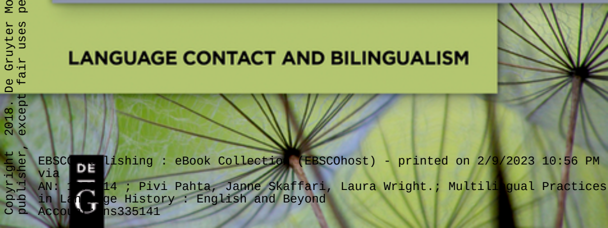
DE GRUYTER
MOUTON

Päivi Pahta, Janne Skaffari,
Laura Wright (Eds.)

MULTILINGUAL PRACTICES IN LANGUAGE HISTORY

NEW PERSPECTIVES

LANGUAGE CONTACT AND BILINGUALISM



Copyright 2018. De Gruyter Mouton. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

EBSC[illegible]lishing : eBook Collectio[illegible] (EBSCOhost) - printed on 2/9/2023 10:56 PM via
AN: 1[illegible]14 ; Pivi Pahta, Janne Skaffari, Laura Wright.; Multili[illegible]gual Practices in La[illegible]ge History : English and Beyond
Accou[illegible]ns335141

Päivi Pahta, Janne Skaffari, Laura Wright (Eds.)

Multilingual Practices in Language History

Language Contact and Bilingualism

Editor
Yaron Matras

Volume 15

Multilingual Practices in Language History

English and Beyond

Edited by
Päivi Pahta
Janne Skaffari
Laura Wright

DE GRUYTER
MOUTON

ISBN 978-1-5015-1381-7
e-ISBN (PDF) 978-1-5015-0494-5
e-ISBN (EPUB) 978-1-5015-0490-7
ISSN 2190-698X

Library of Congress Cataloging-in-Publication Data
A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

© 2018 Walter de Gruyter Inc., Boston/Berlin
Cover image: Anette Linnea Rasmus/Fotolia
Typesetting: RoyalStandard, Hong Kong
Printing and binding: CPI books GmbH, Leck
♾ Printed on acid-free paper
Printed in Germany

www.degruyter.com

Acknowledgements

The idea for this book came up in the process of drafting project applications to the Academy of Finland back in 2011. We are grateful to the Academy for funding two projects that contributed to this work: *Multilingual Practices in the History of Written English* (MultiPract) at the Universities of Tampere and Cambridge in 2012–2016, and *Multilingualism in the Long Twelfth Century* at the University of Turku in 2012–2015. We thank the three universities for enabling the projects to take place, and for the support in organizing the symposium *Historical Code-switching: The Next Step* in Tampere, where the papers included in this volume were first presented and discussed.

We thank the contributors for their co-operation and commitment in the different stages of the book project, and the participants of the Next Step symposium for inspiring discussions. We are indebted to the reviewers who peer-assessed the contributions. We thank Yaron Matras for accepting the volume in the Language Contact and Bilingualism series, Lara Wysong at Mouton for guiding us in the production process, and our editorial assistants Veera Saarimäki, Jukka Tuominen and Vilma Aaltonen for their help in preparing the manuscript.

Tampere, Turku and Cambridge, June 2017
The Editors

Table of contents

Acknowledgements —— v

I Introduction

Päivi Pahta, Janne Skaffari & Laura Wright
1 From historical code-switching to multilingual practices in the past —— 3

Penelope Gardner-Chloros
2 Historical and modern studies of code-switching: A tale of mutual enrichment —— 19

II Borderlands

Herbert Schendl
3 Code-switching in Anglo-Saxon England: A corpus-based approach —— 39

Rita Queiroz de Barros
4 Twentieth-century Romance loans: Code-switching in the *Oxford English Dictionary*? —— 61

Louise Sylvester
5 A semantic field and text-type approach to late-medieval multilingualism —— 77

Simon Meecham-Jones
6 Code-switching and contact influence in Middle English manuscripts from the Welsh Penumbra – Should we re-interpret the evidence from *Sir Gawain and the Green Knight*? —— 97

Janne Skaffari
7 Code-switching in the long twelfth century —— 121

III Patterns

Jukka Tuominen
8 “Trifling shews of learning”? Patterns of code-switching in English sermons 1640–1740 —— 145

Arja Nurmi, Jukka Tyrkkö, Anna Petäjäniemi & Päivi Pahta
9 The social and textual embedding of multilingual practices in Late Modern English: A corpus-based analysis —— 171

Šime Demo
10 Mining macaronics —— 199

Tom ter Horst & Nike Stam
11 Visual diamorphs: The importance of language neutrality in code-switching from medieval Ireland —— 223

Alpo Honkapohja
12 "Latin in recipes?" A corpus approach to scribal abbreviations in 15th-century medical manuscripts —— 243

IV Contexts

Joanna Kopaczyk
13 Administrative multilingualism on the page in early modern Poland: In search of a framework for written code-switching —— 275

Aleksi Mäkilähde
14 Approaching the functions of historical code-switching: The case of solidarity —— 299

Richard Ingham
15 Medieval bilingualism in England: On the rarity of vernacular code-switching —— 319

Laura Wright
16 A multilingual approach to the history of Standard English —— 339

Index —— 359

I Introduction

Päivi Pahta, Janne Skaffari and Laura Wright

1 From historical code-switching to multilingual practices in the past

1 Introduction

Multilingualism and multiculturalism are burning topics in today's societies. Peoples, languages and cultures coming into contact with each other can provoke confusion and concern. However, although the current situation in Europe, for example, tends to be viewed as alarmingly sudden, cultural and language contact and multilingualism are nothing new. Multilingualism in the past was not limited in place and time: we find evidence of it throughout medieval Europe, and in other periods and regions as well. Multilingual societies were composed of multilingual individuals who used more than one language in their daily lives, even within a single utterance. This is manifest in their writing. The surviving written evidence offers us access to code-switching and other multilingual practices of the past, the topic to which this volume – and a growing number of others – is dedicated.

A key term in discussing multilinguals and their communicative practices is code-switching, which has been defined in a number of different ways. We quote Winford (2003: 14): "the alternate use of two languages (or dialects) within the same stretch of speech, often within the same sentence" – and Poplack (1980: 583): "the alternation of two languages within a single discourse, sentence or constituent". Both define code-switching as involving two linguistic codes, although there can be more. Moreover, Winford mentions speech, as code-switching was originally studied as a feature of spoken interaction, whilst Poplack highlights the linguistic, structural context within which the switch takes place. Since historical linguists only work with written records of language use, we have a slightly different emphasis: "the co-occurrence of two or more languages in a single communicative event" (Pahta and Nurmi 2006: 203). Broader definitions have also been made by others – consider Heller (1988: 1): "the use of more than one language in the course of a single communicative episode". The similarities between these definitions, though, hide the variation and variability of

---

Acknowledgement: The authors wish to thank the Academy of Finland (decision numbers 257059 and 258434) for funding the research projects during which this collection was compiled and edited.

DOI 10.1515/9781501504945-001

the phenomenon at hand. Gardner-Chloros has compared it to a chameleon on the one hand (1995: 80) and to an elephant on the other (2009: vii–viii), revisiting the Indian legend of six blind men encountering this “wondrous beast” for the first time and describing completely different parts of it. Similarly, researchers have asked different questions: how are codes switched inside a clause or within a conversation; what constrains switching; when is a non-native word a switch rather than a borrowing; what do switches reveal about the communicators and the community; why does a speaker or writer choose a particular code in a particular context; and where do the boundaries between two codes lie? A very narrow definition would limit the study of code-switching to, for example, the syntax of a bilingual clause; an extremely broad one would incorporate virtually all processes and products of communication involving two or more languages. Yet other approaches take a different starting point for viewing the phenomenon and disregard the notion of codes in the sense of separate languages between which speakers switch; in such approaches, described, for example, by Blommaert (2010: 102), multilingualism

> . . . should not be seen as a collection of ‘languages’ that the speaker controls, but rather as a complex of *specific* semiotic resources, some of which belong to a conventionally defined ‘language’, while others belong to another ‘language’. The resources are concrete accents, language varieties, registers, genres, modalities such as writing – ways of using language in particular communicative situations and spheres of life.

This book considers multilingual communication from a historical perspective. Its title reveals a broad approach: *multilingual practices* is offered as a term which allows researchers to consider a range of topics under the same umbrella, from multilingual societal practice and individual identities and linguistic repertoires to the use of different languages in structuring texts and the syntax of intrasentential code-switching. The phrase is familiar from previous research (e.g. Pahta and Nurmi 2006) although we as editors have not imposed it on the contributors to this volume nor constrained their use of technical meta-language: while differences and similarities between such terms as *code-switching*, *code-mixing*, *foreignism* and others are attended to by individual authors, the chapters in this book are essentially inclusive rather than divisive, with any and all evidence for historical multilingualism deemed to be of interest, whether one-off or conventionalised, or somewhere between these two extremes. This volume stands as an opening up of discussion about historical practices, rather than belonging to this or that camp.

The approach promoted by this volume has multilingualism as a default starting-point: one way or the other, virtually all historical texts are multilingual.

They contain words, phrases and passages originating in a language other than the main language; they may be translations although not necessarily acknowledged as such; their authors may be bilingual from childhood or have become multilingual later.[1] In the medieval European context, for example, practically anyone who could write would have known at least Latin (the lingua franca of the period) and one vernacular language (their mother tongue); and while some of the best-known and most prolific writers are known to have been polyglots (for example, Geoffrey Chaucer), this was not a characteristic of the finest authors alone. This multilingual approach challenges previous ways of looking at older texts, particularly multilingual records. In the past, many scholars have routinely dismissed non-monolingual written language use as simply incompetent (on the assumption that scribes were unable to maintain competence in monolingual Latin), or in the case of juxtaposed languages on the page, as fragmented. Thus, for example, in medieval texts from England which were composed of discrete passages of Medieval Latin, Anglo-Norman French and Middle English, the English portions of the text have been given much more attention than the Latin and French ones; the monolingual expectation has led to looking at one language at a time. Dictionaries in particular sift multilingual texts in this way, extracting words relevant to their concerns but not always mentioning the multilingual context, or, alternatively, not including words relevant to their concerns *because of* the multilingual context. More generally, the non-monolingual background or practices of the writer often go unnoticed, remaining dissociated from the linguistic features of the writer's output or the wider multilingual context of the day.

Embracing a multilingual approach to language history leads the researcher to look beyond the main language of a text and consider what a holistic overview of all the languages in it reveals, about the "grammar" of non-monolingual writing on the one hand or individual identity or societal practice on the other. Multilingual writing and code-switching reflect the linguistic competences and repertoires of individuals and are a response to the expectations of the wider community. There is inevitably an audience implication bound up in multilingual practices, which are interactional, social, varied phenomena. Consequently, the contributors to this volume refer to a variety of frameworks, theories and subdisciplines of linguistics: sociolinguistics, pragmatics, discourse analysis, semantics, lexicography, theoretical and historical linguistics, corpus linguistics, palaeography and codicology.

---

[1] For more on translation-induced language change, see Kranich, Becher, and Höder (2011), and on the translation/code-switching interface, Kolehmainen and Skaffari (2016).

This volume is by no means the first book on multilingual practices of the past (consider, for example, Wenzel 1994; Trotter 2000; Adams, Janse, and Swain 2002; or Hüning, Vogl, and Moliner 2012). Some of our authors also contributed to *Code-Switching in Early English* (Schendl and Wright 2011c), the first collected volume on code-switching in the history of English. Its introductory chapters (Schendl and Wright 2011a, 2011b) provide not only a survey of the field and its key issues but also its history, which started essentially in the 1990s (see also Pahta 2012 and Schendl 2012); there is little need to recapitulate it here. Schendl and Wright (2011b: 34–35) also posit some desiderata for future work on the topic, much of which relates to sources of multilingual data: inclusion of previously dismissed types of material, creating searchable corpora of various genres featuring code-switching, and tracing records of one and the same text-type diachronically. What the field also needs are better descriptions of Anglo-Norman and Anglo-Latin – as England and English have been and continue to be very prominent in historical multilingualism studies, including the present volume – while at the same time it is important to look beyond the borders and coasts of England and aim at a Europe-wide approach, as many of the same phenomena were relevant on the Continent, and elsewhere in the world. Conceptually and theoretically, the desiderata include acknowledging diachronic change and diatopic variation in multilingual practices, applying appropriately adapted theories of present-day multilingualism to mixed-language material of the past and, ultimately, contributing to the study of “the multilingual brain”. The chapters in the present volume show that progress has been made in achieving these goals.

Sections 2 to 5 of this introduction are structured so that they simultaneously discuss prominent issues and problems in the field and provide an overview of the chapters. While the grouping of the chapters highlights certain themes, it is possible to identify other important approaches and analyses in them, and some topics recur in a number of contributions. The contributors analyse evidence for multilingual practices of various types in communities in the past, be they informal and private (equating, insofar as is possible, to present-day informal spoken code-switching), or the codified use of two or more languages in high-register medieval religious or civic documents, or simply the self-conscious and joyful switching between Latin and vernacular languages found in early modern literature. It is hardly surprising that the majority of papers here touch on Latin, as it was the main European language of writing in the historical period. Apart from English and Latin, the other languages discussed include Anglo-Norman French, Irish, Polish, Portuguese, Scots, Spanish and Welsh. Finally, section 6 provides a review of how the chapters meet the desiderata cited above, some concluding remarks, and a call for more research in the field.

2 On code-switching and matrix language

Some of the recurrent concerns in code-switching research relate to terminology. *Code-switching* itself is not a simple or neutral term, despite its frequency. While its use in research on multilingual spoken interaction was established decades ago (e.g. Haugen 1950; Blom and Gumperz 1972; Gumperz 1982), many researchers treat the multilingualism of written texts as clearly different from that of conversations. The crucial difference seems to be spontaneity or lack of planning, which characterises speech far more than writing. To overcome this problem, some have chosen to use code-switching as an umbrella term covering any range of what we call multilingual practices, while others choose to apply a different term altogether, such as *language mixing*, *code mixing*, *language alternation* and *language interaction*. These, however, are not unproblematic either, as even the broadest-looking terms may also have been used in narrower senses. Quite often, code-mixing has been used to refer to intrasentential switching only, while language mixing – which superficially looks transparent and non-technical – has been posited by Auer (1999) as a type of language alternation, intermediate on a continuum from code-switching to fused lects. In sum, there is no consensus on the name of the phenomenon itself.[2]

In addition to terms referring to switching, mixing, borrowing, and so on, there is variation in how the languages involved are labelled. Probably the most widely used and technical of these are *matrix language* and *embedded language*, introduced by Myers-Scotton (e.g. 1993) as part of her Matrix Language Frame model. In the present volume, those contributors who use matrix language as a term do so in its non-controversial sense, meaning *grammatical background*, *syntactic framework* or simply *main language*. An exception is Penelope Gardner-Chloros, who problematises the term and observes that code-switching is produced by individuals who use their repertoire of languages according to circumstance; that what belongs to one language and what belongs to another is not necessarily clear; and that identifying a single primary language in bilingual speech or writing may not be possible. In “Historical and modern studies of code-switching: A tale of mutual enrichment”, Gardner-Chloros further compares the work and interests of “mainstream” sociolinguists working on present-day data and those of historical sociolinguists, pointing out that the former are not very familiar with the research conducted by the latter and that they mainly focus on spontaneous oral production at the expense of other (written) sorts (although the rise of social media is changing that). Historical code-switched data can, however, provide a testbed for code-switching theory, illustrating, for example, the development from an initial code-switch to a widespread, assimilated borrowing. Despite their different challenges, historical and mainstream sociolinguists share a lot of common ground and can benefit from each other's work – in both directions.

---

[2] For more on the terminology, see, for example, Gardner-Chloros (2009: 10–13) and Schendl and Wright (2011b: 23–24).

Overall, the contributors to this volume tend to prefer terms other than *matrix* and *embedded language*. The majority, however, do not shy away from the term code-switching. All of them examine multilingual practices of the past.

3 Exploring borderlands

The second section of the book, "Borderlands", is concerned with grey areas and boundaries on both a concrete and a metaphorical level – geographical border areas and temporally defined groups of texts which have been overlooked thus far, and the conceptually and practically challenging boundary between code-switching and borrowing.

A constant problem in researching short, mainly one-word switches is distinguishing them from *lexical borrowings* or *loanwords*. Both are consequences of language contact and refer to items originating in another language, but the latter are seen as more integrated into the recipient language, more established as part of its lexis and often more frequent. A considerable grey area nonetheless remains, since not all loans are morphologically or phonologically assimilated into the system of the recipient language, nor are all code-switches spontaneous or transient (a term for infrequently used loans, *nonce borrowing* [Weinreich [1953] 1963: 11], occurs seldom in this book). The switching/borrowing issue is specifically addressed in the chapters by Herbert Schendl, Rita Queiroz de Barros and Louise Sylvester, and appears elsewhere as well.

Herbert Schendl considers in "Code-switching in Anglo-Saxon England: A corpus-based approach" the multilingual material in two text-types, homilies and scientific treatises, from the earliest period in the history of written English. Languages pattern differently in the two text-types: homilies contain code-switched quotations and short intrasentential switches (some of them glossed religious terms), whereas treatises contain Latin authorial metacomments in addition to terminology and quotations. Schendl also considers the problem of distinguishing borrowing from code-switching and, after presenting an overview of previous work on the subject, proposes five principles for identifying one-word switches: (i) the author's bilingualism; (ii) lack of morphological integration (unlike loans); (iii) low frequency and restricted distribution; (iv) the stylistic or other effect the switch creates; and in some cases, (v) the explanation accompanying the foreign word. These criteria correspond to some of those proposed by Matras (2009: 110–114); however, Schendl notes that not all Latin word-forms appearing in vernacular texts are code-switches, as he demonstrates with a selection of candidates. Essentially, each potential candidate has to be examined individually; and Schendl concludes that code-switching, leading to lexical innovation and eventually to lexical borrowing, is part of a dynamic process in which individual authors' bilingual resources play an important role.

In "Twentieth-century Romance loans: Code-switching in the *Oxford English Dictionary*?" Rita Queiroz de Barros gives a brief history of the issue of single-word switching versus lexical borrowing, including the problems it poses in lexicography. As a case-study, she analyses how over 1800 Romance-origin twentieth-century words have been treated in the *Oxford English Dictionary* and distinguishes foreignisms (or one-word switches) from borrowings on the basis of such criteria as orthography and pronunciation, the use of glosses, italics or inverted commas, and the presence of un-English morphology. She finds that according to her criteria, over a third of the words considered are foreignisms, although they are usually not labelled as such. The remaining two-thirds, not showing such features, may thus be classified as more assimilated borrowings which more clearly deserve to be included in a monolingual dictionary.

Louise Sylvester has a different approach to short switches/loans in her examination of technical vocabulary. In "A semantic field and text-type approach to late-medieval multilingualism" she considers single-word switches by focusing on whether or not they are general, superordinate terms, or more specialised hyponyms, taking as evidence data from wills, sumptuary laws, petitions and romances included in the *Medieval Dress and Textile Vocabulary in Unpublished Sources* project. Surveying dress and textile terms, she considers three terms occurring in all four text-types and their semantic nuances in various contexts. She then examines other clothing terms only occurring in one text-type. She concludes that as clothing terms are specific to text-type, they may be seen to represent a technolect, a precise and universal technical vocabulary, rather than a specific single language. This brings into question language boundaries: it may not be possible to know whether a given term formed part of the discrete word-stock of Latin, French or English, when there is evidence that it was used in all three. In such circumstances, both code-switching and lexical borrowing turn out to be inadequate terms.

Next, Simon Meecham-Jones takes us to the Welsh–English border in "Code-switching and contact influence in Middle English manuscripts from the Welsh Penumbra – Should we re-interpret the evidence from *Sir Gawain and the Green Knight*?". Language contact between medieval Welsh and Middle English was frequent and lasted for centuries, but written evidence of it is very rare, and Welsh influence on English life and language has been underinvestigated. Meecham-Jones demonstrates that Welsh linguistic features in "monolingual" English texts have been overlooked or misinterpreted, even in texts associated with the border area. He particularly focuses on *Sir Gawain and the Green Knight* and its idiosyncratic vocabulary; many of its obscure words may be of Welsh origin, or potentially represent code-switching. Despite the broad range of research into literary language, there still remains uncharted territory.

The last chapter in this section similarly focuses on a border area often neglected in linguistic studies of medieval English, but a temporal rather than geographical one. In "Code-switching in the long twelfth century", Janne Skaffari discusses the use of Latin, English and French in over a hundred broadly literary (as opposed to documentary) manuscripts currently held in various libraries in England, particularly in Oxford, Cambridge and London, dating from the last decades of the eleventh century to the thirteenth century. The multilingual practices evinced in the material, which is mostly religious, include intertextual, intersentential, intrasentential and also extrasentential or – literally – marginal switching between Latin and English, some of it visually flagged. French plays a very minor role in these manuscripts, which is partly a reflection of the criteria for selecting the material for this study. However, the relative rareness of code-switching between the vernacular languages in written texts (compared to switching between Latin and a vernacular) is a theme visible also elsewhere in this volume.

4 Locating patterns

The next section of the book, "Patterns", draws upon large datasets and corpus linguistic methods. Corpora have thus far not been used extensively in historical multilingualism research; within English historical linguistics, the seminal paper is by Pahta and Nurmi (2006), who consider the long diachrony of code-switching in the history of English writing, using the Helsinki Corpus of English Texts as their data. The chapters in this section both utilise corpora – purpose-built and multipurpose ones – and introduce tools for exploring their multilingual features. Actual multilingual corpora are far rarer than monolingual ones, partly due to the markup problems they pose, but just like individual texts which have been summarily classified as monolingual, the so-called monolingual corpora may also display a wealth of evidence of other languages interacting with the main language (see, for example, the papers in Nurmi, Rütten, and Pahta 2017).

This section starts with another chapter on religious material. In "'Trifling shews of learning'? Patterns of code-switching in English sermons 1640–1740", Jukka Tuominen surveys ten sermons included in the Lampeter Corpus from a variationist point of view. The pattern which emerges from his study is that code-switching to Latin (mostly quotations and terminology) declined over time, with authors born before the 1640s switching more than those born in the latter half of the seventeenth century. Moreover, two nonconformist ministers in particular showed the least code-switching; education may also have played a role here as they, unlike the other authors, were not university educated. A claim by a bishop of the time that the use of foreign languages in English sermons was by 1692 a thing of the past is found to have been prescriptive rather than descriptive: code-switching in sermons continued throughout the period.

Arja Nurmi, Jukka Tyrkkö, Anna Petäjäniemi and Päivi Pahta research a much larger corpus in "The social and textual embedding of multilingual practices in Late Modern English: A corpus-based analysis". They collect evidence of multilingual practices in 18th- and 19th-century English writings from the 34-million-word Corpus of Late Modern English Texts 3.0 and identify multilingual passages using semi-automated discovery methods, particularly *Multilingualiser*, a new tool specifically developed for locating data in multiple languages. The chapter presents a quantitative overview of foreign-language passages – rather than individual words (which are problematic, as seen above) – in written English and of the variety of languages used during this period. It also offers a refined statistical analysis of the sociolinguistic correlations between multilingual practices and the language-external social variables characterising the data, such as the authors' gender, social status, age and education, and the probable target audience of each text. The chapter thus offers baseline evidence of the frequency and types of code-switching in the Late Modern English period and identifies patterns of multilingualism in writing.

Šime Demo similarly examines something other than switching between just two or three languages in "Mining macaronics". He considers early modern macaronic verse, a widespread literary practice that lasted several centuries and included a number of language pairs in Europe, connecting educated literary people across cultures in a fashion resembling today's social media. The dominant language was Neo-Latin (i.e., post-1500 Latin), as used in a self-conscious literary genre, informed by reference to classical works, and predominantly of humorous intent. Demo considers the structural properties of macaronic verse by investigating a corpus of sixty poems made up of eleven language pairs (including such languages as Dutch, Croatian and Portuguese), written between the late fifteenth century and 1969. A pattern or tendency he discovers is that on average, clusters of words from Germanic languages in the macaronic poems are shorter than those from Romance languages, so the size of the vernacular-language cluster seems to be connected to the language family of the embedded-language items. The building of the multilingual corpus is also detailed in the chapter.

Tom ter Horst and Nike Stam, focusing on code-neutral material in two corpora of Medieval Latin/Irish religious writing, show how corpus methods can be employed to examine ambivalence at the fuzzy boundaries between languages. In “Visual diamorphs: The importance of language neutrality in code-switching from medieval Ireland”, they focus on material which is lexically and grammatically sufficient in both of the languages used and ask whether these visual diamorphs trigger a switch. In their corpora, they identify diamorphic lexical borrowings, function words, abbreviated words and also “emblems”, which are more pictorial symbols, such as the Tironian note for “and”, or abbreviation and suspension symbols, which transcended language boundaries. Personal and place names can also be language-neutral. The rather striking pattern in the medieval Irish data is that 75% of the switch-points occur in conjunction with diamorphs, so code-neutral material was indeed strongly operative in triggering or facilitating a switch.

Alpo Honkapohja is also interested in scribal practices and conventions in manuscripts. In “‘Latin in recipes?’ A corpus approach to scribal abbreviations in 15th-century medical manuscripts”, he confronts the medieval European practice of using abbreviation and suspension symbols which serve to obscure morphemes, thereby creating code-neutral material. His data are taken from recipes in five bilingual medical manuscripts. As with Demo’s investigation of macaronic Neo-Latin verse, there is no corpus of medieval abbreviation symbol usage *per se*, so Honkapohja devised his own encoding system in order to count and log the distribution of the various types of abbreviations used in the manuscripts. He asks what kind of linguistic material was abbreviated, and which patterns can be discerned by looking at abbreviation frequencies. He finds differing ratios of abbreviated words in the English part of the corpus (9%) to those in the Latin part (35%), but conversely observes more variation in the use of English abbreviations than in that of Latin ones; regularity characterises the copying of the most technical part of the Latin vocabulary in particular.

5 Considering contexts

Finally, the fourth section, “Contexts”, is concerned with multilingual practices in a variety of communicative, social and historical contexts. While the other chapters do not dismiss the context which gave rise to multilingual practices in text production, the four studies in this section introduce previously overlooked contexts of multilingual language use, or relate a particular context to much broader questions, be they conceptual concerns or issues of language use. The authors review prevailing ideas about historical code-switching and propose refinements to the terminology used in code-switching and multilingualism studies.

In “Administrative multilingualism on the page in early modern Poland: In search of a framework for written code-switching” Joanna Kopaczyk considers switching between Latin, Polish and Scots in seventeenth-century municipal administrative and legal records pertaining to Scottish immigrants working in early modern Poland/Lithuania, and discusses the context of early modern Cracow. She notes that editions and translations often lose or disguise characteristics of multilingual texts, such as the relationship between language and layout or other visual features. Kopaczyk discusses the nature of switch-points – at the macro-genre level, the discourse, clause, word, morpheme, and orthographic levels, with abbreviations sitting between the morphemic and orthographic levels – and builds these levels into a framework which also accounts for the visual aspect of multilingual writing. The chapter thus links historical multilingualism and code-switching studies with recent research combining the verbal content of texts with layout and material context (e.g. Peikola et al. 2017).

The last three chapters discuss multilingual practices in England. The first one contains examples from early modern England and is also concerned with conceptualisation and frameworks: Aleksi Mäkilähde’s chapter “Approaching the functions of historical code-switching: The case of solidarity” contributes towards theory-formation in the field by shedding light on a term which is often used in research into the functions of or motivations behind code-switching, but which is rarely defined explicitly, namely, solidarity. He teases out the nature of the relationship between the speaker and the hearer, with particular emphasis on the power relations and distance of the two, and illustrates the potential of clearly defined and theoretically solid terms by analysing examples from late-seventeenth-century school drama. The school plays display switching between Early Modern English, Latin and Greek and reflect the linguistic practices prevalent in the school context.

Moving to a context frequently discussed in historical code-switching research, Richard Ingham asks in “Medieval bilingualism in England: On the rarity of vernacular code-switching” whether medieval trilingual writing can reveal anything about spoken usage. Elsewhere, he has shown that Anglo-Norman French continued to be spoken in England by bilinguals until the late fourteenth century. He takes as data religious prose and business texts and finds that code-switching in these two domains patterned quite differently, the clergy being proficient at discrete Latin and not mixing the two vernaculars, whereas the business-record keepers used three languages in their texts. In particular, Ingham argues that the combination of [French article + English noun], which frequently occurs in documentary records, is evidence of "snippets of code-switching" akin to present-day spontaneous code-switched speech. Yet, it is only accounts and charters which show this practice, with encroachment into other Anglo-Norman text-types mainly occurring only towards the end of insular French usage in the fifteenth century. He concludes that the data extant provide a skewed vision of the spoken reality: although code-switching between the vernaculars must have been common in spoken communication in medieval England, written evidence of it is rare outside a few administrative text-types which arose from interaction between bilingual speakers.

Finally, Laura Wright considers the relationship between multilingual practices and standardisation in the late medieval English context in "A multilingual approach to the history of Standard English". She asks how the history of Standard English looks if a multilingual perspective is taken, revealing a staggered swap-over from Medieval Latin to proto-Standard English as the predominant language of written record via a sustained period of mixed language writing in the late fourteenth and fifteenth centuries. She rehearses some recent work by scholars in order to show how the twentieth-century view of the development of Standard English has now been superseded. Emerging Standard English can be explained with reference to socioeconomic change, which is also reflected in the mixed-language phase of business writing between two monolingual stages – Latin first, followed by (proto-Standard) English. It is time the results of the multilingual approach to language history were introduced to textbooks and also utilised in research not primarily concerned with historical code-switching.

6 Conclusion

Together, the fourteen chapters in the three sections described above address nearly all of the desiderata set in Schendl and Wright (2011b) for future research on historical code-switching. Obviously, there are no direct answers to the question about the workings of the bilingual brain in this volume, although historical data could and should be used in researching the psycholinguistics of multilingualism, to complement evidence collected from present-day interaction. Nor do the contributors provide improved descriptions of the French and Latin of medieval England; such accounts are not within the scope of this book, but the chapters by Ingham and Wright, in particular, dissect the use of both languages alongside or mixed with English. Meecham-Jones complements our understanding of multilingual practices on the British Isles with his discussion of Welsh-English contact, while ter Horst and Stam focus on Irish and Latin. Many if not most of the authors examine overlooked primary sources and text-types. Sylvester, for example, examines technical vocabulary collected from mostly non-literary texts, and Skaffari focuses on a prominent domain but an under-researched period; Queiroz de Barros uses a continually updated dictionary as her source of potential code-switches. Others use corpora of multiple genres or domains (Nurmi et al.) or of just one (Honkapohja). Schendl discusses material which constitutes the earliest stage of two domains whose multilingual features in later periods have already been objects of research. Short-term diachronic change is addressed by Tuominen, while there is a diatopic dimension in the chapter by Demo, who examines the same bilingual practices in different areas and different language pairs. Kopaczyk adapts theoretical approaches to code-switching, as does Mäkilähde.

In sum, the chapters in this book show that there are multiple ways of viewing concepts such as code-switching and borrowing. The contributors highlight questions of identity, and the many ways in which writers tailored their language to specific interlocutors in different communicative situations. They posit polarities and continua, from individual practices to wholesale language shift, they identify patterns characterising certain text-types, and more than one author confronts the physicality of the evidence: script-switching, use of red ink, positioning on the page, pictorial emblems. First and foremost, this collection emphasises that texts of the past were produced by people with multilingual repertoires and that the communicative practices witnessed in them reflect ongoing and earlier language contact situations. The historical individual deployed their linguistic resources according to the pragmatic constraints of the relevant community at the relevant point in time, often moving fluently from one language to another to express themselves in ways bearing a close resemblance to what today is known as translanguaging, i.e. using their multilingual repertoire as an integrated communication system (see e.g. García and Li Wei 2014). We hope that this volume will serve to stimulate interest in taking a multilingual perspective, as so much historical data extant, from the most ancient to the most recent, is composed of more than one language.

References

Adams, J. N., Mark Janse & Simon Swain (eds.). 2002. *Bilingualism in ancient society: Language contact and the written text.* Oxford: Oxford University Press.

Auer, Peter. 1999. From code-switching via language mixing to fused lect: Toward a dynamic typology of bilingual speech. *International Journal of Bilingualism* 3(4). 309–332.

Blom, J. P. & John J. Gumperz. 1972. Social meaning in linguistic structures: Code-switching in Norway. In John J. Gumperz & Dell Hymes (eds.), *Directions in sociolinguistics: The ethnography of communication*, 407–434. New York, NY: Holt, Rinehart & Winston.

Blommaert, Jan. 2010. *The sociolinguistics of globalization*. Cambridge: Cambridge University Press.

García, Ofelia & Li Wei. 2014. *Translanguaging: Language, bilingualism and education*. Basingstoke: Palgrave Macmillan.

Gardner-Chloros, Penelope. 1995. Code-switching in community, regional and national repertoires: The myth of the discreteness of linguistic systems. In Lesley Milroy & Pieter Muysken (eds.), *One speaker, two languages: Cross-disciplinary perspectives on code-switching*, 68–89. Cambridge: Cambridge University Press.

Gardner-Chloros, Penelope. 2009. *Code-switching*. Cambridge: Cambridge University Press.

Gumperz, John J. 1982. *Discourse strategies* (Studies in Interactional Sociolinguistics 1). Cambridge: Cambridge University Press.

Haugen, Einar. 1950. The analysis of linguistic borrowing. *Language* 29(2). 210–231.

Heller, Monica. 1988. Introduction. In Monica Heller (ed.), *Codeswitching: Anthropological and sociolinguistic perspectives* (Contributions to the Sociology of Language 48), 1–24. Berlin & New York: Mouton de Gruyter.

Hüning, Matthias, Ulrike Vogl & Olivier Moliner (eds.). 2012. *Standard languages and multilingualism in European history* (Multilingualism and Diversity Management 1). Amsterdam & Philadelphia: John Benjamins.

Kolehmainen, Leena & Janne Skaffari. 2016. Multilingual practices in contemporary and historical contexts: Interfaces between code-switching and translation. *Multilingua* 35(2). 123–135.

Kranich, Svenja, Viktor Becher & Steffen Höder. 2011. A tentative typology of translation-induced language change. In Svenja Kranich, Viktor Becher, Steffen Höder & Juliane House (eds.), *Multilingual discourse production: Diachronic and synchronic perspectives* (Hamburg Studies on Multilingualism 12), 11–43. Amsterdam & Philadelphia: John Benjamins.

Matras, Yaron. 2009. *Language contact*. Cambridge: Cambridge University Press.

Myers-Scotton, Carol. 1993. *Duelling languages: Grammatical structure in codeswitching*. Oxford: Clarendon Press.

Nurmi, Arja, Tanja Rütten & Päivi Pahta (eds.). 2017. *Challenging the myth of monolingual corpora* (Language and Computers: Studies in Digital Linguistics 80). Leiden & Boston: Brill Rodopi.

Pahta, Päivi. 2012. Code-switching in English of the Middle Ages. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford handbook of the history of English*, 528–537. Oxford: Oxford University Press.

Pahta, Päivi & Arja Nurmi. 2006. Code-switching in the Helsinki Corpus: A thousand years of multilingual practices. In Nikolaus Ritt, Herbert Schendl, Christiane Dalton-Puffer & Dieter Kastovsky (eds.), *Medieval English and its heritage: Structure, meaning and mechanisms of change* (Studies in English Medieval Language and Literature 16), 203–220. Frankfurt: Peter Lang.

Peikola Matti, Aleksi Mäkilähde, Hanna Salmi, Mari-Liisa Varila & Janne Skaffari (eds.). 2017. *Verbal and visual communication in early English texts* (Utrecht Studies in Medieval Literacy 37). Turnhout: Brepols.

Poplack, Shana. 1980. “Sometimes I’ll start a sentence in Spanish *y termino en español*”: Toward a typology of code-switching. *Linguistics* 18(7–8). 581–618.

Schendl, Herbert. 2012. Multilingualism, code-switching, and language contact in historical sociolinguistics. In Juan M. Hernández-Campoy & Juan Camilo Conde-Silvestre (eds.), *The handbook of historical sociolinguistics*, 520–533. Malden, MA: Wiley-Blackwell.

Schendl, Herbert & Laura Wright. 2011a. Introduction. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English* (Topics in English Linguistics 76), 1–14. Berlin & Boston: Mouton de Gruyter.

Schendl, Herbert & Laura Wright. 2011b. Code-switching in early English: Historical background and methodological and theoretical issues. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English* (Topics in English Linguistics 76), 15–45. Berlin & Boston: Mouton de Gruyter.

Schendl, Herbert & Laura Wright (eds.). 2011c. *Code-switching in early English* (Topics in English Linguistics 76). Berlin & Boston: Mouton de Gruyter.

Trotter, D. A. (ed.). 2000. *Multilingualism in later medieval Britain*. Cambridge: D. S. Brewer.

Weinreich, Uriel. [1953] 1963. *Languages in contact: Findings and problems*, 2nd printing. The Hague: Mouton and Co.

Wenzel, Siegfried. 1994. *Macaronic sermons: Bilingualism and preaching in late-medieval England*. Ann Arbor, MI: University of Michigan Press.

Winford, Donald. 2003. *An introduction to contact linguistics* (Language in Society 33). Malden, MA: Wiley-Blackwell.

Penelope Gardner-Chloros

2 Historical and modern studies of code-switching: A tale of mutual enrichment

1 Introduction

The papers in this volume represent the most recent stage in research carried out by a group of historical sociolinguists (HS) whose centres of interest have increasingly been converging with those of sociolinguists who work on contemporary spoken multilingualism. As Schendl and Wright write in the introduction to their (2011) edited collection, *Code-switching in early English*, historical mixed-language texts have been intensively studied since the 1990s, giving rise to substantial publications which in some cases have transformed the discipline in which they are rooted, and making the study of historical code-switching into a new sub-field of research.[1]

This paper addresses the fact that, despite the appearance of significant publications on historical sociolinguistics (see for example the papers in Hernandez-Campoy and Conde-Silvestre 2012), the traffic has been largely one way, and mainstream or contemporary sociolinguists[2] (MS) have so far failed to profit sufficiently from these new insights. Although many questions of fundamental concern to sociolinguistics can be explored effectively through historical evidence, spontaneous linguistic production overwhelmingly remains the focus of sociolinguistic theory and analysis. To address the question of why the disciplines of historical sociolinguistics and "mainstream" sociolinguistics have remained so separate despite their common interests, one can start by comparing the objects of enquiry, the aims and the methods of both camps.

2 Common interests?

The first question is to what extent historical sociolinguists and contemporary sociolinguists have common interests and common objectives. As Anipa (2012: 186) points out, historical data can be studied synchronically "as snapshots of

---

[1] Skaffari and Mäkilähde (2014: 275) propose the adoption of the term *historical contact linguistics* both for contact linguists working on historical situations as well as historical linguists investigating the role of contact in their source material.

[2] It goes without saying that these terms are just a shorthand. "Sociolinguists working on contemporary material" would be more accurate, but excessively cumbersome.

DOI 10.1515/9781501504945-002

inclusive linguistic repertoires available to past users of a language", i.e. in accordance with an aim which directly parallels that of much contemporary sociolinguistic research. However it is perhaps more significant that HS and MS have a common interest in language change. Thus for example Chirita ([1996] 2011) describes the development of verb-second syntax in German in terms of a historical influence from Latin. In a synchronic and contemporary context, Freywald et al. (2015) describe changes in the verb-second pattern in several Germanic languages as a function of contact with multilingual youth varieties. It can safely be assumed that if a pattern becomes acceptable orally, it will find its way into writing. What the historical study shows is that the reverse may also apply, in that Latin would have influenced German first through the written medium. This example illustrates one way in which the two types of research have the potential to cross-fertilize.

As regards code-switching, Schendl and Wright (2011: 21–22) point out that, where sufficient suitable material is available, the study of historical texts with code-switching has the potential to provide a "decade by decade account of new constructions entering the system and older ones being abandoned". This time-depth is potentially an invaluable resource. Historical sociolinguists are in a strong position to respond to the views expressed by contemporary sociolinguists (Backus 2014; Auer 1999), to the effect that code-switching should not only be studied synchronically but also as an aspect of language contact which can lead to diachronic language change. Auer (1999) for example describes the historical process whereby under specific circumstances, language contact leads first to code-switching, then, under the right conditions, to "code-mixing", and then finally in some cases to "fused lects". Historical texts can also shed light on how to draw a useful distinction between code-switching and borrowing, a question which has continued to preoccupy the MS camp (Poplack and Dion 2012; Myers-Scotton 1992; Gardner-Chloros 2009). A historical approach allows "diachronic hindsight", and comparing earlier and later texts can help establish which features have been permanently adopted. In a similar vein, Muysken (2000: 249) writes: "As the extent of language contact grows, the type of mixing will shift from insertional to either alternational or congruent lexicalization." Illustrating this progression, Trotter (2003) shows how in mixed-language texts dating from the 11th to the late 15th centuries, English was gradually incorporated into French texts first in the noun and verb-root spot, then in noun-phrase-modifier position, and finally in the form of closed class function words. Evidence like this is gold-dust for researchers in the area of code-switching, and one is justified in asking why contemporary sociolinguists, who may understandably lack the linguistic and historical expertise to deal with the original manuscripts, have not taken more advantage of the HS's analyses.

3 Oral and written data

The fact that most MS work essentially with transcripts of oral data as opposed to written texts – i.e. with a different type of written text – has not escaped the attention of HS (Schendl and Wright 2011). It may however be worth looking at this difference in a bit more detail.

Labov's (1972: 208) early principles still carry considerable weight in MS: "The vernacular, in which the minimum attention is paid to speech, provides the most systematic data for linguistic analysis." This principle has been an enduring one and over 40 years later it is still not unusual for research sent for publication to sociolinguistic journals to be criticized or rejected because the data is insufficiently naturalistic and spontaneous. Adaptations of the golden rule are somewhat more acceptable when the object of enquiry is not purely phonological variation.[3] Pronunciation is highly sensitive, mainly in a subconscious way, to the speakers' sense of the formality of a situation[4], and unmonitored performance is therefore the gold standard as regards studies of phonology. It is more acceptable to set up informal interviews and discussions when collecting broader data on a particular variety. This is the case for example with multicultural urban repertoires, where access to settings where these varieties are used entirely spontaneously is problematic for researchers. A recent example is Cheshire et al. (2008, 2011), which traces the emergence of Multicultural London English through an analysis of features from several linguistic levels including discourse and morphosyntax as well as phonology. Milroy had already pointed out in 2002 that sociolinguistics would have to adapt in various ways if variation was to be studied in situations of contact: "It is clear that Labov's basic unit of study, the speech community, is conceived as a construct which does not attend to contacts between communities or to the crosscutting influences entailed by mobility" (2002: 4). Skaffari and Mäkilähde (2014: 254) point to a comparable lack of focus on behalf of some HS on describing language contact in the communities which gave rise to the code-switching texts they study – and, also significant, to a frequent lack of definition of the term *language contact* even in research which deals directly with its consequences, leaving it as a rather indeterminate concept without clear limits (2014: 259).

Indeed until recently, when the social media became an object of study, describing a variety by means of purely written material would have been largely anathema to MS. As Lipski (1982: 192) remarked, "because of the very fact that written documents, particularly those classed as literary, involve not only conscious reflection but also the inherent correction, editing, and rewriting process that accompanies any act of writing, it cannot be claimed that such texts may be used as specimens of naïve, spontaneous linguistic production" (see also Coupland and Jaworski 2009: 5; Lillis 2013). This applies to code-switching as to other aspects of speech, and it is true that written code-switching is likely to differ systematically in certain respects from the oral type. Sebba (2012) questions whether the term code-switching is applicable to the written medium at all, or if it describes a different phenomenon altogether. More specifically, Adams (2003: 304) points out the fact that composing written texts involves a good deal of thought, so written multilingual texts are likely to involve less unmarked code-switching than the spoken variety. Yet Schendl (2000: 77) has shown, using a classification according to parts of speech, that the differences between spoken and written code-switching appear, on the basis of a selection of corpora, to be largely quantitative rather than qualitative.

---

[3] In fact, many aspects of phonology have been studied through written sources. The great vowel shift (Baugh and Cable 2002) is a prime example.

[4] *viz.* Labov's technique of "rapid anonymous observation" in New York stores (1972).

