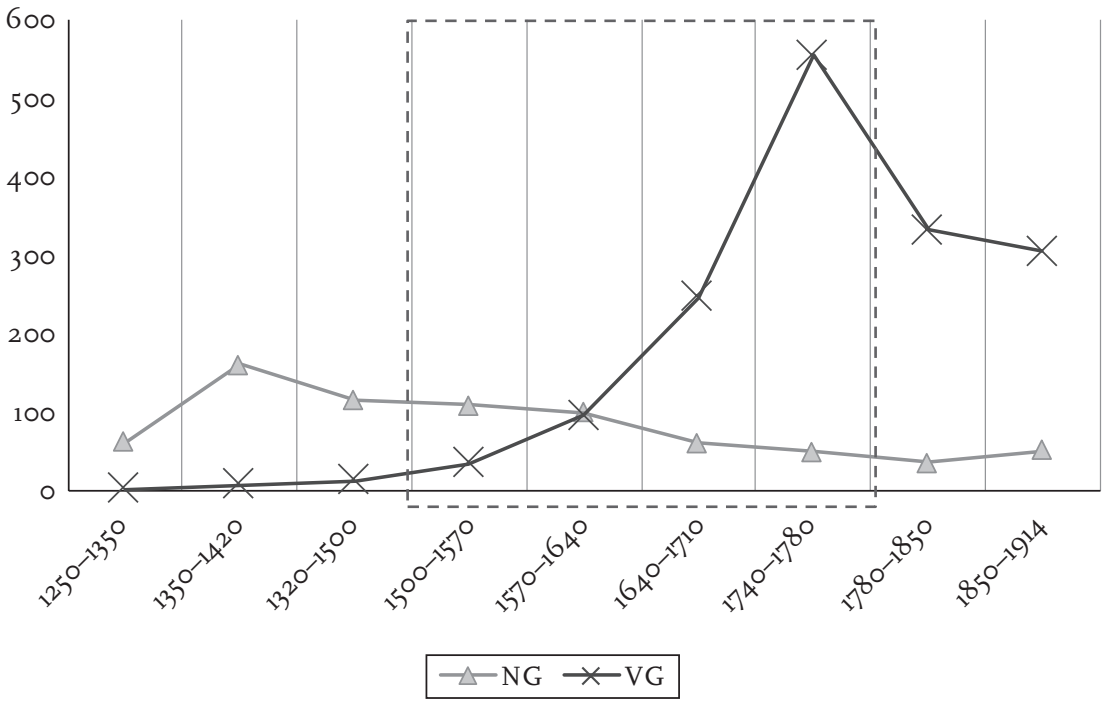
(c1300 (MS a1400) English Metrical Homilies 112/2-4; Tajima 1985: 76)

c. He is so subtill and full of all craft and fleight, that no earthly creature can escape from **being seduced by him.**

(1593, PPCEME; Fonteyn 2019: 44)

Figure 6.1 from Fonteyn (2019: 44) shows the frequencies of nominal and verbal gerunds from the Late Middle to Early Modern period, and clearly demonstrates the sharp increase of verbal gerunds during Early Modern English (which is highlighted with a dashed box in the graph).

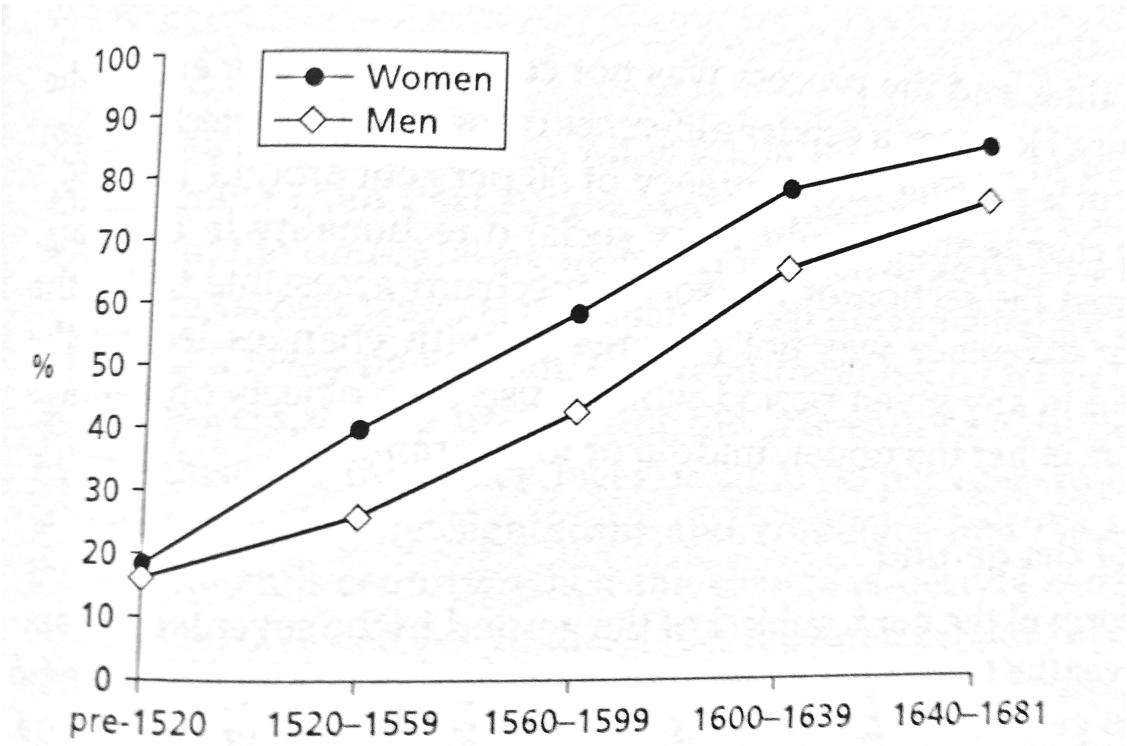
Figure 6.1: Normalized frequencies over 100,000 words of nominal and verbal gerunds from 1250 to 1914



(*NG= Nominal Gerunds, VG = Verbal Gerunds; based on the Leuven English Old to New corpus (Petré 2013) for Middle English (1250–1500), the PPCEME (1500–1710), and the PPCMBE (1710–1914)*)

By measuring the occurrences of gerunds with a bare direct object (as opposed to an *of*-phrase), Nevalainen and Raumolin-Brunberg (2003: 121-2, 196) find that this, like negative declarative *do*, is another change led by women in the data from the *Corpus of Early English Correspondence* (CEEC), as seen in Figure 6.2 (from Nevalainen and Raumolin-Brunberg 2003: 121).

Figure 6.2: Gerunds with bare direct objects in the CEEC



Various theories have been proposed to account for how and why this verbalisation took place, and the debate is not settled yet. Using the comprehensive data on gerunds in Middle English (ME) collected by Tajima (1985), Jack's (1988) account scrutinises the more commonly held theories at the time of his study, concluding that the most likely explanation for the development of a verbal *-ing* form is that it is the result of a combination of influencing factors:

– OE constructions with a genitive noun preceding the verbal noun (e.g., *aligra boca rædinge*, 'reading of holy books' (Jack 1988: 46)) became re-analysed as consisting of a common case noun as argument to a verbal gerund, after the collapse of genitive and common case marking. This resulted in early verbal gerund clauses which had nouns functioning as subjects or objects preceding the gerund, in the ME period. These constructions, Jack (1988: 49-50) argues, cannot be the sole source of verbalisation, e.g., because other nouns than gerunds would also appear with unmarked genitives without becoming re-analysed as a result, but it is likely to have influenced the development in part.

– Infinitives behaved similarly to gerunds, in that they could appear as verbal complements (object, predicates) or subjects, but they were limited in that they could not be complements of prepositions – gerunds were able to fill this gap (Jack 1988: 61-62). This led to an increase in use of gerunds, as they met the demands of the increasing usage of prepositions after the collapse of the nominal case system. However, this theory does not account for why this required verbalisation of gerunds, as nominal gerunds were equally able to fill this gap and favoured prepositional complement positions (e.g., Houston 1989, Fanego 1996: 124-5).
– Influence from French in the ME period may have had an effect on the development, as Old French had similar gerund constructions (*gérondif*), which also appeared as prepositional complement with an adverbial function (Jack 1988: 50-5).

Jack (1988) discusses an additional hypothesis from Dal (1952), who suggests that the *-ing* suffix became reanalysed from a derivational noun-forming suffix to a verbal inflectional suffix. This reanalysis would have come about as a result of the increased productivity of *-ing*, mentioned above, as it no longer only formed abstract nouns. Jack rejects this proposal, apparently largely on the basis that the derivational suffix remains productive in English alongside the verbal inflection: "it follows that the [verbal] gerund cannot be seen simply as a stage reached through fulfilment of a potential inherent in the older verbal noun; the [verbal] gerund was evidently a new form that arose in addition to the older verbal noun [=nominal gerund], and not simply a product of evolutionary change in the verbal noun" (Jack 1988: 43). However, this type of reanalysis finds parallels in other historical changes in English, such as the change from de-verbal nouns to bare infinitives and *to*-infinitives, and from de-verbal adjectives to verbal participles (Los 2016). Indeed, as Los (2016: 274-7) notes, paradigmatic gaps and incomplete changes are a common phenomenon in language change at large, and tolerated by speakers, so this does not seem to be a sufficient reason for why this theory of reanalysis is rejected.

Fanego (2004), de Smet (2008, 2013), and Fonteyn (2019) propose that the verbalisation, and thus early competition between verbal and nominal gerunds, began with bare gerunds – i.e., gerunds with no pre-nominal element – which were ambiguous with respect to their nominal or verbal status. Both Fanego (2004) and Fonteyn (2019) base this analysis on the fact that the verbalisation started

with, and is led by, bare gerunds in prepositional complement position (see also Houston 1989). Fonteyn (2019: 46-7) further observes that, while bare nominal gerunds decline, nominal gerunds with indefinite articles continue to increase in the Early Modern period, which she suggests may be a functional differentiation between the two gerund types: as bare gerunds become verbalised, the non-bare nominal gerunds become further nominalised.

Fanego (2004) identifies structures with temporal or locative adverbs or prepositional phrases (50-a), clausal complements (50-b), or ambiguous adverbs or adjectives (as in (50-c)) as the type of ambiguous structure which could allow either a verbal or nominal interpretation of the gerund, without any surface change until the first clearly verbal structures (as in (49)) appear. In the examples below, originally from Tajima (1985), the gerund clause or phrase is in bold font.

(50) a. Of þ**i comyng at domesday**
"Of the coming at Doomsday"

(c1280; Fanego 2004: 21)

b. þe messanger made anon **asking Whi he made swich leiȝeing**
"the messenger immediately asked the question why he was telling such lies"

(a1300 (MS c1330) Arthour and Merlin 1301?02; Fanego 2004: 20)

c. Vnder þe Monument ȝeo stod wiþoute **wepyng sore**
"she stood close by the sepulchre without weeping bitterly/without bitter weeping"

(c1280 Southern Passion 1874; Fanego 2004: 24)

The clause in (50-c) is, which was first reproduced from Tajima (1985) by Jack (1988: 57), is frequently used as an example to illustrate a case of ambiguity arising from the ambiguous status of 'sore' as an adverb or adjective. However, on a closer look, it appears that 'wiþoute' has been wrongfully interpreted as the preposition *without*, when a more appropriate translation would be 'outside' in this context.[1] Indeed, the immediately following sentence in the Southern Passion

1. Many thanks to Bettelou Los for bringing this to my attention!

reads: 'As ȝeo stoupede a-doun, ȝeo by-heold þer ney Wiþynne þe Monument twey aungeles' (= 'As she stooped down, she beheld there nearby inside the monument two angels'). Thus, this is more likely to be a participial adjunct than an ambiguous gerund in prepositional complement position (which I also infer is Jack's (1988) analysis of this example). Fanego (2004) does not provide another complete example of this type of ambiguous structure, but notes that adverbs derived from adjectives by means of the suffix -*e:*, e.g. *heard* ('hard, severe') > *hearde* 'severely'), and other coalesced adjective and adverb forms, such as *michel* ('much'), would be the type of forms which give rise to this ambiguity.

In the process of verbalisation, gerund structures emerged which had pre-nominal and post-verbal elements, so called "mixed gerunds" or "hybrid" structures. Fanego (1996: 107-8) categorises these hybrid forms into two classes: those with a possesive pre-head (51-a), which are still used in PDE, and "proper" mixed gerunds with determiners as pre-heads (51-b), which have fallen out of use. The proper mixed structures emerged late in the verbalisation process, because they are the most nominal in nature and thus resisted verbalisation the longest (Fanego 2004: 38).

(51) a. [...] to settle notwithstanding **its hanging freely** at the bottom of the string.