In fact, there are different degrees of planning, forethought and spontaneity both in the written and the oral mode. For example, several studies of code-switching have focused on computer-mediated communication (Siebenhaar 2006; Androutsopoulos 2007; Montes-Alcalá 2007), pointing out that material posted on Facebook, Twitter and other social media shows features traditionally associated with speech rather than writing. Historical parallels to these hybrid informal genres are found in certain kinds of letter-writing (Pahta and Nurmi 2009). A broader view of what constitutes sociolinguistic data would include representations of bilingual speech in writing (e.g. Stolt 1964; Timm 1978), and dialogue in novels and plays (e.g. in the comedies of Plautus; cf. Swain 2002: 130–131) – texts which would lose their impact if they were not true to life (Weston and Gardner-Chloros 2015; Gardner-Chloros 2013a). Different types of text, such as court records, epigraphy, business lists and literature involve different degrees of processing by the author (Schendl and Wright 2011: 22–23). Indeed, if sociolinguistic analysis is the attempt to relate linguistic practices to the social characteristics of speakers, then the differences between texts designed for different audiences and with different purposes are an intrinsic part of what sociolinguists need to explain. Le Page (1997: 31–32) had similar thoughts about the sociolinguist's primary duty, which is to describe what "a language" is: "We set out how we saw such a concept evolving from observation of discourse, through the stereotypes denoted by such language-owning names as 'English' or 'French', to that of the most highly abstract and focused Chomskyan 'grammar'; and how actual linguistic behaviour was influenced by the stereotypes as progressively it was named, formalized, standardized, institutionalized, and totemized by a society."

As Lipski (1982: 192) noted, “there is [. . .] a vast amount of linguistic, psychological, and aesthetic information to be obtained from a careful consideration of code-switching in its written form”. This is not information which MS should ignore. Code-switching is not a single, unitary phenomenon either in the written or the spoken mode, it varies as much as any other aspect of production according to text and context, and no single model has so far proved up to the task of providing parameters which apply to all known instances (Muysken 2000). We have noted that the written and spoken varieties manifest the same functions, albeit in different proportions (Schendl 2000: 77). Nurmi and Pahta (2012: 63) also argue that “the study of historical data brings to light many of the same processes and strategies of multilingual communication as are present in present-day societies”. As Adams (2003) suggested, much depends on how marked the language choices are in different spoken or written genres (Barasa 2016; Danet and Herring 2007; Eastman 1992; Gardner-Chloros 2009; Myers-Scotton 1983). Certainly a number of parallels can be found. For example, while Sebba (2012: 11–17) expressed misgivings about the applicability of oral models of code-switching to written language which comprises multimodal features, Machan’s (2011) analysis of Middle English manuscripts shows that some of these features have close parallels in oral devices which frame various parts of an utterance. The – relatively few – accounts of contemporary code-switching in the written medium could provide a way of establishing more specifically how spoken and written code-switching relate to one another in different contexts. For example, McClure’s (1998) comparison of press articles involving English and the respective national languages in Mexico, Spain and Bulgaria, which showed how the type of code-switching encountered in each of these countries reflected the prevalent attitudes towards English, could be followed up with a comparison with spoken code-switching involving English in each of these countries.

Leaving aside the MS’s overwhelming emphasis on the spoken word, MS and HS use similar methodological approaches, though there seems to be a starker division between users of quantitative and qualitative methods among MS. By contrast Nurmi and Pahta (2012) choose to combine their quantitative corpus-based study of code-switching in women’s correspondence between 1400 and 1800 with a qualitative analysis.

HS and MS’s common interest in language change means that various types of time-related studies are found in the latter’s work also, though exactly how to carry out such a study remains a significant topic of discussion (Sankoff 2006; Sankoff and Wagner 2006). Is it best, for example, to compare the speech of younger and older speakers, or, more impractically, to re-sample the same or similar subjects some years later? The time-frame of MS studies is necessarily brief, a matter of a few decades at best (Ashby 2001; Harrington, Palethorpe, and Watson 2000; Labov 1963; Trudgill 1988). Real-time studies of code-switching are even more of a rarity, given the practical problems of collecting spontaneous material from the same or closely comparable bilinguals at separate points in time (Auer 1998; Bentahila and Davies 1983; Gardner-Chloros 2013b; Myers-Scotton 1998).

Last but not least, research like that of Lexander (2012) on multilingual texting in Senegal, or that of Philipp Angermeyer (2012) on Russian-American script alternation (both in Sebba, Mahootian, and Jonsson 2012), or Sarkar and Winer's study (2006) of multilingual rap in Montreal, are helping not only to show the permeability of the boundaries between spoken and written multilingual practices, but also to break down rigid boundaries in the research traditions, to the benefit of all parties.

4 Distinguishing languages and defining the "matrix"

The third issue I will raise here concerns an aspect of the analyses used by both HS and MS when dealing with multilingual material. Both types of researchers have been attracted to the notion of a "matrix language", taken from the work of Myers-Scotton (1983 and ff.). According to Myers-Scotton, the matrix can be identified by various criteria: (1) it is the language which imposes a grammatical framework on a sentence, (2) it is the language which is psycholinguistically most activated in the speaker's brain, and (3) it is the one which is least marked, at a social level, for this type of interaction. Criteria 2 and 3 are difficult to verify in practice, especially for written data but actually for spoken data too; the first would appear to beg the question; finally there are cases where the different criteria do not all point to the same language being the matrix. In previous publications I have outlined the reasons why this concept may mislead us into thinking that reality is tidier than it actually is (Gardner-Chloros 2009: 100–104; Gardner-Chloros and Edwards 2010); not all these arguments can be rehearsed here. Essentially, speakers' linguistic repertoires are unlikely to coincide with the externally defined notion of separate "languages"; and the type of combinations the individual produces vary according to circumstances. It is therefore difficult to assert that the base line usage of an individual can be reduced to specific "languages". As Gafaranga (2007) suggests, we should instead consider what "medium" is chosen by an individual for a specific interaction.

In any case it is unclear what is to be gained, apart from theoretical tidiness, by describing multilingual texts as being made up of a "matrix" language (actually a matrix grammar) and its complement, an “embedded” language. Yet the term “matrix” has passed, with varying levels of success – and often without explaining why it might be helpful – into the vocabulary of many HS and MS. It is used for example by Diller (1997–1998) to examine a body of medieval English drama; yet it has been shown that a variety can be identified as the matrix according to one of the above criteria but not according to another one. For example Rindler Schjerve (1998: 243), who studied code-switching between Italian and Sardinian, found that conversations and utterances which were overwhelmingly in Sardinian from a quantitative perspective, nevertheless violated the proposals of the Matrix Language Frame (MLF) in demonstrating grammatical incursions from Italian – i.e. the “embedded language” could impose its rules on the “matrix language”, which is supposed to be ruled out. Callahan (2002) found the MLF worked well for analysing narrative sections of a sample of Hispanic American literature, but broke down when applied to the dialogue (when one might have expected exactly the reverse since the model was developed for speech).

As Muysken (2000) points out, the model in fact works best in certain types of situation where bilingual proficiency is asymmetric (e.g. colonial or recent migrant settings), where code-switching is mainly insertional rather than alternational, and less well in others. This is illustrated for example by Backus’s (1999, 2009) work on Turkish speakers in the Netherlands.

Let us look next at some practical issues in determining the matrix language in specific cases. For example, both MS and HS analysing multilingual texts face the problem of assigning a long list of words or items to one language or another. This decision is not necessarily rendered easier, in the case of MS, by having access to the recordings, as both loan-words and spontaneous code-switches can be more or less integrated with the words around them, at a phonological as well as morphological level (Gardner-Chloros 2009: 10–11).

(1) **Item** petites **bedes** de **coper** en un **pice** de drap linge

Example (1), from the 14th- and early 15th-century records of the Goldsmiths’ Company (Butterfield 2013: 450), illustrates a type of problem which is also raised by spoken examples of code-switching. Both spoken and written data often include items such as cognates which can be ascribed to either of the varieties, but also various ambivalent – or bivalent – elements, including for example “bare forms” (Halmari 1997: 89–90). Others may simply be ambiguously pronounced or grammatically framed, like *item* and *pice* here. Wright (2002) points out in her analysis of code-intermediate phenomena in medieval business texts that this type of code-switching is the norm for particular writers rather than a sign of poor competence in one of the varieties.

A modern example from Clyne (1987: 754), from the speech of an elderly Dutch immigrant to Australia, shows how comparable issues arise in the analysis of spontaneous speech:

(2) Ja, **in de**, **in de ('in the') big place is** het **a lot**, nou ja ('well let's say'/ 'like'), je kan't ('you know it'), t' is de same als **hier**. Je habt ('you have') **Melbourne en de other places** met **de** ('with the') **high flats and** so ('suchlike'). Dat heb je in Holland ook ('you have that in Holland too'). Maar ('but') 'n, maar **a lot of places nou (now)**, de **same before we go**. D'r is ('there is'), **we go to my sister in Apeldoorn**, en zi hef ('and she has') **de same place** noog ('still').

In this extract, the words *in*, *de*, *is*, *here*, *en de* and *now* are all sufficiently similar in the two languages as to have an ambivalent status in the speaker's idiolect. The ambivalent status may even be deliberate, as in the texts studied by Woolard and Genovese (2007). They describe a "mini-genre" of "bivalent" texts, dating from the 16th to 18th centuries, composed in such a way that they can be read simultaneously as Latin and as various Romance languages, notably Spanish – a way of "punning" with linguistic affiliation rather than with meaning.

Myers-Scotton (2002: 22) argues that the presence of ambivalent items, however common, and whether deliberate or not, should not deflect us from applying the MLF. Instead she allows for the possibility that the matrix itself can be a mixed or "composite" code, i.e. one which is made up of an "abstract frame composed of grammatical projections from more than one variety".

However even a "composite matrix" cannot account for another characteristic of multilingual writing and multilingual speech, i.e. that both contain a degree of variability which one rarely encounters in a standard language. This applies both to historical patterns evidenced in writing, and to contemporary spoken ones. Wright (2002) shows for example that in Anglo-Norman and English mixed texts, adjectives from both languages could be found both as premodifiers and postmodifiers, i.e. following the Anglo-Norman placement rule or the English one.

(3) **wyghtes nouell** ('new weights'): Romance order with English noun and Anglo-Norman modifier

nouell plom ('new lead'): English order with Anglo-Norman noun and Anglo-Norman modifier

nouell led ('new lead'): English order with English noun and Anglo-Norman modifier

(Gybon Maufield [1394] in Wright [2002: 474])

A modern parallel in speech can be found in a study carried out in Strasbourg to assess how code-switching had changed in the last 25 years (Gardner-Chloros 2013b). Due to a general falling off in competence in Alsatian, apparently haphazard variation is found in the dialect – at least in utterances containing code-switching – by the same speaker and within the same conversation, between German and French word-order, e.g. inversion of verb and subject in the first two cases, and no inversion in the second two.

(4) 1. Germanic word-order:

pour me calmer	**muss**	**ich**	noch	lese
in order to calm down	must	I	still	read

‘in order to calm down I still have to read’

les symphonies de Beethoven	**kann**	**i**	alli	uswendi
Beethoven’s symphonies	know	I	all	by heart

‘I know all Beethoven’s symphonies by heart’

2. French word order:

A mon avis **ich hab** d’s selwe problem
‘In my opinion I have the same problem’

Ceci dit **ich fend** doch d’lit heutsodaej [. . .]
‘Having said that I find people nowadays [. . .]’

Sociolinguists generally identify such variation as the synchronic manifestation of language change. Consequently an essentially static model of the grammar[5] cannot, by definition, circumscribe this highly characteristic quality of code-switching.

5 Code-switching as individual expression

Given the amount of variation associated with code-switching even within the speech of a single speaker, should code-switching be seen less as characterizing a society or a community and more as an expression of individual identity? This can be examined at the interface of speech and writing, especially when the individual is not part of a community of code-switchers. It can be seen for example in the correspondence of Vincent van Gogh, using the new multivolume edition of his letters and the accompanying searchable website (Jansen, Luijten, and Bakker. 2009), especially given that we also have some information about his speech habits.

---

[5] Myers-Scotton and Jake (2009) propose a “matrix language turnover hypothesis” which is intended to capture this process of change, but does not solve the problem of reliance on an essentially binary description (see also Bolonyai 1998).

At a certain point in his lifelong correspondence with his brother Theo, Vincent switched over from using their mother-tongue, Dutch, and started using French. This rather clear act of identity followed their living together in Paris for a few years, and coincided with Vincent "coming out" as a (French) artist. His code-switching is the product of his own idiolect rather than representing any community norm, and is usually grammatically unremarkable. But sometimes more dramatic alternation, betraying a troubled state of mind, can be found, as in Letter 184 to his friend Rappard. In this letter, code-switching occurs at all grammatical levels, lexical, inter- and intrasentential, and clearly betrays Vincent's inner turmoil.

(5) De tweede soort maitressen zijn van heel ander kaliber. **Collet montées** – **pharisiennes** – **jésuites!!!** 't Zijn zoodanige **femmes de marbre** – **sphinx** – **vipères glacées** – welke mannen aan zich zouden willen verbinden, geheel en al, eens en voor altijd, zonder van haar kant zich ooit geheel en al zonder slag om den arm, sans arrière pensee over te geven. **Elles sucent le sang**. **Ces maîtresses-là elles *glacent* les hommes**, **et les *pétrifient***.

Doch ik zeide 't was **purement artistique, mon cher** – en dus vergelijk ik de eerste soort maîtressen, **celles qui brûlent**, bij die rigting der kunst die in 't gemeene vervalt.

['The second kind of mistress is of an entirely different disposition. **The strait-laced** – **Pharisees** – **Jesuits!!!** They are the kind of **women of marble** – **sphinx** – **cold vipers** – who would like to bind men to themselves, entirely, for ever and ever, without her ever surrendering wholeheartedly, entirely, unreservedly. **They suck your blood**. **Such mistresses *freeze* men**, **and *petrify* them**.

But I said it was **purely artistic**, **my dear chap** – and so I compare the first kind of mistress, **those who burn**, with that school of art that lapses into generalities.']

(Letter 184 to Rappard, Etten, 12 November 1881)

Furthermore, we have some reported evidence that this written code-switching mirrors that which occurred in his speech at such times. His biographers report that in Vincent's moments of mental anguish, "shouting and gesticulating, he would plunge heedlessly into arguments, pouring out sentences in a wild mix of Dutch, English and French" (Naifeh and White Smith 2012: 511). It is also worth pointing out that in his letters, van Gogh switches not only between languages but also between writing and sketches to complete his meaning, exemplifying the multimodality with which HS are familiar from illuminated manuscripts.

6 CA analysis

At first sight, Conversation Analysis is the MS linguistic framework which is least likely to be relevant in written or historical texts because:

a) Its object is specifically to reveal aspects of conversation as such.
b) It relies largely on classic features of spoken language, such as pauses, repetitions, etc.
c) Practitioners of classic CA explicitly reject the relevance of contextual knowledge, with proponents arguing that the conversation itself – or rather its transcript – contains all necessary clues to guide our interpretation of the linguistic choices (Auer 1998).

By contrast, HS generally value contextual knowledge – though they may lack it in specific instances – in order to understand the implications of the texts they study. So can CA be applied in the case of historical code-switching? There are in fact some intriguing consequences of this method for literary code-switching. As a methodology, CA relies on "contextualization cues" (Gumperz 1982) to interpret what is happening in a given conversation. Such cues for example include pausing, verbal reformulations and repair, and changes in intonation. Unfortunately, apart from some generally rather stylized renditions of pauses or reformulations, this type of conversational detail is generally absent from prose literature, on the basis that it detracts from a fluent reading experience. However, in bilingual contexts, code-switching itself often acts as a contextualization cue, notably in the specific sequencing of languages. It is clear therefore that code-switching provides at least the potential for looking at conversational structure in fiction in a new way.

This recognition of the structural significance of a feature which is not only detectable at aural level, unlike most features used by CA analysts, should increase the potential for applying this type of approach to the written language. Furthermore in at least one influential model, Myers-Scotton and Bolonyai's (2001) Rational Choice Model – the later development of the "Markedness

Model" – it is clear that power relations and other factors outside the conversation itself are necessary to understand the linguistic choices within it.

The writing genre which comes closest to using the norms which conversation analysts usually study is drama. In this, code-switching is represented through the ages. It is found in Roman comedies, notably by Plautus. For example, the use of Greek, noted by Adams (2003), can serve as a form of accommodation by certain characters to others – even when they are not physically present – who are themselves prone to using Greek. Code-switching has also been studied in Middle English mystery and morality plays, where Schendl (2000: 77) shows that switches into Latin serve a variety of socially meaningful functions. Outside the field of drama proper, the Latin and French code-switches in Piers Plowman have been the subject of discussion from a discourse analytic point of view (Machan 1994).

In more modern works, Jonsson (2010) has explored the "local functions" of Spanish–English code-switching in a series of Chicano plays, concluding that it is used "to mark closeness, familiarity, to emphasize bonds, and to include or, on the contrary, to mark distance, break bonds and exclude" (2010: 1296); in short, for many of the functions attested in conversational code-switching.

Finally, the notions of identity, stance and space, which have assumed importance in recent sociolinguistic research, can also be explored in relation to bi/multilingual expression. They are an offshoot of the challenging of fixed boundaries in the linguistic as well as the cultural sphere, and the recognition of a state of affairs which carries a somewhat fashionable label, "superdiversity" (Blommaert and Rampton 2012). Like "translanguaging" (Garcia and Li Wei 2014), this term has arisen in recognition of the fact that our societies are increasingly ethnically, socially and linguistically mixed, that the types of linguistic hybridity which are found are becoming ever more complex, and that speakers who associate themselves with just one or two clear sets of cultural, ethnic or linguistic reference points are gradually becoming a minority.

This view of political and social reality will not sound unfamiliar to linguists working on a range of pre-19th-century material. Indexing one's identity, one's attitude or stance, and defining a discursive space through appropriate and creative mixing of styles and varieties is something for which many historical examples can be found. Adams (2003), for example, shows that Roman funerary inscriptions, far from responding to some pre-ordained convention dictating their form, were often dictated by the deceased during their lifetime, often in a combination of languages designed specifically to represent their cultural and intellectual identity. Similarly Osborne (2012) describes a grave stele from 3rd century BC Athens with inscriptions in Greek and Phoenician and culturally mixed imagery, which he claims represents a kind of deliberate material and linguistic *bricolage* appropriate to the composite identity of the person who commissioned it. Historical examples are just as appropriate as contemporary ones to illustrate the fact that language choices, including the choice to code-switching, not only reflect but also construct relations of power and hegemony between groups (Heller 2011).

7 Conclusion

MS and HS face different challenges and difficulties in interpreting their data. HS depend on what documents are extant and sometimes lack a piece of the puzzle; MS on the other hand may be unable to see the wood for the trees and can scarcely foresee the likely outcomes of change, despite the centrality of language change to the discipline. Despite this, this paper has argued that HS and MS have many common interests and use essentially similar methods. The primacy of oral data for MS is less significant now that the emphasis on phonological variation is no longer as exclusive as it was in the past in sociolinguistics. It has also been argued that understanding how differences between and within genres – written as well as spoken – are represented and come about is itself part of the sociolinguist's task. The study of written multilingual texts – in historical as well as contemporary material – helps us to understand the connections between code-switching and different genres. In many ways it is precisely the variety and variability in code-switching, both oral and written, which makes it interesting. This variety militates against the use of the concept or term "matrix language", which assumes an elusive primacy of one language over another in all bilingual productions.

The systematic study of the relationship between speech and writing is still an underdeveloped science (Carter and McCarthy 1995). A step towards a better understanding of the connections would involve the setting up of comparisons between oral and written code-switching in the same (contemporary) settings. Clarifying how spoken and written code-switching relate to one another in the contemporary period would have potential benefits for our understanding of historical code-switching, for which only written evidence is available. The rise of computer-mediated communication, with its many speech-like qualities, and the current interest in multimodality, also provide a promising avenue for understanding this relationship. But the study of these very modern manifestations of written code-switching should not exclude engagement with its past life, and indeed comparisons between the two.

References

Adams, J. N. 2003. *Bilingualism and the Latin language*. Cambridge: Cambridge University Press.

Androutsopoulos, Jannis. 2007. Bilingualism in the mass media and on the internet. In Monica Heller (ed.), *Bilingualism: A social approach*, 207–230. Basingstoke: Palgrave Macmillan.

Angermeyer, Philipp S. 2012. Bilingualism meets digraphia: Script alternation and hybridity in Russian-American writing and beyond. In Mark Sebba, Shahrzad Mahootian & Carla Jonsson (eds.), *Language mixing and code-switching in writing: Approaches to mixed-language written discourse*, 255–272. New York: Routledge.

Anipa, K. 2012. The use of literary sources in historical sociolinguistic research. In Juan M. Hernandez-Campoy & J. Camilo Conde-Silvestre (eds.), *The handbook of historical sociolinguistics*, 170–191. Oxford: Wiley-Blackwell.

Ashby, William J. 2001. Un nouveau regard sur la chute du *ne* en français parlé tourangeau: s'agit-il d'un changement en cours? *Journal of French Language Studies* 11(1). 1–22.

Auer, Peter (ed.). 1998. *Code-switching in conversation: Language, interaction and identity.* London: Routledge.

Auer, Peter. 1999. From codeswitching via language mixing to fused lects: Toward a dynamic typology of bilingual speech. *International Journal of Bilingualism* 3(4). 309–332.

Backus, Ad. 1999. Evidence for lexical chunks in insertional code-switching. In Brent Brendemoen, Elizabeth Lanza & Else Ryen (eds.), *Language encounters across time and space*, 93–109. Oslo: Novus Press.

Backus, Ad. 2009. Codeswitching as one piece of the puzzle of language change: The case of Turkish yapmak. In Ludmila Isurin, Donald Winford & Kees de Bot (eds.), *Multidisciplinary approaches to code switching*, 307–336. Amsterdam & Philadelpia: John Benjamins.

Backus, Ad. 2014. Towards a usage-based account of language change: Implications of contact linguistics for linguistic theory. In Robert Nicolai (ed.), *Questioning language contact: Limits of contact, contact at its limits*, 91–118. Leiden & Boston: Brill.

Barasa, Sandra. 2016. Spoken code-switching in written form? Manifestation of code-switching in computer mediated communication. *Journal of Language Contact* 9(1). 49–70.

Baugh, Albert C. & Thomas Cable. 2002. *A history of the English language*, 5th edn. London: Routledge.

Bentahila, Abdelali & Eirlys E. Davies. 1983. The syntax of Arabic–French code-switching. *Lingua* 59(4). 301–330.

Blommaert, Jan & Ben Rampton. 2012. Language and superdiversity. *MMG Working Paper* (12–09). http://hdl.handle.net/11858/00-001M-0000-000E-7CFE-9 (accessed 20 May 2016).

Bolonyai, Agnes. 1998. In-between languages: Language shift/maintenance in childhood bilingualism. *International Journal of Bilingualism* 2(1). 21–43.

Butterfield, Ardis. 2013. Fuzziness and perceptions of language in the Middle Ages: Part 3: Translating fuzziness: Countertexts. *Common Knowledge* 19(3). 446–473.

Callahan, Laura. 2002. The matrix language frame model and Spanish/English codeswitching in fiction. *Language and Communication* 22(1). 1–16.

Carter, Ronald & Michael McCarthy. 1995. Grammar and the spoken language. *Applied Linguistics* 16(2). 141–158.

Cheshire, Jenny, Sue Fox, Paul Kerswill & Eivind Torgersen. 2008. Ethnicity, friendship network and social practices as the motor of dialect change: Linguistic innovation in London. *Sociolinguistica* 22. 1–23.

Cheshire, Jenny, Paul Kerswill, Sue Fox & Eivind Torgersen. 2011. Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics* 15(2). 151–196.

Chirita, Diana. [1996] 2011. Latin influence on German word order: A discussion of Behagel's theory. In Irmengard Rauch & Gerald F. Carr (eds.), *Insights in Germanic linguistics II: Classic and contemporary*, 9–26. Berlin & New York: Mouton de Gruyter.

Clyne, Michael. 1987. Constraints on code switching: How universal are they? *Linguistics* 25(4). 739–764.

Coupland, Nikolas & Adam Jaworski. 2009. *The new sociolinguistics reader*. Basingstoke: Palgrave Macmillan.

Danet, Brenda & Susan C. Herring. 2007. *The multilingual internet*. Oxford: Oxford University Press.

Diller, Hans-Jürgen. 1997–1998. Code-switching in medieval English drama. *Comparative Drama* 31(4). 500–537.

Eastman, Carol (ed.). 1992. *Codeswitching*. Clevedon: Multilingual Matters.

Freywald, Ulrike, Leonie Cornips, Natalia Ganuza, Ingvild Nistov & Toril Opsahl. 2015. Beyond verb second – a matter of novel information-structural effects? Evidence from Norwegian, Swedish, German and Dutch. In Jacomine Nortier & Bente A. Svendsen (eds.), *Language, youth and identity in the 21st century: Linguistic practices across urban spaces*, 73–92. Cambridge: Cambridge University Press.

Gafaranga, Joseph. 2007. *Talk in two languages*. Basingstoke: Palgrave Macmillan.

Garcia, Ofelia & Li Wei. 2014. *Translanguaging: Language, bilingualism and education*. Basingstoke: Palgrave Macmillan.

Gardner-Chloros, Penelope. 2009. *Code-switching*. Cambridge: Cambridge University Press.

Gardner-Chloros, Penelope. 2013a. On the impact of sociolinguistic change in literature: The last trilingual writers in Alsace. *Modern Languages Review* 108(4). 1086–1102.

Gardner-Chloros, Penelope. 2013b. Strasbourg revisited: c'est chic de parler français. *International Journal of the Sociology of Language* 224. 143–177.

Gardner-Chloros, Penelope & Malcom Edwards. 2010. When the blueprint is a red herring: Assumptions behind grammatical approaches to code-switching. In Li Wei (ed.), *Bilingualism and multilingualism*, vol. 1, 130–150. Abingdon: Routledge.

Gumperz, John J. 1982. *Discourse strategies*, 2nd edn. Cambridge: Cambridge University Press.

Halmari, Helena. 1997. *Government and codeswitching: Explaining American Finnish*. Amsterdam & Philadelphia: John Benjamins.

Harrington, Jonathan, Sally-Anne Palethorpe & Catherine Watson. 2000. Monophthongal vowel changes in Received Pronunciation: An acoustic analysis of the Queen's Christmas broadcasts. *Journal of the International Phonetic Association* 30. 63–78.

Heller, Monica. 2011. *Paths to post-nationalism: A critical ethnography of language and identity.* Oxford: Oxford University Press.

Hernandez-Campoy, Juan M. & J. Camilo Conde-Silvestre (eds.). 2012. *The handbook of historical sociolinguistics*. Oxford: Wiley-Blackwell.

Jansen, Leo, Hans Luijten & Nienke Bakker. 2009. Van Gogh: The letters. http://vangoghletters.org/vg/ (accessed 29 February 2016).

Jonsson, Carla. 2010. Functions of code-switching in bilingual theatre: An analysis of three Chicano plays. *Journal of Pragmatics* 42(5). 1296–1310.

Labov, William. 1963. The social motivation of a sound change. *Word* 19(3). 273–309.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Le Page, R. B. 1997. The evolution of a sociolinguistic theory of language. In Florian Coulmas (ed.), *The handbook of sociolinguistics*, 15–32. Oxford: Blackwell.

Lexander, Kristin V. 2012. Analyzing multilingual texting in Senegal: An approach for the study of mixed-language SMS. In Mark Sebba, Shahrzad Mahootian & Carla Jonsson (eds.), *Language mixing and code-switching in writing: Approaches to mixed-language written discourse*, 146–169. New York: Routledge.

Lillis, Theresa. 2013. *The sociolinguistics of writing*. Edinburgh: Edinburgh University Press.

Lipski, John M. 1982. Spanish–English language switching in speech and literature: Theories and models. *The Bilingual Review* 9(3). 191–212.

Machan, Tim William. 1994. Language contact in Piers Plowman. *Speculum* 69(2). 359–385.

Machan, Tim William. 2011. The visual pragmatics of code-switching in Late Middle English literature. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English*, 303–333. Berlin & Boston: Mouton de Gruyter.

McClure, Erica. 1998. The relationship between form and function in written national language–English codeswitching: Evidence from Mexico, Spain and Bulgaria. In Rodolfo Jacobson (ed.), *Codeswitching worldwide*, 125–152. Berlin & New York: Mouton de Gruyter.

Milroy, Lesley. 2002. Introduction: Mobility, contact and language change – Working with contemporary speech communities. In Lesley Milroy (ed.), *Mobility, contact, and language change*. [Special issue]. *Journal of Sociolinguistics* 6(1). 3–15.

Montes-Alcalá, Cecilia. 2007. Blogging in two languages: Code-switching in bilingual blogs. In Jonathan Holmquist, Augusto Lorenzino & Lotfi Sayahi (eds.), *Selected proceedings of the Third Workshop on Spanish Sociolinguistics*, 162–170. Somerville, MA: Cascadilla Proceedings Project.

Muysken, Pieter. 2000. *Bilingual speech*. Cambridge: Cambridge University Press.

Myers-Scotton, Carol. 1983. The negotiation of identities in conversation: A theory of markedness and code choice. *International Journal of the Sociology of Language* 44. 115–136.

Myers-Scotton, Carol. 1992. Comparing code-switching and borrowing. *Journal of Multilingual and Multicultural Development* 13(1/2). 19–39.

Myers-Scotton, Carol. 1998. A way to dusty death: The matrix language turnover hypothesis. In Lenore A. Grenoble & Lindsay J. Whaley (eds.), *Endangered languages: Language loss and community response*, 289–317. Cambridge: Cambridge University Press.

Myers-Scotton, Carol. 2002. *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.

Myers-Scotton, Carol & Agnes Bolonyai. 2001. Calculating speakers: Codeswitching in a rational choice model. *Language in Society* 30(1). 1–28.

Myers-Scotton, Carol & Janice Jake. 2009. A universal model of code-switching and bilingual language processing and production. In Barbara E. Bullock & Almeida Jacqueline Toribio (eds.), *The Cambridge handbook of linguistic code-switching*, 336–357. Cambridge: Cambridge University Press.

Naifeh, Steven & Gregory White Smith. 2012. *Van Gogh: The life*. London: Random House.

Nurmi, Arja & Päivi Pahta. 2012. Multilingual practices in women's English correspondence 1400–1800. In Mark Sebba, Shahrzad Mahootian & Carla Jonsson (eds.), *Language mixing and code-switching in writing: Approaches to mixed-language written discourse*, 44–67. New York: Routledge.

Osborne, Robin. 2012. Cultures as languages and languages as cultures. In Alex Mullen & Patrick James (eds.), *Multilingualism in the Graeco-Roman worlds*, 317–335. Cambridge: Cambridge University Press.

Pahta, Päivi & Arja Nurmi. 2009. Negotiating interpersonal identities in writing: Code-switching practices in Charles Burney's correspondence. In Arja Nurmi, Minna Nevala & Minna Palander-Collin (eds.), *The language of daily life in England (1400–1800)*, 27–52. Amsterdam & Philadelphia: John Benjamins.

Poplack, Shana & Nathalie Dion. 2012. Myths and facts about loanword development. *Language Variation and Change* 24(3). 279–315.

Rindler Schervje, Rosita. 1998. Code-switching as an indicator of language shift? Evidence from Sardinian–Italian bilingualism. In Rodolfo Jacobson (ed.), *Codeswitching worldwide*, 221–248. Berlin & New York: Mouton de Gruyter.

Sankoff, Gillian. 2006. Age: Apparent time and real time. In Keith Brown (ed.), *Encyclopedia of language and linguistics*, 2nd edn. Oxford: Elsevier.

Sankoff, Gillian & Suzanne Evans Wagner. 2006. Age grading in retrograde movement: The inflected future in Montreal French. *University of Pennsylvania Working Papers in Linguistics* 12(2). 203–216.

Sarkar, Mela & Lise Winer. 2006. Multilingual codeswitching in Quebec rap: Poetry, pragmatics and performativity. *International Journal of Multilingualism* 3(3). 173–192.

Schendl, Herbert. 2000. Syntactic constraints on code-switching in medieval texts. In Irma Taavitsainen, Terttu Nevalainen, Päivi Pahta & Matti Rissanen (eds.), *Placing Middle English in context*, 67–86. Berlin & New York: Mouton de Gruyter.

Schendl, Herbert & Laura Wright. 2011. Code-switching in early English: Historical background and methodological and theoretical issues. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English*, 15–45. Berlin & Boston: Mouton de Gruyter.

Schendl, Herbert & Laura Wright (eds.). 2011. *Code-switching in early English*. Berlin & Boston: Mouton de Gruyter.

Sebba, Mark. 2012. Researching and theorizing multilingual texts. In Mark Sebba, Shahrzad Mahootian & Carla Jonsson (eds.), *Language mixing and code-switching in writing: Approaches to mixed-language written discourse*, 1–26. New York: Routledge.

Sebba, Mark, Shahrzad Mahootian & Carla Jonsson (eds.). 2012. *Language mixing and code-switching in writing: Approaches to mixed-language written discourse*. New York: Routledge.

Siebenhaar, Beatrice. 2006. Code choice and code-switching in Swiss-German internet relay chat rooms. *Journal of Sociolinguistics* 10(4). 481–506.

Skaffari, Janne & Aleksi Mäkilähde. 2014. Code-switching in historical materials: Research at the limits of contact linguistics. In Robert Nicolaï (ed.), *Questioning language contact: Limits of contact, contact at its limits*, 252–280. Leiden & Boston: Brill.

Stolt, Birgit. 1964. *Die Sprachmischung in Luthers Tischreden: Studien zum Problem der Zweisprachigkeit*. Stockholm: Almqvist & Wiksell.

Swain, Simon. 2002. Bilingualism in Cicero? The evidence of code-switching. In J. N. Adams, Mark Janse & Simon Swain (eds.), *Bilingualism in ancient society: Language contact and the written text*, 128–167. Oxford: Oxford University Press.

Timm, L. A. 1978. Code-switching in *War and Peace*. In Michel Paradis (ed.), *Aspects of bilingualism*, 302–315. Columbia, SC: Hornbeam Press.

Trotter, David. 2003. *Oceano vox*: You never know where a ship comes from: On multilingualism and language-mixing in medieval Britain. In Kurt Braunmüller & Gisella Ferraresi (eds.), *Aspects of multilingualism in European language history*, 15–33. Amsterdam & Philadelphia: John Benjamins.

Trudgill, Peter. 1988. Norwich revisited: Recent linguistic changes in an English urban dialect. *English World-Wide* 9(1). 33–49.

Weston, Daniel & Penelope Gardner-Chloros. 2015. Mind the gap: What code-switching in literature can teach us about code-switching. In Penelope Gardner-Chloros & Daniel Weston (eds.), *Code-switching in literature.* [Special issue]. *Language and Literature* 24(3). 194–212.

Woolard, Kathryn A. & E. Nicholas Genovese. 2007. Strategic bivalency in Latin and Spanish in early modern Spain. *Language in Society* 36(4). 487–509.

Wright, Laura. 2002. Code-intermediate phenomena in medieval mixed-language business texts. *Language Sciences* 24(3–4). 471–489.

II Borderlands

Herbert Schendl

3 Code-switching in Anglo-Saxon England: A corpus-based approach

1 Introduction

Research into English historical code-switching has mainly focused on the Middle English period with its complex linguistic landscape while the Old English period has received little attention. This is surprising since Anglo-Saxon text production, especially in the vernacular, has been studied for a long time and from a variety of perspectives. There is hardly an Old English text, grammatical feature or lexical item that has not attracted scholarly attention. It is also no secret that Old English texts frequently show Latin insertions of various types. Furthermore, the total corpus of Old English has been available in electronic format for some years and most texts are available in reliable editions.

All this should make Old English an ideal target for research into code-switching. The surprising lack of corpus-based studies may result from the short history of historical code-switching research or from the opinion that in the diglossic situation between Latin and Old English mixing of languages hardly occurred (Schendl 2011: 47). It has also been claimed that medieval multilingual texts constitute a different phenomenon than code-switching, a view that has been refuted by recent research into medieval code-switching.

Lack of space does not allow a discussion of extralinguistic aspects of Anglo-Saxon bilingualism and the functions of Latin and Old English, from Alfred's educational programme to the monastic reform of the late tenth century, or of the most important authors and their output of translations and original texts. There has been extensive research into these areas in the last decades, including topics such as the nature of Anglo-Latin bilingualism and bilingual education (e.g. Timofeeva 2010, 2013a; Wieland 2011).

Old English-Latin bilingualism was what Adams (2003: 9) has called "élite bilingualism", restricted to a small class of educated people.[1] Old English was the naturally acquired L1, while Latin was an instructed L2 mainly, though not exclusively, used in writing. A number of bilingual Anglo-Saxon writers, most notably Ælfric, Wulfstan and Byrhtferth, not only produced monolingual texts

---

[1] On the élitist nature of Latin education in the first half of the tenth century see Wieland (2011); see also Timofeeva (2010).

DOI 10.1515/9781501504945-003

in Latin and Old English, but also bilingual texts with clear instances of code-switching.[2]

From the above it is evident that a research survey on Old English code-switching can be rather short. There are some comments on switching in the few bilingual Old English-Latin poems, which, however, lie outside the focus of this paper, which deals with non-literary texts. Among the few studies which acknowledge the existence of code-switching in Anglo-Saxon texts, Timofeeva (2010) deserves mentioning. In a brief section on code-switching, she discusses some examples from poetry and prose, and also points out that in Ælfric's *Grammar* "code-switching is constantly used for pedagogical explanatory purposes" (2010: 18).

Two more data-based studies focus on one specific text-type, namely charters. Schendl (2004) discusses code-switching in early Latin *royal* charters, where switched vernacular words and phrases in the Latin text occur as late as the mid-ninth century, mainly referring to the location and nature of transferred rights and property. After that time, however, the structure of royal charters changed to a monolingual Latin text with an inserted or attached vernacular "boundary clause".

A systematic analysis of *non-royal* charters is Schendl (2011), based on the corpus of 76 land leases issued by Oswald of Worcester. Only few of these leases are in monolingual Latin (9) or English (2), while a majority of 65 show some or even a substantial amount of code-switching between the two languages (Schendl 2011: 59–60). The predominantly intersentential switches partly serve micro-level textual functions (e.g. definition of property and boundaries, sanctions, date). More relevant is the macro-level function of switching, which is linked to the status and domains of Latin and Old English. On this level, code-switching between Latin (as language of authority and church) and the vernacular (as language of lay society) connects two different social and political spheres, namely Oswald's ecclesiastical authority and the members of secular society, whose support Oswald was seeking (Schendl 2011: 87–88).

These studies have confirmed the use of written code-switching as a bilingual textual strategy in Anglo-Saxon England. The focus on one text-type in the two studies on charters, the first diachronic, the second synchronic, has brought to light some interesting regularities and functions of switching in this text-type. The present paper will first extend research to two further text-types, and then address the problematic distinction between code-switching and borrowing.

---

[2] For brief information on these three authors see the respective entries in Lapidge et al. (2014) with further references.

2 Some new research questions on Old English code-switching

The present paper will focus on two research questions:

1. What features does code-switching show in text-types other than charters? This will be dealt with in a mainly descriptive manner on the basis of the text-types *homilies* and *scientific treatises* (section 2.1), which were selected because of their rather frequent use of switching. A close reading of selected texts (for details see 2.1) was supplemented by a systematic search in the electronic *Dictionary of Old English Web Corpus* (*OECorpus*, 2009; see below) to determine the distribution of specific Latin words and constructions in these text-types in the total corpus of Old English.
2. What is the status of single Latin word-forms syntactically integrated in vernacular co-texts in these text-types (and beyond);[3] furthermore, can a corpus-based approach help to classify such syntactically integrated forms as code-switches? This more theoretical question, which also has practical implications, will be discussed on the basis of the total occurrence of some syntactically integrated Latin word-forms in the *OECorpus* (section 2.2).

The *OECorpus* is an electronic corpus of all surviving Old English texts, amounting to more than three million words; it also includes about one million Latin words which appear in close conjunction with Old English texts, mainly glossaries, glossed texts as well as long and short Latin insertions in Old English texts. The *OECorpus* enables single-word and phrase searches (including beginnings or fragments of words), proximity searches and searches for word combinations. Furthermore, the "Word Wheel" provides lists of all Latin and Old English word-forms and their absolute frequencies in the corpus. All search results list the searched item(s) with extensive contexts and references, and, by the use of italics for Latin, provide information on the language of an item.[4]

2.1 Code-switching in homilies and scientific treatises

This section will describe the main functions and patterns of code-switching in a selection of homilies and scientific treatises, without aiming at completeness.

---

[3] "Syntactically integrated" refers to cases like *hi doð* ***absolutionem*** (Scrib 1, Ker) 'they give **absolution**', where the Latin accusative *absolutionem* is governed by the Old English verb form *doð*, but is morphologically non-integrated into the Old English construction.

[4] Two shortcomings of the corpus for our research are the lacking indication of expanded abbreviations and the not always reliable language labels.

There are about 270 Old English homilies,[5] of which those by Ælfric and Wulfstan as well as the Blickling homilies are particularly important. These will form the basis for the analysis,[6] complemented by additional searches in the total *OECorpus* based on findings from these three groups of homilies. Inserted Latin material occurs in all three collections, though its nature and frequency vary. Most frequent are Latin biblical quotations supported by a following vernacular translation or paraphrase; these are often linked to the Latin text by a phrase such as *þæt is* (*on Englisc*) 'that is (in English)', while untranslated quotations are rare.[7] Such code-switched quotations add scriptural authority to the text, but sometimes also have an organisational function, i.e. they can help structure the Old English text (for similar functions of code-switching in Middle English bilingual sermons see Schendl 2013).[8]

(1) *ealswa hit awriten is & gefyrn wæs gewitegod:* ***post mille annos soluetur Satanas****. Þæt is on Englisc, æfter þusend gearum bið Satanas unbunden.* (WHom 5)

'As it is written and was predicted of old: **after a thousand years Satan will be unbound**. That is in English, after a thousand years Satan will be unbound.'

Latin biblical quotations are often incomplete, the missing part being indicated by a metalinguistic Latin comment, such as *et reliqua* 'and so forth' for the final part (frequent in Ælfric and Wulfstan), see (2a),[9] or *usque ad* 'till, up to' for the middle part (6 instances in *Blickling Homilies*), see (2b). The Old English translation immediately following these incomplete Latin quotations, however, is given in full in the texts:

---

5 For first information on Old English homilies, a type of sermons typically interpreting the Gospel lection, see Scragg (2014); Kelly (2003: xxii–xxxi). For code-switching in later medieval and early modern sermons, see also Ingham, ter Horst and Stam, and Tuominen, this volume.

6 All Blickling and Wulfstan homilies in the editions by Kelly (2003) and Bethurum (1957) have been analysed as well as the first fifteen homilies from Ælfric's first and second series (c. 990–995); this amounts to about a quarter of all surviving homilies.

7 Latin titles, not all of which are original, and the Latin gospel text often preceding the vernacular homily will not be taken into account here, but see example (2a).

8 Quotations and text references follow the *OECorpus*, where information on titles and editions is found. In addition to the *OECorpus*, the printed text editions have also been consulted, especially for abbreviations and expansions. Switches into Latin have been marked bold; our marking of Latin elements sometimes differs from that of the *OECorpus*. Translations are by the author, though sometimes based on those provided by the text editions (see references). See Blake (2009), Clemoes (1997) and Godden (1979) for the editions of Ælfric.

9 *et reliqua* is also used to abbreviate the Old English text, see e.g. twice in WHom14.

(2) a. ***DOMINICA SECVNDA POST PASCA*****.** ***Ego sum pastor bonus*****.** ***et reliqua****. Þis godspel þe nu geræd wæs. cwyð þæt se hælend cwæde be him sylfum. Ic eom god hyrde: Se goda hyrde sylð his agen lif for his sceapum*. (ÆCHom I, 17)

‘**Second Sunday after Easter**. **I am the good shepherd**. **And so forth**. This gospel that has been read now, says that the Saviour said himself, I am the good shepherd: The good shepherd gives his own life for his sheep.’

b. *Þa ondswarede he Drihten*, [. . .] *& þus cwæþ,* ***Non est uestrum usque ad potestatem****; Nis þæt eower, he cwæþ, þæt ge witan þa þrage & þa tide þa þe Fæder gesette on his mihte*.[10] (HomS 46, BlHom 11)

‘Then answered the Lord, [. . .] and thus spoke, **It is not for you till power.** It is not for you, he said, to know the times and the seasons that the Father has put in his power.’

More rarely, this metalinguistic information is in the vernacular, see (3a) *& forð oð ende* ‘and so on till the end’ from Ælfric. In (3b), from a fragment of the Benedictine Rule, the inserted metalinguistic comment *oþ hit cume to þisse endunge* ‘till it comes to this ending’ constitutes a code-switch back to the vernacular between the two Latin phrases.

(3) a. *þe we singað [. . .] æt ælcum æfensange.* ***Magnificat anima mea dominum****: & forð oð ende; þæt is min sawul mærsað drihten*: (ÆCHom I, 13)

‘which we sing [. . .] at each evening song: **My soul praises the Lord**: and so forth till the end; that is my soul praises the Lord:’

b. *Geendedum gebede ræde se diacon þis godspel,* ***Turba multa*** *oþ hit cume to þisse endunge****, Ecce, mundus totus post eum abiit*****.** (RegC 1, Zup)

‘This prayer being ended, the deacon should read this gospel, **Much people** till it comes to this ending, **Look, the whole world has gone after him**.’

---

10 The full Latin sentence reads: *Non est vestrum nosse tempora vel momenta, quæ Pater posuit in sua potestate*. Interestingly, *potestatem*, the last word of the incomplete quotation, shows the accusative governed by the metalinguistic preposition *usque ad*, while the underlying full biblical phrase has the ablative *potestate* governed by *in*.

A good example of the organisational function of reduced Latin quotations occurs in the final part of the *Blickling* homily on the "Assumption of Mary", which is structured by the verses of the *Magnificat*; here the beginning of each Latin verse introduces an Old English paraphrase and expansion of the biblical verse, see (4). The first incomplete Latin verse (*Fecit potentiam*) can be read as an intersentential switch which is syntactically and semantically linked to the vernacular paraphrase, while the second one (*Esurientes*) is an isolated single-word switch.

(4) ***Fecit potentiam***, *& he dyde mycle mihte on his <earman>, & he todælde ealle þa þe þær wæron ofermode on heora heortan*, [...]

Esurientes,[11] *& þa wæs Sancta Maria cweþende þæt Drihten ealle þa gefylde on heofona wuldres fægernesse þa þe hie on eorþan leton hingrian & þyrstan for his naman*. (LS 20, AssumptMar)

'**He has shown strength**, and he has done mighty things with his arms, and he has scattered all those that were proud in their heart, [...]

The hungry, and then Saint Mary said that the Lord filled all those with the beauty of heaven's glory that on earth had suffered hunger and thirst for his name.'

The majority of switched quotations are intersentential switches of one or more sentences. Less frequent are syntactically integrated switched phrases such as titles of books (*actus apostolorum*), of prayers and songs (*pater noster; credo in deum*), or religious formulae (*patris et filii et spiritus sancti*; *in modum crucis*). There is also a number of single-word switches, mainly technical religious terms followed by a vernacular translation, such as the seven gifts and the sins of men (*sapientia, consilium*, etc.; *ignorantia, impietas*, etc.), the (arch)angels (*throni*, *principatus*, etc.), and a few religious terms such as *absolutionem* and *discipuli*, whose status will be discussed in section 2.2.

To sum up, code-switching is well-attested in the homilies, but generally restricted to religious, mostly biblical quotations; such "prefabricated" material is well integrated into the vernacular text, but does not presuppose competence in Latin from audience or readership, since vernacular translations generally support the Latin text. Switched phrases and single words are less frequent.

For the second text-type, scientific treatises, two texts with numerous instances of code-switching were chosen, both dealing with topics linked to astronomy, the division of time and the *computus*, i.e. “the medieval science of computation […] concerned mainly with the calendar” (Baker 2014: 121): first, Byrhtferth’s *Enchiridion* (c. 1010), a learned compilation and commentary on these topics with numerous digressions, see the entry on Byrhtferth in Lapidge et al. (2014: 81–82); second, Ælfric’s *De Temporibus Anni* (c. 990 to 995), one of Byrhtferth’s sources, again complemented by searches in the *OECorpus*.

---

[11] The complete Latin verse reads: *Esurientes implevit bonis et divites dimisit inanes* ('The hungry he has filled with good things, and the rich he has sent away empty').

The *Enchiridion* shows a high incidence of Latin material, from switched single words and phrases to full sentences and longer inserted Latin passages. This is well illustrated in the beginning of section II.3 of the *Enchiridion* (Baker and Lapidge 1995: 105–106), where about 25 switches into Latin occur within a page and a half, including 18 single-word switches like *communis*, *puncti*, and a number of switched Latin phrases and short sentences, see (5) and (6). All are supported by a vernacular translation or explanation. In a number of cases, the translation or the wider co-text shows that a Latin term or phrase is not used for lack of an English equivalent, but rather for the prestige of Latin, pedagogical purposes, or to refer to the authorities on which the *Enchiridion* was based.

(5) ***Lunaris annus*** *byð ælce geare* (*þæt ys þæs monan ger*) […] *and* ***communis*** (*þæt ys gemæne ger*) *hyt byð oðre hwile on þam oðrum geare* […] *On twam wisum ys se dæg gecweden:* ***naturaliter et uulgariter***, *þæt ys gecyndelice and ceorlice.* […] ***Vulgaris uel artificalis dies est*** (*þæt byð ceorlisc dæg oððe cræftlic*) *fram þære sunnan anginne þæt heo to setle ga* […]. *Se dæg* […] *hæfð syx and hundnigontig* ***punctos***. *Feower* ***puncti*** (*þæt synt prican*) *wyrcað ane tid on þære sunnan ryne; and forþan ys se prica gecweden forþan seo sunne astihð pricmælum on þam dægmæle.*
(ByrM 1, Baker/Lapidge)

‘**The moon’s year** is every year (that is the moon’s year) […] and **the common** (that is the common year) sometimes comes in the following year […] In two ways is the day spoken of: **in a natural and a vulgar way**, that is natural and vulgar. […] **The vulgar or artificial day is** (that is the vulgar or artificial day) from the sun’s rising until it sets […]. The day […] has ninety-six **points**. Four **points** (that is points) make one hour in the sun’s course; and the point is so called because the sun advances point by point on the sundial.’ (See Baker and Lapidge 1995: 105)

This passage directly continues in (6) with an explicit justification for using Latin for explaining the etymology of Latin *punctus* (*Me ys neod* …), a metalinguistic comment that proves that the author is well aware of mixing two distinct languages.

(6) *Me ys neod þæt ic menge þæt Lyden amang þissum Englisce.* ***Punctus a pungendo dicitur.*** *Forþan ys se prica gecweden forþan he pingð oððe pricað.* [. . .] *Syx and hundnigontig prican beoð on þam dæge, and þa prican habbað* ***minuta*** *twa hund and feowertig.* (ByrM 1, Baker/Lapidge)

‘It is necessary for me to mix that Latin in with this English. **The point is named from pricking**. The point is so called because it pricks or stings. [. . .] There are ninety-six points in the day, and these points have 240 **minutes**.’ (See Baker and Lapidge 1995: 105)

In the text passage following (6), a number of further switched Latin word-forms occur, such as the singular *minutum* taking up the switched plural *minuta* at the end of (6) and the forms *momentum*, *partes*, *zodiacus*, etc.; some of these have been listed as Latin borrowings in the secondary literature, but are good candidates for single-word switches (see section 2.2).

However, Byrhtferth also inserts Latin quotations from scholars, religious authorities and classical poets without providing a direct translation, evidently expecting some Latin competence from most of his readers:

(7) *se mona ys* ***luna uel lucina*** *on Lyden geciged, swa se þeodwita cwæð Virgilius:* ***Casta faue Lucina****. Þa steorran syn gecweden þurh heora stede*, (ByrM 1, Baker/Lapidge)

‘the moon is called **luna or lucina** in Latin, as the learned man Virgil said: **Chaste Lucina, favour** [this birth]. The stars are so called for their stability,’

The second scientific text, Ælfric’s *De Temporibus Anni*, has fewer and mostly shorter Latin switches, mainly biblical quotations with English translations, various short Latin phrases, composite technical terms and phrases referring to dates of the month:

(8) a. *swa swa se witega cwæð:* ***Celi celorum****, þæt is heofona heofonan.* (ÆTemp)
‘as the prophet said, **heavens of heavens**, that is heavens of heavens’

b. *se dæg is gehaten* ***Saltus lune****, þæt is ðæs monan hlyp, forðan ðe he oferhlypð ænne dæg* (ÆTemp)

‘that day is called **moon’s leap**, that is the moon’s leap, because it leaps over one day’

c. *seo lenctenlice emniht is gewislice on* ***duodecima kalendas aprilis****, þæt is on Sancte Benedictes mæssedæg.* (ÆTemp)

‘the spring equinox is correctly on **the twelfth calens of April** [i.e., 21 March], that is on Saint Benedict’s day’

Furthermore, quite a number of single Latin word-forms are embedded in vernacular sentences; some are common terms with established Old English equivalents, such as the four elements (*aer*, *ignis*, *terra*, *aqua*) and the seasons of the year (*ver*, *estas*, *autumnus*, *hiemps*), others are more technical such as the signs of the zodiac (*Aries*, *Gemini*, etc.), and terms like *firmamentum*, *bissextus*, mostly supported by English explanations.

This leads to the second research question mentioned above, namely the status of single Latin word-forms syntactically integrated in Old English constructions; in particular, whether these can be considered as single-word code-switches. This cannot be satisfactorily decided for a single text, but has to be addressed on the basis of all occurrences of a given lexeme in the *OECorpus*. For the following discussion, five lexemes, three from the field of religion and two from science, have been selected, which seem to be likely candidates for code-switches. Such an analysis evidently complements the efforts of lexicographers, who have to decide which morphologically non-integrated Latin words to include in a dictionary (Milfull 2009; Durkin 2014: 129).

2.2 Code-switching or borrowing?

The distinction between code-switching and borrowing has long been a controversial issue in research on language contact and code-switching (Romaine 1995: section 4.6; Gardner-Chloros 2009: 30–33; see also Queiroz de Barros, this volume). These are either regarded as two distinct phenomena or as being "related points on a continuum", the connection between the two being "essentially a diachronic one [which] involves an increase in the usage frequency of new word-forms" (Matras 2009: 110, with references to earlier studies). This view is also taken in the present study. Useful labels for different stages on this continuum are provided by Haspelmath (2009: 38), who restricts the term lexical borrowing to "a completed language change", which starts, like other linguistic changes, as a lexical innovation in the speech of an individual. Such a lexical innovation may – even frequently – originate in utterances that involve code-switching (Thomason 2003: 695; see also Matras 2009: 110) and may spread till it is adopted by a speech community, or it may remain marginal and finally disappear. During the increasing usage frequency of lexical innovations, variation between competing lexical items is to be expected and is well attested in earlier written English. The resulting co-occurrence of different more or less equivalent lexical variants may, as Durkin (2014: 124) complains, complicate matters for the historical lexicographer, but it may also enable us to reconstruct the process of lexical integration, even for remote stages of the language. For Old English, the seemingly chaotic lexical variation in a corpus like the *OECorpus*, which spans several centuries, may in some cases not only mirror regional or stylistic factors, but also the dynamic process of lexical propagation, with fuzzy overlaps of variants.

However, handbook accounts on Old English vocabulary tend to disregard the possibility of single-word code-switches as a mechanism of lexical change and the dynamic process from lexical innovation to lexical borrowing and rather classify all single Latin nouns embedded in late Old English texts, such as *confessio*, *passio(nem)*, or *Zodiacus*, as loanwords or borrowings. Thus Campbell (1959: 219) states that "[v]ery often [. . .] late loanwords retain their Lat[in] declensional endings in whole or in part". Even Kastovsky's (1992: 299) comprehensive survey on Old English vocabulary classifies such single Latin words as loans, irrespective of their "degree of integration". Specialist accounts of Old English loanwords often present more detailed classifications: in his study on learned Latin loanwords in late Old English, Funke (1914) distinguishes between morphologically (and phonologically) integrated loanwords ("Lehnwörter"), non-integrated foreign words or "foreignisms" ("Fremdwörter") and an intermediate group. Though there is wide agreement that non-integrated "foreignisms" form a specific subgroup of late Old English learned words, they are generally subsumed under the term "loanword".[12] Feulner (2000: 36–37) criticises earlier lexical studies for not distinguishing in such cases between "Old English" and "non-Old-English", a distinction which she considers relevant for the inclusion of a word in a dictionary. In this context, she deplores the lack of explicit and consistent criteria for inclusion of Latin words into the new *Dictionary of Old English* (*DOE*), of which letters A to G have been published so far. In her discussion of possible criteria for a finer differentiation, Feulner (2000: 37) refers to interference and integration and once briefly mentions the concept of "switching"; furthermore, she introduces the term "Zitatwort" ('citation word') for inserted "non-Old-English" words from Stolt's (1964) seminal study on language mixing in Martin Luther. However, Feulner does nowhere refer to modern code-switching research.

In a more recent study on late Old English loanwords, Timofeeva (2013b: 167) points out that "it is sometimes very difficult to set apart genuine loanwords from instances of code-switching". However, her detailed analysis of six loanwords does not take up this issue any further.

---

[12] For a detailed survey on research on Old English loanwords from the late nineteenth century to the present see Feulner (2000: 35–44), where problems of classification and methodology are also addressed.

The first study of loanwords referring to code-switching as a possible explanation of single Latin word-forms in Old English texts is Durkin (2014). He discusses five learned Latin words and their partly inconsistent treatment in the *DOE*, including the criteria for their inclusion as headwords and their labelling as Old English forms in the dictionary (Durkin 2014: 124–129). Though Durkin does not explicitly answer the question of "code-switching" or "borrowing" for his examples, he (2014: 129) concludes that "[a] reappraisal of this topic is overdue, taking into account modern perspectives on code switching and mixed-language texts".

Durkin's focus on loanwords makes him emphasise the importance of the *DOE* for such an enterprise. The present section, on the other hand, presents a text-type centred analysis of Old English code-switching with a particular focus on single-word switches. On the basis of the *OECorpus*, all occurrences of selected single Latin word-forms in vernacular co-texts have been collected and analysed for specific formal and distributional criteria and other factors. Most of the criteria used here for classifying a specific Latin word-form as a code-switch are based on Matras (2009: 110–114), for whom the distinction between code-switching and borrowing "involves a bundle of [seven] criteria, each arranged on a continuum" (2009: 113). Thus, for example, his criterion (or "dimension") of bilinguality encompasses the continuum from a bilingual speaker (with different degrees of bilinguality) to monolinguality. Between a prototypical code-switch and a prototypical borrowing, "we encounter fuzzy ground [… which] is largely limited to bilinguals" (Matras 2009: 114).

The following five criteria, which can be applied to a historical corpus like the *OECorpus*, will form the basis for the classification of a Latin word-form as a single-word code-switch.

i. the bilingualism of the author (cf. Matras' "bilinguality continuum"). For many – though not all – Old English texts this can either be inferred from extralinguistic information on an author, the production of texts in different languages by the same author, or from the occurrence of other instances of code-switching in a single text, especially above the level of the single word;
ii. the Latin word-form has to be morphologically non-integrated, i.e. retain its Latin inflectional suffix (cf. Matras' "structural integration" dimension), but it has to be syntactically integrated, i.e. it must conform to the syntactic rules of the Old English construction in which it occurs;[13]

---

[13] Cases where Old English and Latin have identical inflectional suffixes (Feulner 2000: 39–40; Durkin 2014: 124) will not be counted as code-switches.

iii. the word-forms of a specific lexical item should have an overall low frequency of occurrence and restricted textual distribution in the corpus (cf. Matras' "regularity" dimension);
iv. code-switches are more likely to fulfil special textual and stylistic effects than established loanwords (cf. Matras' "functional continuum");
v. a criterion sometimes applied for classifying a Latin word as "not Old English" or a "foreignism" is its occurrence with an English explanation or synonym (Funke 1914: 158–160; Feulner 2000: 40–41). This may, but need not support its status as a code-switch, since definitions of an unknown, frequently technical term also occur in monolingual texts. In our corpus, the first occurrence of a single-word switch which meets the other four criteria is often, but not always supported by a translation or definition; however, the Old English definition of a Latin term may also be a literal translation of a monolingual Latin sentence containing that term, see (9) from Ælfric's *De Temporibus Anni* and his source Bede.[14]

(9)	***Crepusculum***, *þæt is æfengloma* (ÆTemp, B1.9.4) '**Crepusculum**, that is twilight'	*Crepusculum, id est, dubia lux inter lucem et tenebras*, (Bede, Temp)

The remaining part of this paper will discuss the status of five morphologically non-integrated Latin words in Old English co-texts, all of them learned technical terms from the fields of religion and science. These words were selected from two lists compiled for this study, one of single Latin word-forms found in the homilies and scientific treatises analysed for section 2.1; and another list of learned Latin words in late Old English collected from existing vocabulary studies, ranging from Funke (1914) to Durkin (2014). Since this is a study on code-switching in Old English, we were particularly interested in finding clear instances of code-switching on the basis of the five criteria discussed above, but not in unambiguous cases of loanwords with morphological or phonological integration into Old English.

Not all single Latin word-forms in an Old English co-text are equally ambiguous in regard to their status. Thus, the form *esurientes* in (4) refers to a full Latin sentence and is not syntactically integrated into the Old English text, while the syntactically integrated *communis* in (5) refers to the syntactic group *communis annus* (with deletion of *annus*), as the explanation "that is the common year" shows, which is supported by further *OECorpus* data. Both these examples clearly function as code-switches. Similarly, the quoting of Latin terms in both Byrhtferth and Ælfric for which there are well established native words, such as the four elements and the four seasons, can safely be regarded as code-switching in view of their limited distribution in the *OECorpus* and their regular translation. However, in the following five cases, as in many others, a more detailed analysis is necessary.

---

[14] The strict application of the "definition-criterion" for the attribution of language has led to a number of inconsistencies in marking forms as Latin in the *OECorpus* and the *DOE*, see Feulner (2000: 40) and Durkin (2014: 125); in combination with the other criteria it may, however, provide additional support.

confessio: Of the 149 word-forms of this lexeme in the corpus, only one occurs in an Old English co-text, a hybrid verb phrase where the Latin accusative singular *confessionem* is governed by the Old English verb *don*:

(10) *Æfter þam* ***Pater noster*** […] *and æfter þam* ***Credo in deum***, […] *þær hy heora andetnysse don sculon*, […] *and swa* […] *heora* ***confessionem*** *don, þæt is heora andetnesse*. (RegC 1, Zup)

'After the **Pater noster** […] and after the **Credo in deum** […] they shall there do their confession, […] and so […] do their **confession**, that is their confession.'

A number of factors speak for classifying *confessionem* here as a code-switch, though the *OECorpus* does not mark it as Latin: the occurrence of other code-switches in the same text, including another parallel hybrid verb phrase and two Latin names of prayers in the same passage, which is a relatively frequent type of code-switching in the *OECorpus* (criterion i); its full syntactical integration (criterion ii), and its single occurrence in an Old English construction in the whole corpus (criterion iii); furthermore, briefly before and immediately after the switched term, *andetnysse*, the Old English equivalent of *confessio*, occurs: the phrase preceding the switch, *heora andetnysse don*, is semantically and syntactically fully equivalent to the hybrid phrase, while the clause following the switch, *þæt is heora andetnesse*, provides an English explanation of the Latin form (criterion v); Old English *andetnysse* (with numerous spelling variants) is frequent in the *OECorpus*.

As for evidence from historical dictionaries, *confessio* is not given as a headword in the *DOE*, while both the *Middle English Dictionary* (*MED*) and the *Oxford English Dictionary* (*OED*) give the first occurrences for *confession* from the late fourteenth century.

Similar hybrid verb+noun constructions are ***lectionem*** *gefyllan* 'do a reading' (1x, Conf 1.1), ***mandatum*** *gefremman* 'carry out the ritual washing of the feet' (1x, RegC1), ***orationem*** *rædan/cweðan* 'read/say a prayer' (5x, only in Bede), and ***absolutionem*** *don/rædan*, which will be discussed next.

absolutio: This word, also discussed in Durkin (2014: 127), occurs five times in a hybrid verb phrase, with an Old English verb (*don*, *macian*, *rædan*) governing the Latin accusative form *absolutionem*. All instances are marked as Latin in the *OECorpus*. Two of the instances are from Wulfstan, another two from a text closely linked to Wulfstan, dating from around the year 1000.

(11) *& ðonne* ***absolutionem*** *bisceopas ofer hy rædað* (WHom 14)[15]
‘and then the bishops read the **absolution** over them’

The fifth instance of *absolutionem* occurs in a scribbled note in the margin of a late tenth-century manuscript, which looks like a spontaneous commentary by someone reading a passage by Ælfric. Another single-word switch, the noun *penitentes*, occurs shortly before *absolutionem*:

(12) *þonne hi lædað in* ***penitentes****, & hi doð* ***absolutionem*** (Scrib 1, Ker)[16]
‘when they lead in the **penitents**, and they give **absolution**’

I would classify all five instances as code-switches, since they fully or partly meet four of our five criteria (i, ii, iii, v); furthermore, the semantic concept of *absolutio* is covered by well-established native words such as *forgief(ed)nes*, *forlæt(en)nes*, *alætnes*.[17] This classification also corresponds to *DOE’s* decision not to include it as a headword (see also Durkin 2014: 127, who, however, does not explicitly classify *absolutionem* as a code-switch).

As for its treatment in the *OED*, the second edition (1989) has the first attestation of *absolution* from c1200, while the revised entry for *OED3* (September 2009) gives under the heading “Forms” the information “α. OE absolutionem” with two instances of *absolutionem* in Old English co-texts. These are commented upon as being “after the Latin accusative singular”, a wording which seems to imply that they are regarded as Old English words.

processio: The eight Latin word-forms of *processio* in Old English co-texts come from only three texts dating from the early eleventh to the early twelfth century: Ælfric’s third letter to Wulfstan (3x), a fragment of the *Regularis Concordia* (RegC1; 3x) and the *Peterborough Chronicle* (ChronE; 2x).

---

[15] It may be interesting that WHom14 seems to be an (unfinished) outline for a sermon, see note 9; see also the comment on the next example.

[16] *The OECorpus* marks the whole phrase *in penitentes* as Latin. Our translation follows Szarmach (1986: 397, note 26), who evidently interprets *in* as an Old English adverb and *penitentes* as a switched Latin noun.

[17] *absolutio* is not found in the Latin part of the *OECorpus*, but is attested in Anglo-British Latin, see *DMLBS*.

All instances from RegC1 (Zupitza), a text with a number of other code-switches, show the Latin nominative preceded by the pronoun *se(o)*; at its second occurrence the word is explained in Old English:

(13) *seo mare* ***processio****, þæt is þære maran halgunge forðgang* (RegC 1, Zup)
'The greater **procession** that is the advance of the greater consecration'

The instances from Ælfric and Chronicle E (entries 1125, 1131) occur in hybrid prepositional phrases, with the Latin accusative singular *processionem* governed by an Old English preposition, *to* in the instances from Ælfric, *mid* in those from the Chronicle, see (14) and (15).[18]

(14) *Ge sculon on palmsunnandæge palmtwigu bletsian and beran mid lofsange to* ***processionem*** (ÆLet 3, Wulfstan 2)

'On palmsunday, you shall bless palm-twigs and bear (them) with hymns in **procession**'

(15) *and brohten him into cyrce mid* ***processionem****, sungen* ***Te Deum Laudamus****, ringden þa belle*, (ChronE, Irvine, 1125)

'and brought him to church with a **procession**, singing **Te Deum Laudamus**, ringing the bell,'

Though none of the eight forms of *processio* is marked as Latin in the *OECorpus*, their classification as code-switches is supported by their occurrence in authors and texts where switching is also used elsewhere and by the restricted textual distribution compared to the well-attested native synonyms *emb(e)/ymb(e)gang*, *forðgang*. The multiple use of *embegang* to translate Latin *processio* in another text of the *Regularis Concordia* (RegCGl, Kornexl, 4 instances) and the glossing of *illa maior processio* by *seo mare embegang* in RegCGl (Kornexl), as against the hybrid phrase *seo mare* ***processio*** used in RegC1 (Zupitza) with an immediately following explanation in Old English, see (13) above, provide further support for this classification.[19]

---

[18] Old English *to* normally governs the dative, but some clear cases with the accusative occur (Mitchell 1985: §1215); *mid* + accusative is Anglian (Mitchell 1985: §1195), which fits the location of the *Peterborough Chronicle*. Chronicle E also has the first attestations of French *procession, -un* in the annal from 1103 dating from the 1120s (cf. *MED*).

[19] *OED3* (updated June 2007) states "α. OE procession, OE processionem" and comments that these forms are "after the Latin nominative singular [respectively], after the Latin accusative singular", thus, as with *absolution*, implying that these are actually regarded as Old English forms.

Let us finally look at two technical terms from science, more specifically from computus texts, where single-word switches are widespread.

***punctus*:** Of the 31 instances of Latin word-forms of *punctus* in the *OECorpus*, only five are syntactically integrated in an Old English co-text, namely *punctus, -i* (nom.sg./pl.), *punctum, -os* (acc.sg./pl.), all in texts by Byrhtferth (4x) and Ælfric (1x; in his *Grammar*) and marked as Latin in the *OECorpus*. The above criteria clearly support a classification of all five instances as code-switches: the well-documented bilingualism of the two authors, the syntactic integration of the Latin forms and their textual restriction and low frequency (criteria i–iii). The forms also have special textual function, namely the conscious introduction of the Latin technical term for pedagogical reasons (criterion iv); furthermore, the term is twice explained by reference to the native equivalent *prica* (criterion v), as in (16):

(16) *Se dæg* [...] *hæfð syx and hundnigontig* ***punctos****. Feower* ***puncti*** (*þæt synt prican*) *wyrcað ane tid on þære sunnan ryne*; (ByrM 1, Baker/Lapidge) [translation see (5)]

However, only a few pages after the Latin accusative plural in *-os*, Byrhtferth once introduces the morphologically integrated Old English accusative plural form *punctas*. Since *punctas* also occurs once in another, related computus text (Comp 13.2, Henel), it cannot be interpreted as a mistake.[20] This single instance of *punctas* (with Old English inflectional suffix) is evidently a lexical innovation by Byrhtferth, which, in spite of the second attestation of the form, did not spread much, possibly due to the existence of the well-established native lexeme *prica* for the concept: thus, only *prica* is used in ÆTemp (5x) and in another computus text (Comp 9, Cockayne; 12x).

Especially interesting is the lexical variation for the concept in the *Enchiridion*, where Byrhtferth predominantly uses forms of the native word *prica* (26x), but only four times a code-switched form of Latin *punctus*, and once the lexical innovation *punctas*. This lexical variation between a well-established native lexeme, single-word code-switches, and a lexical innovation in a single bilingual text provides impressive insight into the dynamics of the lexical choices and the exploitation of the lexical resources of a bilingual author in the production of a bilingual text in the late Old English period.

---

[20] On the basis of this single occurrence of *punctas*, the editors of the *Enchiridion* list a head-word *punct* m. in their Old English glossary, which is, however, not attested in the *OECorpus* and thus a “ghost-word”.

zodiacus: Kastovsky (1992: 299) lists *circul zodiacus* as a morphologically non-integrated Latin loan referring to Funke (1914: 17), who classifies it as a foreignism. There are two points to be made here: first, *circul* (Latin *circulus*, see *DMLBS*) is indeed an integrated loan in Old English and is also included in the *DOE* (see Feulner 2000: 140–141). It occurs more than fifty times in the *OECorpus* with different Old English inflectional suffixes, but only once in the combination *circul zodiacus*. Thus there is no reason why the syntactic group *circul zodiacus* should be regarded as a loanword. Second, ten instances of Latin word-forms of *zodiacus* are syntactically integrated into an Old English co-text (8x nominative *-us*, 2x accusative *-um*).[21] This together with the following factors speak for classifying *zodiacus, -um* as a code-switched Latin technical term: its limited occurrence in two closely related texts by the bilingual authors Ælfric and Byrhtferth (3x ÆTemp, 7x ByrM 1, Baker/Lapidge); its frequent occurrence with an explanatory phrase (7x); in the *Enchiridion* it is explicitly referred to as a term used by scholars (*þe uðwitan*), see (17), where *zodiacus* is also co-ordinated with three synonymous switched technical terms:

(17) *An circul ys þe uðwitan hatað* **zodiacus** *oððe* **horoscopus** *oððe* **mazaroth** *oððe* **sideralis***, þurh þæne yrnð seo sunne and se mona and þas steorran* ***Saturnus*** *and* ***Iouis***, (ByrM 1, Baker/Lapidge)

‘There is a circle that scholars call **zodiac** or **horoscope** or **mazzaroth** or **sideralis**, through which run the sun and the moon, and the stars **Saturn** and **Jupiter**,’

The use of the Latin word-form in an Old English co-text is sometimes triggered by its preceding occurrence in a Latin co-text, see (18):

(18) ***Partes a partitione circuli zodiaci uocantur.*** [...] *þe beoð on þam circule þe ys* ***zodiacus*** *geciged* [...] ***Zodiacus*** *ys se circul genemned* [...] (ByrM 1, Baker/Lapidge)

‘**Parts are named from the division of the zodiac circle** [...] which are in the circle called **zodiac** [...] **zodiac** is called the circle [...]’

The only criterion which does not meet the requirements for classifying the syntactically integrated forms of *zodiacus* as code-switches is the absence of a default native equivalent or established loanword (criterion iv). However, in a model that works with a bundle of criteria, each of which is seen as representing a continuum (see the above discussion of Matras 2009: 113–114), the fact that four out of five criteria are met places *zodiacus* rather close to the code-switching side of the continuum and does not speak for classifying it as a loanword.

---

[21] All but one of these ten occurrences are marked as Latin in the *OECorpus*.

There is quite a number of similar Latin word-forms embedded in Old English co-texts which seem good candidates for being single-word switches, such as *acolitus* (see Durkin 2014: 124–125), *firmamentum*, *lectio(nem)*, *momentum*, *oratio*, *paradisus*, *passio(nem)*, *terminus*, etc. With some of these, there is also variation between Latin word-forms and morphologically integrated Old English forms as in the case of *punctus*. However, not all Latin word-forms embedded in vernacular texts are necessarily code-switches, as the variation in modern English plural forms like *indexes* vs. *indices* shows (Durkin 2014: 128). The frequency or "regularity" criterion (Matras 2009: 113) is certainly a major criterion in deciding this issue, but only in conjunction with the other criteria discussed above. Extensive further analyses along the lines outlined in this section are needed.

3 Conclusion

Code-switching between Latin and Old English is a widely used bilingual strategy in texts from the Old English period. Previous research on Anglo-Saxon charters as well as the analysis of homilies and scientific treatises presented in this paper have confirmed that a text-type based approach is a promising way to discover similarities and differences of functions and patterns of code-switching in a variety of Old English texts and to establish the nature of code-switching in this period. Such analyses are best carried out within a corpus-based approach, making use of both well-defined sub-corpora of charters, homilies and scientific treatises, as well as the *Dictionary of Old English Web Corpus*, the full electronic corpus of Old English texts.

The descriptive analyses in the first part of this study focused on the main types of code-switching in the two selected text-types, which show both similarities and differences, frequently of a text-specific nature. While Latin sentences, clauses and phrases in vernacular contexts can safely be classified as code-switches, the numerous single Latin word-forms which are syntactically integrated into Old English co-texts are more problematic. Such inserted Latin forms have so far mainly been discussed from a lexicological or lexicographic point of view and have been subsumed under the cover term loanwords. However, the frequent occurrence of such single Latin word-forms in bilingual texts calls for a different approach, namely one that analyses them from the point of view of code-switching. This is exemplified in the second part of the paper, which first discusses criteria for classifying such forms as possible code-switches. On the basis of Matras (2009), five criteria have been chosen as the basis of a detailed analysis of all occurrences of five selected Latin lexemes in Old English co-texts in the total *OECorpus*, with a view of operationalising their classification as code-switches. By this method, the word-forms of four of the five Latin lexemes could be classified as code-switches, while the last one does not fulfil one of the criteria. Since the present paper also follows Matras in considering code-switching and borrowing as forming a continuum and the chosen criteria as equally forming continua, the approach illustrated here can be fruitfully applied to further, less clear cases of switching.

The detailed analysis of the occurrences of the code-switched forms of *punctus* in Byrhtferth's *Enchiridion* has furthermore uncovered an interesting lexical variation between these code-switched forms and the well-established native equivalent *prica*, which clearly dominates in this text. This makes the introduction of a lexical innovation *punctas*, which shows morphological integration into Old English, the more surprising. This case of lexical variation reflects a dynamic process, in which the author clearly and consciously exploits his bilingual lexical resources.

References

Adams, J. N. 2003. *Bilingualism and the Latin language*. Cambridge: Cambridge University Press.

Baker, Peter S. 2014. Computus. In Michael Lapidge, John Blair, Simon Keynes & Donald Scragg (eds.), *The Wiley-Blackwell encyclopedia of Anglo-Saxon England*, 2nd edn., 121–122. Malden & Oxford: Wiley Blackwell.

Baker, Peter S. & Michael Lapidge (eds.). 1995. *Byrhtferth's Enchiridion* (Early English Text Society Supplementary Series 15). London: Oxford University Press.

Bethurum, Dorothy. 1957. *The homilies of Wulfstan*. Oxford: Clarendon Press.

Blake, Martin (ed.). 2009. *Ælfric's De Temporibus Anni*. (Anglo-Saxon Texts 6). Cambridge: D. S. Brewer.

Campbell, Alistair. 1959. *Old English grammar*. Oxford: Clarendon Press.

Clemoes, Peter (ed.). 1997. *Ælfric's Catholic homilies: The first series* (Early English Text Society Supplementary Series 17). Oxford: Oxford University Press.

DMLBS = Latham, Ronald E., David R. Howlett et al. (eds.). 1975–2013. *Dictionary of medieval Latin from British sources*. Oxford: The British Academy & Oxford University Press.

DOE = Cameron, Angus, Ashley Crandell Amos, Antonette diPaolo Healey et al. (eds.). 2007. *Dictionary of Old English: A to G* online. Toronto: Dictionary of Old English Project. http://tapor.library.utoronto.ca/doe/.

Durkin, Philip. 2014. *Borrowed words: A history of loanwords in English*. Oxford: Oxford University Press.

Feulner, Anna H. 2000. *Die griechischen Lehnwörter im Altenglischen* (Münchner Universitäts-Schriften: Texte und Untersuchungen zur englischen Philologie 21). Frankfurt: Peter Lang.

Funke, Otto. 1914. *Die gelehrten lateinischen Lehn- und Fremdwörter in der altenglischen Literatur.* Halle: Niedermeyer.

Gardner-Chloros, Penelope. 2009. *Code-switching*. Cambridge: Cambridge University Press.

Godden, Malcolm (ed.). 1979. *Ælfric's Catholic homilies: The second series* (Early English Text Society Supplementary Series 5). Oxford: Oxford University Press.

Haspelmath, Martin. 2009. Lexical borrowing: Concepts and issues. In Martin Haspelmath & Uri Tadmor (eds.), *Loanwords in the world's languages: A comparative handbook*, 35–54. Berlin & New York: De Gruyter Mouton.

Kastovsky, Dieter. 1992. Semantics and vocabulary. In Richard M. Hogg (ed.), *The Cambridge history of the English language*, vol. I: *The beginnings to 1066*, 290–408. Cambridge: Cambridge University Press.

Kelly, Richard J. 2003. *The Blickling homilies: Edition and translation*. London: Continuum.

Lapidge, Michael, John Blair, Simon Keynes & Donald Scragg (eds.). 2014. *The Wiley-Blackwell encyclopedia of Anglo-Saxon England*, 2nd edn. Malden & Oxford: Wiley Blackwell.

Matras, Yaron. 2009. *Language contact*. Cambridge: Cambridge University Press.

MED = Middle English Dictionary Online. https://quod.lib.umich.edu/m/med/.

Milfull, Inge B. 2009. PULEGE and PSYLLIUM: Old English plant names in p- in the *Oxford English Dictionary*. In Peter Bierbaumer & Helmut Klug (eds.), *Old names – new growth*, 121–143. Frankfurt: Peter Lang.

Mitchell, Bruce. 1985. *Old English syntax*, 2 vols. Oxford: Oxford University Press.

OECorpus = Dictionary of Old English Web Corpus. 2009. Compiled by Antonette diPaolo Healey with John Price Wilkin & Xin Xiang. Toronto: Dictionary of Old English Project. http://tapor.library.utoronto.ca/doecorpus/.

OED = Oxford English Dictionary. http://www.oed.com/.

Romaine, Suzanne. 1995. *Bilingualism* (Language in Society 13), 2nd edn. Oxford: Blackwell.

Schendl, Herbert. 2004. 'Hec sunt prata *to wassingwellan*': Aspects of code-switching in Old English charters. *Vienna English Working Papers* 13(2). 52–68. http://archiv-anglistik.univie.ac.at/fileadmin/user_upload/dep_anglist/weitere_Uploads/Views/0402ALL.pdf (accessed 24 June 2017).

Schendl, Herbert. 2011. Beyond boundaries: Code-switching in the leases of Oswald of Worcester. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English* (Topics in English Linguistics 76), 47–94. Berlin & Boston: De Gruyter Mouton.

Schendl, Herbert. 2013. Code-switching in late medieval macaronic sermons. In Judith A. Jefferson & Ad Putter (eds.), *Multilingualism in medieval Britain (c. 1066–1520): Sources and analysis*, 153–169. Turnhout: Brepols.

Scragg, Donald. 2014. Homilies. In Michael Lapidge, John Blair, Simon Keynes & Donald Scragg (eds.). *The Wiley-Blackwell encyclopedia of Anglo-Saxon England*, 2nd edn., 247–248. Malden & Oxford: Wiley Blackwell.

Stolt, Birgit. 1964. *Die Sprachmischung in Luthers Tischreden: Studien zum Problem der Zweisprachigkeit* (Stockholmer germanistische Forschungen 4). Uppsala: Almqvist & Wiksell.

Szarmach, Paul E. 1986. The earlier homily: *De Parasceve*. In Paul E. Szarmach (ed.), *Studies in earlier Old English prose*, 381–399. Albany, NY: State University of New York Press.

Thomason, Sarah G. 2003. Contact as a source of language change. In Brian D. Joseph & Richard D. Janda (eds.), *The handbook of historical linguistics*, 687–712. Oxford: Blackwell.

Timofeeva, Olga. 2010. Anglo-Latin bilingualism before 1066: Prospects and limitations. In Alaric Hall, Olga Timofeeva, Ágnes Kiricsi & Bethany Fox (eds.), *Interfaces between language and culture in medieval England: A festschrift for Matti Kilpiö* (The Northern World 48), 1–36. Leiden: Brill.

Timofeeva, Olga. 2013a. Anglo-Latin and Old English: A case for integrated bilingual corpus studies of Anglo-Saxon registers. In Paul Bennett, Martin Durrell, Silke Scheible & Richard J. Whitt (eds.), *New methods in historical corpora* (Corpus Linguistics and Interdisciplinary Perspectives on Language 3), 195–204. Tübingen: Narr.

Timofeeva, Olga. 2013b. Latin loans in Old English and Finnish loans in modern English: Can we distinguish statistics from myth? In R. W. McConchie, Teo Juvonen, Mark Kaunisto, Minna Nevala & Jukka Tyrkkö (eds.), *Selected proceedings of the 2012 symposium on new approaches in English historical lexis (HEL-LEX 3)*, 166–176. Somerville, MA: Cascadilla Proceedings Project. http://www.lingref.com/cpp/hel-lex/2012/paper2845.pdf (accessed 24 June 2017).

Wieland, Gernot R. 2011. Bilingual education in Anglo-Saxon England: Alfred to Æthelweard. In Michael Baldzuhn & Christine Putzo (eds.), *Mehrsprachigkeit im Mittelalter: Kulturelle, literarische, sprachliche und didaktische Konstellationen in europäischer Perspektive*, 35–57. Berlin & New York: De Gruyter.

Rita Queiroz de Barros

4 Twentieth-century Romance loans: Code-switching in the *Oxford English Dictionary*?

1 Introduction

Defined by Gumperz (1982: 59) as “the juxtaposition within the same speech exchange of passages of speech belonging to two different grammatical systems or subsystems”, code-switching can assume different forms (cf. for instance Bhatt 2001), result from different motivations (cf. for example Gross 2001) and involve different linguistic units. This paper focuses on the particular case of single-word switching, which, drawing on their own data and on previous reports, Poplack and Dion (2012: 307) consider to constitute the “overwhelming majority of other language material” in both written and spoken discourse.

Single-word switching is known to be related to the linguistic process of lexical borrowing, as both phenomena consist in a contact-induced use of (originally) foreign lexical items. This paper revisits those concepts and the (dis)similarities between them, claiming, as indirectly suggested by Busse and Görlach (2002) and Haspelmath (2009: 42), that single-word switches conform to the traditional concept of foreignism and may therefore become entries in standard lexicographic tools. This conclusion draws upon a case study of the admission and treatment of loanwords in the *Oxford English Dictionary* and calls into question the criterion of dictionary attestation as a means of distinguishing single-word switches from lexical borrowings.

2 Code-switching and lexical borrowing

A review of the literature on the concepts of single-word switching and lexical borrowing shows that they are generally acknowledged as two related phenomena. Though there is an ongoing debate on how distinct they are (Deuchar 2008; Durkin 2009: 173–177), many authors agree that both the processes and results of one-word switching and lexical borrowing differ in their nature. Code-

---

Acknowledgement: The author wishes to acknowledge the support of Centro de Estudos Anglísticos da Universidade de Lisboa while preparing this chapter.

DOI 10.1515/9781501504945-004

switching constitutes “a kind of contact-induced speech behaviour” (Haspelmath 2009: 42), since it is “(ideally) a spontaneous, clearly bounded switch from [. . .] one language to [. . .] another” (Heath 2001: 433); a possible example would be the English item *doorman* in the French sequence “C’est rien que pour ouvrir une porte, être doorman, dire allô” (Poplack and Dion 2012: 291). Lexical borrowing is, instead, “a kind of contact-induced language change”, not necessarily dependent on code-switching (Haspelmath 2009: 42), since its result is “(ideally) a historically transferred form, usually a word (or lexical stem), that has settled comfortably into the target language” (Heath 2001: 433). Lexical borrowing therefore consists in the import of a foreign item into the lexis of a recipient language, as for instance the introduction of Italian *broccoli*, *sonata* or *balcony* into English, in which they have been used since the seventeenth century (Durkin 2014: 371). Furthermore, empirical studies quoted by Poplack and Dion (2011) have revealed that code-switches are governed by the grammar of the source language, while lexical borrowings adapt to the grammar of the recipient language (Poplack and Meechan 1998; Torres Cacoullos and Travis 2010).

Despite the differences just identified and Poplack and Dion’s continued insistence that single-word switches and lexical borrowings should be understood as different phenomena (cf. Deuchar 2008 or Poplack and Dion 2012), most authors do agree that they constitute manifestations of the same contact-induced process. In fact, as even Poplack admitted at an early stage of her research, there seems to be “no unequivocal way of deciding when a lexical item from one language, used during discourse in another [. . .] should be considered a loanword” or “another kind of result of language contact”, as for instance code-switching, incomplete acquisition or interference (Poplack and Sankoff 1984: 99). Single-word switches and lexical borrowings are thus seen by many as mere different points on a single continuum: lexical items are initially taken unchanged from a donor language, i.e. code-switched, and then, if repeated often and widely enough, gradually assume more and more characteristics of the recipient language and become loanwords (cf. e.g. Haspelmath 2009; Myers-Scotton 1993b, 2002, 2006; Thomason 2003; Coetsem 2000; Winford 2009).