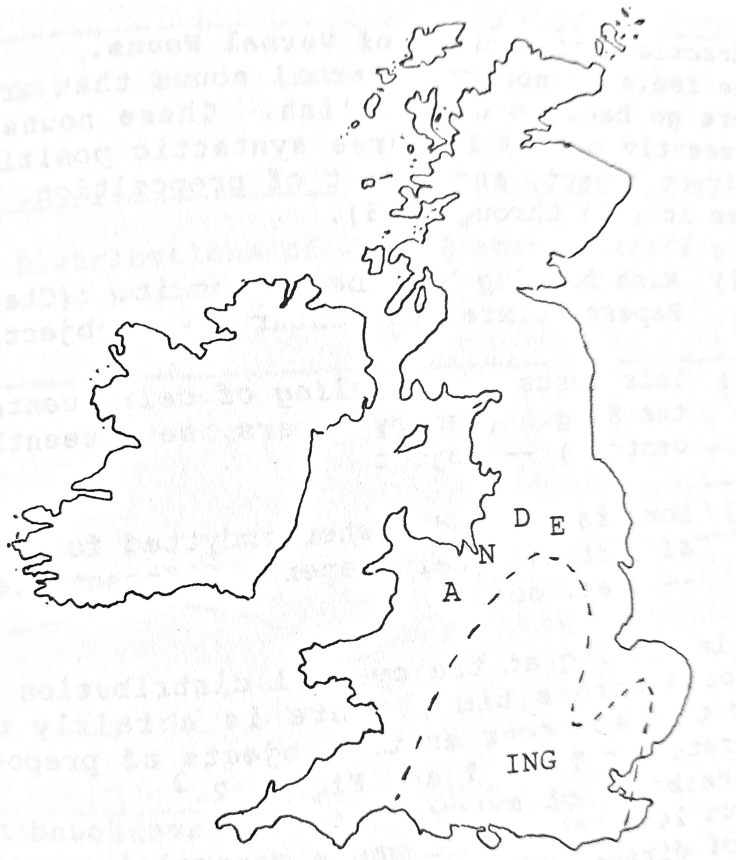
(1675-1676, Boyle *Electricity*; Fanego 1996: 107)

b. [...] for **the inableinge the minister** so profettably to declare the ward

(1599-1601, Hoby *Diary*; Fanego 1996: 107)

In summary, both the structures with genitives preceding nominal gerunds, and the ambiguous examples highlighted by Fanego (2004) could have been influential in catalysing a reanalysis of the nominal gerund into a verb, which may well have been aided by influence from French *gérondif* constructions. This re-analysis would perhaps not have been so readily available, if it had not been for the fact that verbal -*ing* forms were already present in English via the present participle in -*ing*, as a result of the collapse of earlier participial -*and* endings with -*ing* in the thirteenth century. Most dialects of Middle English had -Vnd(e) participial endings, before the -*ing* endings spread outwards from their origin in south-central England (see Figure 6.3).

Figure 6.3: The geographical distribution of participial - *Vnd(e)* (ANDE) and -Vng(e) (ING) in England, 1450.



(*Figure based on Moore, Meech, and Whitehall (1935), reproduced by Houston (1989)*)

Houston (1989) notes that appositive participial clauses have similar discourse functions to prepositional phrases with gerund complements, in that they provide commentary on the place, time, and manner of an event (see (52)). This, Houston (1989) argues, would have led to a conflation of the participles in *-ing* with gerunds in prepositional complement position, facilitating the verbalisation of gerunds.

(52) a. *Prepositional phrase with gerund*

Dr. Parkins, **at his first coming out of Denmarke**, made his braggs that [...]

(Houston 1989: 187)

b. *Appositive participle*

Going to preach, H. Morley of my parish deliv'd mee a note of receipt of my procurations

(Houston 1989: 183)

That the collapse of participial -*Vng(e)* and -*Vnd(e)* and the verbalisation of nominal forms in -*ing* should have a causal relationship (where the former contributed to the latter) is one of the theories rejected by Jack (1988), using evidence from Northern ME: there are early examples (pre-1350) of verbal gerunds in Northern ME texts, at a time when the same texts still had participial -*Vnd(e)*. If participial -*ing* played a role in the verbalisation of gerunds, then this co-occurrence would be unlikely. However, this does not rule out that these forms could have influenced the reanalysis in dialects which already had participial -*Vng(e)* (e.g., Fanego 1996), as this meant that constructions with a bare -*ing* form followed by a direct object would have been an available grammatical option.

Scots retained participial -*and* until as late as the sixteenth-eighteenth century when it increasingly became replaced by -*ing* (e.g., King 1997; Macafee and Aitken 2002) – the timing of this change has led some to assume it is brought on by anglicisation (e.g., Devitt 1989), but it may be that anglicisation reinforced a change that was already underway, initiated by a phonological collapse of /nd/ and /ng/ into /n/ (e.g, Agutter 1990; King 1997; Macafee and Aitken 2002). Thus, the move from -*and* to -*ing* in writing could simply reflect a spelling change, rather than a change in morphology. According to Alexiadou (2011: 148), older speakers of Scots still differentiate between verbal gerunds and participles, as in (53), but she does not provide information about her sources or what dialect this is observed in.

(53) a. *Present Participle*:

He wis aye **stravaigan** aboot.

He was always roaming around,

b. *Gerund*:

He's fond o **stravaigin** aboot.

He likes roaming around.

As regards the verbalisation of *-ing*, the Northern ME examples highlighted by Jack (1988) are attested at a time when there is very little evidence of written Scots, and there was no clear verdict in the literature on when the verbalisation of gerunds took place in Scots until Zehentner (2014). Indeed, in her quantitative study on the *Helsinki Corpus of Older Scots*, measuring normalised frequencies of the investigated forms by 10,000 words, Zehentner (2014) finds that the gap between the emergence of verbal gerunds and the loss of participial *-and* is shorter than what has been assumed; frequencies of verbal gerunds only start to increase significantly towards the end of the sixteenth century, and participial *-and* rapidly declines over the same period. Hence, Zehentner (2014) concludes that its plausible that the two changes may have influenced each other in the case of Scots, and that they should therefore be studied as two parts of the same phenomenon. A straightforward connection between these changes and anglicisation pressures in the time period when they take place has not yet been investigated.

6.3 Research questions and predictions

The aim of this case study is solely exploratory, seeking to investigate the decline of participial *-and* and verbalisation of *-ing* in different contexts, using parsed data. Hence, I ask:

RQ1: What are the proportions of nominal, verbal, and hybrid *-ing*-forms over time in the PCSC?

RQ2: Does the decline of *-and* forms in favour of participial *-ing* take place within the time period covered by the PCSC?

In relation to these research question, the prediction, in line with Zehentner's (2014) findings, is that these two changes occur simultaneously, in the late sixteenth century.

I will also investigate the effect of gender on these changes; the findings by Nevalainen and Raumolin-Brunberg (2003), predict:

A. Female writers are leaders in the rise of verbal gerunds in Scots.

Furthermore, the assumption that the loss of -*and* is an effect of, or enforced by, anglicisation, gives rise to the following, opposite, prediction:

B. Female writers favour participial -*and* for longer than male writers.

6.4 Methodology

A methodological challenge of this study was to identify the gerundial noun, hybrid, and verb forms in the corpus, not only in terms of their syntactic behaviour but also their morphological form. Assuming that -*ing* suffixes do not differ in form depending on whether they combine with nouns or verbs, I used the CS `make_lex` function to extract all forms tagged as VAG, BAG, HAG, or DAG (i.e., present participles and verbal gerunds), manually went through the resulting list of verbs, and put their inflectional endings into a .def file (sorted into one variable for -*ing* suffixes and one for -*and* suffixes). Then, I extracted a list of nouns (N or NS) which ended with any of the suffixes on the lists in the .def file (I added common plural endings to the list of suffixes to capture plural nominal gerunds). I sorted through the resulting list of nouns manually and removed non-gerunds (e.g. *king*). The final list of nouns ending in -*ing* (consisting of ca. 4,900 tokens; there was no item tagged as N or NS with -*and* inflection) was then added to a variable in the .def file. Hence, the lists of verbal suffixes could be used to target verbal gerunds or participles in -*ing* and -*and*, and the list of nouns could be used to target nominal gerunds in the search query.

Constructions described as hybrid forms in the literature – that is, forms with pre-nominal and

post-verbal elements – are parsed in different ways in the PCSC.[2] A small number (n=14) of hybrid cases are tagged as an NP with a noun head and an NP complement, as in (54). Otherwise, the hybrid forms are tagged as an NP, dominating a possessive element or a determiner with an IP-PPL sister (55).

(54) 'The burning the minister of kirkpatriks house'
(NP (D the) (N burning)
(NP-COM (NP-POS (D the) (N minister) (PP (P of) (NP (NPR$ kirkpatriks)))) (N house)))

(55) 'The paying the annualrents'
(NP-OB1 (D the)
(IP-PPL (VAG paying) (NP-OB1 (D the) (NS a*nnual%rents))))

Thus, in order to identify nominal gerunds, verbal gerunds, and hybrid gerunds, the structural definitions which the coding query[3] for this study is based on targets the following structures:

– *Hybrid*:
  1. NPs dominating nominal gerunds which have NP-COM sisters
  2. IP-PPLs dominating verbal gerunds, where the IP-PPL has a determiner or possessive element as sister
– *Noun*:
  All other NPs dominating nominal gerunds
– *Verb*:
  All other IP-PPLs dominating verbal gerunds

2. This is due to an inconsistency in the parsing, which will be rectified before the final version of the corpus is published.
3. The coding query, definition file, and query output are in Appendix 4.

The coding query specifies various other features, such as: the type of *complement* of the gerund (NP, *of*-phrase, PPs where P is not *of*, ADJP, clausal complement), type of *suffix* (*and* or *-ing*), *function* (complement of P, complement of V, or free adjunct), type of pre-nominal element (a possessor, or a determiner), and whether the form has adverbial modification (locative/temporal adverbs, other adverbs, and the negator *not*). Finding instances of gerunds as prepositional complements proved difficult in the case of hybrid structures like (55); CS does not have a function which specifies a mother of the targeted node, so identifying a node based on its mother usually requires workarounds. In cases like (56), the PP grandmother, or prepositional aunt, of the IP-PPL needed to be identified.

(56) 'by my doing things with a seuer[=sure] hand'
(PP (P by)
(NP (PRO$ my)
(IP-PPL (DAG doeing) (NP-OB1 (NS thinges))
(PP (P with) (NP (D a) (ADJ seuer) (N hand))))))

In order to capture these IP-PPLs, I first wrote queries for the other possible positions the IP-PPL could occur in. Then, I manually checked what tokens had not been captured by these first queries, and as they were all complements of prepositions I assigned the remaining tokens as prepositional complements under the *function* column. Just as has been observed in English data, gerunds favour prepositional complement positions, and this is particularly the case with nominal and hybrid gerunds, as seen in Table 6.1.

Table 6.1: Frequencies of gerunds as prepositional complements

	Complement of P (n)	*Total n.*	*Proportion*
Verb	288	2,151	13.4%
Noun	532	862	62%
Hybrid	137	185	74%
Tot:	955	3,198	30%

The low proportion of verbal gerunds in prepositional complement position is likely caused by the fact that participial adjuncts are included in the count; verbal -*ing* forms functioning as adjuncts make up 1,391 tokens.

Finally, I analysed the result of the coding query by (i) binning the data for the analysis of overall proportions of different gerund types by taking the average frequencies within 20-year bins and calculating the 95% confidence interval for each sample (again using the binom package (Dorai-Raj 2014) with the Wilson method), and (ii) fitting the data to a LOESS curve with *year* as a continuous variable, for a complementary analysis of the overall data and to explore the proportions of gerunds and participles in different contexts and functions.

6.5 Results

6.5.1 Verbalisation of gerunds

The result of measuring overall proportions of each type of -*ing* form regardless of function or position can be seen in Figure 6.4 for the binned data, and visualised with LOESS curves in Figure 6.5. In total, there are 2,149 verbal gerunds, and 861 nominal gerunds.