Part of this understanding is the assumption that, at a given point in time, the distinction between instances of one-word switching and lexical borrowing can prove difficult to establish. And though for some scholars this difficulty results primarily from the lack of a proper methodology (e.g. Gardner-Chloros 2009; Winford 2003), most of them admit that single-word switches and lexical borrowings are difficult to tell apart simply because the conceptual distinction between the two notions is, as seen, intrinsically ambiguous. However, according to Myers-Scotton (1993a: 273) this should be considered a “non-issue”, and scholars rely especially on the extra-linguistic traits characterising established lexical borrowings, such as “frequency, diffusion and dictionary attestation” (Poplack and Dion 2012: 311) in order to distinguish instantiations of the two phenomena. As a consequence, dictionaries are crucial reference tools to distinguish one-word switches from lexical borrowings, as explicitly recognised in some studies (cf. e.g. Callahan 2004: 38; Halmari 1997: 188–189; McClure 1998: 132; Myers-Scotton 1993b: 135; Parfenova 2001: 109), so that lexicographic policies can have a decisive impact on the study of code-switching.

3 Lexical borrowing, loanwords and foreignisms

The difficulty surrounding the concept of lexical borrowing is not confined to its affinities with one-word switching, as will be apparent from the review of the concept presented below.

Though various different definitions have been provided to describe it, the majority of linguists tend to consider that lexical borrowing is either or both the process and the result of taking words from a so-called donor language and using them in another, recipient language (e.g. Haugen 1950; Heath 2001; McArthur 2005; Myers-Scotton 2006; Tatsioka 2008). So, as Einar Haugen (1950: 212) insists in his classic treatment of contact-provoked phenomena, at the heart of the definition of borrowing “is the attempted reproduction in one language of patterns previously found in another”.

Since this reproduction can take different forms, lexical borrowing is often divided into different categories. Drawing on Haugen (1950), Winford (2005: 384) distinguishes, among others, the following two major categories: (i) loanwords, comprising direct loanwords (e.g. the French-derived *rendez-vous* in English) and loan blends (e.g. the Pennsylvania German *bassig* made up of the English root *boss* and the German suffix *-ig*); and (ii) loan shifts or loan meanings, including semantic extensions (e.g. *frio* with the meaning of ‘cold infection’ that is used in the Portuguese spoken in North America) and loan translations (e.g. the German *Wolkenkratzer*, derived from the English *skyscraper*).

Of the categories of lexical borrowing presented, to which others might be added (for instance pseudo-borrowings or creations using only foreign morphemes as Japanese *wan-man-ka*, meaning ‘bus with no conductor’ derived from English *one* + *man* + *car* [Winford 2005: 384]), only one is relevant for the purposes of this paper.[1] That is the case of the direct loanword, which, though obviously used to denote words that are moved from one language into another, is not as clear-cut as one might expect. In fact, the direct importation of one foreign lexical item does not always result in an isomorphic item (the phonemic constitution of English *rendez-vous*, mentioned above, is different from the original one) and there is often a gradual adaptation to the recipient language patterns (for example, only recently has the Anglicism *stresse* begun to be represented with a native spelling in European Portuguese).

---

[1] For a thorough review of the concept of pseudo-borrowings cf. Furiassi (2015).

The differences in the degree of nativisation of loanwords can depend on various factors. According to Haspelmath (2009: 42), adaptation can prove indispensable for the word to be usable in the recipient language, since the original item can present "phonological, orthographic, morphological or syntactic properties in the donor language that do not fit into the system of the recipient language". As an example, languages with gender classes cannot avoid assigning each loanword to a gender class, so that it can enter syntactic constructions requiring gender agreement (for instance in Portuguese the word *pen*, derived from English and meaning 'flash-disk', was attributed the feminine gender); and loanwords from languages using writing systems that are different from that of the recipient system have to be transcribed into the new alphabet (a case in point is that of Japanese and Chinese loanwords in English, which have to be respelled in the Latin alphabet). However, the degree of adaptation of a loanword can vary for other less compelling reasons, such as "the age of a loanword, knowledge of the donor language by the recipient language speakers, and their attitude toward the donor language" (Haspelmath 2009: 42).

The recognition of such different degrees of nativisation of loanwords is not new. A well-known antecedent was provided by nineteenth-century German scholars with a puristic bias (Polenz 1967; Krier 1980). They distinguished between *Gastwörter* (loanwords which do not get assimilated into the borrowing language); *Fremdwörter* (which consist of borrowings adapted only in part); and finally *Lehnwörter* (corresponding to those loans which are well-integrated, so that they are very often not felt as borrowings, particularly by those speakers who do not know the language of origin). As the term *Gastwörter* denotes words that remain independent of the supposed recipient language, amounting, therefore, to quotations or citations, it is rarely to be found in the subsequent literature on lexical borrowing; the concepts of *Fremdwörter* and *Lehnwörter*, on the contrary, have proved useful and operative, and have hence been glossed in other languages, such as by the English terms 'foreignism' and 'loanword', respectively. These are frequently used to distinguish between lexical borrowings with different degrees of adaptation to the recipient language.

It should be noted, however, that the theoretical distinction between foreignisms (i.e. borrowings adapted only in part) and loanwords (i.e. loans which are well-integrated, so that they are very often not felt as borrowings) is, once again, not simple. Those terms are sometimes used to label the same items (as happens with *kamikaze* in English, tagged as a loanword in Evans 1997 and as a foreignism in Tuleja 1989), the theoretical distinction between them has been explicitly labelled as "fuzzy" in the literature (Capuz 1997: 87; Mańczak-Wohlfeld 2012: 173, 174), and the effectiveness of the opposition they are meant to state has been questioned, since "it is difficult to specify where the properly foreign begins" (McArthur 2005: 234).

Despite its shortcomings, the distinction between foreignisms and borrowings has not been discarded from the literature (cf. for example Capuz 1997; Podhajecka 2010: 147–148); instead, several criteria have been suggested to indicate when foreignisms have turned into borrowings. Various authors agree on the following: adaptation to the patterns of the recipient language (such as a nativised pronunciation and spelling or the incorporation of host verbal and nominal suffixes), displacement of native synonyms, and frequency of use (cf. e.g. Haugen 1950: 217, 229; McArthur 2005: 234, 361; Onysko and Winter-Froemel 2011: 1550; also implicit in Bloomfield [1933] 1965: 450, 453).

Some of the conditions just presented have already been mentioned in this paper. In fact, adaptation to the characteristics and patterns of the recipient language and frequency of use were indicated in the previous section as criteria for telling apart lexical borrowings and one-word switches. So, foreignisms and one-word switches turn out to be closely related phenomena, as pointed out before by a few authors. Busse and Gorläch (2002) group the two concepts in the same subcategory of English-derived borrowings in German, together with quotation words; and Haspelmath (2009: 43) recognises that "the notion of foreignism is evidently close to that of a single-word switch", the difference between them being that "single-word switches are even less integrated than foreignisms, to the point of not being (clear) members of the language's lexicon". However, as this is no effective way of distinguishing between tokens of one or the other, it is my contention that foreignisms and single-word switches are not simply closely related but overlapping concepts, so that one-word switches can, after all, constitute entries in monolingual dictionaries. This possibility is explored by means of a thorough analysis of the treatment of direct loanwords in the *Oxford English Dictionary* presented in the following sections. If confirmed, it will have an impact on the standard methods of identifying both historical (cf. Wright 2013) and synchronic single-word switches.

4 The *Oxford English Dictionary* and lexical borrowing

The *Oxford English Dictionary* (henceforward *OED*) is rightly seen as the most prestigious dictionary of English. This lexicographic enterprise took its first steps in 1857 and intended to record all the words which were or had been used in English. It bore its first visible fruit in 1884, with the publication of the first installment of the dictionary under the direction of James Murray, and the first edition, entitled *The New English Dictionary on Historical Principles*, was finished in 1928. This work was soon completed by a supplement edited by Onions and Craigie and dated 1933, the year in which the current title of the *OED* was definitely adopted. Together with a second supplement edited by Robert Burchfield from 1972 to 1986, all this work resulted in the publication of the *OED*'s second edition in 1989. This was later converted in a digital resource, and more recently transformed in a third, incomplete and ongoing online edition, which is not only completing but gradually revising the previous entries, for the first time since 1928. At the moment it includes more than 600,000 entries.

Though the *OED* is the most important dictionary of English, its choice as a source to assess the registering of foreignisms in standard lexicographic tools may seem peculiar. In fact, and despite the consensual recognition of its relevance and major achievements (Brewer 2011: 120), the first edition and the 1933 Supplement of the *OED* have been criticised for a typically Victorian Anglo-centrism (Mugglestone 2005: 162; Ogilvie 2012: 14). This is allegedly to be found not only in "racist and sexist definitions" (Brewer 2011: 120), as the often quoted description of *canoe* as "a kind of boat in use among uncivilized nations" (Béjoint 2000; Moon 1989; Mugglestone 2000), but also, and more importantly for the purposes of this paper, by the exclusion of important sources, of both lexical entries and examples, due to a bias towards literary registers (Nevalainen 1999: 337) and the British standard variety (Algeo 1998: 61). Accordingly, Willinsky (1994: 57) described the first edition of the *OED* as Shakespeare's dictionary, since this author's works constitute its largest single source of quotations; and Bailey and Görlach (1982: 4) considered that it "virtually excluded words not in general use in Great Britain and the United States", forgetting all other settings and forms of English. So, although the unusually heterogeneous nature of English vocabulary requires by definition the admission of numerous words imported from foreign sources, this critique suggests that an indefinite but certainly important number of others have been rejected, which the subsequent work led by Robert Burchfield and the revision in progress may not yet have neutralised.

However, this understanding of the *OED* should be toned down. In fact, and in the first place, the main editor of the first edition of the *OED*, James Murray, was well aware of the composite nature of the vocabulary of English and did not eschew the issue of lexical borrowings. His team accepted many into the dictionary's columns. That is the case of the word *senhor*, originally Portuguese. This entry, first published in 1912, is reproduced below in its 1989 version:

> || **senhor**, n.
>
> Pronunciation: /seˈɲor/
>
> Etymology: Pg. *senhor* = Sp. *señor*, It. *signor*, F. *seigneur*:–L. *seniōr-em*, acc. of *senior* SENIOR *a.* and *n.*
>
> In Portuguese use, or with reference to Portuguese: A term of respect placed before the name of a man in addressing him or speaking of him, equivalent to the English ‘Mr.’ Also used without the name as a form of address, equivalent to ‘sir’ in English. Hence, a Portuguese gentleman.
>
> 1795 J. MURPHY *Trav. Portugal* 31, I am sorry, Senhors, (said she,) that you have not rested well.
>
> 1830 *Portugal; or Yng. Travellers* 56 Senhor Maced..turning to Mr. Grey, said, smilingly: ‘This is good policy, is it not, Senhor?’
>
> 1830 *Portugal; or Yng. Travellers* 56 The lively sallies of the Senhor.
>
> 1853 A. R. WALLACE *Amazon & Rio Negro* 198 Though Senhor L. is well acquainted with the river, we here almost lost our way.

As can be seen, *senhor* keeps phonological and spelling traits that are typically Portuguese – the palatal nasal [ɲ] and its spelling <nh> – and is explicitly attached to the Portuguese language and to Lusophone realities, both by the definition of the word (“in Portuguese use, or with reference to Portuguese”) and the examples presented.

Besides accepting such foreignisms or code-switches, James Murray explicitly reflected upon loanwords in the “General Explanations” with which he introduced the first volume of the dictionary (1884). Interestingly, he shared with most of the linguists quoted above the understanding that loanwords should be categorised according to their degree of naturalisation or, in his words, “citizenship in the language” (Murray [1884] 1961: xxix). He divided them into four categories, which he described as naturals (“all native words like *father*, and all fully naturalized words like *street, rose, knapsack*, *gas*, *parasol*”); denizens (“words fully naturalized as to use, but not as to form [. . .] as *aide-de-camp*, *locus, carte-de-visite, table d’hôte*”); aliens (“names of foreign objects, titles, etc. which we require often to use, and for which we have no native equivalents, as *shah*, *geyser*, *cicerone*, *backsheesh*”); and casuals (“foreign words in the same class, not in habitual use”, as *grillo* or *grillino*). But, just as modern linguists insist on the fuzziness of the distinction of loanwords from foreignisms, Murray admitted that the boundaries between the groups identified were not strict, and that loans moved from one to another, with a “constant tendency [. . .] for words to pass from the last to the first” class. That is probably the reason why the categories postulated were never used as labels for the entries included in the dictionary. In the first edition of the *OED*, more exotic items (namely denizens, aliens, and casuals if they "approach, or formerly approached, the position of these" [Murray 1961: xxix]) were simply marked with two small parallel lines, placed before the headword. These lines were called in-house tramlines (Ogilvie 2012: 18), and were meant to mark the entries that were still "alien" or "not yet naturalized". That was the case of *aide-de-camp, cicerone* and *senhor.*

The examples just discussed call into question the image of the *OED* as a Britocentric repository of the English vocabulary and suggest instead a liberal hospitality to borrowings ever since the dictionary's early history. Interestingly, this was precisely the impression of various contemporaries of the *OED*'s first edition. Ogilvie (2012: 53, 75) quotes, among others, the following reviews:

> A particular feature of the C-Cass section is the disproportionately large number of words derived from foreign languages outside the familiar circle of Latin, Greek and French, e.g. caaba, cabaan, caback, caballero, [. . .] cacique, [. . .] &c, not to mention numerous words of Celtic origin imported in comparatively recent times from Scotland and Ireland. (Fennell 1888: 442)
>
> They have been far too liberal in admitting to the columns of an English Dictionary a multitude of words that form no part of the English language. (Anon. 1889: 349)
>
> Queer, outlandish words are quite common in Sir James Murray's dictionary, and this section [tombal–trahysh] has its fair share. (Anon. 1917: 17)

Furthermore, in a quantitative study of the presence of loanwords in the 1933 and 1972–1986 supplements to the *OED*, Ogilvie (2012: 177) provides empirical evidence of the fact that the inclusion of loanwords and of items originating in World Englishes was proportionately similar in both works.

This tendency to liberally admit loanwords into the *OED* was kept in its later editions, as hospitality to foreign influences is more in line with both the global status of English and the non-prescriptive approach of modern lexicography. As a result, the dictionary now contains entries which are surprising at the very least. Stubbs (1988) mentions the cases of *Leibniz* and *hochgeboren* ('well-born'), which both derive from German and date from 1976; to these can be added various other examples, such as *taxe de séjour* (a tax imposed on tourists in France and other countries), imported from French in 1986; *squadra* ("In Italy: a paramilitary squad organized to support and promulgate Fascism; a Fascist cadre")[2], borrowed from Italian and first attested in 1996; or *feijoada* ("black beans stewed with salt pork, sausage, bacon, jerked beef, onions, garlic, and tomatoes"), imported from Portuguese and first attested in 1993. Furthermore, the so-called tramlines used to denote the more exotic items in both the first and second editions of the *OED* were all suppressed, irrespective of the degree of naturalisation of the entry. That is the case of *senhor*, quoted above.

---

[2] Whenever they are delimited by double quotation marks, the glosses of the entries hereafter discussed are quoted from the *OED* online.

Bearing in mind the facts and data presented in the previous paragraphs, there is reason to suspect that the *OED* very often lists not only proper loanwords (such as *sargasso*, from Portuguese, or *aromatherapy*, from French), but also foreignisms or code-switches, that is exotic items used to denote foreign realities. This possibility was investigated by means of the study on the admission and treatment of foreignisms in the *OED* described below.

5 The treatment of foreignisms in the *OED*: A case study

The immediate aims of this study were (i) to determine the number of foreignisms (that is, exotic, non-nativised words) included in the present online version of the *OED*; and (ii) to find out whether such foreignisms are treated differently from loanwords.

Given the size of the dictionary, this study had to focus on a sample of the *OED* online.[3] As the twentieth century was "the period of most vigorous expansion [in lexical importation] since that of the late-sixteenth and seventeenth centuries" (Ayto 2012: n.pg.) and Romance sources have provided a significant number of loanwords at this time, the sample considered is constituted by all the entries in the *OED* originating in modern Romance languages and first attested in English in the twentieth century. It was compiled using the advanced search tools provided by the *OED* online, and amounts to 1813 items.

Each entry on this list was analysed with the purpose of identifying all the foreignisms. This classification was based upon the criteria suggested in the literature to distinguish foreignisms from loanwords – in particular failure to displace native/nativised synonyms (for example *numéro*) and spelling, phonological and/or morphological traits that are not to be found in the recipient language (as *vulgarisation*, pronounced [vylgarizasjɔ̃]). When this proved inapplicable, the referent of the specific entries compiled was used (as in the cases of *Minitel* or *Frelimo*). This procedure allowed for the identification of nine categories of foreignisms in the initial list of words. They are as follows:

1. Items written with non-English graphies, as the <gn> or <zz> sequences in, respectively, *agnolotti* ("a variety of pasta in the shape of small half-moons or squares, filled with minced meat") or *sprezzatura* ("ease of manner, studied carelessness").

---

[3] For other examples and especially other conclusions of this study cf. Barros 2015.

   Excluded are those cases which, though sharing the characteristic mentioned, are too familiar to contemporary speakers of English to be still considered foreignisms. That is the case of *cappuccino*, also imported in the twentieth century.

2. Items whose registered pronunciation includes sounds not integrated in the English phonological system. This is the case of *scugnizzo* [skuˈɲittso] (“in Naples, a street urchin”) or *carte d’identité* [kart didɑ̃tite] (“identity card”).

   Though the members of this class are also usually part of the previous one, the reverse is not true. The pronunciation of *sprezzattura* indicated in the *OED* is [sprɛtsəˈt(j)ʊərə], in which no foreign sound or sound pattern is to be identified.

3. Items displaying marks of foreignism in at least half of their examples, namely italics or inverted commas in their written representation, or a gloss. To this category belong *acqua alta* (an exceptionally high tide which periodically floods Venice), *capo* (“the leader of a branch or ‘family’ of the Mafia”), *panzerotto* (an Italian dish), *couchette* (a “Continental railway carriage in which the seats are convertible into sleeping-berths; such a berth”) or *carretera* (“a main road” in Spain). The examples of this last item, displaying three out of four spellings in italics, are shown below:
   - 1900 Engin. Mag. 19 683 A *carretera*, or macadamized road is the ‘best road’ or route of the ‘fast mail’ in Porto Rico.
   - 1923 Blackwood’s Mag. Aug. 178/1 The extreme provincialism of the Spanish *carretera*.
   - 1924 Blackwood’s Mag. June 804/2 Very pretty they look sailing slowly along the dusty *carretera* to do the day’s marketing.
   - 1958 L. R. Muirhead Blue Guide to Northern Spain (ed. 2) p. cxlii, Apart from the Carreteras Nacionales, however, many of the roads are stony and often under repair.

4. Items presenting grammatical characteristics unknown to English, such as (but not only) adjectives inflected for gender and number. Possible examples are *cabotin* (“a low-class actor”), *scugnizzo* (glossed above), *voyant* (“showy, gaudy, flashy”) and *voulu* (“deliberate”), as shown by the following citations:
   - 1951 L. P. Hartley My Fellow Devils i. 8 He was a *cabotin* sort of character, always playing a lonely part. / 1966 Punch 30 Nov. 826/2 Sarah Bernhardt was a *cabotine*; she excelled in the art of self-advertisement.
   - 1961 ‘W. Haggard’ Arena xi. 91 There was more in Naples than *scugnizzi*.
   - 1906 Punch 18 Apr. 286/3 ‘Anno Domini’..whom one would expect to be smart and *voyante*, is simply the dowdiest, quietest of mice.
   - 1957 Durrell Justine ii. 132 The idea is not spontaneous, but *voulue*.

5. Words constituting non-nativised versions of fully adapted borrowings. *Nuancé*, *problématique*, *rite de passage*, *sensibilité*, *simpliste* and *numéro* are examples of this category, which comes mainly from French. The last example is presented in the *OED* as follows (my emphasis):

   numéro, n.
   Forms: 19– numero, 19– numéro.
   Etymology: < French *numéro*
   = **number n. (in various senses)**.

6. Items denoting untranslatable meanings. That is the case of *pied noir*, defined as "In France: a person of European origin living in Algeria during the period of French rule, *esp.* a French person repatriated after Algeria was granted independence in 1962" ("1977 *Time* 21 Nov. 12/1 Unlike the white settlers of Rhodesia or the French *pieds-noirs* of Algeria, the Afrikaners have no ties to a European motherland"); that is also the case of *retornado*, described as "A Portuguese subject formerly resident in a Portuguese colony who returns to settle in Portugal after colonial independence" ("1987 *Financial Times* 30 Oct. p. vi/4 The country had..to make room for 700,000 'retornados' from overseas – with a further influx still expected from Macao").
7. Words or expressions denoting foreign commercial products, labels or premises. That is the case of *Gitane* or *Gauloise* (cigarettes of French brands), *Minitel* ("a proprietary name for: an interactive videotext system sponsored by the French government, in which computer terminals are linked together via the telephone network. Also: a home computer terminal forming part of this system") or *Monoprix* ("In France: any of a chain of shops selling a range of household and other goods"). Some of such trade terms gained more currency than others (as the two cigarette trademarks), but all remain foreign.
8. Entries constituting the mere translation of English words, which would therefore fit in a bilingual dictionary but only arguably in a monolingual one. This category can be exemplified with *autopista*, *autoroute* and *autostrada*, which are respectively the Spanish, French and Italian translation of motorway; the same happens with *churrascaria*, a word derived from Portuguese and meaning a barbecue restaurant.
9. Items denoting facts, concepts or persons expectable in an encyclopedia, but not in a dictionary of the English language. That is the case of *saudade* (defined as "melancholy, nostalgia, as a supposed characteristic of the Portuguese or Brazilian temperament"), *Duce* (Mussolini), *Falange* ("a Spanish political party, founded in 1933 as a Fascist movement by J. A. Primo de Rivera"), *Frelimo* ("the nationalist liberation party of Mozambique, founded in 1962"), or *Union Cors*e ("a criminal organization controlled by Corsicans, operating in France and elsewhere").

Apart from the classes just identified, the sample included two isolated entries that do not fit in any of the categories listed: the adjective *rondine* ("1923 E. Sitwell Bucolic Comedies 70 Fat blondine pearls Rondine curls Seem.") and the verb *trascine* ("1922 Joyce Ulysses, i. 47 She trudges,... drags, trascines her load"). These words seem to be the product of the creativity of the writers quoted for examples and so have not been considered in the typology above.

The entries integrating one and usually more than one of the nine categories just postulated amount to 663 of the total of 1813 entries. So, the first result to emphasise is that foreignisms represent 37%, i.e. more than a third of the originally Romance items first attested in the twentieth century and admitted into the *OED*. Furthermore, the examples cited, the attestations quoted and the categories identified highlight the fact that foreignisms are indeed code-switches – they are either words which are not nativised at the levels of spelling, pronunciation, morphology or meaning, or mere other-language translations of English words.

As far as the labelling of these foreignisms is concerned, this sample reveals that such a procedure is exceptional. In fact, only a very small minority of the foreignisms identified are labelled as special: 13 out of the total of 663 items are described as "not fully naturalized in English". *Photogenique* is a possible example, as shown below (my emphasis):

> photogenique, adj.
>
> Pronunciation: Brit. /ˌfəʊtəʊ(d)ʒɛˈniːk/, U.S. /ˌfoʊdoʊ(d)ʒɛˈnik/
>
> Forms: 19– photogenique, 19– photogénique.
>
> Etymology: < French *photogénique* PHOTOGENIC adj.
>
> **Not fully naturalized in English**.
>
> = PHOTOGENIC adj. 4.
>
> 1923 *Coshocton* (Ohio) *Tribune* 20 Mar. 1/1 The American man is always handsome, well-dressed and photogenique.
>
> 1948 E. WAUGH *Loved One* 5 Her legs were never photogénique.
>
> 1972 *Times* 12 Dec. 18/ Not very photogenique.

In addition, the criteria for such labelling are neither explained in the introductory materials of the dictionary nor discernible by the analyst. As an example, in this respect *photogenique* parallels *sensibilité*, cited above, as they both constitute less naturalised versions of a corresponding borrowing (*photogenic* and *sensibility*, respectively); but only the first one is described as "not fully naturalized". The signalling of foreignisms in the *OED* proves therefore to be scarce and erratic.

6 Implications for the study of (historical) code-switching

The case study presented in the previous section is no doubt limited. It is based upon a restricted sample and the results obtained may have been intensified by the presence of words imported from French, as the traditional prestige of French among native speakers of English and their familiarity with the language may have increased the receptivity and respect for the original pronunciation, spelling and morphology. However, this analysis demonstrates a liberal admission and treatment of loanwords in the *OED* and especially allows for some conclusions that have an impact on the study of historical code-switching.

In the first place, the twentieth-century Romance loans found in the dictionary and the categories postulated for the non-nativised items definitely support the idea that one-word code-switches and foreignisms are the same. Indeed, main characteristics considered to brand code-switches as different from lexical borrowings are met by the great majority of the foreignisms identified. That is the case of the attested lack of adaptation to the patterns of the recipient language and, given the examples identified, of a predictable low frequency in real language use. As the overlapping of one-word code-switches and foreignisms had been only hinted at so far, this study may contribute to the ongoing discussion on the processes of switching and borrowing.

Secondly, since foreignisms and one-word code-switches turn out to be the same, the analysis presented above proves that code-switches do constitute dictionary entries and that dictionary attestation is therefore less reliable as a tool to distinguish between one-word switches and lexical borrowings than claimed in the literature. A reassessment of the role lexicons play in the identification of (historical) code-switching is therefore in order.

Finally, the case study on the treatment of loanwords described in this paper resulted in the presentation of a theory-dependent but also empirically-driven typology of foreignisms or one-word switches. Though further data may lead to its revision, this new typology can surely be of use to those taking further steps in the field of code-switching.

References

Algeo, John. 1998. Vocabulary. In Suzanne Romaine (ed.), *The Cambridge history of the English language*, vol. 4: *1776–1997*, 57–91. Cambridge & New York: Cambridge University Press.

Anon. 1889. The literature and language of the age. *Edinburgh Review*, April, 339–350.

Anon. 1917. Review of *A New English Dictionary*, *Vol X Tombal-Trahysh*. *The Scotsman*, 20 October, 17.

Ayto, John. 2012. Twentieth century English – an overview. http://public.oed.com/aspects-of-english/english-in-time/twentieth-century-english-an-overview (accessed 4 May 2016).

Bailey, Richard W. & Manfred Görlach. 1982. Introduction. In Richard W. Bailey & Manfred Görlach (eds.), *English as a world language*, 1–6. Cambridge: Cambridge University Press.

Barros, Rita Queiroz de. 2015. "Good policy, is it not, senhor(es)?" The myth of the superiority of English and the treatment of loanwords in the *Oxford English Dictionary*. In Ana Raquel Fernandes, José Pedro Serra & Rui Carlos Fonseca (eds.), *The power of form: Recycling myths*, 213–225. Newcastle upon Tyne: Cambridge Scholars Publishing.

Béjoint, Henri. 2000. *Modern lexicography*. Oxford: Oxford University Press.

Bhatt, Rakesh M. 2001. Code-switching: Structural models. In Rajend Mesthrie (ed.), *Concise encyclopedia of sociolinguistics*, 456–461. Amsterdam, New York, Oxford, Shannon, Singapore & Tokyo: Elsevier.

Bloomfield, Leonard. [1933] 1965. *Language history*. Ed. Harry Hoijer. New York: Holt, Rinehart and Winston.

Brewer, Charlotte. 2011. Johnston, Webster and the *Oxford English Dictionary*. In Haruko Momma & Michael Matto (eds.), *A companion to the history of the English language*, 113–121. Singapore: Wiley Blackwell.

Busse, Ulrich & Manfred Görlach. 2002. German. In Manfred Görlach (ed.), *English in Europe*, 13–36. Oxford: Oxford University Press.

Callahan, Laura. 2004. *Spanish/English codeswitching in a written corpus*. Amsterdam & Philadelphia: John Benjamins.

Capuz, Juan Gómez. 1997. Towards a typological classification of linguistic borrowing. *Revista alicantina de estudios ingleses* 10. 81–94.

Coetsem, Frans van. 2000. *A general and unified theory of the transmission process in language contact*. Heidelberg: Winter.

Deuchar, Margaret. 2008. Evaluating competing models of code-switching: Overcoming incommensurability. Paper presented at the International Conference on Models of Interaction in Bilinguals, University of Bangor, 24–26 October. https://www.yumpu.com/en/document/view/18724237/click-here-for-pdf-centre-for-research-on-bilingualism-in-theory-http://www.bangor.ac.uk/bilingualism/conference/Deuchar.pdf (accessed 20 December 2014).

Durkin, Philip. 2009. *The Oxford guide to etymology*. Chippenham: Oxford University Press.

Durkin, Philip. 2014. *Borrowed words*: *A history of loanwords in English*. Oxford: Oxford University Press.

Evans, Toshie M. 1997. *A dictionary of Japanese loanwords*. Westport, CT: Greenwood Press.

Fennell, C. A. M. 1888. Review of NED Part IV C-Cass. *Athenaeum*, 6 Oct.

Furiassi, Cristiano. 2015. False borrowings and false Anglicisms: Issues in terminology. In Cristiano Furiassi & Henrik Gottlieb (eds.), *Pseudo-English: Studies on false Anglicisms in Europe*, 251–278. Berlin, Boston & Munich: De Gruyter Mouton.

Gardner-Chloros, Penelope. 2009. *Code-switching*. Cambridge: Cambridge University Press.

Gross, S. 2001. Code-switching: Sociopragmatic models. In Rajend Mesthrie (ed.), *Concise encyclopedia of sociolinguistics*, 454–456. Amsterdam, New York, Oxford, Shannon, Singapore & Tokyo: Elsevier.

Gumperz, John J. 1982. *Discourse strategies*. Cambridge: Cambridge University Press.

Halmari, Helena. 1997. *Government and codeswitching: Explaining American Finnish*. Amsterdam & Philadelphia: John Benjamins.

Haspelmath, Martin. 2009. Lexical borrowing: Concepts and issues. In Martin Haspelmath & Uri Tadmor (eds.), *Loanwords in the world's languages: A comparative handbook*, 35–54. Berlin & New York: Mouton de Gruyter.

Haugen, Einar. 1950. The analysis of linguistic borrowing. *Language* 26(3). 210–231.

Heath, J. 2001. Borrowing. In Rajend Mesthrie (ed.), *Concise encyclopedia of sociolinguistics*, 432–442. Amsterdam, New York, Oxford, Shannon, Singapore & Tokyo: Elsevier.

Krier, Fernande. 1980. Lehnwort und Fremdwort im Maltesischen. *Folia Linguistica* 14(1–2). 179–184.

Mańczak-Wohlfeld, Elżbieta. 2012. The status of English lexis in the Polish language. In Zdzisław Wąsik & Piotr P. Chruszczewski (eds.), *Languages in contact 2011*, 165–176. Wrocław: Philological School of Higher Education in Wrocław Publishing.

McArthur, Tom (ed.). 2005. *The Oxford companion to the English language*. Oxford: Oxford University Press.

McClure, Erica. 1998. Form and function in written national language–English codeswitching: Evidence from Mexico, Spain and Bulgaria. In Rodolfo Jacobson (ed.) *Codeswitching worldwide*, vol. 1, 125–150. Berlin & New York: Mouton de Gruyter.

Moon, Rosamund. 1989. Objective or objectionable? Ideological aspects in dictionaries. In Murray Knowles & Kristen Malmkjær (eds.), *Language and ideology*. [Special issue]. *ELR Journal* 3. 59–94.

Mugglestone, Lynda. 2000. An historian not a critic: The standard of usage in the *OED*. In Lynda Mugglestone (ed.), *Lexicography and the OED*, 189–202. Oxford: Oxford University Press.

Mugglestone, Lynda. 2005. *Lost for words: The hidden history of the Oxford English Dictionary*. New Haven & London: Yale University Press.

Murray, James A. H. [1884] 1961. General explanations. In James A. H. Murray, Henry Bradley, W. A. Craigie & C. T. Onions (eds.), *The Oxford English dictionary: Being a corrected re-issue with an introduction, supplement, and bibliography of A new English dictionary on historical principles*, xxvii–xxxiv. Oxford: Clarendon Press.

Myers-Scotton, Carol. 1993a. Building the frame in code-switching: Evidence from Africa. In Salikoko S. Mufwene & Lioba Moshi (eds.), *Topics in African linguistics*, 253–278. Amsterdam & Philadelphia: John Benjamins.

Myers-Scotton, Carol. 1993b. *Duelling languages: Grammatical structure in code-switching*. Oxford: Clarendon Press.

Myers-Scotton, Carol. 2002. C*ontact linguistics: Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.

Myers-Scotton, Carol. 2006. *Multiple voices: An introduction to bilingualism*. Malden, MA: Blackwell.

Nevalainen, Terttu. 1999. Early Modern English lexis and semantics. In Roger Lass (ed.), *The Cambridge history of the English language*, vol. 3: *1476–1776*, 332–457. Cambridge: Cambridge University Press.

OED = Oxford English Dictionary. http://www.oed.com/.

Ogilvie, Sarah. 2012. *Words of the world: A global history of the Oxford English Dictionary*. Cambridge: Cambridge University Press.

Onysko, Alexander & Esme Winter-Froemel. 2011. Necessary loans – luxury loans? Exploring the pragmatic dimension of borrowing. *Journal of Pragmatics* 43(6). 1550–1567.

Parfenova, Olga S. 2001. Linguostatistic study of Bulgarian in the Ukraine. In Rodolfo Jacobson (ed.), *Codeswitching worldwide*, vol. 2, 105–124. Berlin & New York: Mouton de Gruyter.

Podhajecka, Mirosława. 2010. The third edition of the *OED* and lexical transmission: Towards a consistent research methodology. In John Considine (ed.), *Webs of words: New studies in historical lexicology*, 141–190. Newcastle: Cambridge Scholars Publishing.

Polenz, Peter von. 1967. Fremdwort und Lehnwort sprachwissenschaftlich betrachtet. *Muttersprache* 77. 65–80.

Poplack, Shana & Nathalie Dion. 2011. Can borrowing and codeswitching be distinguished? Paper presented at the conference Methods in Dialectology 14, University of Western Ontario, 2–6 August. Abstract at http://www.sociolinguistics.uottawa.ca/shanapoplack/presentations/other.html (accessed 20 December 2014).

Poplack, Shana & Nathalie Dion. 2012. Myths and facts about loanword development. *Language Variation and Change* 24(3). 279–315.

Poplack, Shana & Marjory Meechan (eds.). 1998. Instant loans, easy conditions: The productivity of bilingual borrowing. [Special issue]. *International Journal of Bilingualism* 2(2).

Poplack, Shana & David Sankoff. 1984. Borrowing: The synchrony of integration. *Linguistics* 22(1). 99–136.

Stubbs, Michael. 1988. German loan words and cultural stereotypes. *English Today* 14(1). 19–26.

Tatsioka, Zoi. 2008. Code-switching and English loanwords in the Greek media. *Newcastle Working Papers in Linguistics* 14. 129–144. http://www.ncl.ac.uk/linguistics/research/workingpapers/volume14.htm (accessed 20 December 2014).

Thomason, Sarah Grey. 2003. Contact as a source of language change. In Brian D. Joseph & Richard D. Janda (eds.), *The handbook of historical linguistics*, 687–712. Oxford: Blackwell.

Torres Cacoullos, Rena & Catherine E. Travis. 2010. Variable *yo* expression in New Mexico: English influence? In Susana Rivera-Mills & Daniel Villa (eds.), *Spanish of the Southwest: A language in transition*, 185–206. Madrid & Frankfurt: Iberoamericana / Vervuert.

Tuleja, Tad. 1989. *Foreignisms: A dictionary of foreign expressions commonly (and not so commonly) used in English*. New York & London: Collier Macmillan.

Willinsky, John. 1994. *Empire of words: The reign of the OED*. Princeton, NJ: Princeton University Press.

Winford, Donald. 2003. *An introduction to contact linguistics*. Malden, MA: Blackwell.

Winford, Donald. 2005. Contact-induced changes: Classifications and processes. *Diachronica* 22(2). 373–427.

Winford, Donald. 2009. On the unity of contact phenomena and their underlying mechanisms. In Ludmila Isurin, Donald Winford & Kees de Bot (eds.), *Multidisciplinary approaches to code-switching*, 279–306. Amsterdam & Philadelphia: John Benjamins.

Wright, Laura. 2013. On historical dictionaries and language boundaries. In Liliana Sikorska & Marcin Krygier (eds.), *Evur happie & glorious, ffor I hafe at will grete riches*, 11–26. Frankfurt am Main: Peter Lang.

Louise Sylvester

5 A semantic field and text-type approach to late-medieval multilingualism

1 Introduction

In recent decades medieval multilingualism has been most frequently analysed through examination of various examples of code-switching found in the texts; indeed, research on modern code-switching has provided the main theoretical frameworks for such studies (see e.g. Davidson 2005; Hunt 2000; Schendl 2000, 2012, 2013; see also Schendl and Wright eds. [2011] in which, as one might expect, the term is present in the titles of all but one of the chapters). A particular area of interest, especially for lexical and semantic investigations, has been the single-word switches (see e.g. Trotter 2006, 2010). William Rothwell (2000: 230) has claimed that such code-switches are likely to have occurred because scribes did not know the terminology in Latin or French, though evidence from Schendl and Wright eds. (2011) and Dodd (2012) suggests that this is unlikely to have been the case. The argument has been made that the more technical terms would be the ones most likely to be switched items as it was most necessary for the manual workers to understand them and these would be the people least likely to be familiar with the more prestigious languages (see Wright 1995: 317; Jefferson 2000: 184). Counter-examples which tend to undermine this idea have been adduced in work by David Trotter, who has found widespread switching into Middle English in Anglo-French texts and suggests that it "seems unlikely that the English words are being deployed because of either ignorance of Anglo-French lexis on the part of the scribe, or presumed ignorance of the workmen (this appears plausible in some other contexts, but makes little sense here when this is an account *post factum* of work already carried out)" (Trotter 2010: 53). Herbert Schendl discusses the use of the vernacular for hyponyms in Latin wills, concluding that "even though everyday objects seem to be frequently referred to in the vernacular, there is no clear evidence that this is due to the lack of a Latin term" (Schendl 2013: 51).

This paper adopts a new approach to the unresolved question of which lexical items are most likely to show up as single-word switches, those terms which are the most general in their usage or those which are the most restricted in sense. The texts in the text base which forms the data set for this investigation are almost all monolingual (almost all Middle English or Anglo-French, with a small proportion of Latin texts) but with what appear to be single words in the

DOI 10.1515/9781501504945-005

(other) vernacular; that is, some of the texts contain what look like code-switches, for example, this sentence from the will of Thomas Harpham (dated 1341):

> Item lego Katerinæ filiæ meæ unum **coverlyt** crocei coloris de operibus florum lilii, unum **kyngle**
>
> ['Likewise I leave my daughter Katerina a yellow bedspread with lilies embroidered on it, a girdle']

The will is in Latin, but the emboldened terms are included in the *Middle English Dictionary* (*MED*; *coverlyt* is a borrowing from Anglo-French, and *kyngle* from Old Norse). These kinds of linguistic phenomena appear to fit into two of the categories of code-switching set out by Bullock and Toribio (2009: 3), *congruent lexicalization* ("two languages share a common grammatical structure that can be filled with lexical elements from either language") or *insertion* ("which involves the embedding of a constituent – usually a word or a phrase – in a nested A-B-A structure"). It might be possible to categorize examples like the one quoted above as nonce-borrowings but we do not have the kinds of morphological clues that verbs would offer to invoke structural criteria, and there is no generally accepted distinction between single-word code-switches and nonce-borrowings. Indeed, many researchers consider code-switching and borrowing to be points on a continuum and therefore not distinguished, at least by bilinguals (Schendl and Wright 2011: 24). The lexicological approach taken in this paper makes no assumptions about the classification of the terms it examines as code-switches; rather, it investigates their distribution as phenomena occurring within a multilingual textual culture, focusing on the languages of the texts in which the terms appear and the historical dictionaries in which they are attested, and considering the lexical items within the context of their distribution across a range of text types.

The lexical items belonging to a single semantic domain, DRESS AND TEXTILES, are examined in a range of text types selected and edited for the anthology *Medieval Dress and Textiles in Britain: A Multilingual Sourcebook* (Sylvester, Chambers, and Owen-Crocker 2014).[1] Four text types were selected: wills, sumptuary laws, petitions, and romances, with total word counts for each text type roughly though not entirely evenly balanced: the work involved in transcribing and editing the manuscripts means that we have fewer petitions in the text base.

---

[1] The sourcebook was the main output of the project *Medieval Dress and Textile Vocabulary in Unpublished Sources* and the textual extracts and translations are taken from it. Where existing editions were used, their conventions with regard to the expansion of abbreviations were followed. We are grateful to the Leverhulme Trust for funding for the project between 2009 and 2012.

The excerpts from the wills consist of 6,289 words; those of the sumptuary legislation, 7,189 words; 2,781 of the petitions; and 8,116 of the romances. The distribution of the vocabulary was investigated under the following headings: (i) Terms shared across all four text types [3 items]; (ii) Terms shared across two or all three of the administrative document types (but not the literary text type) [approx. 40 items]; (iii) Terms shared across the romance and at least one other text type [approx. 115 items]; (iv) Terms restricted to one text type [approx. 350 items].[2] The occurrences of lexical items which appear under (i) and (iv) are discussed below. Code choice is examined mainly by recourse to the database of the *Lexis of Cloth and Clothing Project* (*LexP*).[3] Our intention in creating the *LexP* database was to collect the terminology in which cloth and clothing was named and described in all the languages in use in medieval Britain. This means that for each lemma in the database, evidence taken from all the relevant historical dictionaries is arrayed below the headword, so our knowledge of the language(s) of the texts in which the terms occur in texts in our text base can be set alongside the information about the language(s) in which the terms are attested (in dictionary citations) throughout the medieval period.

2 Dress and textile terms in most general use

Three cloth and clothing terms appear in all four text types: *kirtle*, *pelure* and *fur* (*Oxford English Dictionary* [*OED*] headwords are given here: the terms occur in a range of forms in the medieval texts), and each is discussed below. The small number of terms in this category is probably a result of differences between the text types. The general terms *array* (outfit, attire, clothing) and *garnementes* (articles of clothing, garments) occur in petitions and romances but are not found in wills or sumptuary laws, almost certainly because in both these text types exact specifications are crucial (see Sylvester 2011, 2016). The first term, *kirtle*, appears to contradict the idea of the importance of particularity of reference. The definition of *kirtle* in the *LexP* database indicates that the term has a range of meanings:

---

[2] The numbers are not exact because it is not always clear whether we are dealing with the same word attested in different languages, or different words.

[3] The database can be found at http://lexissearch.arts.manchester.ac.uk/ (26 January 2016). We are grateful to the AHRC for the funding between 2006 and 2011 of *The Lexis of Cloth and Clothing in Britain c. 700–c. 1450: Origins, Identification, Contexts and Change* project. The project's website includes details of the dictionaries from which data and definitions were drawn: http://lexisproject.arts.manchester.ac.uk/ (27 January 2016).

> *Garment*, general term for an upper-body garment of various descriptions: a man or boy's tunic, (later) a woman's gown; originally short but with varying lengths; on occasion described as made of wool or fur. The term kirtle could be used to describe a simple tunic with a belt gathering (found, for example, in Biblical translations). Kirtles were worn by both sexes [. . .] and were also worn by nuns. In the fourteenth century, the term typically applied to a tunic (or a longer gown) worn alone by working men and women, or under a surcoat, cote-hardie or gown; also carried figurative, abstract and anatomical senses: a covering or protection; etc.

A second sense for *kirtle* is listed in *LexP*, though overlaps with the first meaning are noted: "specifically, a woman's body garment: often an outer garment worn over a smock or under a mantle, gown or pilch". The *LexP* database lists the term as O[ld]E[nglish], M[iddle]E[nglish], A[nglo]F[rench], L[atin], and O[ld]Sc[ots]. This means that *kirtle* is cited in the dictionaries of those languages. It may also imply that if the term appears in an English, Anglo-French or Latin (or Old Scots) text, it is a native term or a borrowing but not a code-switched item, depending on the policies of the relevant dictionaries (see Trotter 2010; Wright 2011), though its etymology is Old English.

Although it is worth keeping in mind the fact that meanings are not stable, and *kirtle* becomes polysemous and undergoes slight semantic shift during the medieval period, the application of each of its senses in the text base appears to be quite precise. The term appears in two wills, one piece of sumptuary regulation, one petition and the response it elicited, and four romance texts. The wills are both by women and are in Middle English with Latin titles and closing formulae. In the will of Isabel Gregory dated 1431–1432, we find:

> also I be-quethe to Ione my dowter, a blew goune and a grene **kyrtyll**, and a schete
> ['Also I bequeath a blue gown and a green kirtle and a sheet to my daughter Joan'][4]

Other legacies of cloth and clothing in this will include a table-cloth, bed, mattress, sheets, cushions, cloaks, and gowns. In Margaret Asshcombe's will, dated 1434, the term appears as follows:

> also y be-quethe to Clemens, the woman that kepes me, a gowne of Musterdevylers, & a **kyrtell** of musterdevylers with grene sleues, & an hode of blak of lure, & an hod of blewe. [. . .] alsoo y bequ*e*the to Aneys Copursmy*th* a combe of yverie, & to Aneys hir' mayden', a russet **kyrtell** furred with lambe.
>
> ['Also I bequeath to Clemens, my housekeeper, a musterdevillers gown and a kirtle of musterdevillers with green sleeves and a hood of black cloth of Lierre, and a blue hood. [. . .] Also I bequeath to Aneys Copursmyth an ivory comb, and to Aneys her maid a russet gown trimmed with lambswool.']

---

[4] Translations are taken from Sylvester, Chambers, and Owen-Crocker (2014) and are my own unless specified.

This will also includes the bequest of a cloth to be painted and turned into two altar cloths for the church of Saint Mary Staining (though the clergyman has to pay for the painting of the cloth). In this context, the *kirtle* bequests look quite downscale, one is left to the housekeeper and one to a maid, the latter kirtle is made of russet, an inexpensive woollen cloth, and is only trimmed with lambswool.[5]

The context of the term's appearance in the second text type, sumptuary legislation, is the set of injunctions on the dress and ornamentation of female religious imposed by the Bishop of Lincoln in 1441 in response to accusations made against the (evidently) opulently dressed prioress. In an earlier examination she had confessed to wearing all except for cloth of Rennes, though she claimed that she only wore a fur cap "of estate" (status) on medical grounds. In the injunction the term appears thus:

> Also we enioyne yowe [. . .] that none of yow, the prioresse ne none of the couente, were no vayles of sylke ne no syluere pynnes ne no gyrdles herneysed wyth syluere or golde, ne no mo rynges on your fyngres then onne, ye that be professed by a bysshope, ne that none of yow use no lased **kyrtels**, but butonede or hole be fore, *ne* that ye use no lases a bowte your nekkes wyth crucyfixes or rynges hangyng by thayme.
>
> ['Also we enjoin you [. . .] that none of you – not the prioress nor any of the convent – wear veils of silk, nor silver pins, nor girdles decorated with silver or gold, nor any more rings on your fingers than one (you that are professed by a bishop), nor that any of you use any laced kirtles, but buttoned or closed at the front, nor that you use any laces about your necks with crucifixes or rings hanging on them'] (Chambers)

In this text type, as with the citations in the *LexP* database, the DRESS AND TEXTILES items appear to match the language of the text and we do not find evidence of code-switching. The context here suggests that *kirtle* is a more elevated item than is indicated in the wills: the other (forbidden) items include silk veils, silver pins, and rings and other items mentioned are accompanied by rich descriptors such as the "gyrdles herneysed wyth syluere or golde". This is also true of the kirtles which must not be laced but buttoned or not require any fastening.

In the third text type, the petitions, we find a reference to the term *kirtle* in a list of clothing, jewellery, accessories, and money in a petition for the king's pardon made on behalf of Agnes Balle who was being held in prison in London accused of stealing the items, and in Agnes Balle's royal pardon also issued in 1403. The petition is written in Anglo-French with an endorsement in Latin and the pardon in (heavily abbreviated) Latin:

---

[5] We know that russet is not considered a high-end textile from the sumptuary legislation of 1363 which states that "charetters, charuers, chaceours des charues, bovers, vachers, berchers, porchers, deyes et touz autres gardeins des bestes, batours des bledz et toutes maneres des gentz d'estat de garceon entendantz a husbonderie, et toutes autres gentz qe n'eient quarante solidees de biens, ne de chateux a la value de quarante soldz, ne preignent ne usent nul manere des draps sinoun blanket et russet."

> une toge de colo*ur* de blue furrez […] **kertill** de russet pric' de xij d. un smok p*ric*' de vj. d. un broche d'argent d'orez p*ric*' de xij d. vne burse pric'[…] p*ric*' ij d. deux kercheves pric' xij d. vn chaperon pric' xij d. & iij s. iiij d. en monoie
>
> ['an overgarment of blue cloth, a furred (lined or trimmed) […] kirtle of russet worth 12d. a smock worth 6d. a brooch of gilded silver with silver worth 12d., a purse worth […] worth 2d., two kerchiefs worth 12d. a hood worth 12d., and 3s. 4d. in money'] (Chambers)

> unam togam blodii coloris furratam cum cuniculis p*recii* […]x solido*rum* & octo denario*rum* unum **kyrtyll** de russeto p*re*cii duodecim denario*rum* unam camisiam p*re*cii sex denario*rum* unu*m* monile argenti de aurat' p*re*cii duodecim denario*rum* unam bursam p*re*cii qua*tu*or denario*rum* septem anulos auri p*re*cii vigniti solido*rum* unu*m* cordon' p*re*cii duo*rum* denario*rum* duo velamina p*re*cii duodecim denario*rum* unu*m* capucinum p*re*cii duodecim denario*rum* & tres solidos & qu*at*uor denarius in pecunia numerate
>
> ['a blue coloured tunic (*toga*) furred with rabbits, worth *6* shillings and eight pence, a kirtle of russet worth twelve pence, a chemise worth six pence, a gilded silver necklace worth twelve pence, a purse worth four pence, seven gold rings worth twenty shillings, a "cordon" worth two pence, two veils worth twelve pence, a hood worth twelve pence, and three shillings, four pence in money'] (Chambers)

Table 1: The dress and textile terms as found in Agnes Balle's petition and pardon

AF plea	L pardon
toge	togam
kertill	kyrtyll
smok	camisiam
	cordon'
deux kercheves	duo velamina
chaperon	capucinum

We may wonder why *smok*, a single-word switch or borrowing appears in the Anglo-French text and not *chemise* (or a variant): perhaps the switch to *kertill* inspired the continuation with *smok*. Note that in Middle English the meanings appear not to be entirely synonymous as in the following reference in the *MED*: "Hire chemise smal and hwit … and hire smoc hwit" (cited from Homilies in Cambridge, Trinity College B.14.52). All the other terms in the pardon are attested in Latin only or in all three languages.

The fourth text type, the romances, yield five occurrences of *kirtle* in *Emaré*, *Sir Landevale*, *Sir Launfal*, and *Sir Gawain and the Green Knight*; thus, it appears in four Middle English romances, two of which are translations (at varying distances) of an Anglo-French lay (Marie de France's *Lanval*). There are two contexts in which a *kirtle* is mentioned in *Emaré* and looking at them together seems to underline the wide range of connotations which may attach to this garment term:

> The lady that was gentyll and small
> In **kurtull** alone served yn hall,
>   Byfore that nobull kyng (ll. 391–393)
>
> ['The lady, who was graceful and slender, served in hall before the king wearing just her kyrtle']
>
> Tomorowe thou shall serve yn halle,
> In a **kurtyll** of ryche palle,
>   Byfore thys nobull kyng (ll. 847–849)
>
> ['Tomorrow you will serve in hall wearing a tunic of rich cloth before this noble king']

In the first example, "In kurtull alone" signals that Emaré is not dressed as a noble woman would be, but is perhaps dressed appropriately for someone who is serving. Her son, however, is going to be serving in the hall because Emaré wishes him to be seen: although he will wear a kirtle, it is made of a fine precious cloth. In *Sir Landevale*, which has been dated to the first half of the 14th century (Shepherd 1995), the two most beautiful women the hero has ever seen walk out of the woods, dressed as follows:

> **Kyrtyls** they had of purpyl sendell,
> Small i-lasid, syttyng welle (ll. 55–56)
> ['They had gowns of purple sendal tightly laced and beautifully fitted']

Sir Landevale is a close translation of the Anglo-French *Lanval*, a text dated to the 1160s (Laskaya and Salisbury 1995), and here "kyrtyls" translates the term *bliauz*. During the 12th century the term *bliaut* signifies an elaborate dress and a costly fabric in texts from England and France, and it continues to carry these senses in Britain into the fifteenth century. The word *bleaunt* was a native form of Old French *bliaut* and was popular in Middle English romances referring to the material or to garments generally worn by aristocratic characters. This suggests that "kyrtyl" was perhaps not the obvious choice of term for this poem. The later *Sir Launfal* of Thomas Chestre, which has been dated to the late 14th century (Laskaya and Salisbury 1995), employs the same term as that in the earlier version in English at this moment in the narrative:

> Har **kerteles** wer of Indesandel,
> Ylased smalle, jolif, and well –
> Ther myght noon gayer go (ll. 229–231)
>
> [‘Their gowns were made of indigo silk and were laced tightly and well. No one could have been more beautifully turned out’]

Here a context of wealth is indicated by the accompanying furs and jewels. The only occurrence of *kirtle* in *Sir Gawain and the Green Knight* (dated to the late 14th century; Tolkien and Gordon [1967] 1978: xiii]) comes when Lady Bertilak offers her girdle to Sir Gawain:

> “I schal gif yow my girdel, þat gaynes yow lasse.”
> Ho la3t a lace ly3tly þat leke vmbe hir sydez,
> Knit vpon hir **kyrtel** vnder þe clere mantyle (ll. 1829–1831)
>
> [‘“I shall give you my belt that is of less profit to you.” She lightly took hold of a cord that was tied around her waist, looped about her kirtle under the fine mantle’]

In giving the girdle Lady Bertilak must divest herself of it, and so Gawain is offered a glimpse of the kirtle that she wears beneath her mantle. Here *kirtle* seems to stand for a generalized item of clothing and a garment that is somewhat intimate.[6]

The other two terms that are shared across all four text types are partial synonyms, they are *fur* and *pelure*. The latter looks more specialized to modern eyes, but in the medieval period both terms signify fur (in the modern sense), although the term *pelure* appears to have both a broader semantic range (it can be used for clothing in general) and a narrower one (specifically budge).

As we might expect, given their presence in all four text types in our period, the *LexP* database shows that both *pelure* and *fur* appear in Latin, Anglo-French and Middle English (the noun form of *fur* is also attested in Old Scots and Welsh). The attestations come from all the relevant dictionaries and the texts are in the expected languages, except for the citation from the *Anglo-Norman Dictionary* (*AND*) which contains a number of code-switched items. In the text base the differing usages of the two terms across the wills are striking. There are references to *fur* (including derivations of the base forms) in nine wills. The earliest example occurs in the will of Thomas Harpham dated 1341 which has the following:

---

[6] In addition to “kyrtel”, the poem includes the term “bleaunt”, which appeared to stand in the place of kirtle in the Anglo-French lay. “He were a bleaunt of blew þat bradde to þe erþe” (l. 1928).

> meliorem supertunicam meam cum capucio ejusdem sectæ **fururatam**, nomine mortuarii. [...]. Item lego Agneti filiæ meæ tunicam meam de taune cum longo collobio et capucio **fururatis** ejusdem coloris
>
> ['my best surcoat with matching hood similarly furred, in the name of my death. [...] Likewise I leave my daughter Agnet my orange-brown tunic with a long shirt and furred hood of the same colour.']

As is evident from this example, the wills contain lists of garments trimmed with fur or sometimes, contrastingly, untrimmed. Across the nine wills, the contexts for mentions of *fur* include a number of morphological variants deriving from the different languages in which the term is attested. The attestations are as follows:

> meliorem supertunicam meam cum capucio ejusdem sectæ **fururatam**
> longo collobio et capucio **fururatis** ejusdem coloris
> unum courtby de blueto cum **furura**
> my **ffurre** of Calabre
> worstede Goune with þe **ffurre**
> my **furre** of Fycheux
> a gowne of blew worsted **furred** wit þrotes and polles of Martrons
> a gowne of gray russet **furred** wit Ionetis and wylde Catis
> a gowne of grene frese, in ward, &c, **furryd** with blak Lambe
> a **furre** of beuer and oter medled
> and I gown **furred** with gret menyvere
> a gown **furred** with Cristy gray
> a gown **furred** with Besshe
> all my gownes off clot*h* off gold and off sylke, with-outyn þe **ffurres**
> **furres** þat my lady my moder knowes off
> a blake goun **furred** v1 funes
> a newe **fure** of martirs
> all þe **ffurrurs** þat I haue, be sould
> a blake cote, **furryd**
> a gounne of russet, **furrit**
> a russet kyrtell **furred** with lambe
> myn purple goune **furred** with martrons
> my scarlet goune **furred** with martrouns
> a gowne of Blake, **furred** wit*h* ficheux
> gowne of Russet **furred** wit*h* blak
> a gowne of Rede Damaske **furred** wit*h* martrens
> gowne of Russet medley **furred** wit blak
> a sangwen' gowne **furred** wit*h* martres
> a scarlet gowne **furred** wit*h* foynes
> a gowne of scarlet wit*h* slyt slyues **y-furred**
> my best gowne of the lyuere of Skynners craft, bot*h* **furre** and cloth.
> my secunde best **furre** and gowne
> and clothis of silke, with-oute **ffurrereur**, euerychon

The variety of constructions in which variants of *fur* can appear seems striking: sometimes the garment featuring fur is mentioned (e.g. “longo collobio et capucio fururatis ejusdem coloris”), in some cases the animal from which the fur comes is named (e.g. “a gowne of gray russet furred wit Ionetis and wylde Catis”), sometimes even the body part of the animal is included in the descriptor (e.g. “a gowne of blew worsted furred wit þrotes and polles of Martrons”). These citations indicate that in this period the expression for “trimmed with *x* fur” frequently includes the description “fur” in the deverbal adjective indicating “trimmed” or “lined” rather than attaching it to the name of the donor animal (thus, “furred with *x*”). In other cases, however, the nominal is used, e.g. “furres þat my lady my moder knowes off”: this seems to designate fur garments. This range of possible usages presumably means that the term is likely to appear more frequently than any semantic equivalent.[7]

By contrast, *pelure* appears in only one will, that of Roger Flore or Flower Esq., dated 1424–1425. The will is written in Middle English with Latin title and closing formula.

> Also I wul þat my gownes for my body, þe whic*h* ben **ffurred** whit*h* **pelure**, be dalt amongis my childre, to ilke after here degre and age, so þat Thomas and Anneys haue four’ of þe best. And I wul þat þe remenaunt of my cloþes for my body be dalt amonges my seruauntes
>
> [‘Also I wish my gowns that I wear which are furred to be shared among my children, to each according to their rank and age, so that Thomas and Anneys have four of the best. And I wish the remainder of the clothes that I wear to be shared among my servants’]

Given its unusual context and usage (there are no other examples of the construction “furred with pelure” in the text base), this looks like a mistake: perhaps the testator intended the animal of origin to be indicated here. This idea is underscored by the fact that *pelure* appears almost exclusively in Anglo-French texts in the administrative text types.

Both terms appear in the sumptuary legislation in texts in Anglo-French, Middle English and Old Scots. In the Act of October 1363 (see Ormrod 2005), *furred* seems to denote simply “lined”. The most common formulation in the Parliamentary Petition of September 1402 is *furre de* (or variants of this): thus here the term seems to denote the actual fur of the animal. *Pelure* appears in four sets of documents and has such a wide reference that in the 1363 Act it means “clothes” in general rather than fur:

---

[7] Compare Present-day English which only allows expressions such as “lined/trimmed with mink” where the type of fur is specified.

> seignurs et chivalers, femmes povres et autres apparaile des dames, povres clercs **pellure** come le roi et autres seignurs
>
> ['lords and knights, poor and other women wear the dress of ladies, and poor clerks wear clothes like those of the king and other lords'] (Chambers)

The examples of *fur* in the petitions occur in the Agnes Ball documents. *Pelure* appears in a petition to the king by Robert de Montfort (c. 1322) for the reinstatement of a livery that had been granted for life but had fallen into arrears when the giver's estate was forfeit to the crown, and the giver executed. The petition indicates that the clothing should be commensurate with the petitioner's status.

> A no*str*e seign*eur* le roi si le plet & a son conseil prie Robert de Mountfort clerk' qe de ceo qe sire Beralmeu de Badlesmere qe ad forfez ceo qe en li est dever n*ost*re seign*eur* le roi le estote tenu par bone obligacion ensele de son sele en demy drap de colur couenable pur clerk'destat oveqes la **pelure** pur surcot clos & pur surcot overt oveqes un chap*er*on de menuueyr
>
> ['Robert de Mountfort, clerk, prays our lord the king, if it pleases him, and his council, that whereas Lord Bartholomew de Badlesmere, who had forfeited that which our Lord the King gave him through good obligation sealed with his seal – one half cloth of colour suitable for a clerk's estate, with the material for a "closed" surcoat and for an "open" surcoat, and with a hood of miniver'] (adapted from Chambers' translation)

Here *pelure* should be translated "material" for two surcoats suggesting that in parliamentary Anglo-French the term had the more general sense as well as its specific (though perhaps not very specific) meaning "fur".[8] The second example comes from a petition by Thomas and Margaret de Beauchamp between 1394 and 1399 asking the king to make good on a grant he had apparently made to the couple that included items of bedding and clothing:

> et autre de v*ost*re g*ra*ciouse s*eignour*ie graunt al dit Margarete sa vesture **pellure** & tout l'aparaill & biens p*ur* son corps & sa chambre
>
> ['and also by your gracious lordship grant the said Margaret her clothing, fur and all that belongs to them and goods for her person and her chamber'] (Chambers)

Although the petition contains a number of terms in Middle English, the part relating to Margaret's clothing appears to be entirely in Anglo-French, with "pelure" here signifying "fur" (Sylvester, Chambers, and Owen-Crocker 2014).

---

[8] There has been some discussion about this among the compilers of the text base: it seems possible that in some contexts *pelure* means "material". In the translation of the Act of October 1363, adapted from the translation by Mark Ormrod (2005), "Edward III, Parliament of October 1363", the term is translated as "fur/clothing". In the petition to the king by Robert de Montfort, the term was originally translated as "material" because this seemed to be the sense here and because it is used in this way in the *Dialogues of Saint Gregory* (c. 1212) in the *AND*; however, another member of the editorial team believes that the term *pelure* only signifies fur.

Fur and *pelure* occur in the same two romances, *Sir Launfal* and *Sir Gawain and the Green Knight*. In the former we find:

> Har manteles wer of grene felvet,
> Ybordured wyth gold, ryght well ysette,
> **Ypelured** wyth grys and gro (ll. 232–234)
>
> ['Their cloaks were made of green velvet bordered with gold beautifully adorned, and edged with grey and white fur']
>
> Her matyll was **furryd** wyth whyt ermyn,
> Yreversyd jolyf and fyn –
> No rychere be ne might (ll. 946–948)
>
> ['Her cloak was trimmed with white ermine with the edges turned back, prettily decorated. It was impossible that there could be a costlier one']

In both instances the meaning seems to be "trimmed with" or "lined with". In *Sir Gawain and the Green Knight*, we find the following examples:

> A meré mantile abof, mensked withinne
> With **pelure** pured apert, þe pane ful clene
> With blyþe blaunner ful bry3t, and his hod boþe (ll. 153–155)
>
> ['a fine mantle on top, adorned inside with a trimmed fur lining in full view, the elegant trimming with beautiful ermine, all white, and his hood too']
>
> And þenne a meré mantyle watz on þat mon cast
> Of a broun bleeaunt, enbrauded ful ryche
> And fayre **furred** wythinne with fellez of þe best,
> Alle of ermyn in erde, his hode of þe same (ll. 878–881)
>
> ['a fine mantle was thrown over him of a rich brown fabric, richly embroidered and beautifully lined with fur, with the best pelts, the best ermine in the world with a hood to match']
>
> Bot þe lady for luf let not to slepe,
> Ne þe purpose to payre þat py3t in hir hert,
> Bot ros hir vp radly, rayked hir þeder
> In a mery mantyle, mete to þe erþe,
> Þat watz **furred** ful fyne with fellez wel pured (ll. 1733–1737)
>
> ['But the lady could not sleep for love, nor could the intention fixed in her heart falter, so she got up quickly, took herself there in a fine mantle that reached the ground and was richly lined with trimmed pelts']
>
> He were a bleaunt of blwe þat bradde to þe erþe,
> His surkot semed hym wel þat softe watz **forred**,
> And his hode of þat ilke henged on his schulder,
> Blande al of blaunner were boþe al aboute (ll. 1928–1931)
>
> ['He wore a long blue over-tunic that reached to the ground, his surcoat was softly furred and suited him well and his hood, made to match, hung on his shoulder. Both were trimmed with ermine']

> His cote wyth þe conysaunce of þe clere werkez
> Ennurned vpon veluet, vertuus stonez
> Aboute beten and bounden, enbrauded semez,
> And fayre **furred** withinne wyth fayre **pelures** (ll. 2026–2029)
>
> [‘his tunic with the beautiful embroidered emblem, embellished on the velvet with jewels with special powers, set with gold and adorned with embroidery along the seams, and gorgeously lined inside with beautiful fur’]

Here we may note that the requirements of the pattern of alliteration would suggest the use of a term other than “fur” in the second half in line 2029. In *Sir Launfal* the Middle English allows the use of *pelure* as a verb, and in *Sir Gawain and the Green Knight* with the deverbal adjective *pured* so that mantles can be described as furred with a specific kind of fur and trimmed with fur respectively. It seems as if it is only the differing grammatical constructions which the two terms are allowed in the different languages which keep them apart as well as appearing to influence their frequency of use.

3 Dress and textile terms appearing in only one text type

The question of the possible link between specificity of reference and code choice requires examination of terms which are at the other end of the scale, in terms of their distribution. The text base includes a large number of terms which appear in only one of the four text types. For this study, I focus on terms falling into two specific semantic sub-groups: FUR (containing names of particular furs to set against the two general terms for fur which appear in all four text types), and TEXTILES (containing names of particular textiles). Classified by text type these yielded six types of fur in the wills (though one phrase had one element which appeared in a number of other text types); six from the sumptuary laws; and one each from the petitions and the romances. The textile names group offered eight terms that were unique to the wills; sixteen in the sumptuary laws (including six phrases in which one element appears in other text types); one from the petitions; and seven from the romances. The information given about the language attribution in the *LexP* database was checked for each of the terms, and this was compared with the language of the text in the text base in which the term appeared. Of the *fur* terms, only four were attributed by the *LexP* database to a single language, one in each text type grouping (see Table 2). The other terms all appear in texts in at least two, and in many cases three of the three languages under consideration. Checking the records collected in the *LexP* database confirmed that there were dictionary citations for the terms in

Table 2: Fur terms unique to specific text types

Wills	**Sumptuary laws**	**Petitions**	**Romances**
beuer al: AF, ME, L; l: ME	*bugee* al: AF, L, ME; l: AF	*cuniculis* al: L; l: L	*blaunner* al: ME; l: ME
calabre al: L, ME; l: ME	*chat* al: AF; l: AF		
ficheux, fycheux al: AF, ME; l: ME	*conil/conyng* al: AF, L, ME; l: AF		
*gret menyvere** al: AF, L, ME; l: ME	*gopil/gupille* al: AF, L; l: AF		
Ionetis al: AF, L, ME; l: ME	*letuce, letueses, letuse, letuses, letyce* al: AF, L, ME; l: AF, ME		
wylde catis al: ME; l: ME	*sabelo, sables sable* al: AF, L, ME; l: L, ME, AF		

* = terms which appear in more than one text type but which feature in compounds or phrases that appear in only one text type
al = language(s) in which the term is attested
l = language(s) of the text(s) in which the term appears in the text base

the language suggested by the particular historical dictionary in which the attestations appeared, i.e. the *MED* included references where the language of the citation was Middle English rather than, say, Latin.[9]

The results obtained for the textile names are very similar (see Table 3): for the wills, all but three terms were attested in at least two languages and of those three the base language of the texts in which the terms appeared match the language to which the terms are attributed. The petitions yielded only one term unique to this text type, namely *kerseys*. It is attested in Latin, Anglo-French and Middle English and appears in a text whose base language is Anglo-French. The *AND* labels the term "M.E." and the *OED* suggests that its origin may relate to the eponymous Suffolk. Above the definitions, the *OED* includes four citations, the first in Latin and the next three in Anglo-French and the *LexP* database includes a number of citations from *The Parliament Rolls of Medieval England* in Anglo-French. All but one of the terms unique to the romances are attested in at least two languages, and the exception is a term which appears in a text whose base language matches that of its attribution. The terms unique to the sumptuary laws are again similar: only two of the sixteen lexical items are

---

[9] As is now quite well known, this is not always the case and the historical dictionaries differ in their policies. An example is provided by the citations gathered for *ionetis* which include, from the *MED*, "iij ourles de genet' pro mantell' Regine" and "A gowne of gray russet furred wit Ionetis and wylde Catis".

Table 3: Textile terms unique to specific text types

Wills	**Sumptuary laws**	**Petitions**	**Romances**
camelino, kamelyn al: AF, ME, L; l: L	*bustian* al: AF, L, ME; l: ME, AF	*kerseys* al: AF, ME, L; l: AF	*alexandrine* al: AF, ME; l: AF, ME
canuas al: AF, L, ME; l: ME	*chamelet* al: AF, L, ME; l: ME		*bys* al: AF, L, ME; l: ME
carde al: L, ME, AF; l: ME	*d'aignel, dagnell (= agnel)* al: L, AF; l: AF [note that English *lamb* appears in two text types, wills and sumptuary laws]		*chesil* al: AF, L, ME; l: AF
crosse werk al: ME; l: ME	*fustian* de Napuls* al: AF, L, ME, OSc (*fustian*); l: AF		*cyprus* al: AF, ME; l: ME
frees al: AF, ME, L; l: ME	*fustian* of napuls* al: AF, L, ME, OSc (*fustian*); l: ME		*indesandel* al: L, AF, ME; l: ME (*sendal* L, AF, ME; *inde* L, AF, ME)
musterdevylers al: ME; l: ME	*laine* al: L, AF; l: AF		*paile* al: L, ME, AF; l: ME, AF
serico al: L; l: L	*lawn* al: AF, ME, OSc; l: OSc, ME, AF		*tolouse* al: ME; l: ME
worsted al: ME, AF, L; l: ME	*lienge tele* al: AF; l: AF		
	sarcenet al: L, ME; l: ME		
	scarlet cloth engreyned engrain* al: ME, AF l: ME (*scarlet* = L, AF, ME)		
	scarlet drap en graine en graine* al: AF, ME; l: AF (*scarlet* = L, AF, ME) l: AF (*scarlet* = L, AF, ME)		
	soie al: L, ME, AF; l: AF		
	umple al: ME; l: ME, AF		
	velwetmotley al: AF, ME (both parts); l: AF		
	velewet uppon velewet* al: ME, L, AF; l: ME		
	velewet sur velewet* al: L, AF, ME; l: AF		

* = terms which appear in more than one text type but which feature in compounds or phrases that appear in only one text type)
al = language(s) in which the term is attested
l = language(s) of the text(s) in which the term appears in the text base

attested in only one language, and in one case, the base language of the text matches the language of attestation. There is one case, *umple*, which looks like a switched or borrowed term.

The one example of a term where the language of the text in which the term appears does not map onto the language(s) of attestation derived from the *LexP* database is the term *umple*. This term is listed as attested in Middle English only according to the *LexP* database because it is not included in the *AND* as a headword though it does appear in two citations and is evidently in Anglo-French documents cited in the *MED*. The sumptuary laws in which the term appears are the Act of 1463 which is in Middle English and the Anglo-French version drawn from it: assuming that the scribe was competent in both languages, it seems possible to conclude that the importation of a Middle English term into an Anglo-French document is a switch.

In terms of the period in which we find all this vocabulary shared across Latin, Anglo-French and Middle English, this seems to vary according to the text type. The "fur" terms in the wills are all attested in the first half of the 15th century, as is the one example from the petitions. The one *fur* term unique to the romance text type dates from the late 14th century, as do all but one of the fur terms in the sumptuary laws. Most of the latter appear in the Act of 1363, though many continue in use into the mid-15th century. The dates of usage of the textile terms unique to specific text types also seem to vary according to the text type. The two earliest examples from wills are attested in the mid-13th and mid-14th centuries, the others all date from the late 14th and early 15th centuries. The textile terms unique to the sumptuary laws almost all date from the mid-15th century. The one term from the petitions is from the late 14th century, and terms unique to the romances are generally earlier than we might expect, some dating from the 1160s and only one later than the 14th century. The terms with the most specific reference in this sample, then, are mostly shared across a number of languages but the data here suggests that this process which has been termed merging or the production of a mixed language, but about which we need to discover more, differed according to text type. The textual culture of the romances seems to be the forerunner here, though more data is needed to test this idea properly.

4 Conclusion

The focus of studies of multilingualism in medieval Britain has been on the phenomenon of code-switching. The lexical data on which this discussion has focused offers a view of a different kind of multilingualism in play in this period.

The gathering together of citations of lexical items from the semantic field of DRESS AND TEXTILES from a range of text types examined against the data in the *LexP* database has allowed us to see how much of the content-laden vocabulary was shared by Anglo-French, medieval Latin and Middle English. This supports Tony Hunt's (2011) claim that as far as lexis is concerned, attempting to draw clear distinctions of language identity are futile, since such distinctions were certainly blurred in medieval England:

> Mutual borrowing and the idea of a shared vocabulary are immediately evident from a glance at the D[ictionary of]M[edieval]L[atin from]B[ritish]S[ources], the MED, and the AND. The dictionaries include hundreds of shared terms, so attention to their occurrence in all three languages is necessary for a full understanding of any of these lexical items. (2011: 63–64)

The data and discussion offered here may seem to point to the suggestion made by Trotter (2011: 157) that it would be wrong to assume that what are now perceived as separate languages were regarded as separate or even separable in the Middle Ages. These ideas are echoed in recent discussions in contemporary sociolinguistics on languaging and repertoire. Jørgensen et al. (2011), for example, introduce their discussion of what they term "polylanguaging" as follows:

> It is a widely held view that language as a human phenomenon can be separated into different "languages", such as "Russian", "Latin", and "Greenlandic". [...] the recently developed sociolinguistic understanding [is] that this view of language can not be upheld on the basis of linguistic criteria. "Languages" are sociocultural abstractions which match real-life use of language poorly. (2011: 23)

Blommaert and Rampton (2011: 3–4) observe that there is now a sizeable body of research on ideologies of language that "denaturalizes the idea that there are distinct languages" and suggests that named languages such as English or German are "ideological constructions" (see also e.g. Makoni and Pennycook 2007; Blackledge and Creese 2010). These and similar discussions, however, relate to speakers and not writers. The data that medievalists have is written, and is in general not accidentally scribbled down texts, but documents and literary writing, possibly the most consciously composed genres that exist. We need to be cognizant of what that means for this period. I would argue, however, that while the multilingual resource offered by the database of the *LexP* indicates that many words are attested across more than one language in the writings of medieval England, we do not have evidence that medieval speakers and writers regarded all languages as one repertoire or that they did not know to which language certain words belonged. Rather, we witness the creation of something approaching a technolect in which nouns, in particular, were not ported across linguistic barriers, but could be used because they would be readily understood by users and because they had accreted both a precision (making them worth using) and a universality (making them easily comprehensible). These deployments do not appear to be the same as borrowings, in that the terms seem quickly to put down roots in the non-origin language, nor are they exactly code-switches even though it may look as if the terms are filling gaps perceived by speakers used to working with more than one language. Trotter's work has shown that in many cases of what have been labelled code-switches, there is a perfectly good word in the lexicon of the base language of the text, and this is borne out by studies of modern code-switching behaviour (see e.g. Gal 1988; Myers-Scotton 1988; Romaine 1989; Gardner-Chloros 2009). It does not appear to be their foreignness that forms any part of the decision to include these terms in the texts.

Investigating the distribution of the vocabulary from one semantic field, here DRESS AND TEXTILES, within a range of texts across different text types offers a rationale for assessing the presence of technical lexis that avoids the dependence on pre-defined genres for a definition of technicality. It seems that the creation of a technical lexis in this area is subject to the competing aims of precision and functionality, and that a technical lexis developed in the field of DRESS AND TEXTILES in text types such as the sumptuary laws and the petitions (which were likely to appear in versions in both Middle English and Anglo-French), and the romances (the Middle English versions of which were in many cases translations from Anglo-French). In the sumptuary legislation documents of the mid-15th century, it is noteworthy that the two texts, one in Middle English and one in Anglo-French, mirror each other very closely. It is clear that more work needs to be done on the distribution of vocabulary across text types in order to gain a more precise sense of how technical terminology was developed and used within the multilingual textual culture. Attention also needs to be paid to the levels of the semantic hierarchy at which lexical items appear in the different text types. Both of these approaches offer information vital to our understanding of the relationships between specificity/breadth of meaning, code choice and text type in the medieval period.

References

AND = Anglo-Norman Dictionary. http://www.anglo-norman.net/.

Blackledge, Adrian & Angela Creese. 2010. *Multilingualism: A critical perspective*. London: Continuum.

Blommaert, Jan & Ben Rampton. 2011. Language and superdiversity. *Diversities* 13(2). 1–20.

Bullock, Barbara E. & Almeida Jacqueline Toribio. 2009. Themes in the study of code-switching. In Barbara E. Bullock & Almeida Jacqueline Toribio (eds.), *The Cambridge handbook of linguistic code-switching*, 1–17. Cambridge: Cambridge University Press.

Davidson, Mary Catherine. 2005. Discourse features of code-switching in legal reports in late-medieval England. In Janne Skaffari, Matti Peikola, Ruth Carroll, Risto Hiltunen & Brita Wårvik (eds.), *Opening windows on texts and discourses of the past*, 343–351. Amsterdam & Philadelphia: John Benjamins.

Dodd, Gwilym. 2012. Trilingualism in the medieval English bureaucracy: The use – and disuse – of languages in the fifteenth-century privy seal office. *Journal of British Studies* 51(2). 253–283.

Gal, Susan. 1988. The political economy of code choice. In Monica Heller (ed.), *Codeswitching: Anthropological and sociolinguistic perspectives*, 245–264. Berlin: Mouton de Gruyter.

Gardner-Chloros, Penelope. 2009. *Code-switching*. Cambridge: Cambridge University Press.

Hunt, Tony. 2000. Code-switching in medical texts. In D. A. Trotter (ed.), *Multilingualism in later medieval Britain*, 131–147. Cambridge: D. S. Brewer.

Hunt, Tony. 2011. The languages of medieval England. In Michael Baldzuhn & Christine Putzo (eds.), *Mehrsprachigkeit im Mittelalter*, 50–68. Berlin & New York: Walter de Gruyter.

Jefferson, Lisa. 2000. Language and vocabulary of the fourteenth- and early fifteenth-century records of the Goldsmiths' Company. In D. A. Trotter (ed.), *Multilingualism in later medieval Britain*, 175–211. Cambridge: D. S. Brewer.

Jørgensen, J. N., M. S. Karrebæk, L. M. Madsen & J. S. Møller. 2011. Polylanguaging in superdiversity. *Diversities* 13(2). 23–37.

Laskaya, Anne & Eve Salisbury (eds.). 1995. *The Middle English Breton lays*. TEAMS edn. Kalamazoo, MI: Medieval Institute Publications.

LexP = Lexis of Cloth and Clothing in Britain Project. http://lexisproject.arts.manchester.ac.uk/.

Makoni, Sinfree & Alastair Pennycook. 2007. *Disinventing and reconstituting languages*. Cleveland: Multilingual Matters.

MED = Middle English Dictionary. http://quod.lib.umich.edu/m/med/.

Myers-Scotton, Carol. 1988. Code switching as indexical of social negotiations. In Monica Heller (ed.), *Codeswitching: Anthropological and sociolinguistic perspectives*, 151–186. Berlin: Mouton de Gruyter.

OED = Oxford English Dictionary. http://www.oed.com/.

Ormrod, W. M. (ed.). 2005. Edward III: Parliament of October 1363, text and translation. In Chris Given-Wilson, P. Brand, A. Curry, R. E. Horrox, G. Martin, W. M. Ormrod, & J. R. S. Phillips (eds.), *The Parliament rolls of medieval England, 1275–1504*. Scholarly Digital Editions. http://www.sd-editions.com/PROME/.

Romaine, Suzanne. 1989. *Bilingualism*. Oxford: Blackwell.

Rothwell, William. 2000. Aspects of lexical and morphosyntactical mixing in the languages of medieval England. In D. A. Trotter (ed.), *Multilingualism in later medieval Britain*, 213–232. Cambridge: D. S. Brewer.

Schendl, Herbert. 2000. Linguistic aspects of code-switching in medieval English texts. In D. A. Trotter (ed.). *Multilingualism in later medieval Britain*, 77–92. Cambridge: D. S. Brewer.

Schendl, Herbert. 2012. Literacy, multilingualism and code-switching in early English written texts. In Mark Sebba, Shahrzad Mahootian & Carla Jonsson (eds.), *Language mixing and code-switching in writing: Approaches to mixed-language written discourse*, 27–43. London: Routledge.

Schendl, Herbert. 2013. Multilingualism and code-switching as mechanisms of contact-induced lexical change in Middle English. In Daniel Schreier & Marianne Hundt (eds.), *English as a contact language*, 41–57, notes 331–333. Cambridge: Cambridge University Press.

Schendl, Herbert & Laura Wright. 2011. Code-switching in early English: Historical background and methodological and theoretical issues. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English*, 15–45. Berlin & Boston: Mouton de Gruyter.

Schendl, Herbert & Laura Wright (eds.). 2011. *Code-switching in early English*. Berlin & Boston: Mouton de Gruyter.

Shepherd, Stephen H. A. (ed.). 1995. *Middle English romances*. New York & London: Norton.

Sylvester, Louise. 2011. Linguistic choices relating to dress and textiles across different text types produced in medieval Britain. In Emma Volodarskaya & Jane Roberts (eds.), *The second international conference 'Language, Culture and Society in Russian/English Studies', 25–26 July, 2011: The proceedings*, 289–293. London: University of London.

Sylvester, Louise. 2016. The language of textiles in Anglo-Saxon wills. In Maren Clegg Hyer & Jill Frederick (eds.), *Text, intertext, textiles: Essays in honor of Gale R. Owen-Crocker*, 27–41. Woodbridge: Boydell.

Sylvester, Louise M., Mark C. Chambers & Gale R. Owen-Crocker (eds.). 2014. *Medieval dress and textiles in Britain: A multilingual sourcebook*. Woodbridge: Boydell.

Tolkien, J. R. R. & E. V. Gordon (eds.), Norman Davis (rev.). [1967] 1978. *Sir Gawain and the green knight*. Oxford: Clarendon Press.

Trotter, David. 2006. Language contact, multilingualism, and the evidence problem. In Ursula Schaefer (ed.), *The beginnings of standardization: Language and culture in fourteenth-century England*, 73–90. Frankfurt: Peter Lang.

Trotter, David. 2010. Language labels, language change, and lexis. In Christopher Kleinhenz & Keith Busby (eds.), *Medieval multilingualism: The francophone world and its neighbours*, 43–61. Turnhout: Brepols.

Trotter, David. 2011. Death, taxes and property: Some code-switching evidence from Dover, Southampton, and York. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English*, 155–189. Berlin & Boston: Mouton de Gruyter.

Wright, Laura. 1995. A hypothesis on the structure of macaronic business writing. In Jacek Fisiak (ed.), *Medieval dialectology*, 309–321. Berlin & New York: Mouton de Gruyter.

Wright, Laura. 2011. On variation in medieval mixed-language business writing. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English*, 191–218. Berlin & Boston: Mouton de Gruyter.

Simon Meecham-Jones

6 Code-switching and contact influence in Middle English manuscripts from the Welsh Penumbra – Should we re-interpret the evidence from *Sir Gawain and the Green Knight*?

1 Code-switching and Medieval Welsh: An exception amongst European medieval linguistic practices?

The recognition of historical code-switching as a significant research topic in its own right has caused an appraisal of the concurrent use of multiple linguistic systems in medieval Western Europe. This concern with the features and practices resulting from language contact and medieval multilingualism has encouraged critics to reimagine how medieval writers thought about language and the distinctions between languages, and the extensive surviving corpus of code-switching in medieval texts has played a crucial role in this attempt at rethinking how medieval writers understood the concept of language(s). It is not just that code-switching is considered as being more prevalent and prominent in a significant variety of medieval literary works and texts of record (and probably more various in the forms in which it is found) than is generally found in later written sources. More importantly, switches of language which earlier critics had often adduced as evidence of the limitations and fallibilities of medieval scribes are increasingly recognised as capable of being read as sophisticated expressions of scribal volition, creating pragmatically significant distinctions within the text, which may inspire and deserve interpretation.

Within this developing understanding of medieval linguistic contact across a wide range of European languages, the generally accepted lack of such evidence from sources showing contact between written English and written Welsh in the pre-modern era seems ever more anomalous. Should this apparent absence be interpreted as proof that contact between these two languages was governed by a set of rules different from those manifest in, for example, the contact of medieval English with Latin and continental and insular forms of French? If countenanced, such a conclusion would seem to question how far the patterns

DOI 10.1515/9781501504945-006

observable in other areas of medieval code-switching can be understood as demonstrating general principles about language contact rather than specific, though interesting, pragmatic examples.[1] Alternatively, is it possible that a combination of scholarly and ideological factors have led to such evidence as there is of medieval English/Welsh contact having been obscured?

Certainly, the work of the antiquarians and scholars who began the systematic processes of recovering earlier English material in the eighteenth and nineteenth centuries, from Warton and Ritson to Furnivall and Skeat, often reveals an influence inspired by contemporary ideas of nation and national character, which were to be redeemed and validated by the development of a vernacular corpus which paralleled and supported the development of the national state. Hanna (2008) describes the influences as balancing distinct forms of political and cultural "anxiety":

> But the EETS editors, as any cursory survey of prefaces to their editions will copiously demonstrate, engaged in the project out of an interest in national glory and devotion to English culture. Caught between, on the one hand, a belief in the retarding effects on "English" nationhood exerted by the Norman Yoke and, on the other, a considerable anxiety about the glorious national past unearthed by German philological scholarship, they saw Englishness (or English-only) projectively as the foundation upon which coming Elizabethan glory would rest (Hanna 2008: 132).

Such cultural anxieties are no less visible in the works of French, German and Italian critics and philologists. The inclusion of materials from other linguistic systems therefore seemed archaic (in delaying the "inevitable" rise of the vernacular), better understood as an example of scribal error, through which copyists "myswrote"[2] the texts in front of them.

Once the presumption of the "naturalness" of the desire to create only monolingual texts is questioned, critics and readers are forced to engage with the different priorities and procedures inherent in the act of writing within complex (and necessarily contingent) conditions of multilingualism. The medieval corpus of mixed language texts offers a subject for study of singular interest, since the range of languages sampled extends beyond the Romance continuum of Late Latin derived codes, with attendant problems of how to ascertain whether emergent forms represent Late Latin and early vernacular creations or the absorption through the Romance tradition of words derived from other linguistic systems. Given the greater syntactical and phonological distance between Old English and some of the forms it uses and absorbs from other language-systems in contact-derived phenomena, the probability of error in selecting between closely similar (and closely related) forms is reduced, and it seems more certain that scribes consciously chose to draw attention to their bilingual or multilingual capabilities. It is not surprising, then, that code-switching has been implicated as a probable major contributing factor to the enormous lexical expansion of Middle English.

---

[1] See for example Muysken (2000), discussed by Gardner-Chloros (2009).

[2] "And for ther is so greet diversitee / In English and in wryting of our tonge, / So preye I god that noon miswryte thee, / Ne thee mismetre for defaute of tonge." (Geoffrey Chaucer, *Troilus and Criseyde* book V, 1793–1796. Text from *The Riverside Chaucer*, Benson 1987.)

For Hanna, thinking about the difficulty of writing a monolingual English text in the thirteenth century, the practice of code-switching reflects both the difficulty and the relative unimportance (to the medieval scribes themselves) of distinguishing between items which modern editors might seek to mark as "English", "Latin" and "French" (whether continental or insular), noting that, "just as English appears within Latin in the Lambeth 260 exemplum book, Latin is not totally foreign to the in-column presentation of the English texts of the MS" (Hanna 2008: 132).