Figure 6.4: Proportions of verbs, nouns, and hybrid gerunds, out of total *-ing* forms

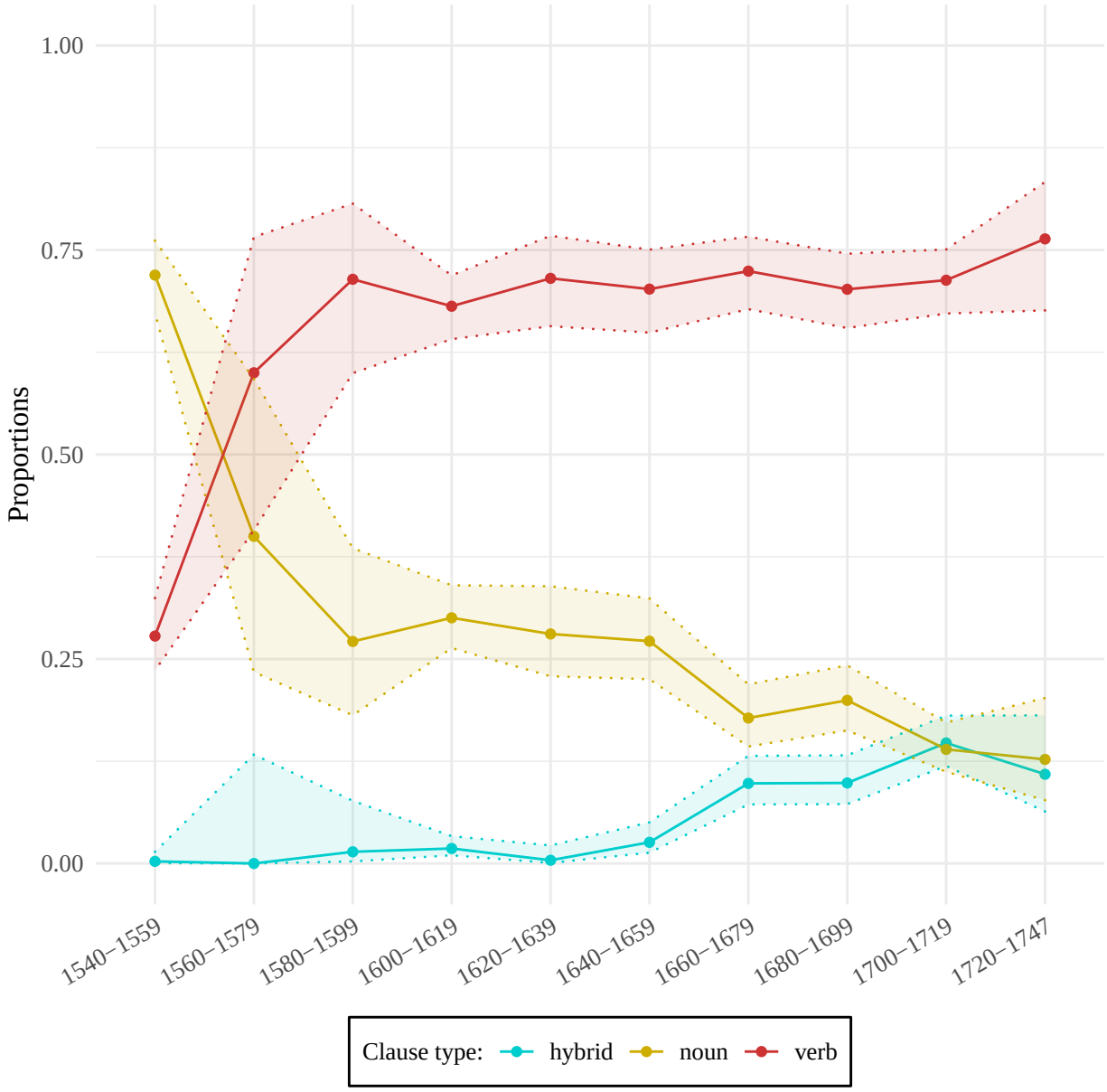
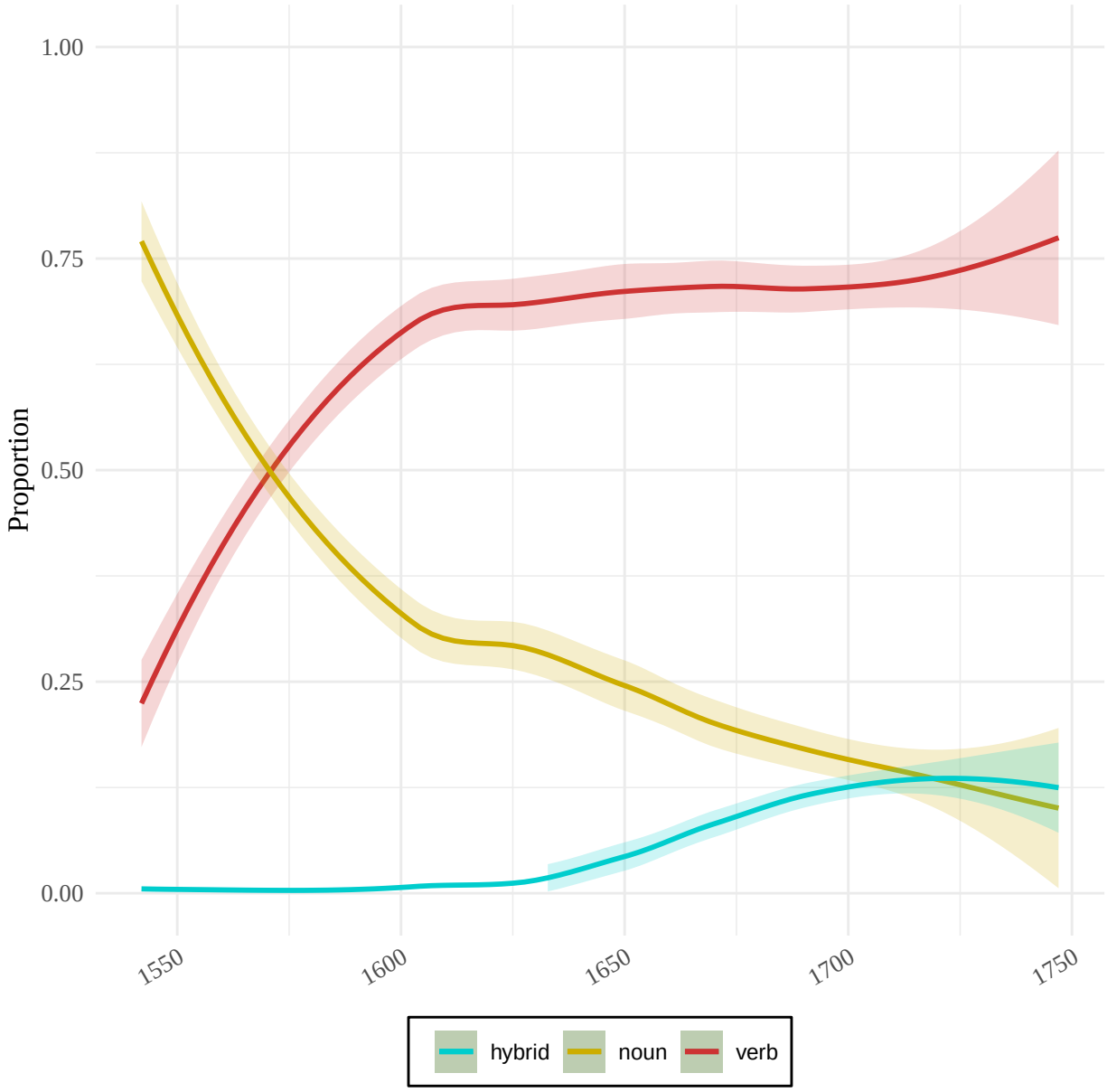


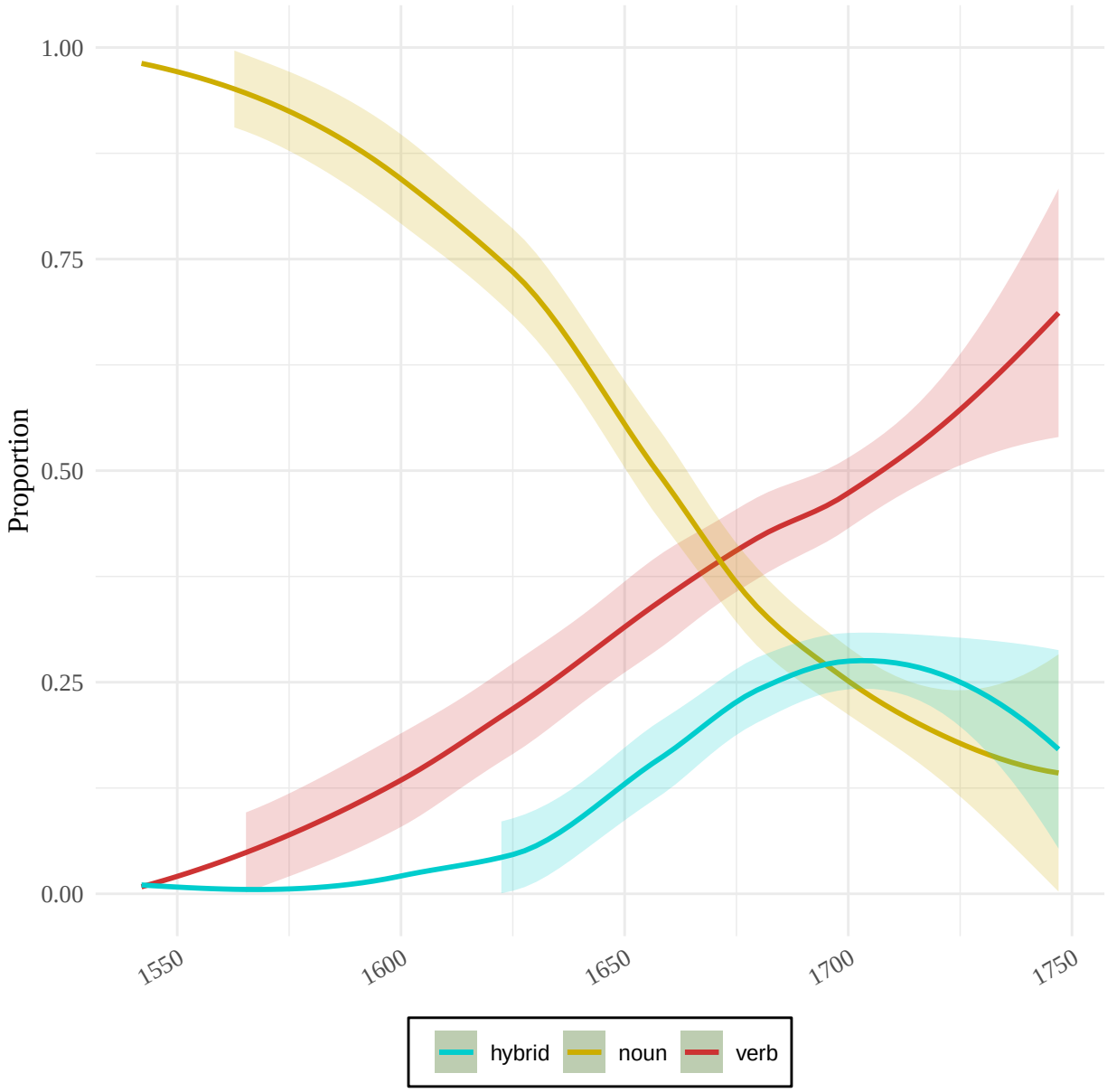
Figure 6.5: Proportions of verbs, nouns, and hybrid gerunds, out of total *-ing* forms (LOESS curves).



There appears to be a sharp shift taking place in the latter half of the sixteenth century, when verbal gerunds increase as nominal gerunds decline. There are only 14 cases of hybrid gerunds in total before 1650, after which their occurrences increase until they appear as frequently as nominal forms after 1700.

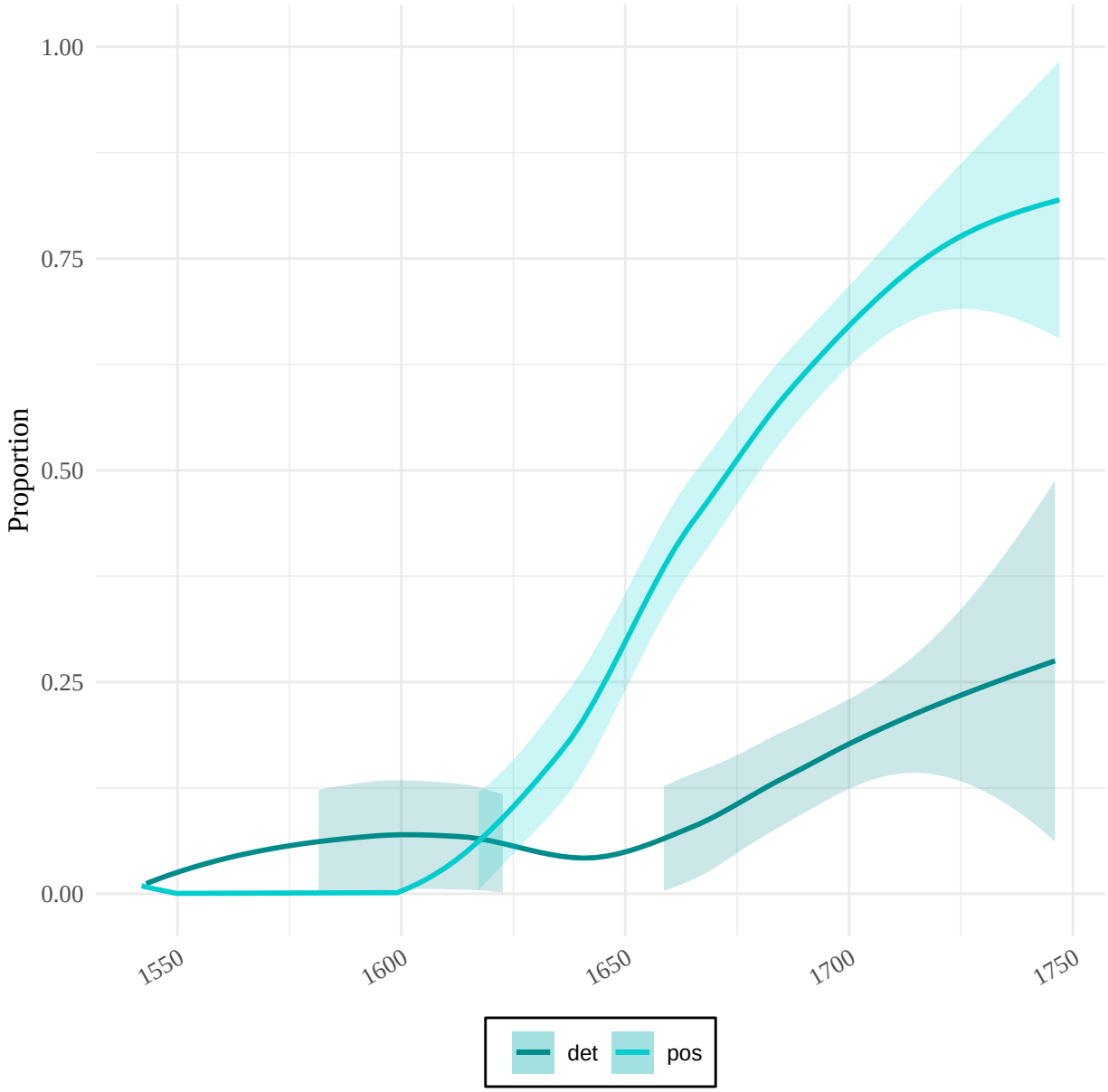
Figure 6.6 measures proportions of gerunds that are complements of prepositions – in this function, the shift happens ca. 100 years later than in the overall data, around 1675. Here, it also appears that the verbalisation of gerunds happens less rapidly. Nominal gerunds in prepositional complement position constitutes ca. 61.7% (531/861), and the same proportion for verbal gerunds is only 13.3% (286/2,149).