Hanna (2008: 132) diagnoses the scribe(s) as traversing "a particularly fluid and perturbed linguistic frontier, one commonplace in the English Middle Ages but largely invisible, both to those who concern themselves only with metropolitan canonical poetic texts (read 'Chaucer') and to those who investigate only texts in carefully edited printed editions". This apparent lack of fear in delineating, and, at times, crossing that boundary has paradoxically been masked in modern scholarly editions by the discomfort of later critics who, in Hanna's (2008: 162) view, "deliberately sanitized to reduce, if not obliterate, signs of polylingual usage".

Where Hanna (2008) describes a porous linguistic border, Butterfield (2009) goes further in seeking to deny the discreteness of medieval British linguistic systems which, in her view, had been encapsulated in the metaphorical imagining of languages as if nation states. But Butterfield's interest lies less in general principles of linguistic change, and more on the specifics which characterise the relationship of Middle English with insular and continental French. In drawing attention to the immensity of the debt she believes Middle English poetic culture owes to Francophone culture, she collapses the distinction between them:

> Sharing a language and a literature crucially confuses a clear sense of distinction between English and French. The world and work of words from the period is thickly coloured by histories of meaning, speaking, and writing that are both English *and* French, that is homely *and* foreign (Butterfield 2009: xxi).

Her conclusion leads her close to seeming to question the existence of Middle English, arguing for the essential unity of Middle English and the variant forms of (regionalised) French:

> In order to think through the nature of the Anglo-French relationship we need to take account of this vital fact, that the two cultures shared a language for four centuries, indeed, one might argue, that they share it still (Butterfield 2009: xx).

The distinction she makes is an interesting one: two *cultures* sharing one language rather than one culture, partially identified and partially distinguished by the use of two languages. If one accepts the model of two cultures sharing a language, the language in question should be considered French not English so it is scarcely surprising that Butterfield posits that Middle English

> could be defined precisely as a form of French. For writers and readers of both languages (who were likely to know Latin as well) 'English' is not therefore a single concept that works merely in polarity with French; it contains and is contained by French in a subtle, constantly changing, and occasionally antagonistic process of accommodation (Butterfield 2009: 99).

There is unquestionably an element of rhetorical over-statement in her formulation, which in its attempt to synthesise an eye-catching conclusion makes no concession either to the changing nature of Middle English diachronically from c. 1100 to 1400 or the gradually declining role of insular French as a vernacular in England in that period. It scarcely seems likely that fourteenth-century speakers of versions from more distant corners of "Francophonie" such as Savoy or Hainault would have identified Middle English as an alternative dialect of French. Similarly, Rothwell (e.g. 2000, 2001, 2004) has demonstrated the uneven usage of French in medieval England, with its concentration in particular areas (and presumed absence elsewhere) which must have contributed to the extensive dialect variation encountered in fourteenth-century Middle English sources. In contrast, Chaucer's familiarity and engagement with French literary tradition (which forms the starting point of Butterfield's argument) was exceptionally broad and deep, and requires to be considered a wholly untypical response. It can scarcely be matched even in the work of John Gower (though he also knew the French tradition intimately and wrote elegantly in Anglo-Norman), and must be considered to be of a different order of magnitude from that displayed in other English poetic traditions of this period. The evidence from *Piers Plowman* (from the West Midlands), or that from the widely-circulated poem *Prik of Conscience* (from the North of England), does not support Butterfield's theory of English literary culture being decisively influenced by a total immersion in French culture. Nonetheless, Butterfield's interpretation of the literary evidence, like Hanna's more restricted reading, is revealing in its focus on the lack of stability in the boundaries and conventions of Middle English. In both circumstances, extensive code-switching proves to be an important consequence of the porous and diffidently policed performance of the language, albeit a consequence which inevitably exacerbates that changeability. The work of Townend (2002), Dance (2011, 2014), Poussa (1982) and others (see e.g. Bibire 2001; Bjorkman 1900, 1902; Burnley 1992) on the absorption of Scandinavian elements into Middle English has highlighted another set of circumstances in which developing Middle English proved highly permeable to the assumption of many features from another linguistic system, though the thinness of the surviving manuscript record makes it impossible to determine how extensive a role code-switching played in the process.

Yet French (in its various forms), Latin and Old Norse/Danish were not the only languages with which Middle English was in contact. From the Norman conquest of England to the Battle of Bosworth in 1485 (and unquestionably before that, also, and after, up to the present day) life in medieval Wales and the Welsh March was characterised by the frequency and inevitability of language contact between English and Welsh. The conquest of Wales was a long-drawn out process, from the early Norman land seizures of the 1070s to the formal declaration of the annexation of all Welsh territory in the Statute of Rhuddlan in 1284. The Normans' favoured practice was to identify the most fertile agricultural land – until the Edwardian conquest, the wilder and less fertile regions were ignored, and territory outside English control was classified as "pura Wallia". To protect the land they had seized, the new owners built castles, and these defensive settlements were populated by migrant English speakers who alone had the right to live within the castle walls, while the towns were serviced by Welsh-speaking servants and agricultural workers who were obliged in live in designated "Welshries" outside the city walls. In practice, the management of the large and lucrative Marcher land-holdings was made possible primarily through the efforts of an emerging Welsh gentry class, which simultaneously managed the land while continuing to self-identify as Welsh. One famous example of what might seem to modern observers an apparent contradiction was the conduct of the Vaughan family of Hergest, who served with distinction in a variety of government functions, and fought for their (English) sovereigns in France and in the Wars of the Roses.[3] Nonetheless, they also took pains to maintain their role in Welsh culture, and are recorded as active patrons of bards and musicians.[4] Nor were commercial and cultural links confined to those in positions of power and influence. The enormous financial rewards derived from the colonial adventure flowed through the crossing points of the major border towns – Shrewsbury, Chester and Ludlow among them – ensuring a constant supply of Welsh-speakers in contact with English speakers – farmers and drovers selling produce, Welsh mercenaries to fight in France (Chapman 2015) and the English civil wars. It should not be forgotten, also, that the practice of Welsh law continued. Though the Welsh legal system derived from jurisprudential principles wholly distinct from English Common Law, its use was frequently invoked by English landowners and citizens in Wales and the March when they felt its application would benefit their interests. All these factors served to complicate what might have appeared to be a linguistically-segregated pattern of settlement within Wales, which, allied to the crucial economic importance of the March controlling the Welsh economy means that all of medieval Wales and the Marcher Lordships, and perhaps even some distance beyond that, should be considered the "Welsh Penumbra" – that is, an area where the English and Welsh languages were in daily and intimate contact – in this case over a period of several hundred years.

---

[3] Griffiths (2004).

[4] "The heir, Watkin Vaughan, maintained the tradition which made Hergest a resort for the greatest Welsh bards of the fifteenth century. For three generations Welsh culture found a home at Hergest. There were preserved the 'Red Book of Hergest,' which is now at the Bodleian Library, Oxford, and the 'White Book of Hergest,' the collection of Welsh prose and verse (believed to have been largely transcribed by Lewis Glyn Cothi) which was lost in the Covent Garden fire of 1808." Jones (1959).

It is striking, though, that code-switching from Welsh into English is absent from Schendl and Wright's (2011) summary of the practice in medieval Britain:

> During the medieval period, written code-switching (at first between Medieval Latin and Old English, and then after 1066 between those languages and Anglo-Norman) was both a normal phenomenon within the British Isles, and Europe-wide (Schendl and Wright 2011: 33).

In the field of code-switching, as in so many other areas of the development of English, it seems that one major site of language contact has not yet been accommodated within the prevailing discourse.[5] When the length of time of sustained contact between the two languages is considered, it may seem the more puzzling that linguistic historians have been content to accept that this contact, like the contact between Brythonic/Welsh and Old English over more than five hundred years, has left no traceable evidence. Probably the most widely used history of the English language concedes that this lack of evidence is surprising:

---

[5] There has been considerable research considering how far it is possible to distinguish between code-switching, "borrowing" and other contact-derived phenomena (see e.g. Romaine 1995; Gumperz 1982; Appel and Muysken 1987; Myers-Scotton 1992; Matras 2009). Since the prevailing view is to deny the existence of such practices involving Medieval Welsh and Middle English, this essay will use the term code-switching as an inclusive term which represents all manifestations of contact-induced variation.

> Nothing would seem more reasonable than to expect that the conquest of the Celtic population of the British Isles by the Anglo-Saxons and the sudden mixture of the two peoples should have resulted in a corresponding mixture of their languages (Baugh and Cable 1993: 72).

Nonetheless, its final conclusion reaffirms the orthodoxy that:

> Outside of place names, the influence of Celtic upon the English language is almost negligible (Baugh and Cable 1993: 74).

The explanation offered, though, is inevitably unconvincing in its impressionistic vagueness:

> The relation of the two peoples was not such as to bring any considerable influence on English life or on English speech. The surviving Celts were a submerged people (Baugh and Cable 1993: 74–75).

In recent years, linguists have sought to overturn this orthodoxy, proposing a number of features of the development of English which might owe their origin to Celtic contact, and this approach has become known as the "Celtic hypothesis" (see Filppula, Klemola, and Pitkänen 2002). The hypothesis draws together a number of imperfectly explained features where the development of English has diverted from its Germanic peer-languages; from the alternative systems of the verb *to be* (which Tolkien had suggested as showing Celtic influence) to such features as the rise of the internal possessor construction, the development of the progressive, the Northern Subject Rule and the rise of periphrastic *do*. Summarising the theory, Hickey (2012: 506) notes that although "the Celtic hypothesis now goes back over a century [… it] has occupied the periphery of English historical scholarship". It is striking, though, how scepticism about the existence of contact phenomena, on which the theory depends, has been subject to re-evaluation, whereas the presumption that there is almost no lexical evidence which might be adduced in support of the hypothesis remains all but unscratched.

In the most recent survey of Celtic lexical influence Durkin (2014) reaffirms the reported absence of lexical evidence, providing graphs to show that, based on the revision so far completed for the third edition of the *Oxford English Dictionary* (*OED*), Welsh (unlike Irish) is not even one of the twenty-five most productive sources for borrowing. Given the intimate and extended contact between the languages over more than one and a half millennia, one might consider the fact that Welsh is deemed to have had less influence on English than Malay, Hawaiian and Scots Gaelic as puzzling, and requiring a more rigorous explanation than, for example, Baugh and Cable's (1993) metaphor of a people and language submerged.

Interestingly, Durkin (2014) proves amenable to the possible plausibility of the Celtic hypothesis because, in his view, it need provide no denial of the absence of lexical contact evidence:

> Although they are often presented collectively as the Celtic hypothesis, these suggestions are nearly all independent of one another, having in common simply the idea that there may have been Celtic substratal influence on English. For our purposes, the interesting question is how this would be likely to correlate with lexical borrowing. According to the current prevailing model in contact linguistics, perhaps rather surprisingly, extensive structural borrowing and minimal lexical borrowing are actually very compatible in situations where speakers have switched from a low-prestige language into the newly adopted high-prestige language: [. . .] various aspects of their [i.e. Celtic-to-English shifters'] linguistic behaviour (including perhaps aspects of their use of basic function words such as *be* or *do*) may have been shaped by their background as speakers of Celtic languages learning English. Thus, even if all of the suggestions [. . .] do correctly identify Celtic influence on English (which many scholars would consider very unlikely), this still does not necessarily conflict with the observation that English shows little lexical borrowing from Celtic languages (Durkin 2014: 90).

In taking this stance he might be seen to be in line, for example, with the conclusions of Laker and Russell (2011):

> However, in the last decade scholars have reasoned that the dearth of Brittonic loanwords in English that initially led some scholars to dismiss the possibility of Brittonic influences on English is not decisive. Instead, many linguists now argue that, similar to contemporary situations of adult language learning, foreign accent and grammar will be the most perpetual reminders of first language encroaching on the second language (Laker and Russell 2011: 111).

Benskin (2011: 182) also refuses to dismiss the theory, though he hints that he had expected to: “But Celtic origin remains a possibility, whereas at the outset of this inquiry it looked not to be.”

Similarly Mufwene (2001) concedes the likelihood of Celtic influence, arguing from general principles of language contact, and supported by the analogy of other linguistic systems, in which such influence has been better documented and more recently observed. This apparent shift in fortune for the Celtic hypothesis, however limited in scope it may prove to be, does not require and need not suggest any challenge to the accepted wisdom that “there are only a few loanwords from Celtic”. Thomason, for example, is clear that:

> It also is not justified to argue against a contact explanation on the ground that if there are no loanwords, there cannot be any structural interference either [. . .] often, in cases of shift-induced interference, there are few or no loanwords from the shifting group's original language in the target language, and in these cases there may well be structural interference without any significant lexical interference (Thomason 2001: 92).

We can see, then, the history of Celtic language influence on English being interpreted into a particular pattern of contact-induced change, a pattern which (conveniently) is compatible with the presumed absence of lexical borrowing.

There are inevitably problems in applying a formulation which conflates the history of Celtic influence into a single phenomenon, producing a single model of shift-induced interference which must be presumed to have endured for one and a half millennia. The truth is that there were major regional and temporal differences between the relative strengths of Brythonic/Welsh and English speakers at different stages of this contact. Though Durkin (2014: 76) concludes confidently that "the number of generally accepted early Celtic borrowings in English is very small", so it must be worth investigating how much room for research and reflection is conceded by the qualifying phrase "generally accepted". The scope of this study is to test Durkin's conclusion by examining the evidence of linguistic borrowing from Welsh into English in two texts written by scribes in the Welsh Penumbra (defined above) in the later Middle Ages, at a time when the two language systems existed alongside each other in that area (as elsewhere in medieval Wales) within a relatively stable multilingual relationship.

2 Problems in methodology

Certainly, there are cultural, historical and methodological reasons why the presence of Welsh lexical items might have been underestimated, or even overlooked. The first is simply a lack of familiarity. More than fifty years have gone by since, in an O'Donnell lecture, Tolkien (1983) declared:

> I would say to the English philologists that those who have no first-hand acquaintance with Welsh and its philology lack an experience necessary to their business (Tolkien 1983: 163).

Nonetheless, much of the recovery of medieval English sources was carried out by philologists whose knowledge of Frisian and Gothic far exceeded their knowledge of Irish and Welsh, a training at times clearly traceable in their conclusions. It was a difficulty acknowledged by Skeat (1885: xxiv) in the preface to his *Etymological Dictionary* in which he noted that Celtic influence was a "particularly slippery subject to deal with". This difficulty was increased by the evidence of language-contact between Brythonic and Latin in the Romano-British period, raising the difficulty of proving whether the adoption of lexical items into English showed evidence of language contact between Old English and Latin or Old English with

Brythonic which had received some terms from Latin.[6] But Skeat draws attention to another problem, the relative lack of evidence for Brythonic/Middle Welsh compared to the substantial corpus of writings in Old English. Skeat (1885: xxiv) laments the "want of definite information on its older form in a conveniently accessible way". His remarks draw attention to the disastrously poor survival of manuscripts of vernacular Welsh writing. There is a lack of manuscripts that can be dated before 1250–1350, even including such manuscripts as the *Book of Aneirin* (Cardiff MS 2.81) and the *Book of Taliesin* (NLW MS Peniarth 2), the contents of which are generally agreed to be much older than the manuscript witnesses. In Daniel Huws' (2000) survey of the early Welsh manuscripts, he concludes that:

> NLW Peniarth 9 (the Welsh Charlemagne cycle) with its colophon dated 1336 and the Book of the Anchorite (Oxford, Jesus College 119) with its colophon dated 1346 are the two earliest dated Welsh manuscripts. [. . .] There are approaching forty manuscripts in Welsh which can be supposed earlier than 1346. There is no room to believe that any of them were written before about the middle of the thirteenth century (Huws 2000: 212).

The reasons for this paucity of textual record remain obscure. Maybe the loss of early Welsh manuscripts was exacerbated by difficulties of interpretation. Huws (2000: 212) raises the question whether "one reason for the non-survival of pre-1250 Welsh vernacular books [. . . was] that they were largely in Insular script". But the most likely explanation for the absence of Welsh records is that Welsh manuscripts were another casualty of the recurring warfare of English incursions from Harold Godwinson to Edward I and Henry IV. There is no obvious reason why the English Marcher Lords would value or preserve manuscripts in Welsh, but there are many reasons why they might wish to destroy them, while perhaps seeking to demoralise the opposition. Underlying the English conquest of Wales was an extensive and complex establishment of a myth of justification, which explained England's rights in terms other than merely the superiority of force (Meecham-Jones 2008: 9). Implicit in this was the denial of the legality of prior Welsh arrangements, which may have been supplemented by the destruction of written sources such as wills, charters and other documents of historical record, or those which evidenced land ownership.

One element of this justification involved the discourse of racial inferiority – that is the repetition of the idea (sometimes overtly, but more often through imagery) that the Welsh people were ethically as well as physically inferior to the English race, which is why the Welsh required the English to rule them. The existence of sophisticated literary manuscripts would seem to have been incompatible with such presumptions, which was both ironic and inconvenient given the stylisation and complexity of Welsh poetic culture in this period. One recurring feature of anti-English discourse in Wales was the repetition of claims of the ritual burning of Welsh books, sometimes associated with the mythological figure of Ysgolan. Whether the myth owed anything to distant memories of planned "bibliocausts", in Schwyzer's (2004: 81) term, is uncertain but as well as the destruction of the seats of the Welsh royal families, during the conflicts a number of Welsh cathedrals and monasteries were set alight, reducing the chances of early manuscripts in their collection surviving. It is a striking fact that the earliest surviving textual evidence for written Welsh is to be found in manuscripts that are known to have been preserved in libraries outside Wales for most of the past millennium, such as the Lichfield Gospels/Book of St Teilo and the Cambridge Juvencus (Cambridge University Library MS Ff.4.42), which contain ninth- (and perhaps eighth-century) and ninth/early tenth-century marginalia in Welsh respectively. As Huws (2000: 38) notes, "every pre-1100 Welsh book, and every fragment of one, has been cared for since before 1600, if not since the end of the Middle Ages, outside Wales".

---

[6] Tolkien (1983: 176–179) notes that although English philologists have expressed an interest in "the Latin loan-words in Old English" he believed that "in this matter the importance of the Welsh evidence is not yet fully recognized".

In considering the unevenness of manuscript survival, it is important to remember that every category of vernacular manuscripts produced in Wales remains unaccounted for. Ceridwen Lloyd-Morgan, Charles-Edwards and others have demonstrated how the code-switching of French terms into Welsh poetry demonstrates that "Welsh literary circles were aware of the most recent developments in French literature" (Lloyd-Morgan 2008: 165), but no manuscripts written in French have been identified as being of Welsh production – which need not deny their existence. Furthermore medieval Wales had a substantial English-speaking population in the "planted" garrison towns, as well as many whose job was to mediate between a Welsh-speaking population and an English and/or Anglo-Norman government. It is inevitable that, particularly in the later Middle Ages, both these classes would have created an audience for English manuscripts – the first as a means of validating their Englishness despite living in Wales, the latter as facilitating their contact with the English population as well as signaling their acceptance of the political status quo. So far, scholarship has been extremely cautious in ascribing the production of any texts in English, French or Latin to Welsh scriptoria in the absence of additional external marks demanding explanation (which generally means marginalia in Welsh or inscribed Welsh personal names). But even when such additional evidence is present, there seems to be a reluctance to propose production, as opposed to later ownership or usage, to Wales.[7]

Such caution may be understandable, but it has created extensive problems of focus in English medieval dialectology. Once again, Wales proves to be a missing piece in the proposed solution. When compiling the *Linguistic Atlas of Late Mediaeval English* (*LALME*), the editors included a mere nine manuscripts from Wales. This is clearly far too few to provide even an initial profile of the forms of Middle English found in Wales, or even one district of Wales. In describing English in medieval Wales, Llinos Beverley Smith (2000: 9) speculated that, due to settlement by incomers from different parts of England, the English language in Wales must have been a "mixing bowl of dialects". She was reluctant to speculate on whether medieval Wales produced a distinctive form of English, as had happened in Ireland: "It is difficult to be certain whether a Cambro-English comparable to the Irish-English identified in late medieval Ireland ever existed" (Beverley Smith (2000: 11). But this question is hard to answer, in part, because it has never been posed as part of a sustained enquiry. The very small corpus of Middle English so far associated, even at the most speculative level, with Wales, does not permit a statistically significant pattern to be determined. But the scrutiny of how apparent evidence in other texts has been "explained away" provides grounds for caution at the very least.

First, the absence of early Welsh manuscripts has had serious consequences for attempts to trace Welsh influences on English. Due to the much greater number of surviving Old English manuscripts, and their greater antiquity, ascertaining whether a word derives from Welsh or English is made more difficult. A striking example occurs in the text of *Beowulf*, which survives in a single manuscript dated around 1010 A.D. but which has generally been supposed to be (in part or whole) centuries older than the manuscript record. Presented in *Beowulf* with the image of a candle "weaxen wonna leg", the *OED* interprets the adjective *wan* as representing an Old English form "Old English wann (won), dark, gloomy, black". The *OED* expresses no concern that no parallel is "found in any of the other Germanic languages", suggesting an association with Old Germanic *winnan* ('to strive, toil, suffer', etc.) The possibility of a relationship to the Welsh form *gwan* ('weak') which undergoes a frontal mutation to *wan* (familiar in contemporary Wales, for example, in the often seen road sign *pont wan* 'weak bridge') is considered but the dictionary concludes that "association of the two words in later (Middle English and Mod. E) periods is more probable than ultimate connection." This may be a surprising conclusion for,

---

7 Marx (2008) considers the difficulties raised by MSS NLW Peniarth 12, NLW Peniarth 50, NLW 3041 and NLW 21608.

although a code-switch into Brythonic/Welsh might initially seem unlikely in *Beowulf*, it is striking that the *OED*'s list of illustrative quotations for the hypothesised English term *wan* draws heavily on texts and manuscripts associated with the Welsh border country such as *Hali meidenhad*, *Piers Plowman*, the *Harley Lyrics* and Trevisa's *Polychronicon*.

Elsewhere, texts from this canon of Welsh Penumbra manuscripts provide more generally, perhaps universally, accepted evidence of contact influence. One of the most significant, and one of the most accomplished of these texts, is the early thirteenth-century religious text *Ancrene Wisse*. It has long been accepted that the text contains a number of Welsh words, such as *cader* 'chair' and *baban* 'baby'. This lexical influence was noted by Tolkien (1929), playing a role in his distinguishing of "AB language", a dialect of early Middle English evidenced in Cambridge, Corpus Christi College MS 402 (which contains *Ancrene Wisse*) and Oxford, Bodleian Library MS Bodley 34 which contains *Hali meidenhad* amongst other texts), which also contribute some code-switched or "borrowed" items (*keis* 'henchman', *genou* 'kiss'). Yet the appearance of these items in the AB dialect has never led to suggestions that any of these texts were written in English-speaking areas in Wales or by writers or scribes who were (at least to some degree) bilingual English–Welsh speakers, which might seem surprising. To accommodate the language contact, scholars have sought to place their composition within an area classified within *LALME*, but close enough to the Welsh border to permit contact – in this case, western Herefordshire. The same location has been proposed for MS Peniarth 12 (discussed by Marx 2008), and it is striking how large a number of texts has been adjudged to be written in forms of "West Midland" dialect, leading to a profusion of texts ascribed to Herefordshire, Shropshire and Cheshire. Herefordshire has also been suggested as the place of origin for the *Harley Lyrics*, as well as important early examples of the English *Brut* tradition such as Oxford, Bodleian Library MS Rawlinson B 17. The placing of so many texts significant in the development of Middle English into this small geographical dialect-area perhaps invites scepticism, providing too easy a solution for the limited appearance of code-switched items in material from the Welsh March. However, the existence of this large corpus of manuscripts written, if not in English-speaking Wales, then in the Welsh Penumbra in the late Middle Ages offers an intriguing resource for code-switching research.

3 Assessing the textual evidence

A lyric from Llanover MS B.5 (now NLW MS 16031), a sixteenth-century manuscript associated with Glamorgan, provides the earliest well-attested example of a writer experimenting with an integrated model of code-switching between the two languages. Due to the richness of its agricultural land, Glamorgan had experienced early conquest and extensive colonisation by English-speaking settlers from the end of the eleventh century. It is interesting that this lyric seems to be the work of an English poet code-switching into Welsh, rather than the reverse:

> I have nothyng mor y ddywedyd
> But consyder wrth fy mhenyd
> In my hart mae saeth yn pido
> Come in hast yddy thynny oddyno.
>
> By my troth hyn wy n y geiso
> What you promise oy gywyro
> I will stand heb ddowt f anwyled
> In my words er gwaetha r hollfyd (Llanover MS B.5)

Evidence earlier than the early modern period remains elusive. Given the highly literary nature of the surviving evidence (e.g. poetry written in strict metres in Welsh and alliterative forms in English, but relatively little informal or domestic material, particularly in Welsh) if code-switching is found in late medieval texts from the Welsh March, we might expect it to show contact between literary traditions, rather than contact in the speech community, which must also have been occurring, in line with Schendl's (2013: 42) argument that "code-switching was a widely accepted textual strategy in late fourteenth- and early fifteenth-century England". But it seems that the manuscripts identified as being from the Welsh March provide evidence of the development of a written tradition of English (see e.g. Revard 1981, 2000) alongside, but apparently distinct from, the continued flourishing of a tradition of writing in Welsh.[8]

Clearly manuscripts from the Welsh Penumbra provide the most apt resource for questioning this presumption. A preliminary study of the romance *Sir Gawain and the Carl of Carlisle*, written in the third quarter of the fifteenth century close to the Welsh border in Shropshire or Herefordshire (Huws 1996) in NLW MS Brogyntyn II.1 does indeed show some evidence of language-contact with Welsh, none of which is noted by its editors Kurvinen (1951) and (more recently) Hahn (2000) (see Meecham-Jones 2012). A revealing example of the failure to consider Welsh influences occurs in Kurvinen's (1951) glossing of a description of the Carl's daughter, near the end of the poem:

---

[8] Charles-Edwards (2000) points to the probability of professional secular scribes working in Welsh in the fourteenth century. The existence of such a class presupposes the patronage of a class or classes willing and able to sustain such a profession.

> The Carllus douȝhtter forthe was brouȝht
> That was so feyr and bryȝht.
> As gold wyre schynyde her here.
> Hit cost M. li [a thousand pounds] and mar,
> Her aparrell pertly pyȝhte. (lines 416–420)

Kurvinen's (1951: 140) interpretation of "pertly pyȝhte" as 'skillfully created' seems to draw on the *Middle English Dictionary's* (*MED*) derivation of the adverb from the Latin participle *apertus* ('open/public/free; uncovered/exposed/opened; frank/clear/manifest; cloudless'), but there is no obvious connection with its suggested meaning in this text. Hahn's (2000) suggested 'beautifully adorned' seems plausible in context, but also some distance from its supposed Latin source. Neither show any familiarity with the Welsh adjective *pert* which translates as 'beautiful' or 'pretty', a term which is predominantly used in relation to female attractiveness. It is hard not to feel that Kurvinen (1951) and Hahn (2000) have been led astray by the appearance of what seems to be the same word twice in *Sir Gawain and the Green Knight* (henceforth *SGGK*), (probably) written in a nearby or neighbouring district, a century or slightly less before. But the two appearances in that text do not refer to a woman's attractiveness, and both are capable of being interpreted as deriving from *apertus*, even if the possibility of a derivation from Welsh cannot be dismissed:

> And spekez of his passage, and pertly he sayde 544
>
> 'As is pertly payed Þe chepez Þat I aȝte' 1941 (Davis 1967)

But if Kurvinen (1951) and Hahn (2000) expected to interpret *the Carl of Carlisle* through the scholarship on *SGGK*, it seems equally plausible to expect any pattern of correspondence to work in the opposite direction, also. Interestingly, *SGGK* is listed by the *OED* in the illustrative examples for the word *loop*, a term the dictionary accepts as being of Celtic origin, even if "the form is more readily derived from O[ld] N[orse]". It is a surprising word to see in this context, since the forms it is shown to be related to derive from Scots Gaelic (*lūb*) and medieval Irish (*lūbaid*, *lūbtha*) rather than Middle Welsh. The existence of an unrelated alternative term (*dolen*) in Middle Welsh does not preclude language contact, particularly since the record of Middle Welsh is derived primarily from literary texts, with their particular conventions and formalities (and it is possible that any form associated with *lūb* might have been considered to be of the wrong register, or have been found only in certain dialects of Middle Welsh). Nonetheless the listing of this word must provoke curiosity as to whether there exists a class of words found in *SGGK* which are probably or certainly derived from Welsh, but which have been overlooked or disregarded by editors.

Immediately, certain challenges demand to be faced. Assigning the poem's composition, or even copying, to the Welsh Penumbra is itself contentious. For McIntosh (1963: 5), the application of isoglosses led him to declare with certainty that the poem's place of origin could be determined "only [. . .] with reasonable propriety in a very small area either in SE Cheshire or just over the border in NE Staffordshire". That would place the composition close to the Welsh border, but probably towards the very edge of inescapable daily contact with Welsh. But many scholars have expressed reservations about the methodology and conclusions of *LALME*'s placing of the poem, a dissatisfaction summarised, for example, by Putter and Stokes (2007):

> Our own view is that LP 26 misrepresents the poet's language in two ways. In the first place, the profile contains many errors and omissions; in the second place, it does not discriminate with sufficient care between scribal and authorial usage, and between normal usage and metrically constrained usage. As a consequence of these weaknesses, LP 26 cannot in our view be trusted to provide a precise localization of the language of the scribe, let alone of the poet (Putter and Stokes 2007: 470).

Though the availability of *LALME* isoglosses has encouraged attempts to move the poem's production further north or east into Lancashire or Cheshire, the failure to map equivalent isoglosses for English settlements in medieval Wales need not preclude the hypothesis that such a setting might have provided the social and creative matrix within which *SGGK* was composed. The garrison towns established in Wales were populated with migrants from the Midlands who could be expected to bring with them linguistic features from their home regions, some of which must have played a role in the development of distinctive dialects of Middle English used within Wales.

For this initial study, to test if there is any lexical evidence which might support such a hypothesis, the fourth fit of *SGGK* was chosen, which comprises about twenty per cent of the text. Initially, the results might seem disappointing, as there are few, if any, words present which can only be explained as being derived from Welsh. There are some terms which demand attention, and which perhaps illustrate the *de minimis* approach of editors to the problem. We have, for example, two appearances in this fit of the term *crag*, which Davis (1967), at least, is happy to link to Middle Welsh. The *OED* is more circumspect, fretting whether the term is closer to Welsh *carreg* than *craig* and noting that the word is "apparently of ancient use in the local nomenclature of the north of England and Scottish Lowlands". This ancient use is scarcely surprising, given that the survival of toponymic features is precisely where one might expect to find evidence of displaced languages. Nonetheless, it seems less plausible that the poet (or scribe) of *SGGK* should be influenced by distant place names than by a word which may have been used very commonly within his linguistic community. Unfortunately, the survival of records from the fourteenth century does not permit us to speculate how far fieldnames and the names given to geographical features in the Welsh Penumbra reflected Welsh usage which may have been obscured or erased by the subsequent Anglicisation of the region, and/or the habits and preferences of Anglophone mappers, lawyers and clerics, for example during the creation of Ordnance Survey maps and the enclosure of common land. We have better records of the names and locations of parishes, and these provide clear evidence of the familiarity of *carreg* and *crug* ('rocky outcrop') in Marcher place-names. Due in part to the instability of the Powys/English border, medieval Shropshire contained a number of parishes which bore Welsh names, and several of which straddled the border. Within the parish of Llanymynech were the Shropshire settlements of Llwyntidman/Llwyntidmon and Trepennal, but also Carreghwfa/Carreghofa across the border in Powys. The Shropshire parish of Alberbury (also recorded as Llanfihangel yng Ngheintyn) included Crugion in Powys (now known as Criggion). Mills (2011) in the *Dictionary of British Place Names* is happy to accept Crug as a productive constituent of CrugHywel (which survives in Anglicised form as CrickHowell) and Crickadon in Powys, but feels obliged to suggest an Old English root for Crockton in Worcestershire. For Cricklade in Wiltshire he (2012) suggests OE *gelid* 'difficult river crossing' with "an obscure first element, possibly Celtic creig 'rock'". It seems clear that there would have been no need for the poet of *SGGK* to have travelled beyond the Welsh March for him to have encountered these Welsh terms being frequently used as toponymic terms.

Highlighting the poet's use of the word *mulne* 'mill':

as water at a mulne 2203

raises similar problems. The *OED* traces the origin of the term to post-classical *molina*, but in the examples offered by the *MED* it becomes clear that the process of losing the *n* is well-advanced by the middle of the fourteenth century. Though the *MED* lists *mulne* as a "southern, South-Midlands" variant, all trace of the *n* is absent, for example, in the work of Trevisa, writing in Gloucestershire in the last quarter of the fourteenth century. It seems plausible, then, to speculate either that the *SGGK* poet or scribe intended the Welsh word *melin* 'mill', confident that it would be understood, or that he exercised his profession within a linguistic community in which frequent contact with the Welsh word had retarded the change-in-progress of the dropping of the *n*, which was predominant in Eastern forms.

One of the most notable characteristics of the text of *SGGK* is the very high occurrence of terms rarely recorded elsewhere, or in some cases found only in London, British Library MS Cotton Nero A.X. It must be worth considering whether these rare or unattested items might show contact influence from Welsh. Thus, Davis (1967) cites four appearances of the adjective *wynne* which he glosses as 'delightful, lovely.' It is a reading echoed by the *OED* that suggests that the word represents an adjectival use of the Old English *wyn* 'joy, pleasure, delight, bliss'. It is striking, though, that the *OED* appears to believe that the term is confined to Middle English alliterative verse, while their examples of this adjectival use are drawn almost exclusively from texts from the Welsh Penumbra – primarily MS. Cotton Nero A.X but also the *Harley Lyrics*, and Laȝamon's *Brut*. Reading the final occurrence of *wynne* in the text, it becomes clear how imprecise such an identification must be considered:

> Ho wayned me vpon Þis wyse to your wynne halle 2456

It would be easy to translate the term using the Welsh adjective *gwyn*, which experiences a frontal mutation to *wyn*. Although it primarily functions as a colour term, for white or silver, the semantic range of the term is extended to a range of favourable connotations which permit possible readings such as 'good, shining, well-favoured' or 'blessed' which explains its popularity as a term of praise used as the first element in compound Welsh forenames – such as Gwynfor, Gwyneira and Gwenhwyfar (Guenevere). This lexical spread would allow the Welsh-derived term to describe the hall as plausibly as the *OED*'s rather speculative derivation, a fact which proves to be true also for the other appearances in the text – making sense in context for "wynne is the wort" (518), "wynne worschip" (1032) and "Þe wynne golde" (2430). The apparent recorded absence of the term *wynn* in other dialect areas of English surely means that linguists need to take seriously the possibilities either that the term is derived from Welsh or that its appearance was inspired or influenced by the poet or scribe's conscious or unconscious recognition of the similarity of lexical possibilities between English and Welsh usage.

Admittedly, the terms so far suggested as showing Welsh influence in *SGGK* might be characterised as infrequent or unfamiliar terms, but one of the characteristics of *SGGK* which has drawn attention is the idiosyncrasy of its vocabulary. This has been explained primarily as a result of its use of the alliterative romance form, which creates a demand both for words which can be alliterated and words which can be rhymed. Occasional code-switching from Welsh might well prove useful for the former, though rarely so for the latter purpose. Such a strategy should not surprise – quite apart from the strict demands imposed by the alliterative form, the poet(s) of *SGGK* and *Pearl* is clearly a poet highly sensitive to the phonological nuances of his verse. Putter (2011) has shown how both poems demonstrate a restrained pattern of code-switching into French for particular pragmatic effects. Unquestionably the use of Welsh forms in a Marcher context could have been used to add nuance to the text in ways which are now difficult or impossible to reconstruct.

Faced with the sheer volume of unfamiliar vocabulary glossed in *SGGK*, it is hard for a researcher investigating Celtic influence on this lexis not to feel like Gawain himself, venturing into “a forest ful dep, þat ferly watz wylde” (741). Many of these terms are listed as obscure, and have been presumed to be a part of a literary vocabulary associated with the alliterative tradition and chivalric values. But if modern English no longer needs terms like *hende* and *gome*, it is important to remember that over 600 years Welsh vocabulary has also been subject to lexical change, in particular dialect advergence, furthered by the processes of standardisation of the written form set in motion by the early modern Bible translations. Though this process may have been less extensive than that experienced by Middle/Early Modern English, it was still substantial. Further, the evidence is somewhat partial, surviving primarily in poetry which was written to conform to the complex and highly stylised rules of *barddoniaeth*. It seems inevitable, then, that many inappropriate or insufficiently mellifluous words were excluded from these texts, and probably also words which had only local, dialectal currency.

Allowing for such difficulties, it becomes easier to understand why recovering a Welsh contact element in Middle English poetry has been seen as a fraught endeavor. Nonetheless, there is a wealth of distinctive phenomena in *SGGK* which demands consideration. One prominent example might be the word *gryndel*:

> Bolde burne, on Þis bent be not so gryndel’ 2338

Davis (1967) glosses this as ‘fierce’ while the *OED* offers ‘fierce, angry’. It is potentially an important diagnostic key since, as well as appearing in *SGGK*, it is found in another of the MS. Cotton Nero A.X poems, *Patience*:

> Be noȝt so gryndel good man 524

It is not recorded elsewhere. The *OED* records its origins as being unknown but suggests a comparison with “Old Norse *grimd* fierceness, <grimm-r grim adj.>”. Stanley (2001: 92) notes this interpretation of the word as meaning ‘fierce’ or ‘angry’ when commenting on Klaeber’s placing of the word in a footnote while considering the origin of the name Grendel in *Beowulf*. Davis (1967) moves in a different direction. He notes the appearance in line 312 of the term *gryndel-layk* (glossed as ‘fierceness’) which he links to the Old Norse *grindill* ‘storm’ with the suffix *leikr*. *Leikr* is affixed to adjectives to form nouns of quality – the *OED* notes its popularity in the *Ormulum*, but also notes that no such formulations have survived into modern English. From this noun, Davis (1967) suggests a back-formation to the adverb *gryndelly* line 2299, and to the adjective *gryndel*. There are problems with this interpretation. Though Davis accepts *gryndellayk* he notes an alternative textual reading *gry dellayk* which seems less compatible with his suggested solution. Also, the applicability of a noun from storm in context (during the Green Knight’s railing at King Arthur’s court) stretches credibility.

There is an alternative possibility. From Medieval Welsh we might draw the noun *grym* which can be translated as ‘force, vigour, power’. The range of connotations is a little different from the Old Norse suggestions, and perhaps more ethically neutral, but they are at least as plausible in context. Accounting for the *del* suffix poses more problems, but it could perhaps represent a form of the Welsh *dull* – which translates as ‘method, manner, style’ and could perhaps in context stretch to ‘mood’ or ‘demeanour’. It is important, also, to remind ourselves that if the suggested derivation of *gryndel* from Welsh *grym* is hypothetical, it is no more hypothetical than the linkage with Old Norse *grimd*. Also, if the origins of *gryndel* are wholly or even partly derived from Welsh, then it may be plausible that *gry dellayk* is not related to *gryndel* and requires separate explanation. *Gryndel* is merely one example of a large category of insufficiently explained terms from MS Cotton Nero A.X which appear to show contact influence – whether from Norse, or Welsh, or conceivably both (as well as French). Determining which should be a worthy quest for scholars “bi craftes wel lerned”.

4 Conclusion

An examination of the fourth fitt of *SGGK* shows that there is limited but striking evidence of words derived from Welsh, particularly amongst the vocabulary which is distinctive to the poem and rarely, if ever, found elsewhere. The relative infrequency of code-switching suggests that this does not represent a casual contact phenomenon drawn from a wide-spread spoken bilingualism, though the highly stylised requirements of Middle English alliterative verse (and of Welsh models of literary composition using devices of *cynghanedd*[9]) may be responsible for restricting the appearance of code-switched or borrowed items.

---

[9] *Cynghanedd*: “An intricate system of alliteration and rhyme in Welsh poetry” (*OED*).

Nonetheless, examination of the fourth fitt of *SGGK* suggests that code-switching from Welsh may be a highly self-conscious literary technique, being used to establish a particular cultural or political point which has become obscured by the critical non-recognition of what is being done. The evidence from the fourth fitt of *SGGK* also suggests that the presumption that there is no code-switching from Welsh into Middle English may not survive more rigorous study.

References

Appel, Rene & Pieter Muysken. 1987. *Language contact and bilingualism*. London: Edward Arnold.

Baugh, Albert C. & Thomas Cable. 1993. *A history of the English language*. London: Routledge.

Benskin, Michael. 2011. Present indicative plural concord in Brittonic and early English. *Transactions of the Philological Society* 109(2). 158–185.

Benson, Larry D. (ed.). 1987. *The riverside Chaucer*. Boston: Houghton Mifflin.

Beverley Smith, Llinos. 2000. The Welsh and English languages in late medieval Wales. In D. A. Trotter (ed.), *Multilingualism in later medieval Britain*, 7–24. Cambridge: D. S. Brewer.

Bibire, Paul. 2001. North Sea language contacts in the early Middle Ages: English and Norse. In Thomas R. Liszka & Lorna E. M. Walker (eds.), *The North Sea world in the Middle Ages: Studies in the cultural history of north-western Europe*, 88–107. Dublin: Four Courts Press.

Bjorkman, Erik. 1900–1902. *Scandinavian loan-words in Middle English*, 2 vols. Halle: Karras & Niemeyer.

Burnley, David. 1992. Lexis and semantics. In Norman Blake (ed.), *Cambridge history of the English language*, vol. II: *1066–1476*, 409–499. Cambridge: Cambridge University Press.

Butterfield, Ardis. 2009. *The familiar enemy: Chaucer, language and nation in the Hundred Years War*. Oxford: Clarendon Press.

Chapman, Adam. 2015. *Welsh soldiers in the Later Middle Ages 1287–1422*. Woodbridge: Boydell & Brewer.

Charles-Edwards, T. M. 2000. *Early Christian Ireland*. Cambridge: Cambridge University Press.

Dance, Richard. 2011. “Tomarȝan hit is awane”: Words derived from Old Norse in four Lambeth homilies. In Jacek Fisiak and Magdalena Bator (eds.), *Foreign influences on medieval English* (Studies in English Medieval Language and Literature 28), 77–127. Frankfurt am Main: Peter Lang.

Dance, Richard. 2014. Getting a word in: Contact, etymology and English vocabulary in the twelfth century. *Journal of the British Academy* 2. 153–211.

Davis, Norman (ed.). 1967. *Sir Gawain and the green knight*. Oxford: Oxford University Press.

Durkin, Philip. 2014. *Borrowed words: A history of loanwords in English*. Oxford: Oxford University Press.

Filppula, Markku, Juhani Klemola & Heli Pitkänen (eds.). 2002. *The Celtic roots of English*. Joensuu: Joensuu University Press.

Gardner-Chloros, Penelope. 2009. *Code-switching*. Cambridge: Cambridge University Press.

Griffiths, R. A. 2004. Vaughan family (*per. c.*1400–*c.*1504). *Oxford dictionary of national biography*. Oxford University Press. http://www.oxforddnb.com/view/article/48656 (accessed 6 October 2017).

Gumperz, John J. 1982. *Discourse strategies* (Studies in Interactional Sociolinguistics 1). Cambridge: Cambridge University Press.

Hahn, Thomas. 2000. Gawain and popular chivalric romance in Britain. In Roberta L. Krueger (ed.), *The Cambridge companion to medieval romance*, 218–234. Cambridge: Cambridge University Press.

Hanna, Ralph. 2008. Lambeth Palace Library MS 260 and the problem of English vernacularity. *Studies in Medieval and Renaissance History* 20. 131–199.

Hickey, Raymond. 2012. Early English and the Celtic hypothesis. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford handbook of the history of English*, 497–507. Oxford: Oxford University Press.

Huws, Daniel. 1996. MS Porkington 10 and its scribes. In Jennifer Fellows, Rosalind Field, Gillian Rogers & Judith Weiss (eds.), *Romance reading on the book: Essays on medieval narrative presented to Maldwyn Mills*, 188–207. Cardiff: University of Wales Press.