Figure 6.6: Proportions of verbs, nouns, and hybrid gerunds, out of total *-ing* forms as prepositional complements (LOESS curves).



Finally, taking a closer look at only the hybrid forms, it can be seen in Figure 6.7 that the proper mixed gerunds, with a determiner as pre-nominal element, are the first mixed forms to emerge, and they are not declining at the end of the period under investigation. Unsurprisingly, the hybrids with a possessive pre-nominal element are far more frequent when they do emerge. The total number of hybrid forms is 75, whereof 24 are proper mixed gerunds and 51 have a possessive pre-nominal element.

Figure 6.7: Proportions of pre-nominal determiners and possessors, out of total hybrid -*ing* forms (LOESS curves).



(57) gives examples of a proper mixed gerund and a hybrid gerund with a possessive pre-nominal element from the PCSC.

(57) a. **the bringing it to a good issue** is the greatest obligation you can put on your affectionat Cousing

(PCSC ID: 1432_66_M1700; Kenneth Mackenzie, 1700)

b. befor yow wer Certan off **your vndertaking journey for Loundone**

(PCSC ID: 1115_10_M1650; John Gordon, 1660)

The investigation into hybrid structures with NP-COM complements also uncovered examples of non-gerundial nouns followed by bare noun complements instead of prepositions (58).

(58) a. I may be of **use [to?]** ye government before the pretender does Come

(PCSC ID:1368_2_M1700; Simon Fraser, 1714)

b. I haif maid delygent lawbo*uris% to haif **Intelligens [of]** quhat is ye causs*s% y=t= Inglis me*n% is fawuorit & ye auctorite no=t= obeyit nor s*er%uit

(PCSC ID:354_10_M1540; Henry Stewart, 1548)

c. & haif cawsit ye wardens to meit schortly for **redres*s% [of]** m=master pamer hand

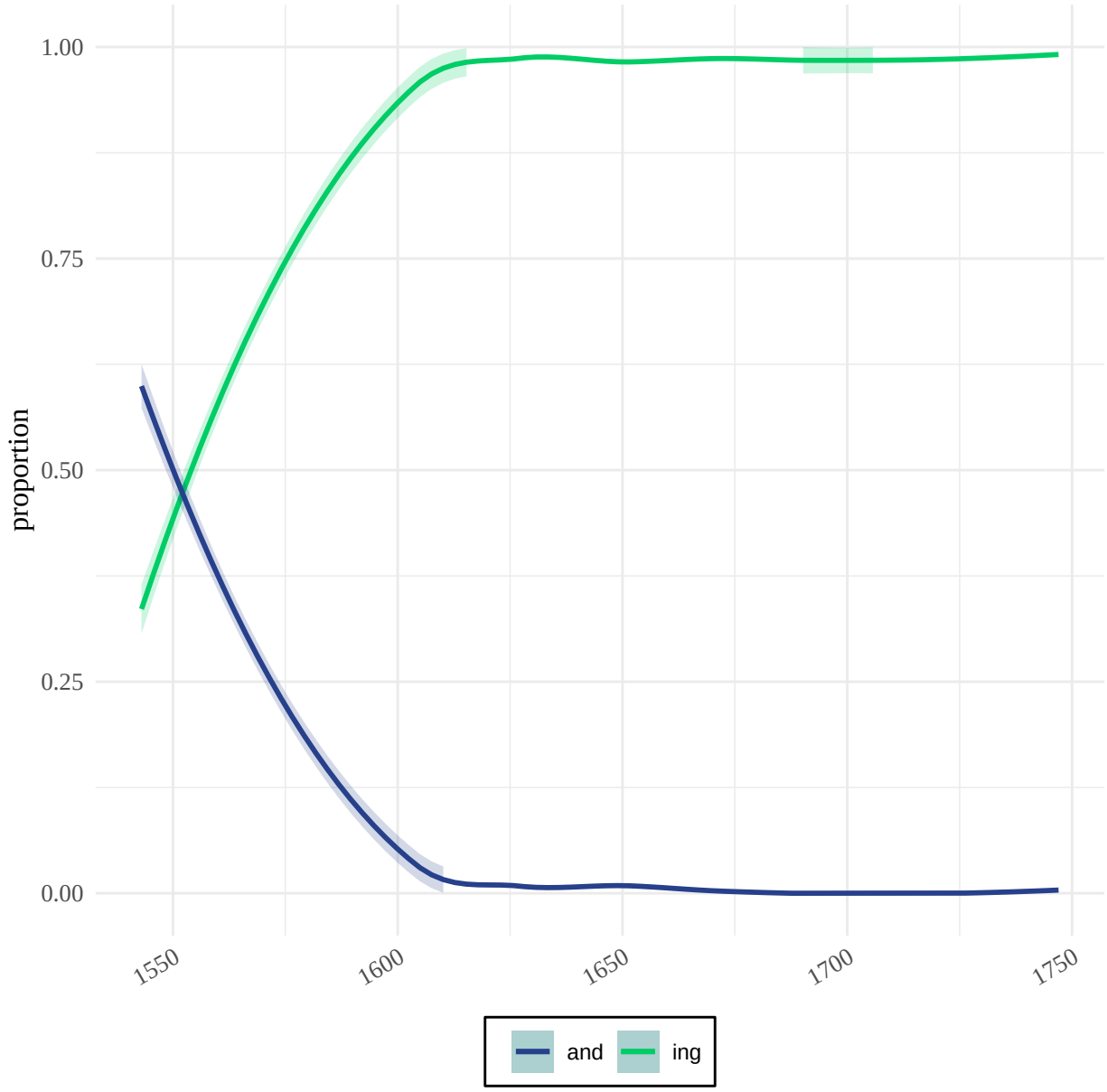
(PCSC ID:360_16_M1540; Gilbert Kennedy, 1552)

However, all but one of these (58-a) occur before 1600, so it does not seem like this is evidence of extension of a grammar which allowed NP complements of N while hybrid forms were on the rise. It is perhaps more likely that the pre-1600 example are residual from a system which employed case rather than prepositions to express genitive relationships between nouns, as they all seem to correspond to *of* phrases in PDE.

6.5.2 The merger of participial -*and* and -*ing*

In measuring the proportions of participial -*and* and -*ing*, it was only possible to look at forms classified as verbs, as there are no nouns or hybrids with an -*and* suffix. The result can be seen in Figure 6.8. The total number of verbs in -*and* and -*ing* is 2,109, whereof 163 have -*and* and 533 have -*ing* before and including 1618. It should be recalled that the change witnessed here could reflect a change in spelling practices, as the verb ending represented by both these forms may have been pronounced as /n/.

Figure 6.8: Proportions of verbs in *-and* and *-ing* (LOESS curves).



Nearly all verbs in *-and* are free adjuncts (156/166; they are either daughters of another clause, or independent roots), and there is only one example of an *-and* form as complement of a preposition, given in (59). However, while *twychand* is participial in form, it mostly functions as a preposition itself in older Scots[4], so this is not a typical example of a verbal gerund.

4. "Tuichand prep.". Dictionary of the Scots Language. 2004. Scottish Language Dictionaries Ltd. Accessed 28 Jul 2022 <http://www.dsl.ac.uk/entry/dost/tuichand_prep>

(59) for as twychand zow=r= l[=lordship] [..]

(PCSC ID: 424_80_M1540, Alexander Gordon, 1547)

Thus, in the PCSC the -*and* endings appear to fall out of use completely by 1650; after 1618, there are only three clauses with verbal -*and*, as seen in (60).

(60) a. I am no to blame to seik yis deweteis **restand** yis fourtene zeiris.

(PCSC ID: 652_16_M1600; William Douglas, 1642)

b. all sowmes thatt ver **resta*n%d awand** to my Father the tyme of his disceass

(PCSC ID: 780_144_M1600, John Erskine, 1643)

c. Conserning thes Muneys that is dew be my Lord Mellwell wnto me **Restand awnd** owt off his land*is% within the parochine off Kirkaldie ffor the Crope 1649

(PCSC ID: 1168_63_M1650; Charles Seton, 1650)

Since all of these examples are with the verb 'restand', and two of them contain the construction 'restand owand' ("remains owed", in relation to outstanding rent or other debts in the examples in (60)), it may be that -*ing* is resisted in these formulaic contexts.

6.5.3 Summary

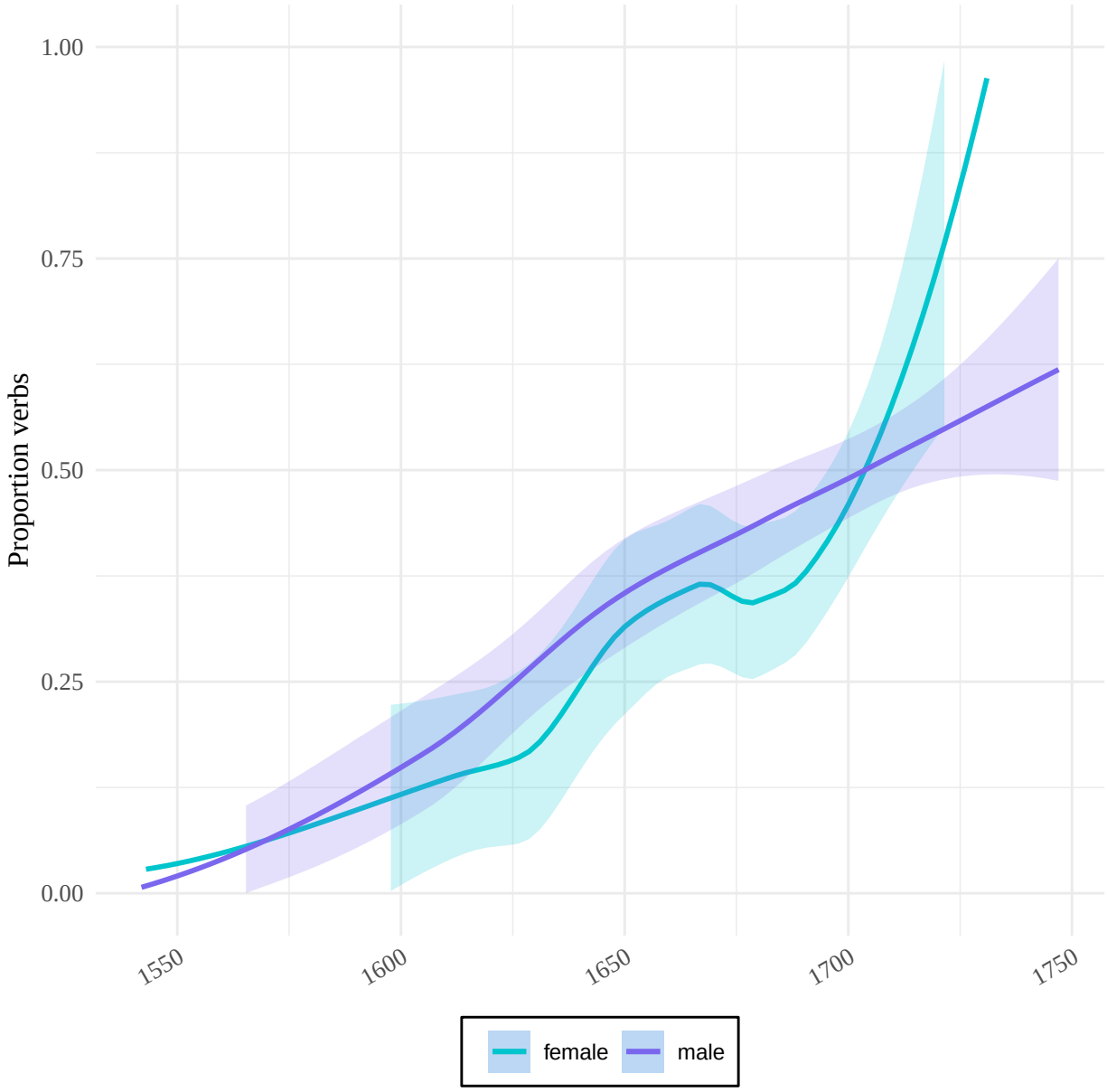
At a first glance, it appears that the findings by Zehentner (2014) are upheld in this data: at the end of the sixteenth century, the overall proportions of verbs in -*ing* increase rapidly while nouns in -*ing* and present participles in -*and* decline. However, on only measuring proportions of gerunds that are in prepositional complement position, the picture changes; the overhaul of nominal gerunds by verbal ones happen about 100 years later than what is seen in the overall data, simultaneously with hybrid structures peaking in their frequencies. Given that nominal gerunds decline earlier in the overall data than in the data for only gerunds that are prepositional complements, it appears

that the verbalisation of nominal gerunds does not start with gerunds that are complements to prepositions. The *-and* suffix appears to be restricted to typically participial contexts, and the verbalisation of nominal gerunds does not lead to, e.g., hybrid forms in *-and* while the suffix *-and* is still in use.