Huws, Daniel. 2000. *Medieval Welsh manuscripts*. Cardiff: University of Wales Press and the National Library of Wales.

Jones, Evan David. 1959. Vaughan family, of Hergest, Kington, Herefords. *Dictionary of Welsh biography*. http://yba.llgc.org.uk/en/s-VAUG-HER-1400.html (accessed 6 October 2017).

Kurvinen, Auvo (ed.). 1951. *Sir Gawain and the Carl of Carlisle in two versions* (Annales Academiæ Scientiarum Fennicæ 71). 1–229.

Laker, Stephen & Paul Russell. 2011. Languages of early Britain: Introduction. *Transactions of the Philological Society* 109(2). 109–112.

LALME = McIntosh, Angus, M. L. Samuels & Michael Benskin (eds.). 1986. *A linguistic atlas of Late Mediaeval English*, 4 vols. Aberdeen: Aberdeen University Press.

Lloyd-Morgan, Ceridwen. 2008. Crossing the borders. In Ruth Kennedy & Simon Meecham-Jones (eds.), *Authority and subjugation in writing of medieval Wales*, 139–174. New York: Palgrave Macmillan.

McIntosh, Angus. 1963. A new approach to Middle English dialectology. *English Studies* 44. 1–11.

Marx, William. 2008. Middle English texts and Welsh contexts. In Ruth Kennedy & Simon Meecham-Jones (eds.), *Authority and subjugation in writing of medieval Wales*, 13–26. New York: Palgrave Macmillan.

Matras, Yaron. 2009. *Language contact* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

MED = Kurath, Hans & Robert Lewis (eds.). 1956–2001. *Middle English Dictionary*. Ann Arbor, MI: University of Michigan Press.

Meecham-Jones, Simon. 2008. Introduction. In Ruth Kennedy & Simon Meecham-Jones (eds.), *Authority and subjugation in writing of medieval Wales*, 1–11. New York: Palgrave Macmillan.

Meecham-Jones, Simon. 2012. Language contact in the Welsh Penumbra: Evidence in NLW MS Brogyntyn II.I. In Jane Roberts & Emma Volodarskaya (eds.), *The second international conference 'Language, Culture and Society in Russian/English Studies', 25–26 July, 2011: The proceedings*, 50–74. London: University of London.

Mills, A. D. 2011. *The dictionary of British place names*. Oxford: Oxford University Press.

Mufwene, Salikoko S. 2001. *Ecology of language evolution*. Cambridge: Cambridge University Press.

Muysken, Peter. 2000. *Bilingual speech: A typology of code-switching*. Cambridge: Cambridge University Press.

Myers-Scotton, Carol. 1992. Comparing codeswitching and borrowing. *Journal of Multilingual and Multicultural Development* 13(1–2). 19–39.

OED = Oxford English Dictionary Online. www.oed.com.

Poussa, Patricia. 1982. The evolution of early Standard English: The creolization hypothesis. *Studia Anglica Posnaniensia* 14. 69–85.

Putter, Ad. 2011. Code-switching in Langland, Chaucer and the Gawain poet. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English*, 281–302. Berlin & Boston: Mouton de Gruyter.

Putter, Ad & Myra Stokes. 2007. The 'Linguistic Atlas' and the dialect of the 'Gawain' poems. *Journal of English and German Philology* 106(4). 468–491.

Revard, Carter. 1981. Three more holographs in the hand of the scribe of MS Harley 2253. *Notes and Queries* 28(3). 199–200.

Revard, Carter. 2000. Scribe and provenance. In Susanna Fein (ed.), *Studies in the Harley manuscript: The scribes, contents and social contexts of British Library MS Harley 2253*, 21–109. Kalamazoo, MI: Medieval Institute Publications.

Romaine, Suzanne. 1995. *Bilingualism* (Language in Society 13), 2nd edn. Malden, MA: Blackwell.

Rothwell, William. 2000. Aspects of lexical and morphosyntactical mixing in the languages of medieval England. In D. A. Trotter (ed.), *Multilingualism in later medieval Britain*, 213–232. Cambridge: D. S. Brewer.

Rothwell, William. 2001. Stratford Atte Bowe re-visited. *Chaucer Review: A Journal of Medieval Studies and Literary Criticism* 36(2). 184–207.

Rothwell, William. 2004. Henry of Lancaster and Geoffrey Chaucer: Anglo-French and Middle English in fourteenth-century England. *The Modern Language Review* 99(2). 313–327.

Schendl, Herbert. 2013. Multilingualism and code-switching as mechanisms of contact-induced change in later Middle English. In Daniel Schreier & Marianne Hundt (eds.), *English as a contact language*, 41–57. Cambridge: Cambridge University Press.

Schendl, Herbert & Laura Wright. 2011. Code-switching in early English: Historical background and methodological and theoretical issues. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English*, 15–45. Berlin & Boston: Mouton de Gruyter.

Schwyzer, Philip. 2004. *Literature, nationalism and memory*. Cambridge: Cambridge University Press.

Skeat, Walter W. 1885. *Concise etymological dictionary of the English language*. Oxford: Clarendon Press.

Stanley, Eric G. 2001. 'A Very Land-fish, Languagelesse, a Monster': Grendel and the like in Old English. In K. E. Olsen & L. A. J. R. Houwen (eds.), *Monsters and the monstrous in medieval northwest Europe*, 79–92. Leuven: Peeters.

Thomason, Sarah G. 2001. *Language contact*. Edinburgh: Edinburgh University Press.

Tolkien, J. R. R. 1929. Ancrene Wisse and Hali Meiðhad. *Essays and Studies* 14. 104–126.

Tolkien, J. R. R. 1983. English and Welsh. In Christopher Tolkien (ed.), *The monsters and the critics and other essays*, 162–197. London: George Allen.

Townend, Matthew. 2002. *Language and history in Viking Age England: Linguistic relations between speakers of Old Norse and Old English* (Studies in the Early Middle Ages). Brepols: Turnhout.

Janne Skaffari

7 Code-switching in the long twelfth century

1 Introduction

In the field of historical code-switching research, English historical linguists have been particularly active, exploring medieval and modern data, but as the history of English provides an enormous range of material, it may not be surprising that they have thus far overlooked at least one important sub-period in the English Middle Ages. This is the period immediately following the Norman Conquest of 1066, when England became increasingly trilingual. This chapter is devoted to the long twelfth century, from the late eleventh century until well into the thirteenth. Although the concept of "long century" seems to have been used less frequently by historical linguists than by historians, there are several recent publications in which a "long twelfth century" appears with reference to the history of English between, approximately, the Norman Conquest and the second quarter of the thirteenth century (e.g. Skaffari 2009; Faulkner 2012b; Kwakkel 2012; cf. Treharne 2012).[1] This designation is particularly appropriate here: we encounter several languages in the material dated to this period, which makes a label such as Early Middle English inadequate, particularly as some of the English is not Middle but, rather, Old English – even if copied around 1200.

The objective of the present chapter is to provide a linguistically orientated overview of code-switching and other multilingual practices in material from the long twelfth century, with English as one of the languages involved. The growing community of historical code-switching researchers has virtually ignored this period, in spite of its "endemic multilingualism" (Trotter 2003: 84) and the recent efforts of scholars working on medieval language and literature to bring this period into focus (particularly Treharne 2012 and Faulkner 2012b). To fill the

---

Acknowledgment: The author wishes to acknowledge the Academy of Finland for supporting the *Multilingualism in the Long Twelfth Century* project (decision number: 257059).

---

[1] Kwakkel (2012: 79) associates this period with the "Twelfth-Century Renaissance", dating it to 1075–1225; Skaffari (2009), following the sub-periodisation employed in the *Helsinki Corpus of English Texts*, does not consider primary sources from after 1250; and Faulkner (2012b: 275) "provocatively" chooses to start his long twelfth century with a pre-Conquest date, 1042, the beginning of Edward the Confessor's reign. Treharne's (2012) work on a period similar to ours starts from the earlier of the eleventh-century conquests, the Viking one by Cnut, and continues until 1220.

DOI 10.1515/9781501504945-007

gap, the extant material needs to be described and examined, and approaches originating in research on other periods tested further. Only the first few steps can be taken here, as the material and phenomena are more diverse and copious than is often expected of this period.

This chapter first provides overviews of research on multilingualism in medieval England and of the extant texts and manuscripts from the long twelfth century. The bulk of the chapter is concerned with some of the multilingual features in these manuscripts, seeking to identify forms, functions and patterns of code-switching. These mostly involve switching between English and Latin, but it is also important to consider French.[2] The findings are summarised in the concluding remarks, emphasising also the variedness of multilingual practices in this under-researched but highly important transitional period.

2 Research on multilingual practices in medieval England

As the Norman Conquest continues to be treated as essentially the greatest turning point in the history of pre-modern Britain, with well-known – although anything but fully researched – implications for government, culture and language, it does not seem unreasonable to expect that the long twelfth century differs from the preceding Anglo-Saxon era in terms of types and levels of multilingualism and the ways in which multilingual features appear in the written record. While the Norman Conquest certainly promoted French into a prominent High language in England – to use sociolinguistic terminology and follow a traditional view – it did not as such introduce a new language into England: there were French-speakers and activity in French on the British Isles before 1066, particularly in Edward the Confessor's time; Trotter (2009: 162), for example, questions the Conquest as a "watershed". At the same time, another European High language, Latin, strengthened its position as the main language of writing in England. While English was Low by comparison, it remained the first and only language of the majority. It also, within a few generations, became a language worth acquiring by the Normans' descendants; there is a continuing debate on when, how and at what rate this happened (see below for some studies on French–English bilingualism and language shift).

---

[2] The varieties of French often discussed in English historical linguistics have been called Anglo-Norman, Anglo-French and insular French (see Wogan-Browne 2009 for more). The present chapter normally refers to French only, without further qualifications.

Rather than post-Conquest materials, research into multilingual writing practices, code-switching and code choice has focused on texts from the later medieval period. Such studies (e.g. Wenzel 1994; Pahta 2003; Machan 2011; Trotter 2011; Wright 2013) have both broken new ground in multiple ways and applied and refined concepts and methods originally introduced in research into contemporary conversational code-switching. More work is also being done on code-switching involving Early or Late Modern English (e.g. Tuominen and Nurmi et al., this volume). By contrast, there has been fairly little research into code-switching in Anglo-Saxon England (but see Schendl 1996–this volume; Timofeeva 2010). In a rare diachronic study, Pahta and Nurmi (2006: 205) have, nonetheless, estimated that the quantities of code-switching in Old and Middle English are quite similar, since the Old English text excerpts of the *Helsinki Corpus of English Texts* contain 1.8 code-switched units per 1000 words on average, and the Middle English sub-corpus only slightly more (2.2). In the light of this overall assessment, it is hardly reasonable to assume that the transitional period bridging Old and Middle English would be completely devoid of code-switching.

A search for code-switching research on the long twelfth century yields little more than fleeting mentions. In *Code-Switching in Early English* (Schendl and Wright 2011c), there is one chapter on Old English and many on Late Middle English, while none focus on the intermediate period. For Middle English scholars, an obvious place to start locating earlier language mixing would be *Ancrene Wisse*, the best-known religious text of the period: it has indeed been mentioned by Schendl (1996: 53), and Trotter (2003: 85), in his re-evaluation of its vocabulary, discusses bilingualism and hybridisation, regarding code-switching as “concomitant” with and a “constituent” of multilingualism. Skaffari (2009) refers to code-switching occasionally in his discussion of lexical borrowing into Early Middle English (e.g. subsection 4.2.4 on the switching–borrowing issue and the case study in chapter 7). Faulkner (2012b) calls for a multilingual perspective on literature in post-Conquest England but does not directly discuss code-switching, and Treharne (2012), analysing manuscripts containing English, contemplates the multilingual situation but does not mention code-switching as such. Considerably more attention has indeed been paid to other outcomes of language contact at this time, and the question of multilingualism itself: the importance of French in the medieval linguistic landscape (e.g. Wogan-Browne et al. 2009; see also Ingham 2012), the influx of French-derived vocabulary (see e.g. Durkin 2014: 223–280 for a recent dictionary-based account and Rothwell 1998 for the importance of insular adoption in lexical borrowing from French), and the extent and duration of English–French bilingualism in the upper classes (e.g. Berndt 1976; Short 1992) have figured prominently in linguistic work on post-Conquest

England. The relative novelty of French in England should not, however, conceal the continued presence and increasing importance of Latin at this time (see e.g. Hunt 1991; Morgan and Thomson 2008: 25).

3 The manuscripts

The temporal gap identified above does not by any means apply to just historical code-switching research. In English historical linguistics, there seems to have been a tendency to avoid focusing on this "difficult" period alone since it provides relatively small datasets to examine. While having to cope with scanty evidence understandably attracts few linguists, more material exists than is often acknowledged: for example, Treharne (2012: 98–101) lists some eighty manuscripts "with main texts in English" from between the 1050s and c. 1100 – ample evidence that English did not stop being written immediately after the Norman Conquest. A great deal more material is nonetheless available from the Late Middle English period, and also the preceding Old English offers substantial quantities of data, especially if compared to other European vernaculars of the same date.[3] However, many of the Old English texts are actually preserved in somewhat later, post-Conquest manuscripts. They are not just copies made out of antiquarian interest but material which was still utilised in post-Anglo-Saxon England and therefore compiled into new, relevant selections (see Irvine 2000: 42–43; Swan 2006: 156; Faulkner 2012a: 182; Treharne 2012: 96–98 et passim). When compared to this material transmitted from the Old English period, original compositions from post-Conquest England are relatively rare until at least the first half of the thirteenth century.

An excellent resource for exploring English in post-Conquest sources is *A Linguistic Atlas of Early Middle English, 1150–1325* (*LAEME*), and the *Catalogue* of its sources (Laing 1993). The contents of the manuscripts are divided into several categories in the *Catalogue*: literature, documents, glosses and mixes of any of the above. The present chapter is part of a research project drawing on manuscripts dated in the *Catalogue* to a potentially pre-1250 date, of which there are almost 270. While this may seem a large number, it has to be admitted that it conceals the variedness of the material, as it does not distinguish between a book containing several English texts and, for example, a Latin manuscript with some vernacular lyrics scribbled on a single folio: either type counts as one item in this account. Nevertheless, both types serve as viable examples of medieval writing practices in multilingual England. To complement the *Catalogue*, I have also consulted another irreplaceable resource, the *Production and Use of English Manuscripts 1060 to 1220* project site (edited by Da Rold et al., 2010–2013; abbreviated here *EM*), with information about over two hundred manuscripts. Partly overlapping with the temporal scope of the *Catalogue* and therefore its list of manuscripts, *EM* helps in locating additional material down to the decade of the Norman Conquest and also provides more detailed descriptions of the manuscripts than the *Catalogue* aims to do.[4] Most of the long-twelfth-century material is preserved in the repositories of London, Oxford and Cambridge: in both *EM* and the pre-1250 coverage of the *Catalogue*, the share of the manuscripts held at the repositories of these three cities is approximately 84 per cent.

---

[3] As Treharne (2012: 104) points out, "all of the English manuscripts and documents that have survived from England to *c*.1200 and beyond are remarkable testimony to a vernacular textual tradition that is rarely paralleled elsewhere in the medieval world".

The present chapter is not intended to be an exhaustive survey of all potentially relevant manuscripts but has for practical reasons a narrower focus. My main objective has been to examine manuscripts in which the English component is very broadly classifiable as literature (e.g. Laing 1993: 5–6), including all types of religious writing.[5] The predominant domain of vernacular English literature at the time was religion (e.g. Treharne 2012: 12), as exemplified by compilations of Old English homilies as well as many of the newer texts. Moreover, the examples provided and manuscripts mentioned in this chapter come from a specific period, that is, from between the last quarter of the eleventh century and the second of the thirteenth. Thus, material dated in the catalogues to, for example "C13", is excluded from closer analysis here; although the temporal boundaries of the long century are malleable according to the researcher's needs (see e.g. Brett 2015: 1), they should not be too liberal.

Overall, I have by the time of writing consulted over 110 long-twelfth-century manuscripts.[6] Fifty-four of them contain literary material specifically from the period 1075–1250; these are the sources of the present study. In two thirds of these manuscripts, Latin is the dominant language. While attention has been paid to texts with English as an embedded language as well as material with Old or Early Middle English as the matrix, all of the linguistic examples cited below have an English matrix, though some manuscripts with Latin as the main language are also discussed. The material includes some of the best-known vernacular texts of the period, *Ancrene Wisse* being a prime example; as Schendl and Wright (2011a: 5, 8–9) note, it is useful for scholars to return to much-researched texts and genres equipped with the novel approach of code-switching.

---

[4] From a code-switching studies point of view, the problem with both resources is that while they identify the presence of English in manuscripts containing also material in other languages, their main focus does not lie on multilingualism, nor can they provide information about instances of intrasentential code-switching, especially single words on the switching–borrowing cline, which appear at the "bottom level" (see 4.4 below). The latter point is a major reason why a quantification of the code-switching sequences in the manuscripts is not possible here.

[5] For example, just under one half of the London/Oxford/Cambridge material is classified as literary in the *Catalogue*.

[6] The information provided in catalogues cannot replace viewing the code-switch in its physical manuscript context.

4 Code-switching: Towards locating levels and patterns

What is evident from the two sections above is that while the sources containing English cannot compete with later Middle English manuscripts in terms of quantity of extant material or range of domains, it is larger than has often been acknowledged. Its multilingual features have not been explored as such, despite the evidence that code-switching involving English is attested both earlier and later in medieval England. Section 4 endeavours to start such exploration by proceeding from the largest textual units down towards the smallest grammatical or linguistic ones, starting with multilingualism within manuscripts (4.1) and followed by code-switching represented visually on the manuscript page (4.2), utilised within texts (4.3) and, finally, occurring within sentences and clauses (4.4).[7] The emphasis is on English–Latin code-switching; French is discussed separately in Section 5. Below, each of these levels is discussed briefly and illustrated with some manuscripts and linguistic examples from the period. Accounts of code-switching at different levels have been published before (see e.g. Pahta 2003: 199–206), and also appear in the present collection (see Kopaczyk, this volume). The present division is based on observable forms, not potential functions of or motivations for code-switching. However we define the levels at which code-switches can be observed, overlap and multiple interpretations are possible, as will also become evident.

4.1 Manuscript level

Whole manuscripts, regarded here as the topmost level at which code-switching takes place, are often not included at all in historical code-switching. Schendl and Wright (2011b: 23–24), for example, dismiss "the frequent occurrence of different monolingual texts collected in a single manuscript [. . .], though such 'intertextual' switching 'often resembles the situational code-switching typical of diglossia', and equally constitutes an important testimony for the widespread multilingualism of the literate part of medieval society". A broader and more inclusive definition of code-switching – or, simply, multilingual practices – would, however, also include monolingual texts appearing in one and the same manuscript but written in different languages, that is, intertextual switching. Even if the authors or scribes had not had any reason to code-switch within the texts they were composing or copying, a subsequent user of the material found it useful to collate the texts, producing one manuscript with material in more than one code, which would then be encountered by a reader. The production of such a book would not be motivated by the monolingual expectation of some intended readership, but by the commissioner's or copyist's desire to own, read or produce a book united by a single genre, the same type of subject matter, or some such factor related to his or her needs and abilities (see e.g. Voigts 1996: 819 and Honkapohja, this volume, for examples of medical manuscripts).[8]

---

[7] In addition, some claim that a switch point may be located within a word, but this level is not considered here.

A good example of a manuscript with a vernacular text poised between Latin materials is Oxford, Bodleian Library, MS Digby 4. After nearly a hundred folios of Latin, it contains *Poema Morale*, an important Early Middle English poem surviving in several manuscripts, on ff. 97r–110v, followed by two short texts in Latin. Another example of language-mixing at the manuscript level is Worcester, Cathedral Library, MS Q 29, in which an English-language Nativity sermon appears between Latin sermons and a Latin text on computus, beginning near the bottom of the second column on f. 136v and continuing until the top part of the second column on f. 137r; it appears to be an integral text on these folios, not a later addition loosely connected to the co-text. Within the English sermon, there is also some code-switching into Latin, which underlines the fact that a single source may of course contain multilingual features at more than one level.[9]

A third example, London, British Library, MS Cotton Claudius D iii, contains Latin calendars, some verse material in French and the Benedictine Rule, in which each Latin chapter is followed by an Old English one, reiterating the content of the preceding chapter. As this particular text is an adaptation of the Rule for the use of nuns (Laing 1993: 73), the presence of English seems analogous to the use of the vernacular – rather than (only) Latin – in other religious texts for women, *Ancrene Wisse* being the most famous example from this period. The Latin-English Rule does, however, differ from the two manuscripts described above and from Cotton Claudius D iii itself as a complete manuscript: in the Rule, the passages in two languages can be regarded as part of a single bilingual whole (for code-switching at text level, see 4.3), whereas genuine intertextual code-switching takes place between separate texts contained within the same book.

---

[8] Applying this idea may be challenging: for example, a codex containing manuscript material produced c. 1200 may have been bound much later and therefore hardly represents long-twelfth-century multilingual practices from the producers' point of view.

[9] Stanley (1961) has edited the sermon.

In the examples of intertextual code-switching above, the monolingual units appear consecutively in the manuscript, but this is not the only form of language-mixing at this level. Often margins have been put to good use by contemporary scribes or subsequent annotators, whose chosen language may not be that of the body text; little scraps of English, for example, appear on the last folios of Latin manuscripts. These could potentially be regarded as extrasentential switches, in both the literal and technical sense of the word. As a code-switching term, "extrasentential" is associated with tag switching, the use of such embedded-language units as interjections, which do not have a clear grammatical role in sentence structure; like Poplack's (1980: 589) tags, they "may be inserted almost anywhere". Material written in the margins does not, however, have to constitute a separate text, as will be seen in Section 5 below, with French rather than Latin appearing with English.

4.2 Page level

The page level refers to what the reader observes on the manuscript page or opening as scribal output; in addition to the verbal content, it also encompasses higher and lower level visual elements (see Varila et al. 2017), namely, the layout of the manuscript (see Parkes 1991 for *mise-en-page*) and the choice or appearance of the script. While it was typical of English manuscripts until well after the Conquest that separate scripts were employed for Old English and Latin – Insular and Caroline, respectively (e.g. Treharne 2000: 25) – the visual aspect has not been considered in historical code-switching research until quite recently. Machan's "visual pragmatics" (2011) is related to such concepts as "visual prosody" (Meurman-Solin 2013) and "pragmatics on the page" (Carroll et al. 2013). Essentially, the visual approach proposes that linguistic forms are not the sole means of conveying a message to the recipient by the producer, but that the verbal content is to be considered also in its material and visual context. Machan himself examines the use of red ink in connection with Latin code-switches in late medieval manuscripts of the English *Piers Plowman* and some other works. Other methods of marking a change from one language to another include the use of a different script or the same script in a different size; today we of course often see the same function performed by the use of italics.

Many of the manuscripts examined here are visually quite modest. In London, British Library, MS Stowe 34, red underlining does appear in connection with Latin quotations and other, shorter switches into Latin.[10] The few cases in which the Latin is not underscored may simply be oversights by the rubricator, who was probably not one of the two main scribes whose work the manuscript is (as indicated in *LAEME*). This is illustrated by (1): while the rubricator has underlined the first quotation from a psalm, which is followed by an English translation, he has missed the following one, from another psalm.

(1) Of hem sade ðe ꝑphete. **Fuerunt m̄ lacrime me panes die ac nocte** [underlined in red]. mine teares he sade me waren bred daiȝ ⁊ niht swa gode hie þouhten. Of oþres kennes teares he sade **lac̄mis meis stratuū meum rigabo** [not underlined]. Ich scal watrien min bedd mid mine teares. (London, British Library, MS Stowe 34, f. 47r)[11]

'Of them said the prophet: *Fuerunt mihi lacrimae meae panes die ac nocte*, "My tears," said he, "were my bread day and night; so good they seemed [to be]." Of other kinds of tears he said: *Lacrimis meis stratum meum rigabo*, "I shall water my bed with my tears."' (Holthausen 1888: 146)

Highlighting the Latin in a vernacular text – which may not be motivated primarily by the desire to pick out the embedded language but to mark important components of the discourse, such as quotations from authoritative sources (see also Machan 2011: 314) – is not unusual or unexpected in medieval manuscripts.[12] Visual flagging adds weight to the selected words or passages and also helps the reader – even the researcher, centuries later – to locate them, but there are plenty of examples in the manuscripts with little or no visual marking of quotations or other instances of code-switching. In manuscript culture, any type of decoration was always linked with cost, in time, effort and ink, so even if the motivation existed, there may not have been enough resources. It is typical of vernacular manuscripts of this period that they look modest, according to Treharne (2012: 97): they were meant to be read, not just looked at.

---

[10] *Vices and Virtues*, the only text the manuscript contains, is rich in code-switches. For more information, see Skaffari (2016).

[11] The code-switched units are bolded.

[12] The visual appearance of English within Latin texts and passages seems to be less conspicuous.

Language-mixing can also be interlinked with how entire pages are laid out. An example from the more elaborate end of the range of manuscripts produced during this period, the Eadwine Psalter (Cambridge, Trinity College, MS R.17.1, from the middle of the twelfth century) is a large and intricate book, containing the psalms in different Latin versions and glossed in three different languages. In a recent analysis, Treharne (2012: 173) discusses the Psalter in terms of multilingualism on the page, or "the praxis of a multimedia multilingualism", explaining the position and use of English in the book. Whereas the most prominent position on the page is given to the preferred Gallicanum version of the Psalms, the adjacent versions are in smaller script and accompanied by interlinear glosses in the vernaculars of England – English and French. Treharne points out that the version glossed in English is the one that is placed closest to the important Gallicanum Psalms; English is thus less marginal than French, which appears at the edge of the page.

4.3 Text level

Language-mixing may also be observed within individual texts, the next level at which code-switching takes place. Much of code-switching here can simply be labelled as intersentential or interclausal. Code-switching at the text level is not, however, uniform, nor is it only comprised of clauses or sentences embedded in the body text.

An easily observable practice in medieval manuscripts is the use of the embedded language in a discourse-organising or text-structuring function: for instance, headings may be in Latin although the body text is in English, and similarly the incipits and explicits. There is plenty of evidence of this practice in later texts, but the long twelfth century also provides a variety of examples, such as the typical closing formula occurring in *Ancrene Wisse* (2).

(2) Ase ofte as ʒe habbeð ired eawiht her on, greteð þe leafdi wið an Aue, for him ꝥ swonc her abuten. Inoh meaðful ich am þe bidde se lutel. ¶**Explicit**. Iþench o þi writere i þine beoden sumchearre, ne beo hit ne se lutel. Hit turneð þe to gode, ꝥ tu bidest for oðre. (Cambridge, Corpus Christi College, MS 402, f. 117v)

'Whenever you have read anything in this book, salute our Lady with a Hail Mary for the man who worked on it. I am moderate enough in asking so little. *The End*. Remember your scribe sometimes in your prayers, no matter how little. It will benefit you if you pray for others.' (Millett 2009: 165)

In (2), the Latin word *explicit* is preceded by a red-and-blue paraph. It does not conclude the whole text but instead is followed by a short colophon, in the same hand as the text above it, which offers the last piece of instruction in this penitents' manual.

At this level, code-switching does not exclusively appear at boundaries of various sections within a text but also within passages of text. There is also variation in length, from short clauses to longer units used intersententially. Example (3) is from *Sawles Warde* – a member of the *Ancrene Wisse* group of Early Middle English religious prose – and features a biblical quotation from Matthew.

(3) Her of ha herieð godd ⁊ singeð a unwerget eaū iliche lusti in þis loft songes. as hit iwriten is. **Beati q̄ habitant**. ⁊ **c̄**. Eadi beoð þeo laūd. þe i þin hus wunieð. (Oxford, Bodleian Library, MS Bodley 34, f. 80r)

'Thereof they praise God, and ever unwearied, ever alike joyful, they sing this song of praise [as it is written], *Beati qui habitant &c*. – Blessed are those, O Lord, who dwell in thine house.' (Morris 1868: 262, 264)

The Latin switch is short, a truncated quotation serving as a cue pointing to an authoritative source.[13] There are numerous similar cases, but they may not have the same type of inter-passage text-organising function as discussed above. The use of another language does, however, help to identify them as quotations and therefore as rhetorically distinct from their immediate co-text.

There is overlap between this level and the previous one, as the embedded units may be visually flagged. This we see in (2) with the paraph before the Latin word, while the quotation in (3) does not stand out from the vernacular text; MS Bodley 34 is not visually striking and has very little colour (red). Embedded words and phrases may also be flagged at the clause level.

4.4 Clause level

The clause level is where code-switching is intrasentential (intraclausal), comprising phrases or individual words. A familiar concern of code-switching researchers is whether – and how – intrasentential single-word switches can be distinguished from loanwords.[14] In the present material, there are plenty of examples of this form of code-switching, often featuring names and terms of various types – a not untypical type of code-switching in writing (see e.g. Matras 2009: 107; Pahta and Nurmi 2006: 213–214; international names may also be viewed as diamorphs; see ter Horst & Stam, this volume). A few examples will suffice here, presented without co-text for the sake of brevity. Part 1 of *Ancrene Wisse* deals with the daily devotional life of the anchoress and is therefore rich in references to Latin prayers: Cambridge, Corpus Christi College, MS 402 contains on a single page (f. 6r) "Placebo" ('I shall please') twice and eight other instances of very short switches (one or two words). The late entries of the *Peterborough Chronicle* occasionally mention hymns such as "Te deum laudamus" (Oxford, Bodleian Library, MS Laud Misc. 636, f. 88r). In London, British Library, MS Stowe 34, both vices and virtues are typically introduced with their Latin names, essentially religious terminology, although unlike personal and place names as well as the songs and prayers mentioned above, they are usually accompanied by English translations or explanations (4).

---

[13] Schendl, this volume, calls such quotations "incomplete". In another manuscript containing *Sawles Warde*, London, British Library, MS Royal 17 A xxvii, the same quotation is longer: "Beati qui habitant in domo tua dn̄e" (f. 9v).

[14] See, for example, Queiroz de Barros and Schendl, this volume, for a discussion of the topic.

(4) AN hali mihte is icleped **fides recta**. ꝥ is rihte ȝeleaue hie is angīn of alle cristendome (London, British Library, MS Stowe 34, f. 8v)

'One holy virtue is called *fides recta*, that is, right belief. It is the beginning of all Christianity.' (Holthausen 1888: 24)

The Latin term is followed by the corresponding English phrase informing the reader, who may not have understood much Latin, of the meaning of the switch. Diller (1997–1998: 510) calls this "English support" (see also [1] and [3] above).

The examples above appear uninflected in the sentences where they are attested – they may be used in subject or similar positions or be difficult to adjust grammatically if they are full Latin structures containing internal inflection (e.g. "Te Deum laudamus") – but Latin inflection is occasionally visible, as at a few points in MS Stowe 34 (example 5).

(5) ðe hali apostel nāneð ðese þrie haliȝe mihtes togedere. **Fidem. Spem. Karitatē**. And seggeð ꝥ þies ðe hatte **karitas** is heiȝest ⁊ betst of ðese þrie. (London, British Library, MS Stowe 34, ff. 11v–12r)

'The holy apostle calls these three holy virtues together, *fidem, spem, caritatem*, and says that this which is called *caritas*, is the highest and best of these three.' (Holthausen 1888: 34)

Here all three Latin nouns, underlined in red ink in the manuscript, are inflected in the accusative case, being governed or controlled by the English verb for 'call, mention'. Similar examples are found in some of the manuscripts consulted. English itself had by this time lost much of its inflectional morphology.

In addition to Latin words and phrases embedded in English sentences, there are French-derived lexical items whose status in the English vernacular is not always obvious. While there are loanwords which seem to have become relatively well-established in Early Middle English (such as *grace* '(God's) grace, mercy', *pes* 'peace' and *serven* 'serve, work for'; see Durkin 2014: Chapter 12 and Skaffari 2009: 153–168 for more examples and analysis), some items of French origin might be viewed as intrasentential code-switching, leading into lexical borrowing in some cases. Furthermore, there are occasional longer stretches of French in some of the manuscripts, which is discussed below.

5 Code-switching into French

As mentioned under 2 above, it is natural to expect changes in multilingual practices to have taken place in England after 1066 and for them to become visible in material produced at that time. The most striking difference one can anticipate must be associated with French, whose strengthened presence cannot be overlooked in studies of spoken and written communication in post-Conquest England. At the societal and cultural macro level, the growing importance of French is reflected in the production and dissemination of Anglo-Norman texts in the twelfth and thirteenth centuries (e.g. Lodge 1992: 81); it has been noted by Short (1992: 229) that "French literature begins, to all intents and purposes, in twelfth-century Anglo-Norman England". However, while this body of Anglo-Norman literature, comprising historical writing, hagiography, scientific texts, verse, and so forth, is certainly a remarkable feature of text production in England at this time, it has not been explored within the present project.[15] Some observations can nonetheless be made on the basis of the other sources consulted.

Individual English texts and manuscripts suggest that the contact with French did not have much impact at the earliest stage. In his exploration of the third edition of the *Oxford English Dictionary*, Durkin (2014: 257–264) finds that the most remarkable peaks in lexical borrowing from French (his "French only" category) appear in the first half of the fourteenth century, whether we consider the greatest leap upwards from the previous half-century or the largest share of

---

[15] The amount of English in post-Conquest Anglo-Norman manuscripts "remains to be worked out" (Da Rold and Swan 2011: 261, footnote 14).

French loans of all new vocabulary (that is, in absolute or relative terms); the fifty years following 1200 pale in comparison with these peaks, although they do clearly surpass the preceding fifty years. Examining the use of almost three hundred French-derived words in Late Old and Early Middle English text excerpts, Skaffari (2009: 192–195) identifies a statistically significant increase in the number of loanwords from the second half of the twelfth century to the first half of the thirteenth. This diachronic change also shows when texts *copied* from Old English sources between 1150 and 1250 are compared with those presumably originally *composed* within this period. In either case, there is much less evidence of French influence in the older material, although that, too, postdates the Norman Conquest by approximately one hundred years.

Unlike lexical borrowing, code-switching between English and French in post-Conquest England has received little attention. The coexistence of English and Latin in written material, discussed at length above, is perhaps unsurprising given the role of Latin as a High language associated with religion, learning, administration and the law – as a medieval "language for specific purposes" in multiple domains. While Latin was a ubiquitous feature of medieval culture, it is improbable that the use of French could have been rare and restricted to monolingual settings only, be they conversation at the highest levels of society or monolingual literature composed for the Anglo-Norman elite. It is noteworthy, however, that in the *Helsinki Corpus of English Texts*, Pahta and Nurmi (2006: 205–206) find as few as three French switches per 100,000 words of Middle English text, each of them on average 14 words; switches to Latin are over sixty times more common in the Middle English sub-corpus. This alone suggests that while code-switching was an established practice in medieval English writing, switching between the vernaculars was not particularly frequent in texts. The long twelfth century does not seem to have been an exception to this.[16]

French words, phrases and clauses do appear in the sources examined for the present chapter, but in a clear minority of cases. *EM* contains a helpful index of manuscripts containing material in French, with 28 entries. Judging by the index, the earliest records of French in manuscripts containing English were glosses. Marginalia also appear, and towards the end of *EM*'s coverage, there are some French-language verses and proverbs.[17]

---

[16] Code-switching between vernaculars is also discussed by Ingham, this volume.

[17] Unfortunately the descriptions provided in *EM* do not always specify where in the manuscript the French appears, as is the case with London, British Library, MS Cotton Faustina A ix, which is listed in the index of manuscripts with French, but the description of its contents, appearance and history (Treharne 2010–2013) does not mention either French or Anglo-Norman.

One of the earliest switches into French in the material considered here is in Cambridge, University Library, MS Ii. 1. 33. This manuscript, dated to the second half of the twelfth century, is a collection of Ælfric's homilies, saints' lives and a translation of parts of *Genesis* (Laing 1993: 46; Da Rold 2010–2013). It is essentially an example of Old English rewritten in the twelfth century, to echo the title of the ground-breaking collected volume by Swan and Treharne (2000), but, as is often the case, it is not a monolingual manuscript: it contains not only occasional Latin incipits, additions and annotations but also two instances of French, in the bottom margins of ff. 70v and 120r. The first of these appears in Ælfric's homily on St Andrew and reads (6):

(6) Icest auint en achaia. dunt plusur unt ^{oi} parler. / dedenz la cite d^{e} patras que u^{us} auez oi numer. (Cambridge, University Library, MS Ii. 1. 33, f. 70v)[18]

'This happened in Achaea, of which many have heard speak, [/] within the city of Patras, which you have heard named.' (Traxel 2004: 77, note 99)

As this example is a marginal addition, it is not self-evident that it should be considered a code-switch at all, although we certainly see multilingualism at work on a manuscript page here. In his book on the manuscript, Traxel (2004: 77–79) devotes three pages to this addition but does not use the term "code-switching"; other recent work on this manuscript (Irvine 2000: 54–55; Swan 2000: 78–80; Treharne 2012: 165) does not mention the French at all, with the exception of Da Rold and Swan (2011). While the addition on f. 70v is not embedded inside or between English clauses – it is truly extrasentential (see 4.1) rather than intra- or intersentential – it does appear "within one communicative event" (Schendl and Wright 2011b: 23), constituted by reading this particular homily. When reading this page, the reader is unlikely to disregard entirely the addition, which is clearly visible in the margin and which, like the longer English addition equally visible on the same page, contributes to the narrative on St Andrew; both supplement information about his death but do not repeat each other's content.[19] Moving from the recipient to the producer, there is more to suggest that the addition is no accident: the manuscript is probably the work of two main scribes according to *EM* (Da Rold 2010–2013), or possibly four

---

[18] The manuscript is available for viewing in the Cambridge Digital Library. The French addition appears at the bottom of the folio at http://cudl.lib.cam.ac.uk/view/MS-II-00001-00033/154.

[19] The "multilingual experience" of an individual reader or recipient may be quite different from that of the scribe or producer. The definitions of code-switching tend to focus more on production than reception – an important and intriguing issue but not, however, within the scope of this chapter. Da Rold and Swan (2011: 263–264) also touches on these issues, not so much with respect to code-switching but manuscript production.

(Traxel 2004: 37–59), but the scribe responsible for the French here also wrote most of the rest of the manuscript, although not the body text on this particular folio. The date of the addition could therefore be relatively close to that of the body, not decades or centuries later as some additions or corrections often are. Moreover, as both the English and the French addition were written on lines specifically ruled for them (Da Rold 2010–2013), they were not mere scribbles. The motivation behind code selection is not, however, easy to discern here: Traxel (2004: 78) suggests that the shorter addition may have been copied from an unknown French manuscript, and Da Rold and Swan (2011: 263) propose that the marginalia were additional material for readers to use in writing or preaching. Despite its language and position in the margin, the French text is part of the same message as the rest of the page. It is therefore not intertextual, manuscript-level code-switching but can be placed at the page and text levels.

Another example of French in an early English text is less controversial but still appears in the periphery of a text rather than at its centre. In one of its three manuscripts – London, British Library, MS Royal 17 A xxvii – *Sawles Warde* ends in a colophon, with both Latin and French in the English text (7):

(7) **Par seinte charite** biddeð a **pater nost̄** for iohan ꝥ þeos boc w̄t. Hwa se þis writ haueð ired. Ant crist him haueð swa isped. Ich bidde **ꝑ seinte charite**. þet ȝe bidden ofte for me. Aa **p̄r n̄r**. ant **aue marie**. þet ich mote ꝥ lif her drehen. Ant ure lauerd wel icwemen. I mi ȝuheđe ⁊ in min elde. þet ich mote iħu crist mi sawle ȝelden. AMEN. (London, British Library, MS Royal 17 A xxvii, f. 10v)[20]

‘*Par seinte charite* [for the sake of charity], pray a *Pater noster* for John, who wrote this book. Whoever has read this writing and Christ has so prospered him, I pray, *par seinte charite*, that you pray often for me a *Pater noster* and *Ave Maria*; so that I may so lead my life and well please our Lord, in my youth and in my old age, that I may yield my soul to Jesus Christ. Amen.’ (modernised from Morris 1868: 266)

As observed by Skaffari (2003: 92), the French phrase is an example of code-switching, “par seinte charite” used twice in the scribe’s personal conclusion to the text. It can easily be seen as an example of tag switching, which is underlined by its appearance in two different positions in the clauses, intensifying the scribe’s entreaties. Just as in (6), the French switches do not appear as core content.

---

[20] Some of the visual detail is lost in presenting this and other examples in print, but this is unlikely to be essential from the code-switching point of view.

Overall, the role of French in written code-switching does not seem to be very prominent at this time. The reason for this preliminary conclusion, however, depends firstly on the focus of my research, which lay on material with English as the matrix or the major embedded language, which led to the exclusion of insular French manuscripts. Secondly, the distinction between the two vernaculars may have been less relevant than the one between Latin and vernacular, as proposed by David Trotter (e.g. 2011: 182). This does not mean that the vernaculars were equal; on the contrary, the Anglo-Normans regarded English as an uncouth, barbaric language, as noted by Treharne (2012: 151), who otherwise stresses the value and versatility of English as a language of writing.

6 Concluding remarks

This study set out to provide an overview of code-switching in English sources from the long twelfth century. Although there has been a notable gap in the coverage of multilingual practices at this time, there is absolutely no reason to overlook this period as an object of historical code-switching research. Familiar as the main features of the post-1066 linguistic situation may seem to be, there is much more that we should know and understand about multilingual writers, readers and texts. The body of extant material is not as small as has been previously suggested, but there is plenty of variation in how much code-switching it contains, and at which levels. Latin is the dominant language in the majority of the sources, and its use also persists within the vernacular English texts in the domain of religion, whereas the role of French is remarkably small in the manuscripts examined.

We have discussed four levels at which code-switching can be observed in medieval (and other) sources – manuscript, page, text and clause – corresponding to intertextual, visually flagged, interclausal and intraclausal switching, as well as the somewhat problematic case of extrasentential code-switching, the “other” language not appearing within the text but, rather, marginalised on the edges of the page. These are the levels at which multilingual features are observable, but it is worth noting that the same manuscript may contain examples of code-switching at multiple levels, and that visual flagging coincides with features at other levels: it is a way of highlighting inserted or alternating words, phrases or passages, and not independent of them. Moreover, the inclusion of the visual aspect serves as a reminder of the usefulness of consulting original manuscripts instead of simply relying on editions or catalogues.

The current study has only been able to draw on some of the relevant manuscripts from the period. Another limitation is that while selecting sources on the basis of their descriptions in catalogues can provide a bird's eye view of the material, it may nonetheless obfuscate what takes place at the lower levels. This applies perhaps most of all to intraclausal single-word switching, with its risk of confusion with lexical borrowing. It is possible to overcome these limitations in future research by including more of the extant manuscripts in a principled manner but at the same time pausing to examine at least some of them with much greater attention to the minutest of details. More work is still needed to place the long twelfth century firmly on the map of historical code-switching research.

References

Primary sources

Cambridge, Corpus Christi College, MS 402. http://parkerweb.stanford.edu/parker/actions/zoom_view.do?ms_no=402&page=1R.
Cambridge, University Library, MS Ii. 1. 33. http://cudl.lib.cam.ac.uk/view/MS-II-00001-00033/1.
London, British Library, MS Cotton Claudius D iii
London, British Library, MS Royal 17 A xxvii
London, British Library, MS Stowe 34
Oxford, Bodleian Library, MS Bodley 34
Oxford, Bodleian Library, MS Digby 4
Oxford, Bodleian Library, MS Laud Misc. 636
Worcester, Cathedral Library, MS Q 29

Secondary sources

Berndt, Rolf. 1976. French and English in thirteenth-century England: An investigation into the linguistic situation after the loss of the Duchy of Normandy and other continental dominions. In Heinrich Scheel (ed.), *Aspekte der anglistischen Forschung in der DDR: Martin Lehnert zum 65. Geburtstag*, 129–150. Berlin: Akademie-Verlag.

Brett, Martin. 2015. Introduction. In Martin Brett & David A. Woodman (eds.), *The long twelfth-century view of the Anglo-Saxon past*, 1–9. Farnham: Ashgate.

Carroll, Ruth, Matti Peikola, Hanna Salmi, Mari-Liisa Varila, Janne Skaffari & Risto Hiltunen. 2013. Pragmatics on the page: Visual text in late medieval English books. *European Journal of English Studies* 17(1). 54–71.