6.5.4 Gerunds and Gender

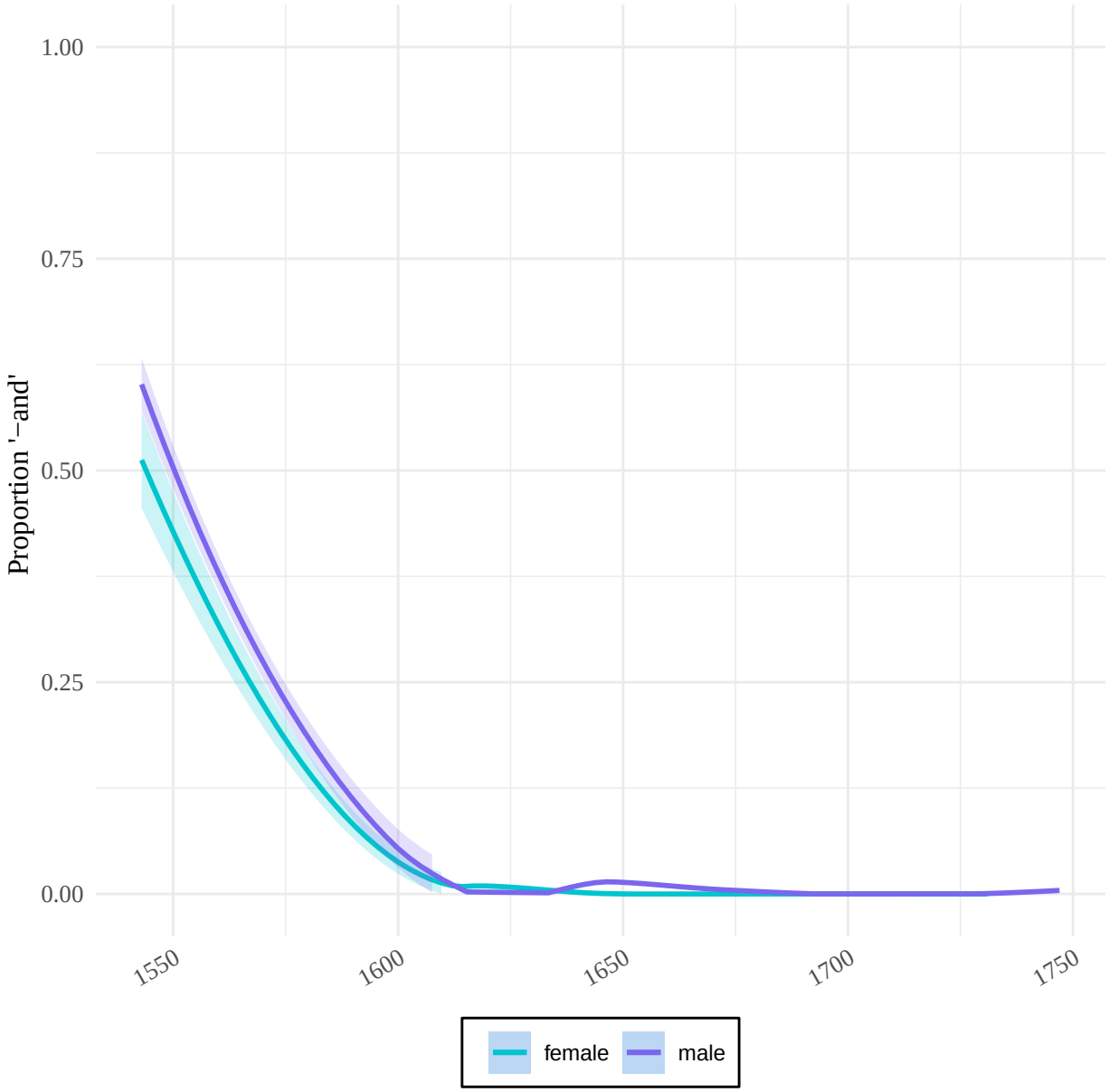
The effect of gender on the verbalisation of *-ing* forms was investigated by looking only at proportions of verbal gerunds in prepositional complement function. This also makes the results more comparable to the findings by Nevalainen and Raumolin-Brunberg (2003), who only include gerunds that are complements of prepositions in their study. The results, in Figure 6.9, indicate a similar time trend in the rise of verbal gerunds between male and female writers until 1700, when women's usage increases more rapidly then that of men. Aggregating the data over the entire period, women use slightly higher proportions of verbal gerunds (93/290, compared to 190/659 for men), but this difference is not significant ($p = 0.315$)

Figure 6.9: The rise of verbal gerunds in -*ing*, stratified by writer gender (LOESS curves).



Stratifying the proportions of participial -*and* by gender (Figure 6.10) shows slightly less usage of -*and* by the female writers (8/182, compared to 158/574 for men). Measuring the proportions until 1625, this difference is significant ($p < 0.0001$)

Figure 6.10: Proportions of verbs in *-and*, stratified by writer gender (LOESS curves).



6.6 Discussion and conclusion

The findings of this study brings more clarity to the rise of verbal *-ing* forms in Scots. As regards Research Question 2, *Does the decline of -and forms in favour of participial -ing take place within the time period*, the results show that the loss of *-and* is largely completed by 1618 (with the exception of a few, possibly formulaic, occurrences in the 1640s). As regards RQ1, *What are the*

proportions of nominal, verbal, and hybrid ing-forms over time in the PCSC?, it was shown that the verbalisation of *-ing* only appears to be simultaneous with the decline of *-and* when testing data in all contexts where *-ing* appears, which is likely due to the *-ing* forms increasing in participial contexts as the *-and* forms decline. However, when testing the proportions in a position typical for gerunds, prepositional complement position, the decline of nominal *-ing* happens ca. 100 years after the decline of *-and*. The overall data indicates a decline of nominal *-ing* in other positions before this time, which suggests that the verbalisation did not start with prepositional complements as has been found for English. Similar to English, the hybrid forms emerge at a late stage in the change, at the same time as verbal complements of prepositions overtake nominal complements of prepositions, which indicates that this is when the more crucial reanalysis of *-ing* forms take place. While this case study has only scratched the surface of the complex topic of *-ing* forms in Scots, future research should seek to uncover the finer details of the various contexts these forms emerge in.

Finally, as regards the predictions regarding the role of women in the rise of *-ing* forms in Scots, it was found that (i) women and men use verbal *-ing* at similar levels until around 1700, when the usage of women increases, and (ii) women use significantly less *-and* forms, in favour of *-ing* during the period of decline of participial *-and*. These results do not show the same clear leading role for women in the verbalisation of *-ing* as has been found for English, but it seems that women are adopting participial *-ing* quicker than men, which would be unexpected if the decline of *-and* is an anglicisation outcome. Thus, neither prediction A or B holds up in these results.

Chapter 7

Discussion: Is syntactic change in Transition Scots an anglicisation outcome?

7.1 Syntactic outcomes of language contact

7.1.1 Determining the likelihood of contact-induced change

The findings of the case studies presented in Chapters 4-6 all give evidence of syntactic change taking place during the transition period of Scots. Thomason, in an attempt to broadly define what kinds of linguistic changes count as contact-induced, states:

> "any linguistic change that would have been less likely to occur outside a particular contact situation is due at least in part to language contact" (Thomason 2001: 62).

This definition is particularly tricky to apply when it comes to such closely related languages as Scots and English, as it is notoriously difficult to discern what changes would have occurred independently of the contact between these two languages, after they split from their shared ancestor (i.e., developed through *drift*). Therefore, we cannot assume that changes are induced by contact, or a result of borrowing, without considering other factors than, simply, change towards similar outcomes. Poplack and Levey build on Thomason's (2001) definition, by specifying more clearly how a particular feature could be determined to be contact-induced.

> "A candidate for contact-induced change in a contact variety is present in the presumed source variety and either 1) absent in the pre-contact or non-contact variety, or 2) if present (e.g., through interlingual coincidence), is not conditioned in the same way as in the source, and 3) can also be shown to parallel in some non-trivial way the behavior of a counterpart feature in the source" (Poplack & Levey 2010: 398).

Thus, if the features under investigation can be shown to be present in English (as the source variety) and not have been present in Scots before the contact event took place, it is more likely to be contact-induced. The criteria in 2) and 3) above, if my interpretation is accurate, can perhaps be better understood as relating to *shared grammaticalisation*, as described in Robbeets and Cuyckens (2013); if a candidate feature for contact-induced change is identical to the source language feature in terms of its distribution and function, it is likely to be inherited. On the other hand, if the feature differs between the languages involved in the contact, it is more evidence for either a contact-induced change, or different developments through *drift*. This diagnostic has been applied to assess the extent of contact-induced transfer or shared inheritance in common features of Old Frisian and Old English (by Colleran 2017). In Colleran's (2017) study, and in many of the studies which have given rise to the discussed criteria, it is unknown whether the proposed source language originated a feature or whether it was part of a shared ancestor of both languages. In the case of English and Scots in the Early Modern period, we have a reasonably good idea of which of the investigated features were common to both languages before the contact scenario under investigation took place, but it is less obvious whether the development in Scots have come about through contact or *drift*. Pa-Tel provides a related diagnostic which will be useful for us to help establish whether a feature is contact-induced or not:

> "if two languages, known to be in contact, exhibit a similar pattern, but only in one of them are intermediate stages in the development of said pattern attested, that language is the source of the change. The language which only attests to the final result is more likely the borrowing language" (Pa-Tel 2013: 316).

That is, if a candidate feature is adopted into the receiving language in its fully grammaticalised form, then it is more likely to be a transferred feature.

A final important diagnostic for whether contact-induced change is likely is circumstantial evidence, in the form of social context: Thomason and Kaufman state that "It is the sociolinguistic history of the speakers, and not the structure of their language, that is the primary determinant of the linguistic outcomes of language contact" (1988: 35), which is demonstrated by that linguistic constraints on language-induced change which has been proposed in the literature have often

found a counter-example in data from languages in contact. Hence, it is crucial to understand the social context in which these syntactic changes in Scots took place in order to assess whether they feasibly could be contact-induced.

Thus, for this discussion of whether the syntactic changes identified in this thesis are outcomes of language contact, I will use the following criteria as a diagnostic for contact-induced change:

1. Is the social context such that contact-induced syntactic change is likely to take place?
2. Is the timing such that contact-induced change is feasible – that is, are the features under investigation present in English before they emerge in Scots, and do they emerge in Scots during the relevant contact period?
3. Do the features show similar developmental patterns in Scots as in English?

If 1 and 2 are found, we can be more confident that a change is contact-induced. If 3 is found, it may be evidence that a *drift* process, rather than contact, may have caused the change. The first criteria (*social context*) will be investigated in Section 7.1.2, the second (*timing*), and the third (*shared grammaticalisation*) in 7.1.4.

7.1.2 Social context

A social predictor of contact-induced language change is *intensity* of contact, which is measured by the length of contact, the numerical advantage of one group of speakers over the other, the socio-political dominance of one group over the other, and the level of bilingualism in the contact community (Thomason and Kaufman 1988: 67, Thomason 2001: 66). As Scots and English probably always were mutually intelligible, the level of bilingualism is difficult to gauge – the situation between speakers of Scots and English in Scotland in the fourteenth-sixteenth centuries may be compared to the situation between speakers of Old English (OE) and Old Norse (ON) in the Danelaw; at the time of the viking settlement of England, OE and ON had only been diverging from their shared North-West Germanic ancestor for about 300 years, and, as demonstrated by Townend (2002), there is plenty of evidence for that the Danelaw was a bilingual society but that the individuals were not bilingual, and that the languages were similar enough that mutual intelligibility is likely. At some point, however, ON ceased to be spoken in England as the speakers

of ON shifted to speaking English. The contact between ON and OE is argued to have caused significant structural change in English, such as the loss of verb-second word order, which was likely induced by imperfect learning (Kroch et al. 2000). Trudgill (1986) also demonstrates that even in situations of dialect contact, where the varieties in question would be highly typologically similar and mutually intelligible, there is still likely to occur speaker accommodation, leading to levelling and simplification despite the original varieties being structurally similar. Thus, regardless of whether the speakers in contact in Scotland were truly bilingual or not, contact between very similar languages could lead to structural change. In fact, less intense contact is needed for structural transfer to take place between typologically similar languages, than for languages that are more typologically divergent (Thomason 2001: 71).

In Chapter 2, I suggested that Scots-English contact pre-1560 could be described as a convergence contact situation, and that the type of contact then changed as English became socio-politically dominant – here, 1560 will be used as a break-off point for when the Transition period starts, to facilitate the discussion, but the shift in the late sixteenth century had, of course, not such a clean starting point. Thus, in the sense of Thomason and Kaufman (1988), the intensity of the contact situation would have increased at this point, which led to the increased borrowing from English into Scots we see in this period (see Chapter 2.2.3-2.2.4). Before 1560, it appears that Scots syntax is impervious to changes in word order happening in English, as V-to-I is retained and *do*-support does not develop at the same time as it does in Scots. This gives more evidence for that the contact situation between English and Scots was less intense before 1560, further supported by that the nature of those borrowed features which we do see imply that the contact before 1560 ranks low, perhaps at level 2 on Thomasson and Kaufman's (1988: 74-5) borrowing scale ("minor phonological, syntactic, and lexical features [...] that cause little to no typological disruption"). The evidence presented in the background chapter (2.2.3-2.2.4) suggest that the intensity of contact, according to Thomason and Kaufman's (1988) criteria, would have increased dramatically with the new socio-political dominance of English after 1560, and particularly after the Union of Crowns in 1603. Thus, the social context in the sixteenth to eighteenth century appears to be one in which contact-induced structural change could take place

7.1.3 Timing

As regards criterion 2, *timing*, we can make the following observations based on the case study findings:

– In the case of the *Northern Subject Rule* (NSR), the transferred structure would be Standard English (StE) agreement. The findings in Chapter 2 confirm that StE agreement largely replaces the NSR pattern during the seventeenth century, and that StE agreement had regularised in English before this point.
– *Do* support was selected for investigation based on that it emerged in Scots during the Transition period, and that this took place ca. 200 years after the emergence in English. Indeed, this is corroborated by the findings in Chapter 5.
– According to the findings of Chapter 6, the verbal -*ing* forms emerges before 1540; the curve suggests a starting point around 1500. The curve also indicates that the decline of participial -*and* in favour of -*ing* starts in the earlier sixteenth century. In English, the verbalisation of -*ing* and decline of -*and* begins in the thirteenth century

Thus, the change from NSR to StE agreement and the rise of *do*-support fits the *timing* criteria for contact-induced change. The rise of verbal -*ing* is not as straightforward; while verbal -*ing* appeared in English ca. 300 years before the point when we speculate that it emerged in Scots, the emergence in Scots does not start in the relevant contact period, i.e. the Transition period. However, the change seems to happen more rapidly during the Transition period.

7.1.4 Shared grammaticalisation

The third criterion warrants more speculative discussion, as all three features would need to be studied in more detail in terms of their syntactic behaviour and semantic function than what this thesis has been able to do, in order to determine their similarity to their English counterparts. The three features will again be assessed in turn:

- StE agreement appears to be adopted in all relevant contexts, i.e. where it differs from the NSR, at the same time, evidenced by the similar decline of *-(i)s* inflection with plNP and non-adjacent NSR subjects. There are still competing options in Scots, and some dialects exhibit *subject type sensitivity* in S-V agreement, but I judge this to be a case of residual grammar competition between the NSR and StE agreement, rather than differences within the StE agreement system.
- Affirmative and negative declarative *do* appears at the same time in Scots, but the regulation of *do* could be analysed as following similar grammaticalisation stages as the regulation of *do* in English, i.e. an intermediate *do* stage. The findings suggest that, while the intermediate *do* stage is not identical in Scots and English (although, we cannot be sure as the sample sizes are too small to be conclusive with respect to some features), it is not the post-1575 English *do* that has been transferred into Scots, despite the feature emerging in Scots post-1575. If we indeed analyse this intermediate *do* as a true intermediate stage, then this change would fit the *shared grammaticalisation* criterion, meaning that the change is likely resulting from *drift* and not English influence. However, as noted in Chapter 5.2.2, the nature of intermediate *do* as described in Ecay (2015) is not necessarily that of an actual intermediate stage from causative *do* to dummy *do*, but it could be a third variant of *do* which is spreading northward into Scots from English, and the later emergence of this type of *do* could then be the result of a time lag.
- The results for the rise of verbal *-ing* are the most difficult to assess based on the findings in Chapter 6. There is an indication that the verbalisation of gerunds follows a different trajectory in Scots than in English, but more analysis is needed to determine what differences there are between the developments.

Out of the three features, it is the change from NSR to StE agreement which best fits a contact-induced change analysis based on the *shared grammaticalisation* criterion. Whether *do*-support fits the criterion depends on how we analyse the so-called "intermediate" *do* auxiliary (as truly intermediate, or as a different *do*), and the findings regarding the rise of verbal *-ing* are too inconclusive to judge their similarity to English.

7.1.5 Conclusions

To summarise this discussion, it seems that there is some evidence for that contact played a role in inducing or influencing the syntactic changes in Scots investigated for this thesis. It has been concluded that the social context and timing of the change speaks in favour of the decline of the NSR and rise of *do*-support being contact-induced changes. The rise of the NSR also meets the *shared grammaticalisation* criterion, but it is not clear whether *do*-support does. The fact remains that a theory of an independent development of Scots *do* does not give a satisfactory explanation for why this development, and the loss of verb-raising which facilitated the regulation of *do*, takes place nearly 200 years later in Scots than in English, during the period of intense anglicisation of Scots. For now, I conclude that the social context and timing of the rise of Scots *do* suggests that it is a contact-induced changed, acknowledging that the fact that the auxiliary shows intermediate *do* qualities compromises such an analysis if it is the case that this is truly an intermediate stage in the grammaticalisation of *do*. Finally, the findings with respect to the *timing* of the rise of verbal *-ing*, in the pre-1560 period, and the uncertainty regarding its similarities to English verbal *-ing*, means that no confident conclusion can be drawn regarding the origin of this change in Scots.

7.2 The role of gender

A secondary investigation was made for each case study, which sought to uncover what the role of women was in the the trajectory of the changes investigated. The background to this investigation was that Scottish women have been found to resist anglicisation pressures in other areas of grammar. This contradicts the usually assumed principles of the role of women in language change, as formulated by Labov (2001: 261-93), which predict that women adopt new features, whether they be changes from above or below, and thus be leaders of change. A question to raise, then, is whether the behaviour of Scottish women with respect to new grammatical systems can be used as an indicator of whether those systems are transfers from English. A case in favour for

this hypothesis, would be that the best candidate for contact-induced change out of the features investigated, the decline of the NSR, is also the only feature where the female writers have a clear preference for the outgoing structure. As regards *do*-support and the rise of verbal -*ing*, women favoured the incoming grammar in both cases, where a significant difference could be observed.

7.3 Final conclusions and outlook

With this thesis, I have sought to fill some of the gaps in our knowledge of sixteenth to eighteenth century Scots syntax, and the nature of contact between Scots and English of that time. In pursuing this research aim, I have created a resource which will facilitate future studies on Transition Scots syntax, and facilitate comparative studies with English and other languages: the *Parsed Corpus of Scottish Correspondence* (PCSC). I have demonstrated the suitability of the PCSC for quantitative investigations of diachronic syntactic change through case studies on the NSR, *do*-support, and the rise of verbal -*ing*, and have explored the possibility for sociolinguistic research using the PCSC through my investigation of the effects of writer gender on these syntactic changes. In doing so, I have showcased a range of approaches to mine the PCSC data and retrieve relevant results for each case study. Finally, I assessed the contact situation between English and Scots in Scotland during the Scots Transition period, and whether the syntactic changes uncovered for each of the case studies can be ascribed to Anglo-Scots contact.

The main findings of this thesis reveal the transformative nature of Scots syntax in the sixteenth to eighteenth century, as the language undergoes dramatic changes in its subject-verb agreement system through the decline of the NSR and the rise of *do*-support, and further rearrangement in the verbal paradigm through the rise of verbal -*ing* in both participial and gerundive function. The decline of the NSR was found to be a likely candidate for contact-induced syntactic change, but the origin of the rise of *do*-support and verbal -*ing* could not be concluded until further investigation is carried out.

There are plenty of avenues for further research opened up by the findings of this research. These case studies have only scratched the surface of the potential of the PCSC, and particularly as

regards sociolinguistic research – I have, for example, not included location or social class as factors in any of the changes investigated. There is also scope for more fine-grained analysis of syntactic processes involved in the development of each of the features investigated for the case studies. A more in-depth analysis of theories of mechanisms in language change would also benefit from considering the impact of frequency in grammaticalisation processes (Bybee 2003), or the Tolerance Principle (Yang 2016) in diachronic change. I hope that the PCSC and the findings of this thesis can serve as ground work for future studies on the areas described.

Bibliography

Adger, D., & Smith, J. (2010). Variation in agreement: a lexical feature-based approach. *Lingua*, *120*(5), 1109-1134.

Agutter, A. (1988). Standardisation in Middle Scots. *Scottish Language*, 1-9.

Agutter, A. (1990). Restandardisation in Middle Scots. In S. Adamson, V. Law, N. Vincent, & S. Wright (Eds.), *Papers from the 5th International Conference on English Historical Linguistics.* Amsterdam/Philadelphia: John Benjamins Publishing.

Aitken, A. J. (1979). Scottish speech: a historical view with special reference to the standard English of Scotland. In A. Aitken & T. MacArthur (Eds.), *Languages of Scotland* (p. 85-118). Edinburgh: Chambers.

Aitken, A. J. (1984). Scots and English in Scotland. In P. Trudgill (Ed.), *Language in the British Isles.* (p. 517-532). Cambridge: Cambridge University Press.

Aitken, A. J. (1985). A history of Scots. In M. Robinson (Ed.), *The concise Scots dictionary* (p. ix-xvi). Aberdeen: Aberdeen University Press.

Aitken, A. J., Bratley, P., & Hamilton-Smith, N. (1984). *Older Scottish texts: the Edinburgh DOST corpus.* Oxford Text Archive. [https://ota.bodleian.ox.ac.uk/repository/xmlui/handle/20.500.12024/0701].

Alexiadou, A. (2011). Gerund types, the present participle and patterns of derivation. In C. Maienborn & A. Wöllstein (Eds.), *Events arguments: Foundations and applications* (p. 139-152). Berlin/Boston: Max Niemeyer Verlag.

Auer, P. (2005). Europe's sociolinguistic unity, or: a typology of European dialect/standard constellations. In N. Delbecque, J. van der Auwera, & D. Geeraerts (Eds.), *Perspectives on variation* (p. 7-42). Amsterdam: Mouten de Greuyter.

Bald, M. (1926). Anglicisation of Scottish printing. *The Scottish Historical Review*, *23*(90), 107-115.

Beal, J. (1997). Syntax and morphology. In C. Jones (Ed.), *The Edinburgh history of the Scots language* (p. 335-377). Edinburgh: Edinburgh University Press.

Beck, J., Ecay, A., & Ingason, A. K. (2011). *Annotald, version 1.3.10.* https://annotald.github.io.

Biberauer, T., & van Kemenade, A. (2011). Subject positions and information-structural diversification in the history of English. *Catalan Journal of Linguistics*, *10*(1), 1-53.

Bobaljik, J. D. (2002). A-chains at the PF-interface: Copies and 'covert movement. *Natural Language and Linguistic Theory*, *20*(2), 197-267.

Bolinger, D. (1977). *Meaning and form.* London: Longman.

Borin, L., Forsberg, M., & Roxendal, J. (2012). Korp - the corpus infrastructure of språkbanken. In *Proceedings of lrec 2012* (p. 474-478). Istanbul: ELRA.

Börjars, K., & Chapman, C. (1998). Agreement and pro-drop in some variaties of English. *Linguistics*, *36*, 71-98.

Brinton, L. (1996). *Pragmatic markers in English: Grammaticalization and discourse functions.* Berlin/New York: Mouton de Gruyter.

Brinton, L., & Bergs, A. (2017). *Middle English.* Berlin/Boston: Mouton de Gruyter.

Buchstaller, I., Corrigan, K., Holmberg, A., Honeybone, P., & Maguire, W. (2013). T-to-R and the Northern Subject Rule: questionnaire-based spatial, social and structural linguistics. *English Language and Linguistics*, *17*(1), 85-128.

Bybee, J. (2003). Mechanisms of change in grammaticization: The role of frequency. In B. D. Joseph & R. D. Janda (Eds.), *The handbook of historical linguistics* (pp. 602–623). Oxford: Blackwell.

Chan, H. S., & Tan, H. (2009). Maybe, perhaps, I believe, you could: Making claims and the use of hedges. *The English Teacher*, *31*, 98-106.

Colleran, R. (2017). *Keeping it in the family: Disentangling contact and inheritance in closely related languages* (Unpublished doctoral dissertation). University of Edinburgh.

Corrigan, K. (1997). *The syntax of South Armagh English in its socio-historical perspective.* (Unpublished doctoral dissertation). National University of Ireland.

Dal, I. (1952). Zur entstehung des englischen participium praesentis auf -ing. *Norsk Tidsskrift for Sprogvidenskap*, *16*, 5-116.

Dawson, J. (1997). *Clan Campbell letters 1559-1583.* Scottish History Society, 5th ser., vol. 10. Edinburgh: Scottish History Society.

de Haas, N., & van Kemenade, A. (2015). The origin of the Northern Subject Rule: subject positions and verbal morphosyntax in older English. *English Language and Linguistics*, *19*(1), 49-81.

de Smet, H. (2008). Nominal gerunds in 16th-century English: the function of the definite article. *Folia Linguistica Historica*, *28*(1), 77-113.

de Smet, H. (2013). *Spreading patterns: Diffusional change in the English system of complementation.* Oxford: Oxford University Press.

de Haas, N. (2011). *Morphosyntactic variation in Northern English: The Northern Subject Rule, its origins and early history.* Utrecht: LOT.

Denison, D. (1985). The origins of periphrastic do: Ellegård and Visser reconsidered. In R. Eaton, O. Fischer, W. F. Koopman, & F. van der Leek (Eds.), *Papers from the 4th International Conference on English Historical Linguistics* (p. 45-60). Amsterdam/Philadelphia: John Benjamins Publishing.

Denison, D. (1993). *English historical syntax.* London: Routledge.

Devitt, A. J. (1989). *Standardizing written English: diffusion in the case of Scotland, 1520-1659.* Cambridge: Cambridge University Press.

Dossena, M. (2013). Ego documents in Scottish corpora: The contribution of nineteenth-century letters and diaries to the study of language history. In *Language in Scotland* (p. 91 - 111). Leiden, The Netherlands: Brill.

Dow, E. W. (1926). *Atlas of European history*. New York: H. Holt and Co. Accessed from: [https://archive.org/details/atlasofeuropeanhistory1926/page/n47/mode/2up].

Ecay, A. (2015). *A multi-step analysis of the evolution of English do-support* (Unpublished doctoral dissertation). University of Pennsylvania.

Ellegård, A. (1953). *The auxiliary do: The establishment and regulation of its use in English.* Stockholm: Almqvist and Wiksell.

Eunson, J., Murray, L., & Ipsos MORI Scotland. (2009). *2011 census question testing – the language question.* https://www.scotlandscensus.gov.uk/media/ljqhbh5w/2011-census-language-question.pdf.

Fanego, T. (1996). The gerund in Early Modern English: Evidence from the Helsinki Corpus. *Folia Linguistica Historica*, *30*, 97-152.

Fanego, T. (2004). On reanalysis and actualization in syntactic change. *Diachronica*, *21*(1), 5-25.

Fernández Cuesta, J. (2011). The Northern Subject Rule in first-person-singular contexts in early Modern English. *Folia Linguistica Historica 32*, 89-114.

Fischer, O., & van der Wurff, W. (2006). Syntax. In *A history of the English language* (p. 109-198). Cambridge: Cambridge University Press.

Fonteyn, L. (2019). *Categorically in language change: the case of the English gerund.* Oxford/New York: Oxford University Press.

Galves, C., de Andrade, A. L., & Faria, P. (2017). *the Tycho Brahe Parsed Corpus of Historical Portuguese.* http://www.tycho.iel.unicamp.br/ tycho/corpus/texts/psd.zip.

Gardela, W. (2017). *A study of Gan, Can and Beginnen in the Northern English and Scots of the late fourteenth and the fifteenth centuries* (Unpublished doctoral dissertation). The University of Edinburgh.

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

Görlach, M. (1991). *Introduction to Early Modern English.* Cambridge: Cambridge University Press.

Görlach, M. (2002). *A textual history of Scots.* Heidelberg: Unversitätsverlag C. Winter.

Gotthard, L. (2019). Why do-support in Scots is different. *English Studies*, *100*(3), 1-25.

Graham-Campbell, J., Hall, R., Jesch, J., & Parsons, D. N. (2001). *Vikings and the Danelaw.* Oxbow books.

Haeberli, E. (2002). *Features, categories and the syntax of A-positions. crosslinguistic variation in the Germanic languages.* Dordrecht: Kluwer.

Harris, J. (1984). Syntactic variation and dialect divergence. *Journal of Linguistics*, *20*(2), 303-327.

Haugen, E. (1966). Dialect, language, nation. *American Anthropologist*, *68*, 922-35.

Henry, A. (1995). *Belfast English and Standard English: Dialect variation and parameter setting.* Oxford: Oxford University Press.

Hickey, R. (2012). Early English and the Celtic hypothesis. In T. Nevalainen & E. Closs Traugott (Eds.), *The oxford handbook of the history of English* (p. 497-507). Oxford: Oxford University Press.

Hock, H. H., & Joseph, B. D. (1996). *Language history, language change, and language relationship: An introduction to historical and comparative linguistics.* Berlin/New York: Mouton de Gruyter.

Hofman, K. (2019). Approaching Transition Scots from a micro-perspective: The Dunfermline corpus, 1573-1723. In R. Alcorn, J. Kopaczyk, B. Los, & B. Molineaux (Eds.), *Historical dialectology in the digital age.* Edinburgh: Edinburgh University Press.

Houston, A. (1989). The English gerund: syntactic change and discourse function. In R. W. Fasold & D. Schiffrin (Eds.), *Language change and variation* (p. 173-196). Amsterdam/Philadelphia: John Benjamins Publishing.

Ihalainen, O. (1994). The dialects of England since 1776. In *The cambridge history of the English language. volume 5: English in Britain and overseas. origins and development* (p. 197-274). Cambridge: Cambridge University Press.

Jack, G. B. (1988). The origins of the English gerund. *NOWELE. North-Western European Language Evolution*, *12*, 15-75.

Jamieson, E. (2015). *An investigation of verb raising in the Shetland dialect of Scots* (Unpublished master's thesis). University of Edinburgh.

Jonas, D. (2002). Residual V-to-I. In D. Lighfoot (Ed.), *Syntactic effects of morphological change* (pp. 251–270). Oxford: Oxford University Press.

Joseph, J. (1987). *Eloquence and power. the rise of language standards and standard languages.* London: Pinter.

Kayne, R. (2000). Microparametric syntax: Some introductory remarks. In J. R. Black & V. Motapanyane (Eds.), *Parameters and universals* (p. 3-9). Oxford: Oxford University Press.

King, A. (1997). The inflectional morphology of Older Scots. In C. Jones (Ed.), *The Edinburgh history of the Scots language* (p. 156-181). Edinburgh: Edinburgh University Press.

Klemola, J. (2000). The origins of the Northern Subject Rule. In H. L. Tristram (Ed.), *The celtic Englishes ii* (p. 329-346). Heidelberg: Unversitätsverlag C. Winter.

Kniezsa, V. (1997). The origins of Scots orthography. In C. Jones (Ed.), *The edinburgh history of the scots language* (p. 24-46). Edinburgh: Edinburgh University Press.

Kopaczyk, J. (2013). Rethinking the traditional periodisation of the Scots language. In R. M. Millar & J. Cruickshank (Eds.), *After the storm: Papers from the Forum for Research on the Languages of Scotland and Ulster* (p. 233-260). University of Aberdeen.

Kroch, A. (1989). Reflexes of grammar in patterns of language change. *Language variation and change*, *I*, 199-244.

Kroch, A., & Santorini, B. (2021). *The Penn-BFM parsed corpus of historical French, version 1.0.* https://github.com/beatrice57/mcvf-plus-ppchf.

Kroch, A., Santorini, B., & Delfs, L. (2004). *The Penn-Helsinki Parsed Corpus of early Modern English, release 3.* University of Pennsylvania. [https://www.ling.upenn.edu/hist-corpora/PPCEME-RELEASE-3/index.html].

Kroch, A., Santorini, B., & Diertani, A. (2016). *Penn parsed corpus of modern British English, 2nd edition, release 1.* University of Pennsylvania. [https://www.ling.upenn.edu/hist-corpora/PPCMBE2-RELEASE-1/index.html].

Kroch, A., & Taylor, A. (2000). *The Penn-Helsinki Parsed Corpus of Middle English, 2nd edition, release 4.* University of Pennsylvania. [https://www.ling.upenn.edu/hist-corpora/PPCME2-RELEASE-4/index.html].

Kroch, A., Taylor, A., & Ringe, D. (2000). The Middle English verb-second constraint: a case study in language contact and language change. In S. C. Herring, P. van Reenen, & L. Schøsler (Eds.), *Textual parameters in older languages* (p. 353-391). Amsterdam/Philadelphia: John Benjamins Publishing.

Labov, W. (2001). *Principles of linguistic change, volume 2: Social factors.* Malden, MA: Blackwell Publishers.

Lass, R. (1992). Phonology and morphology. In N. Blake (Ed.), *The Cambridge history of the English language: Volume ii* (p. 23-155). Cambridge: Cambridge University Press.

Light, C. (2011). *Parsed corpus of early New High German.* University of Pennsylvania.

Los, B. (2009). The consequences of the loss of verb-second in English: information structure and syntax in interaction. *English Language and Linguistics*, *13*(1), 97-125.

Los, B. (2015). *A historical syntax of English.* Edinburgh: Edinburgh University Press.

Los, B. (2016). The genesis of non-finite forms. In A. Bannink & W. Honselaar (Eds.), *From variation to iconicity: Festschrift for Olga Fischer on the occasion of her 65th birthday* (p. 265-280). Amsterdam: Pegasus.

Lowe, C. (1999). *Angels, fools and tyrants. Britons and Anglo-saxons in southern Scotland.* Edinburgh: Canongate Books with Historic Scotland.

Macafee, C. (1992/1993). A short grammar of Older Scots. *Scottish Language*, *11/12*, 10-36.

Macafee, C. (1997). Older Scots lexis. In C. Jones (Ed.), *The edinburgh history of the Scots language.* Edinburgh: Edinburgh University Press.

Macafee, C., & Aitken, A. J. (2002). *A history of Scots to 1700.* [http://www.dsl.ac.uk/about-Scots/history-of-Scots].

Martineau, F., Hirschbühler, P., Kroch, A., & Morin, Y. C. (2021). *the MCVF corpus, parsed, version 2.0.* https://github.com/beatrice57/mcvf-plus-ppchf.

McCafferty, K. (2003). The Northern Subject Rule in Ulster: How Scots, how English? *Language Variation and Change*, *15*(1), 105-139.

McClure, D. (1994). English in Scotland. In R. Burchfield (Ed.), *The Cambridge history of the English language, vol. 5. English in Britain and overseas : Origins and development* (p. 21-93). Cambridge: Cambridge University Press.

McColl Millar, R. (2012). *English historical sociolinguistics.* Edinburgh: Edinburgh University Press.

McColl Millar, R. (2020). *A sociolinguistic history of Scotland.* Edinburgh: Edinburgh University Press.

McIntosh, A. (1989). Present indicative plural forms in the later Middle English of the North Midlands. In A. McIntosh, M. L. Samuels, & M. Laing (Eds.), *Middle English dialectology: Essays on some principles and problems.* Aberdeen: Aberdeen University Press.

Meurman-Solin, A. (1993a). Periphrastic do in sixteenth- and seventeenth century Scots. In A. Meurman-Solin (Ed.), *Variation and change in early Scottish prose* (p. 259-276). Helsinki: Suomalainen Tiedeakatemia.

Meurman-Solin, A. (1993b). *Variation and change in early Scottish prose.* Helsinki: Suomalainen Tiedeakatemia.

Meurman-Solin, A. (1995). *The Helsinki corpus of Older Scots.* Helsinki: Department of Modern Languages, University of Helsinki. [https://varieng.helsinki.fi/CoRD/corpora/HCOS/index.html].

Meurman-Solin, A. (1997). Differentiaion and standardisation in Early Scots. In C. Jones (Ed.), *The Edinburgh history of the Scots language* (p. 3-23). Edinburgh: Edinburgh University Press.

Meurman-Solin, A., & VARIENG. (2016). *Manual to the Helsinki Corpus of Scottish Correspondence 1540-1750 (Scotscorr).* Available at: https://www.kielipankki.fi/wp-content/uploads/ScotsCorr_Manual_2016.pdf.

Meurman-Solin, A., & VARIENG. (2017). *Helsinki Corpus of Scottish Correspondence (1540-1750)* [text corpus]. FIN-CLARIN: http://urn.fi/urn:nbn:fi:lb-201411071.

Miller, J. (1993). The grammar of Scottish English. In J. Milroy & L. Milroy (Eds.), *Real English: The grammar of English dialects in the British Isles* (p. 99-138). Harlow: Longman.

Moessner, L. (1997). The syntax of Older Scots. In C. Jones (Ed.), *The edinburgh history of the Scots language* (p. 112-155). Edinburgh: Edinburgh University Press.

Montgomery, M. B. (1994). The evolution of verb concord in Scots. In A. Fenton & D. A. McDonald (Eds.), *Proceedings of the third international conference on the languages of Scotland* (p. 81-95). Edinburgh: Canongate Academic and The Linguistic Survey of Scotland, School of Scottish Studies, University of Edinburgh.

Montgomery, M. B. (1997). Making Transatlantic connections between varieties of English. *Journal of English Linguistics*, *25*(2), 122-141.

Montgomery, M. B. (2013). The linguistic landscape of eighteenth-century South Argyll, as revealed by Highland Scot emigrants to North Carolina. In J. M. Kirk & I. Macleod (Eds.), *Scots: Studies in its literature and language* (p. 183-210). Amsterdam: Rodopi.

Moore, S., Meech, S. B., & Whitehall, H. (1935). *Middle English dialect characteristics and dialect boundaries* (Vol. 13). Ann Arbor: University of Michigan Publications. Language and Literature.

Murison, D. D. (1979). The historical background. In A. Aitken & T. MacArthur (Eds.), *Languages of Scotland* (p. 2-13). Edinburgh: W R Chambers.

Muysken, P. (2012). Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism: Language and Cognition*, *16*(04), 709-730.

Nevalainen, T. (2006a). Historical sociolinguistics and language change. In A. van Kemenade & B. Los (Eds.), *The handbook of the history of English* (p. 558-588). Blackwell Publishing.

Nevalainen, T. (2006b). *An introduction to Early Modern English*. Edinburgh: Edinburgh University Press.

Nevalainen, T., & Raumolin-Brunberg, H. (2003). *Historical sociolinguistics: Language change in Tudor and Stuart England*. London/New York: Routledge.

Nevalainen, T., Raumolin-Brunberg, H., Keränen, J., Nevala, M., Nurmi, A., & Palander-Collin, M. (1998). *the corpus of early English correspondence*. Department of Modern Languages, University of Helsinki. [https://varieng.helsinki.fi/CoRD/corpora/CEEC/index.html].

Nurmi, A. (1999). *A social history of periphrastic do* (Vol. 56). Helsinki: Société Néophilologique.

Nurmi, A. (2011). The rise and regulation of periphrastic do in negative declarative sentences: a sociolinguistic study. In D. Kastovsky & A. Mettinger (Eds.), *The history of English in a social context: A contribution to historical sociolinguistics* (p. 339-362). New York: de Gruyter Mouton.

Pa-Tel, N. (2013). Contact or inheritance? Criteria for distinguishing internal and external change in genetically related language. *Journal of Language Contact*, *6*(2), 313-28.

Petré, P. (2013). *LEON: Leuven English old to new, version 0.3.*

Pietsch, L. (2005a). 'some do and some doesn't': Verbal concord variation in the north of the British Isles. In B. Kortmann, T. Herrmann, L. Pietsch, & S. Wagner (Eds.), *A comparative grammar of british English dialects. agreement, gender, relative clauses* (pp. 125–210). Berlin: Mouton de Gruyter.

Pietsch, L. (2005b). *Variable grammars: Verbal agreement in northern dialects of English*. Berlin/Boston: Max Niemeyer Verlag.

Pollock, J.-Y. (1989). Verb movement, Universal Grammar and the structure of IP. *Linguistic Inquiry*, *20*.

Poplack, S., & Levey, S. (2010). Contact-induced grammatical change: A cautionary tale. In *Language and space – an international handbook of linguistic variation: Volume 1 – theories and methods.* Berlin: Mouton de Gruyter.

Postma, G. (2010). The impact of failed changes. In A. Breitbarth, C. Lucas, S. Watts, & D. Willis (Eds.), *Continuity and change in grammar* (p. 269-302). Amsterdam/Philadelphia: John Benjamins Publishing.

Poussa, P. (1990). A contact-universals origin for periphrastic do with special consideration of Old English-Celtic contact. In S. Adamson, V. Law, N. Vincent, & S. Wright (Eds.), *Papers from the 5th international conference on English historical linguistics* (Vol. 65). Amsterdam/Philadelphia: John Benjamins Publishing.

R Core Team. (2021). *R: A language and environment for statistical computing, Vienna, Austria. https://www.r-project.org/.* R Foundation for Statistical Computing.

Randall, B. (2000/2013). *Corpussearch 2: A tool for linguistics research.* http://corpussearch.sourceforge.net/CS.html.

Rissanen, M. (2008). Syntax. In R. Lass (Ed.), *Cambridge history of the English language* (Vol. 3, p. 187-331). Cambridge: Cambridge University Press.

Robbeets, M., & Cuyckens, H. (2013). *Shared grammaticalization: With special focus on Transeurasian languages.* Amsterdam/Philadelphia: John Benjamins Publishing.

Rodríguez Ledesma, M. (2013). The Northern Subject Rule in first-person singular contexts in fourteenth-fifteenth-century Scots. *Folia Linguistica Historica*, *34*, 149-172.

Rodríguez Ledesma, M. (2017). The Northern Subject Rule in the Breadalbane collection. *English Studies*, *15*, 1-17.

Romaine, S. (1982). *Socio-historical linguistics: its status and methodology*. Cambridge: Cambridge University Press.

Santorini, B. (2021). *The Penn parsed corpus of historical Yiddish.* https://github.com/beatrice57/penn-parsed-corpus-of-historical-yiddish.

Shaw, M., & de Smet, H. (2022). Loan word accommodation biases: Markedness and finiteness. *Transactions of the Philological Society*, 1-1.

Shoemark, P., Kirby, S., & Goldwater, S. (2017). Topic and audience effects on distinctively Scottish vocabulary usage in Twitter data. In J. Brooke, T. Solorio, , & M. Koppel (Eds.), *Proceedings of the workshop on stylistic variation,* (p. 59-68). Copenhagen: Association for Computational Linguistics.

Smith, J. (2000). 'you ø na hear o' that kind o' things': Negative do in Buckie Scots. *English World Wide*, *21*(2), 231-259.

Smith, J., Adger, D., Aitken, B., Heycock, C., Jamieson, E., & Thoms, G. (2019). *The Scots Syntax Atlas.* University of Glasgow. https://Scotssyntaxatlas.ac.uk.

Sorace, A. (2000). Gradients in auxiliary selection with intransitive verbs. *Language*, *76*(4).

Stein, D. (1985). Discourse markers in Early Modern English. In R. Eaton, O. Fischer, W. F. Koopman, & F. van der Leek (Eds.), *Papers from the 4th international conference on English historical linguistics* (p. 283-302). Amsterdam/Philadelphia: John Benjamins Publishing.

Stuart-Smith, J. (2004). Scottish English: Phonology. In B. Kortmann & E. Schneider (Eds.), *A handbook of varieties of English* (Vol. 1, p. 47-67). Mouton de Gruyter.

Tajima, M. (1985). *The syntactic development of the gerund in Middle English.* Tokyo: Nan'un-do.

Taylor, A., Mitchel, M., & Santorini, B. (2003). the Penn Treebank: an overview. *DOI: 10.1007/978-94-010-0201-1_1*.

Taylor, A., Nurmi, A., Warner, A., Pintzuk, S., & Nevalainen, T. (2006). *York-Helsinki parsed corpus of early English correspondence.* University of York and University of Helsinki. Distributed through the Oxford Text Archive. [https://ota.bodleian.ox.ac.uk/repository/xmlui/handle/20.500.12024/2510].

Taylor, A., Warner, A., Pintzuk, S., & Beths, F. (2003). *the York-Toronto-Helsinki parsed corpus of Old English prose.* https://www-users.york.ac.uk/ lang22/YcoeHome1.htm.

The University of Edinburgh. (2008). *A Linguistc Atlas of Older Scots, phase 1: 1380–1500.* http://www.lel.ed.ac.uk/ihd/laos1/laos1.html.

Thomason, S. G. (2001). *Language contact: An introduction.* Edinburgh: Edinburgh University Press.

Thomason, S. G., & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics.* Berkeley: University of California Press.

Thoms, G. (2017). *Handout 4: The amn't gap.* (Presented at *Linguistic variation and English dialect syntax: EGG summer school.* Olomuc, Czechia.)

Thoms, G., Adger, D., Heycock, C., & Smith, J. (2019). Syntactic variation and auxiliary contraction: the surprising case of Scots. *Language*, *95*(3).

Tieken-Boon van Ostade, I. (1990). The origin and development of periphrastic auxiliary do: A case of destigmatisation. *NOWELE. North-Western European Language Evolution*, *16*(1).

Tortora, C., Santorini, B., Blanchette, F., & Diertani, C. (2017). *The audio-aligned and parsed corpus of Appalachian English, version 0.1.* www.aapcappe.org.

Townend, M. (2002). *Language and history in Viking age England: linguistic relations between speakers of Old Norse and Old English.* Turnhout: Brepols Publishers.

Trudgill, P. (1986). *Dialects in contact.* Oxford: Blackwell.

Truswell, R., Alcorn, R., Donaldson, J., & Wallenberg, J. (2017). *A Parsed Linguistic Atlas of Early Middle English.* https://github.com/rtruswell/PLAEME_current.

Truswell, R., Alcorn, R., Donaldson, J., & Wallenberg, J. (2019). A Parsed Linguistic Atlas of Early Middle English. In R. Alcorn, B. Los, J. Kopaczyk, & B. Molineaux (Eds.), *Historical dialectology in the digital age* (p. 19-37). Edinburgh: Edinburgh University Press.

van der Auwera, J., & Genee, I. (2002). English do: on the convergence of languages and linguists. *English Language and Linguistics*, *6*, 283-307.

van der Wal, M. J., & Rutten, G. (Eds.). (2013). *Touching the past: Studies in the historical sociolinguistics of ego-documents.* Amsterdam/Philadelphia: John Benjamins Publishing.

van Eyndhoven, S., & Clark, L. (2019). The <quh->–<wh-> switch: an empirical account of the anglicisation of a Scots variant in Scotland during the sixteenth and seventeenth centuries. *English Language and Linguistics*, *24*(1), 211-236.

van Eyndhoven, S., Gotthard, L., & Filgueira, R. (2021, October). *'Scots for the masses'? exploring the use of Scots in 19th century digitised chapbooks.* (Paper presented at the ISLE 6 conference, Ludwig Maximilian University of Munich (Online))

van Gelderen, E. (2008). Negative cycles. *Linguistic Typology*, *12*, 195-243.

van Kemenade, A. (2000). Jespersen's cycle revisited: Formal properties of grammaticalization. In S. Pintzuk, G. Tsoulas, & A. Warner (Eds.), *Diachronic syntax: Models and mechanisms* (p. 51-75). Oxford: Oxford University Press.

van der Horst, J. M. (1998). Doen in Old and Early Middle Dutch: A comparative approach. In A. van Leuvensteijn, I. Tieken-Boon van Ostade, & M. J. van der Wal (Eds.), *DO in English, Dutch, and German. History and present-day variation.* (p. 53-64). Münster: Stichting Neerlandistiek VU / Nodus Publikationem".

Walkden, G. (2016). the HeliPaD: a parsed corpus of Old Saxon. *International Journal of Corpus Linguistics*, *21*, 643-674.

Wallenberg, J., Ingason, A. K., Sigurdsson, E. F., & Rögnvaldsson, E. (2011). *Icelandic parsed historical corpus (IcePaHC), version 0.9.* http://www.linguist.is/icelandic_treebank/Icelandic_Parsed_Historical_Corpus_(IcePaHC).

Warner, A. (2002). What drove DO? In C. Kay, S. Horobin, & J. Smith (Eds.), *New perspectives on English historical linguistics* (Vol. I, p. 229-242). Amsterdam/Philadelphia: John Benjamins Publishing.

Warner, A. (2005). Why DO dove: Evidence for register variation in Early Modern English negatives. *Language Variation and Change*, *17*, 257-280.

Woolf, A. (2007). *From Pictland to Alba: 789-1070.* Edinburgh: Edinburgh University Press.

Yang, C. (2016). *The price of linguistic productivity: how children learn to break the rules of language*. Cambridge, Massachusetts: MIT Press.

Zehentner, E. (2014). From phrase to clause: On the development of present participle and verbal noun in Middle Scots. *VIEWS: Vienna English Working Papers*, *23*.

Zimmerman, R. (2014). *The parsed corpus of Middle English poetry.* https://pcmep.net/.