Da Rold, Orietta. 2010–2013. Cambridge, University Library, Ii. 1. 33. In Orietta Da Rold, Takako Kato, Mary Swan & Elaine Treharne (eds.), *The production and use of English manuscripts*

1060 to 1220. University of Leicester. http://www.le.ac.uk/english/em1060to1220/mss/EM.CUL.Ii.1.33.htm (accessed 26 April 2015).

Da Rold, Orietta & Mary Swan. 2011. Linguistic contiguities: English manuscripts 1060–1220. In Elizabeth M. Tyler (ed.), *Conceptualizing multilingualism in medieval England, c. 800–c. 1250* (Studies in Early Middle Ages 27), 255–270. Turnhout: Brepols.

Diller, Hans-Jürgen. 1997–1998. Code-switching in medieval English drama. *Comparative Drama* 31(4). 506–537.

Durkin, Philip. 2014. *Borrowed words: A history of loanwords in English*. Oxford: Oxford University Press.

EM = The Production and Use of English Manuscripts 1060 to 1220. 2010–2013. Edited by Orietta Da Rold, Takako Kato, Mary Swan & Elaine Treharne. University of Leicester. http://www.le.ac.uk/ee/em1060to1220 (accessed 22 September 2014).

Faulkner, Mark. 2012a. Archaism, belatedness and modernisation: "Old" English in the twelfth century. *Review of English Studies* 63(259). 179–203.

Faulkner, Mark. 2012b. Rewriting English literary history 1042–1215. *Literature Compass* 9(4). 275–291.

Holthausen, Ferd[inand] (ed.). 1888. *Vices and virtues, being a soul's confession of its sins, with reason's description of the virtues: A Middle English dialogue of about 1200 A.D.* (Early English Text Society Original Series 89). London: N. Trübner & Co.

Hunt, Tony. 1991. *Teaching and learning Latin in thirteenth century England*, 3 vols. Cambridge: D. S. Brewer.

Ingham, Richard. 2012. *The transmission of Anglo-Norman: Language history and language acquisition* (Language Faculty and Beyond 9). Amsterdam & Philadelphia: John Benjamins.

Irvine, Susan. 2000. The compilation and use of manuscripts containing Old English in the twelfth century. In Mary Swan & Elaine M. Treharne (eds.), *Rewriting Old English in the twelfth century* (Cambridge Studies in Anglo-Saxon England 30), 41–61. Cambridge: Cambridge University Press.

Kwakkel, Erik. 2012. Biting, kissing and the treatment of feet: The transitional script of the long twelfth century. In Erik Kwakkel, Rosamond McKitterick & Rodney Thomson (eds.), *Turning over a new leaf: Change and development in the medieval manuscript* (Studies in Medieval and Renaissance Book Culture), 79–125. Leiden: Leiden University Press.

LAEME = A Linguistic Atlas of Early Middle English, 1150–1325. Compiled by Margaret Laing and Roger Lass. Edinburgh: © 2007 University of Edinburgh. http://www.lel.ed.ac.uk/ihd/laeme1/laeme1.html.

Laing, Margaret. 1993. *Catalogue of sources for* A Linguistic Atlas of Early Medieval English. Cambridge: D.S. Brewer.

Lodge, R. A. 1992. Language attitudes and linguistic norms in France and England in the thirteenth century. In P. R. Coss & S. D. Lloyd (eds.), *Thirteenth century England IV: Proceedings of the Newcastle Upon Tyne conference 1991*, 73–83. Woodbridge: Boydell.

Machan, Tim William. 2011. The visual pragmatics of code-switching in late Middle English literature. In Herbert Schendl and Laura Wright (eds.), *Code-switching in early English* (Topics in English Linguistics 76), 303–333. Berlin & Boston: Mouton de Gruyter.

Matras, Yaron. 2009. *Language contact*. Cambridge: Cambridge University Press.

Meurman-Solin, Anneli. 2013. Visual prosody in manuscript letters in the study of syntax and discourse. In Anneli Meurman-Solin and Jukka Tyrkkö (eds.), *Principles and practices for the digital editing and annotation of diachronic data* (Studies in Variation, Contacts and

Change in English 14). Helsinki: Research Unit for Variation, Contacts, and Change in English. http://www.helsinki.fi/varieng/journal/volumes/14/meurman-solin_a/ (accessed 3 July 2013).

Millett, Bella (transl.). 2009. *Ancrene Wisse – guide for anchoresses: A translation based on Cambridge, Corpus Christi College, MS 402*. Exeter: University of Exeter Press.

Morgan, Nigel & Rodney M. Thomson. 2008. Language and literacy. In Nigel Morgan & Rodney M. Thomson (eds.), *The Cambridge history of the book in Britain*, vol. II: *1100–1400*, 22–38. Cambridge: Cambridge University Press.

Morris, Richard (ed.). 1868. *Old English homilies and homiletic treatises of the twelfth and thirteenth centuries II* (Early English Text Society Original Series 34). London: Kegan Paul, Trench, Trübner & Co.

Pahta, Päivi. 2003. On structures of code-switching in medical texts from medieval England. *Neuphilologische Mitteilungen* 104(2). 197–210.

Pahta, Päivi & Arja Nurmi. 2006. Code-switching in the *Helsinki Corpus*: A thousand years of multilingual practices. In Ritt Nikolaus, Herbert Schendl, Christiane Dalton-Puffer & Dieter Kastovsky (eds.), *Medieval English and its heritage: Structure, meaning and mechanisms of change*, 203–220. Frankfurt: Peter Lang.

Parkes, M. B. 1991. The influence of the concepts of *ordinatio* and *compilatio* on the development of the book. In Mary A. Rouse & Richard H. Rouse (eds.), *Scribes, scripts and readers: Studies in the communication, presentation and dissemination of medieval texts*, 35–69. London: The Hambledon Press.

Poplack, Shana. 1980. “Sometimes I’ll start a sentence in Spanish *y termino en español*”: Toward a typology of code-switching. *Linguistics* 18(7–8). 581–618.

Rothwell, W. 1998. Arrivals and departures: The adoption of French terminology into Middle English. *English Studies: A Journal of English Language and Literature* 79(2). 144–165.

Schendl, Herbert. 1996. Text types and code-switching in medieval and Early Modern English. *VIEWS* 5(1&2). 50–62.

Schendl, Herbert & Laura Wright. 2011a. Introduction. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English* (Topics in English Linguistics 76), 1–14. Berlin & Boston: Mouton de Gruyter.

Schendl, Herbert & Laura Wright. 2011b. Code-switching in early English: Historical background and methodological and theoretical issues. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English* (Topics in English Linguistics 76), 15–45. Berlin & Boston: Mouton de Gruyter.

Schendl, Herbert & Laura Wright (eds.). 2011c. *Code-switching in early English* (Topics in English Linguistics 76). Berlin & Boston: Mouton de Gruyter.

Short, Ian. 1992. Patrons and polyglots: French literature in twelfth-century England. *Anglo-Norman Studies* XIV: *Proceedings of the Battle Conference 1991.* 229–249.

Skaffari, Janne. 2003. Lexical borrowings in Early Middle English religious discourse: A case study of Sawles Warde. In Risto Hiltunen & Janne Skaffari (eds.), *Discourse perspectives on English: Medieval to modern* (Pragmatics & Beyond New Series 119), 77–104. Amsterdam & Philadelphia: John Benjamins.

Skaffari, Janne. 2009. *Studies in Early Middle English loanwords: Norse and French influences* (Anglicana Turkuensia 26). Turku: University of Turku.

Skaffari, Janne. 2016. Code-switching and vernacular support: An Early Middle English case study. *Multilingua* 35(2). 203–226.

Stanley, E. G. 1961. An inedited Nativity sermon from Worcester. *English and Germanic Studies* 7. 53–79.

Swan, Mary. 2000. Ælfric's *Catholic Homilies* in the twelfth century. In Mary Swan & Elaine M. Treharne (eds.), *Rewriting Old English in the twelfth century* (Cambridge Studies in Anglo-Saxon England 30), 62–82. Cambridge: Cambridge University Press.

Swan, Mary. 2006. Old English textual activity in the reign of Henry II. In Ruth Kennedy & Simon Meecham-Jones (eds.), *Writers of the reign of Henry II: Twelve essays*, 151–168. New York & Basingstoke: Palgrave Macmillan.

Swan, Mary & Elaine M. Treharne (eds.). 2000. *Rewriting Old English in the twelfth century* (Cambridge Studies in Anglo-Saxon England 30). Cambridge: Cambridge University Press.

Timofeeva, Olga. 2010. Anglo-Latin bilingualism before 1066: Prospects and limitations. In Alaric Hall, Olga Timofeeva, Ágnes Kiricsi & Bethany Fox (eds.), *Interfaces between language and culture in medieval England: A Festschrift for Matti Kilpiö*, 1–36. Leiden: Brill.

Traxel, Oliver M. 2004. *Language change, writing and textual interference in post-Conquest Old English manuscripts: The evidence of Cambridge, University Library, Ii.1.33*. Frankfurt am Main: Peter Lang.

Treharne, Elaine M. 2000. The production and script of manuscripts containing English religious texts in the first half of the twelfth century. In Mary Swan & Elaine M. Treharne (eds.), *Rewriting Old English in the twelfth century* (Cambridge Studies in Anglo-Saxon England 30), 11–40. Cambridge: Cambridge University Press.

Treharne, Elaine. 2010–2013. London, British Library, Cotton Faustina A. ix. In Orietta Da Rold, Takako Kato, Mary Swan & Elaine Treharne (eds.), *The Production and Use of English Manuscripts 1060 to 1220*. University of Leicester. http://www.le.ac.uk/english/em1060to1220/mss/EM.BL.Faus.A.ix.htm (accessed 26 April 2015).

Treharne, Elaine. 2012. *Living through Conquest: The politics of early English, 1020–1220*. Oxford: Oxford University Press.

Trotter, David. 2003. The Anglo-French lexis of the *Ancrene Wisse:* A re-evaluation. In Yoko Wada (ed.), *A companion to Ancrene Wisse*, 83–101. Cambridge: D. S. Brewer.

Trotter, David. 2009. "Stuffed Latin": Vernacular evidence in Latin documents. In Jocelyn Wogan-Browne with Carolyn Collette, Maryanne Kowaleski, Linne Mooney, Ad Putter & David Trotter (eds.), *Language and culture in medieval Britain: The French of England c.1100–c.1500*, 153–163. Woodbridge: York Medieval Press.

Trotter, David. 2011. Death, taxes and property: Some code-switching evidence from Dover, Southampton, and York. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English* (Topics in English Linguistics 76), 155–189. Berlin & Boston: Mouton de Gruyter.

Varila, Mari-Liisa, Hanna Salmi, Aleksi Mäkilähde, Janne Skaffari & Matti Peikola. 2017. Disciplinary decoding: Towards understanding the language of visual and material features. In Matti Peikola, Aleksi Mäkilähde, Hanna Salmi, Mari-Liisa Varila & Janne Skaffari (eds.), *Verbal and visual communication in early English texts* (Utrecht Studies in Medieval Literacy 37), 1–20. Turnhout: Brepols.

Voigts, Linda E. 1996. What's the word? Bilingualism in late-medieval England. *Speculum* 71(4). 813–826.

Wenzel, Siegfried. 1994. *Macaronic sermons: Bilingualism and preaching in late-medieval England*. Ann Arbor, MI: University of Michigan Press.

Wogan-Browne, Jocelyn. 2009. General introduction: What's in a name: The "French" of "England". In Jocelyn Wogan-Browne with Carolyn Collette, Maryanne Kowaleski, Linne Mooney, Ad Putter & David Trotter (eds.), *Language and culture in medieval Britain: The French of England c.1100–c.1500*, 1–13. Woodbridge: York Medieval Press.

Wogan-Browne, Jocelyn with Carolyn Collette, Maryanne Kowaleski, Linne Mooney, Ad Putter & David Trotter (eds.). 2009. *Language and culture in medieval Britain: The French of England c.1100–c.1500*. Woodbridge: York Medieval Press.

Wright, Laura. 2013. Mixed-language accounts as sources for linguistic analysis. In Judith A. Jefferson, Ad Putter & Amanda Hopkins (eds.), *Multilingualism in medieval Britain (c. 1066–1520): Sources and analysis* (Medieval Texts and Cultures of Northern Europe 15), 123–136. Turnhout: Brepols.

III Patterns

Jukka Tuominen

8 “Trifling shews of learning”? Patterns of code-switching in English sermons 1640–1740

1 Introduction

> The impertinent Way of dividing Texts is laid aside, the needless setting out of the Originals, and the vulgar Version, is worn out. The trifling Shews of Learning in many Quotations of Passages, that very few could understand, do no more flat the Auditory. *Pert Wit* and *luscious Eloquence* have lost their Relish. (Burnet 1692: 216; emphasis original)

In *A Discourse of the Pastoral Care*, bishop Gilbert Burnet (1643–1715) looked back with approval at the way preaching in the Church of England had changed within his lifetime. In the past, Burnet claimed, ministers had paraded their knowledge of Hebrew and Greek, the original languages of the Bible, and of the Latin Vulgate; contemporary pulpit oratory, by contrast, was characterized by clarity, naturalness and a return “to the plain Notions of simple and genuine Rhetorick” (1692: 216).

Burnet’s comments are not a neutral account of recent developments but an ideologically colored statement that needs to be viewed against the various rifts in his church and society. Writing soon after the accession of William and Mary in 1689, at a time of religious and political ferment in which he was intimately involved, Burnet would have had an interest in presenting the opponents of the new regime’s policies as hopelessly out of touch with the needs of the common parishioners (cf. Spurr 2006: 203–206). The words nevertheless also suggest a noticeable change in preaching. Modern scholars have followed the views of Burnet and other contemporaries and used the decline of foreign-language elements as an important characteristic of a new English sermon style in the latter half of the seventeenth century (Mitchell 1932: 310–312; Davies 1961: 65–67; Lessenich 1972: 25–27). These accounts, however, are qualitative and typically based on the works of prominent divines of the time. A corpus of sermons by lesser-known writers offers the opportunity to reassess this traditional view in broad quantitative terms and in a diachronic perspective.

Religion has been identified as a domain favoring language contact and mixing (e.g. Schendl 1996; Spolsky 2003; Pahta and Nurmi 2006), and it provides a

---

Acknowledgements: The research reported here has benefited at various stages from insightful comments by Terttu Nevalainen, Arja Nurmi, Richard Ingham, Jukka Tyrkkö, the editors of this volume, and two anonymous reviewers. Any errors and infelicities in the article, however, are solely attributable to the author. Financial support from the Academy of Finland-funded project “Multilingual Practices in the History of Written English” is gratefully acknowledged.

DOI 10.1515/9781501504945-008

wealth of source texts in a range of genres whose communicative settings remain relatively constant, enabling comparisons over time (Kohnen 2010). While multilingual sermons and homilies are known from England from the Old English period onwards (Iglesias-Rábade 1996; Schendl 1996; Pahta and Nurmi 2006; see also Schendl and Ingham, this volume), and especially so-called "macaronic" sermons from the fourteenth and fifteenth centuries have attracted scholarly attention (e.g. Wenzel 1994; Schendl 2000a, 2000b, 2013; Halmari and Regetz 2011; Fletcher 2013), the development of code-switching practices in English religious texts in the seventeenth and eighteenth centuries has largely remained unstudied. In their survey of code-switching in English writing from ca. 700 to 1710 based on the *Helsinki Corpus*, Päivi Pahta and Arja Nurmi (2006) find that sermons are consistently among the genres with most switches. However, the period 1640–1710 is represented in their source corpus by extracts from only three sermons, all from the 1670s (see Kytö 1996). The almost complete absence of foreign-language elements in sermons by Thomas Twining (1734–1804) at the end of the eighteenth century (Nurmi and Pahta 2010) reinforces the view of a gradual decline in code-switching in sermons over time (Pahta and Nurmi 2006). This study seeks to test and supplement the results of these earlier studies with a more extensive dataset of complete sermon texts, enabling a more detailed analysis, and to extend the research into decades which so far have not been examined. Moreover, the study addresses for the first time the differences in code-switching practices between the various religious communities represented in the material.

Specifically, the aim of this article is to provide an empirical, sociolinguistically oriented description of general structural and functional patterns of code-switching in English sermons from the period 1640–1740, using a sample of ten sermons from the *Lampeter Corpus of Early Modern English Tracts* (Schmied, Claridge, and Siemund 1998; for details see Claridge 1999/2003). Code-switching is understood here as "the use of more than one language in the course of a single communicative episode" (Heller 1988: 1), and the analysis is carried out in a variationist framework: differences in individuals' conscious or unconscious choice to switch languages are assumed to correlate with independent language-external variables in the specific socio-cultural and linguistic context (cf. Weinreich, Labov, and Herzog 1968; Tagliamonte 2012).

2 The socio-cultural context of multilingual sermons in early modern Britain

2.1 Sermons in performance and print

Thomas Kohnen (2010: 539) views sermons as "a prototypical genre of religious instruction" defined by the communicative setting where a clergyman, a “specialist” member of the community with detailed theological knowledge and ecclesiastical power, addresses a congregation of laypeople. Although this communicative setting has remained relatively constant over centuries, other genre characteristics of sermons and the linguistic forms given to them are subject to change. The production and reception of texts are both influenced by the shared genre expectations of a particular time and cultural context (Taavitsainen 2001: 140–141). These expectations condition authors’ linguistic choices, but they also guide the way contemporary and later readers, including present-day scholars, interpret a text and its specific features, such as the use or non-use of code-switching (Machan 2011: 312).

A key question for any study of early modern sermons is the relationship between the written text available to the modern scholar and what was (presumably) delivered from the pulpit. It is usually impossible to determine to what extent the written version of a sermon represents its oral performance (Kohnen 2010: 539; for divergent views on this issue see e.g. Raymond 2003: 222–224 and Green 2006: 249). The final, printed form can be seen as a layered end-result of a process of composition, where the various stages may each have left their mark. Factors affecting the outcome include the circumstances of text production, i.e. how the sermon was prepared first for oral delivery and then for publication; the conventions associated with or limitations imposed by the publication format; and, in particular, the two target audiences whose needs and expectations the sermon-writer sought to address: the specific congregation who first heard the sermon, and the potentially much wider readership of the published version.

Whether the sermon was written in full before preaching or delivered using only an outline of headings or other notes (Davies 1975: 141–142; Green 2006: 237, 248), it could be altered afterwards for publication. In practice, this usually meant adding new material (Green 2009: 28). While features particular to the oral delivery, such as gestures and prosody, were lost, the written form also had its own characteristic features, including paratextual elements such as dedications, prefaces, and marginalia (see Genette 1997; Smith and Wilson 2011). Publication also raises the question of whose linguistic choices and competence the final product represents: even in sermons – like the ones studied here – where the printed text is an authorized version, presumably prepared and approved by the author himself, influence from a typesetter or a house style cannot be ruled out (cf. Pahta 2004: 80). Like the relationship between the oral and written versions of the sermon, the extent of such influence is usually impossible to determine. Moreover, the sermons selected for publication give a skewed picture of the totality of contemporary preaching: the pulpit oratory of “an average minister on a routine Sunday to a typical rural congregation” usually did not find its way into print (Green 2006: 238).

The sermons studied here were published as pamphlets, which as a publication form are characterized by relatively short length and stitching rather than binding. They were easier, faster and cheaper to produce than books, and consequently accessible to a wider readership, whom writers would have taken into account in designing their texts (Raymond 2003: 5–8, 80–82; Groeger 2010: 37–43). Furthermore, so-called occasional sermons marking a special event, typically preached before a congregation unfamiliar to the minister, are prevalent among those published, and have been characterized as having a "this-worldly", often polemical focus on current issues (Green 2009: 16–18, 27). Such features would not necessarily have been as important in sermons published in book form, e.g. in the collected works of a clergyman, which had a scholarly air and could probably only be afforded by the wealthiest members of the reading public. Pamphlet buyers, by contrast, are likely to have included both clergy looking for models or interested in the work of their colleagues, and lay readers ranging from the gentry to the "middling sort". Sermon pamphlets could also be bought or commissioned by individuals or organizations for free distribution. Apart from commercial considerations, the motivations for publishing a sermon might thus include commemorating a specific occasion, disseminating religious or political ideas, and even self-promotion or self-aggrandizement (Green 2006: 238–239, 251–252; Rivers 2009: 579, 592–593). While the number of individual titles published in the seventeenth and eighteenth centuries was vast, most pamphlets had limited print runs of 250–1,500 copies, and only works by prominent divines were likely to be reprinted (Raymond 2003: 80; Green 2006: 252). In a society with widespread illiteracy, the primary target audience for published sermons was relatively small and mainly composed of male readers who enjoyed a degree of economic security and had at least some formal education.

2.2 Functions of code-switching in sermons

The challenges of positioning published sermons on the continuum from spoken to written language also affect how multilingual elements in them are viewed. Several studies of code-switching in historical written material make a distinction between "free" switching, which is similar to the spontaneous spoken usage of fluent bilinguals, and "prefabricated chunks" (e.g. Schendl 2000b: 72; Nurmi and Pahta 2004: 430; see also Nurmi et al., this volume), i.e. elements which do not result from the writer's own creative language use but derive from other texts or are highly conventionalized. Religious writing in particular appears to favor switching that follows the established communicative practices of the domain: in medieval and early modern sermons, switching typically occurs for technical terms, quotations, references, or textual organization (Pahta and Nurmi 2006: 213–218; cf. Wenzel 1994). The categories partly overlap, and given the polysemous nature of multilingualism, switches are often open to several simultaneous functional interpretations ranging from the context-specific to the interpersonal and community levels (cf. Gardner-Chloros 2009: 42–43). Different religious groups have their own traditions and views of the role of the sermon in worship, which in turn are realized in differing linguistic practices. For example, it has been suggested that one function of code-switching in early modern sermons is to deliberately exclude the majority of their hearers and readers, emphasizing the role of the established church and its clergy as keepers of sacred truth. Frequent and visually prominent use of the learned languages may also be a way to emphasize the mediated and textual nature of the printed sermon as opposed to its oral performance (Ettenhuber 2011: 49–50). This seems especially clear when code-switching occurs in paratextual elements, which clearly involve additions to the prior oral version.

2.3 Languages and clerical education

The languages other than English used in Britain and Ireland in the seventeenth and eighteenth centuries form three distinct groups with different social meanings and contexts associated with each. The other, mainly Celtic vernaculars forming the first group were important for ministers working in multilingual or non-English-speaking parishes, but their impact on the clergy as a whole must have been quite limited. The second group consists of the modern foreign languages such as French, German and Italian, which were needed for international communication and were increasingly seen as part of a well-rounded education. Of particular importance in the religious domain, however, was the third group comprising the three learned languages, Latin, (Classical) Greek and Hebrew.

While Latin remained the *lingua franca* of international scholarship and a cornerstone of educated life throughout the early modern period, its status was gradually waning with the vernacularization of religion in Protestant communities and on a pan-European level in fields such as science, particularly from the second half of the seventeenth century onwards (Burke 1991). As for Greek and Hebrew, the Protestant goal of making the Bible accessible for lay readers in translation paradoxically served to emphasize their value, as it was – at least in theory – crucial that pastors be able to understand, transmit and explain the precise meaning of Scripture (cf. Trueman 2011: 61–62). Greek was also the language of the eastern tradition of Christian theology, and, like Latin, provided access to the culture and learning of the pre-Christian classical world.

The linguistic competence that enabled the clergy to use several languages was acquired through education. By the beginning of the period studied here the Anglican clergy was more or less entirely university-educated (O'Day 1982: 109, 139). The years of exposure from grammar school or private tutoring to university must have given graduates a thorough grounding in the learned languages, especially Latin, but the outcome may often have been receptive rather than productive competence (Clarke 1959: 66–67; Burke 1991: 32, 38; Trueman 2011: 62 has a more positive view). There must have also been considerable differences between individuals in both aptitude and achievement. Although the early modern humanist educational ideal was equal competence in the three learned languages, in practice Latin had by far the most varied curriculum while Greek and especially Hebrew were much more limited to the religious sphere (O'Day 1982: 68–69).

Education was closely tied to the established church, and dissenters or nonconformists, i.e. Protestant groups such as Presbyterians, Baptists and Quakers, were barred by law from the English universities. This led to the founding of "dissenting academies" with a more practical course of study, including subjects such as modern languages and science (O'Day 1982: 212–214). The Baptists in particular had a strong anticlerical and antiacademic bent, but while their ministers typically lacked formal theological training, literacy nevertheless gave them access to Scripture and consequently elite status and power within the congregation (Mullett 1994: 193, 198, 201).

3 Material and methods

3.1 The *Lampeter Corpus* sermons

The material for this study comes from the *Lampeter Corpus*, a computer-readable collection of non-literary prose texts that were published as pamphlets in Britain or Ireland between the years 1640 and 1740 (Schmied, Claridge, and Siemund 1998).[1] The corpus comprises 120 texts, grouped into six domains with two texts for each decade, selected with the aim of reflecting the variety of subject matter and text types found in pamphlets. The texts are mostly first editions and are in principle included in the corpus in their entirety. Importantly for a study of code-switching, the presence or absence of languages other than English in the texts was not a criterion in the selection process. The ten sermons used here are identified in the corpus by a marker consisting of RelA ("A" text in the domain Religion) and the year of publication. They were published between the years 1642 and 1730 and add up to 108,540 words (Claridge 1999/2003: 4–6, 38–44).

---

[1] The corpus is available on the ICAME Collection of English Language Corpora CD-ROM (second edition, 1999) and online through the University of Oxford Text Archive (http://ota.ox.ac.uk/desc/2400); the texts can also be downloaded from the *Lampeter Corpus* website (http://www.tu-chemnitz.de/phil/english/sections/linguist/real/independent/lampeter/lampr.htm; both sites accessed November 30, 2014).

The sermons are by different authors with varying educational and religious backgrounds, and they also differ in terms of topic and occasion. The authors' background is summarized in Table 1. Eight were ministers in the established Protestant Churches of England and Ireland, and had attained at least the degree of Master of Arts at university. The remaining two authors, the Baptist John Piggott and the Presbyterian Samuel Wright, seem not to have had a university education. Even the Anglican authors exhibit some variation in their religious affiliation, which is likely to have affected their style of preaching. Nine of the ten authors are known to have published at least one other text in addition to the sermons studied here, and most had progressed through several appointments in different parts of England or Ireland. These factors suggest a degree of ambition and ascending career paths (cf. Barrie-Curien 1993: 94–95). However, the authors do not rank among the most prominent and lastingly influential clergymen of their day. Although seven of them have entries in the *Oxford Dictionary of National Biography*, none, for example, appears to have been the subject of an extensive modern study.

In contrast to the authors, the sermons exhibit great variation within the confines of their shared genre. Apart from differences in topic, the length of the texts varies considerably. Many of the sermons marked special occasions, and their original audiences, as far as their composition can be determined, were quite diverse.[2] The material thus offers several angles for analysing variation in the authors' multilingual practices. Given the small size of the dataset, however, caution is needed in assessing the possible influence of these overlapping extra-linguistic variables on the authors' code-switching.

---

[2] Although the composition of the original audience is taken into account in the analysis as a factor likely to have had an effect on code-switching practices, the material is used here as it stands in its published form. A thorough investigation of the complex issues involved in the relationship of the oral and printed versions, including that of the "sermon proper" and paratextual additions, as well as ways in which authors may have accommodated their language for the hearers and readers of the sermon, is beyond the scope of this article.

Table 1: Background information on the authors and audiences of the ten sermons in the *Lampeter Corpus* (compiled from the corpus files; *Oxford Dictionary of National Biography*; Wood 1691: 890; Ivimey 1823: 565–567; Gordon 1897: 184; Claridge 1999/2003: 38–44; Schaffer and Stewart 2005: 45, 54 n. 51)

	Author:				**Sermon:**	
Text marker	**name**	**date of birth**	**education**	**religious affiliation**	**length (words)**	**audience**
RelA1642	Thomas Hill	1600s	Cambridge	Church of England (Puritan)	18,845	House of Commons
RelA1653	Joseph Sedgwick	1613/14	Oxford; Cambridge	Church of England (moderate Puritan/ latitudinarian?)	9,464	academics + others
RelA1669	Richard Sherlock	1612	Oxford; Trinity College, Dublin	Church of England	4,719	bishop, clergy (+ others?)
RelA1679	Henry Jones	1605	Trinity College, Dublin	Church of Ireland	16,991	incl. Lord Lieutenant of Ireland
RelA1682	Thomas Pittis	1636	Oxford	Church of England	8,016	lawyers
RelA1696	John Piggott	not known	not known	Baptist	8,389	London Baptists (?) (mixed?)
RelA1708	John Waller	c. 1673	Cambridge	Church of England (?)	6,793	incl. gentry and clergy
RelA1711	Samuel Wright	1683	nonconformist academy at Attercliffe	Presbyterian	9,337	mourners for Dr. Francis Upton (mixed?)
RelA1721	Joseph Trapp	1679	Oxford	Church of England (High Church)	13,853	Sons of the Clergy association
RelA1730	Arthur Bedford	1668	Oxford	Church of England	12,133	mixed London congregation?

3.2 Identifying code-switched passages

The *Lampeter Corpus* uses a variant of the SGML coding system to indicate many features of the original pamphlets, such as information on their structure, layout, and typography, including the use of the Greek and Hebrew scripts. Each text also has a header which provides some information on the text and its author (see Claridge 1999/2003: 11–37).

Elements which the compilers have considered foreign are coded in the corpus by special tags which include the attribute *lang* to specify the language used. Eleven languages in addition to English are identified in the corpus as a whole; three of them, Latin, Greek and Hebrew, occur in the ten sermons studied here. The compilers' criterion for deciding whether borderline cases could be considered English was ultimately whether the item is marked in the *Oxford English Dictionary* (*OED*) as "not naturalized, alien" (Claridge 1999/2003: 25).[3]

The data for this study were collected by searching for the string "lang=" in the PDF files of the ten sermons, but the compilers' judgment was not followed indiscriminately, and the number of switched passages used in the analysis below therefore differs slightly from the raw data based on what is coded as foreign in the corpus. Since the aim is to compare the authors' language choices, items which derive from the printing conventions of the period, such as the imprimatur and the word *Finis* at the end of a publication, were excluded from the analysis. Single-word items were in principle accepted as potential code-switches, but cases deemed ambiguous were excluded as borrowings if they showed clear English morphology or were attested in the *OED* at least fifty years before their use in the corpus. Foreign-derived proper nouns were also excluded with two exceptions: names that occur in a longer string in the foreign language, and book titles. Bivalent forms such as abbreviations (including abbreviated names) and numerals were considered foreign when part of a passage with some clearly identifiable foreign material. In other words, passages with an identifiable foreign element were considered foreign throughout up to the next unambiguously English word.

3.3 Analytical approaches

The approaches used in the study generally follow Nurmi and Pahta (2004) and Pahta and Nurmi (2006). The number and length in orthographic words of switched passages in Latin, Greek and Hebrew in each sermon were calculated and related both to the decade in which each text was published (real time) and to the authors' decades of birth (apparent time) to analyse diachronic developments. The frequencies and switch lengths were also examined in terms of the religious affiliation and education of the author and the composition of the original audience, to the degree this was possible with the available information. The results of these tests are reported below in section 4.1.

---

[3] The symbol || is used in the second edition of the *OED* to indicate words which are not naturalized in English in terms of form, inflection or pronunciation (see Preface to the Second Edition (1989), section "Main words", I.1). The symbol is no longer used in the third edition entries in the online version of the *OED* (for an extensive discussion, see Queiroz de Barros, this volume).

The analysis of the functions of code-switching in the sermons in section 4.2 proceeds from the categories identified by Pahta and Nurmi (2006), although their typology is here modified and developed on the basis of the *Lampeter Corpus* data. The starting point was identifying potential quoted passages.[4] Passages that were determined to derive from or refer to other texts were also grouped by the date of the source. Code-switched expressions which could not be traced to a single source but were deemed to belong to a specialized field were classified as terms.

4 Code-switching in the *Lampeter Corpus* sermons

4.1 Frequency and length of switches

Using the criteria discussed above, the material contains a total of 232 passages where the language is Latin, Greek or Hebrew instead of English. Illustrative examples for each language are given in (1)–(3). The direction of switching is nearly always from English into one of these languages and back to English, as in the examples below; there are only four instances of switching from Latin to Greek or vice versa.

(1) Whereupon St. Gregory: ***Quam reverendi sunt Pastores optimi Sanctæ Ecclesiæ*** – *how reverendly to be esteemed are the Pastors of holy Church*, who whilst they faithfully serve the Lord in the Execution of their function, they are so closely joyn'd unto him [...] (RelA1669: 11)[5]

---

[4] Potential source texts were identified by using the full-text databases *Thesaurus Linguae Graecae*, *Library of Latin Texts*, *Patrologia Latina Database*, *Early English Books Online* and *Eighteenth Century Collections Online* as well as the Google search engine.

[5] In the examples **bold** typeface indicates the code-switched segment. If the original text has a gloss it is indicated by *italics*; otherwise, a translation is provided in brackets. The use of the Greek and Hebrew scripts follows the typography in the original pamphlets, as indicated in the corpus coding. In example (1) the gloss omits the adjective *optimi*, 'the *best/excellent* pastors'.

(2) [. . .] you have a good invitation to it, Rev. 21.7,8. **ὁ νικῶν** *he that is overcomming*, not he that hath overcome, shall inherit all things, [. . .] (RelA1642: 55–56)

(3) So much the Originall word משלי rendred, *Proverbs* seemes to import; raigning commanding sentences, being so full of wisdome gravity and authoritie; [. . .] (RelA 1642: 2)

Switches occur in all ten sermons, but the range of languages and the frequency of switching vary considerably. The data are summarized in Table 2. The highest frequency of switching is found in the earliest sermon, RelA1642, which alone contains nearly a half of all the switches in the material. In terms of languages, Latin clearly dominates, accounting for most or all of the switches in nine of the ten sermons. RelA1653 stands out in having no switches into Latin. While six texts have switches into Greek, they are relatively frequent only in the two earliest ones. Switches into Hebrew occur only in RelA1642, and their effect on the overall picture is negligible.

Table 2: Number and frequency (/1,000 words) of code-switched passages in the *Lampeter Corpus* sermons

Sermon	**Total**		**Latin**		**Greek**		**Hebrew**	
RelA1642	110	(5.8)	86	(4.6)	22	(1.2)	2	(0.1)
RelA1653	10	(1.1)	0	–	10	(1.1)	0	–
RelA1669	12	(2.5)	11	(2.3)	1	(0.2)	0	–
RelA1679	57	(3.4)	53	(3.1)	4	(0.2)	0	–
RelA1682	18	(2.2)	17	(2.1)	1	(0.1)	0	–
RelA1696	2	(0.2)	2	(0.2)	0	–	0	–
RelA1708	5	(0.7)	5	(0.7)	0	–	0	–
RelA1711	1	(0.1)	1	(0.1)	0	–	0	–
RelA1721	5	(0.4)	5	(0.4)	0	–	0	–
RelA1730	12	(1.0)	11	(0.9)	1	(0.1)	0	–
Total	232	(2.1)	191	(1.8)	39	(0.4)	2	(0.0)

The diachronic distribution of the data suggests two distinct periods. With the exception of RelA1653, the frequency of switching remains above the average of 2.1 switches per 1,000 words up to the 1680s, but then drops below it for the last 5 decades. The significance of temporal change is supported by apparent-time analysis. In Figure 1, the sermons are arranged by decades in the order of the authors' date of birth; in the case of Hill (RelA1642) and Waller (RelA1708), this is estimated from the date of graduation (cf. Claridge 1999/2003: 35).[6] Again, the texts appear to form two groups, with the decades of the Civil Wars and the Interregnum as the watershed: the authors who use code-switching more frequently were all born in the period 1600–1639, whereas the authors of the later sermons with much less frequent switching were born after 1660. Moreover, apart from the exceptionally low frequency of switching in RelA1653, the frequencies of switching form a cline, with the sermons by younger authors constantly showing less switching than those by authors born in the previous decade.

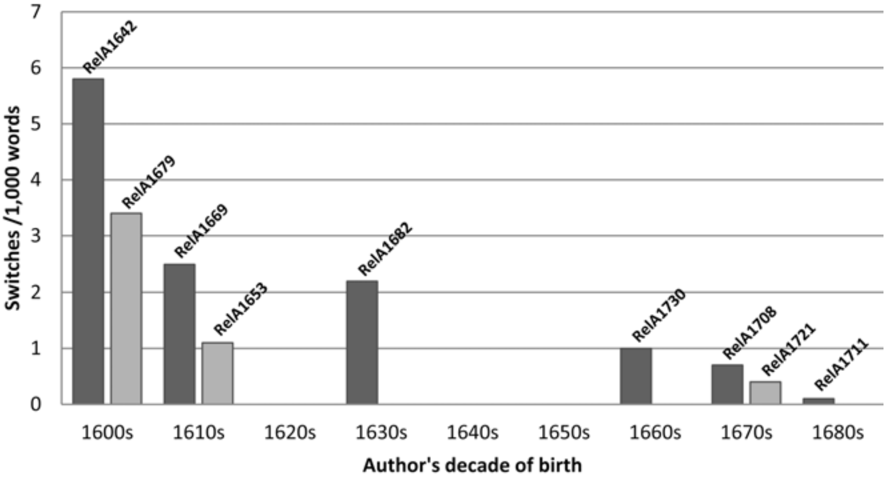


Figure 1: Frequency of code-switching in the *Lampeter Corpus* sermons according to their authors' (probable) decade of birth. RelA1696 excluded due to lack of data

However, time is not the only variable which differentiates the sermons. The lowest frequencies are in the two texts by the nonconformist ministers Piggott and Wright, suggesting a link with the educational background or religious affiliation of the authors. These two texts are also among the minority with no switches into Greek.

Other factors, such as the occasion of the sermon, the make-up of the audience present at the time of delivery, and the intended readership of the published text, are also likely to have played a role, but they are not easy to quantify. In general terms, the audiences of the five earlier sermons seem to have come from higher social strata than those of the five later ones (see Table 1).

---

[6] Piggott (RelA1696) is excluded from the analysis since there is practically no information on his background.

Contemporary evidence suggests that clergymen were careful to take the composition of their audience into account in preparing their sermons. A congregation of high-ranking, well-educated people would have been accustomed to hearing quite refined pulpit oratory and able to understand and appreciate such elements as quotations in Latin and Greek (Davies 1975: 158).

Turning to the length of the switched passages, the switches in the material range from single words to phrases, clauses, sentences and longer strings of text. The longest switch is a Latin passage of 46 words in RelA1708, and four other texts contain Latin switches of over 20 words. Table 3 shows the range of switch lengths in the sermons alongside three measures of central tendency for all code-switched passages and for switches into Latin. Looking at the material in its entirety, there seems to be a link between the frequency of switching into a foreign language and the average length of those switches: Latin switches, which occur most frequently, are also the longest. Greek switches are on the average considerably shorter at 2.4 words, and the two switches into Hebrew are only single words. As with frequency, the overall average length of switches in the texts is largely determined by the Latin switches.

The longest individual switches and consequently the widest range of switch lengths tend to be found in the texts which also have a relatively high frequency of switching. The data do not suggest clear patterns based on real or apparent time, or on the religious affiliation and education of the authors. Two texts stand out as exceptional. In RelA1653 all the switches are into Greek, and the low average switch length of 2.7 can thus be explained by the complete lack of the typically longer Latin switches in the sermon. The figure is, however, in line with the overall average length for Greek. The unusually high average of 15.0 in RelA1708, on the other hand, is largely due to the 46-word switch in this sermon; the average length of the other switches in the text is only 7.3 words.

As the Greek and especially the Latin switches in the material show a wide range of switch lengths, the average (i.e. mean) switch length does not necessarily indicate what is typical. The median and mode values given in Table 3 vary from text to text, but both are usually clearly lower than the average switch length, which is more affected by the relatively few very lengthy switches. The median and mode of the Greek switches are both 1, i.e. more than half of them (21 of the 39 switches) are single words. Overall, one- to three-word switches constitute slightly more than a half of all switches. Comparing the distribution of switch lengths and the frequency of switching in the texts indicates that frequent switches also tend to be relatively short, although the relationship between the two measures is not a simple inverse one. Such patterns are not readily identifiable when only average switch length is used.

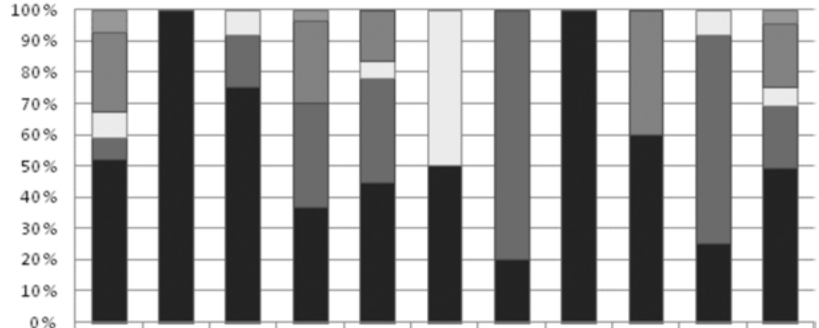
Table 3: Length in words of code-switched passages in the *Lampeter Corpus* sermons. ASL = average switch length, % = percentage of switches with the mode value; NA = not applicable

	All switches					**Latin**			
Sermon	**ASL**	**median**	**mode**	**%**	**range**	**ASL**	**median**	**mode**	**%**
RelA1642	5.8	3	1	26.4	41	6.9	4	2	22.1
RelA1653	2.7	2	1	40.0	10	–	–	–	–
RelA1669	8.9	5	5	25.0	37	9.4	5	5	27.3
RelA1679	5.2	3	2	33.3	23	5.5	3	2	32.1
RelA1682	7.4	6.5	5	22.2	21	7.2	6	5	23.5
RelA1696	7.5	7.5	NA	–	3	7.5	7.5	NA	–
RelA1708	15.0	8	NA	–	42	15.0	8	NA	–
RelA1711	6.0	NA	6	100.0	0	6.0	NA	6	100.0
RelA1721	7.0	9	NA	–	11	7.0	9	NA	–
RelA1730	6.2	5.5	5	25.0	13	6.6	6	5	27.3
Total	6.1	3	2	20.3	45	6.9	5	2	21.5

The *Lampeter Corpus* results are similar to the ones reported by Pahta and Nurmi (2006: 210) for the early modern sermons of the *Helsinki Corpus*. For the entire early modern period (1500–1710 in the corpus), the frequency is 3.5 switches per 1,000 words of text and average switch length 8.9 words. For the years 1640–1710 where the two corpora overlap, the frequency is 4.7 with average switch lengths for individual sermons ranging from 2.5 to 9.7 words in the *Helsinki Corpus*; the frequency for the *Lampeter Corpus* sermons from these decades is 2.9. The difference is at least partly explained by the high incidence of switching (9.4) in one of the three *Helsinki Corpus* sermons for this period; in the other two, the frequency of switching is only 0.4 and 1.6 (Arja Nurmi, personal communication, April 11, 2012). The figures may also partly result from different criteria used in coding foreign segments in the corpora and in deciding what to include in the analysis. Furthermore, the inclusion of the full published sermons in the *Lampeter Corpus* is likely to even out possible peaks and troughs in switching in different parts of the texts; the *Helsinki Corpus* data, by contrast, come from shorter extracts.

4.2 Functions of code-switching

Nearly all of the switches in the data fall into two general functional categories: intertextuality and terminology. Intertextual switches are understood here as quotations and identifiable paraphrases, commonplaces, and source references. Intertextual and terminological switches are related to the formal communicative situation and established discourse practices of sermons as a genre. Together these types cover over 95 percent of the switches in the data (see Figure 2). Most of the remaining switches are single words or two-word phrases which, like many of the intertextual switches and terms, are used to introduce a distinction or to highlight a particular expression whose meaning is then explained in detail. Only two short switches of three and four words, both in RelA1642, do not fit into any of these types, and even they may not be genuine examples of "free" switching by the author but rather cases that merely could not be identified as prefabricated intertextual switches with the tools used. Characteristic features of the most common functional types are presented below.



	RelA 1642	RelA 1653	RelA 1669	RelA 1679	RelA 1682	RelA 1696	RelA 1708	RelA 1711	RelA 1721	RelA 1730	All texts
▨ other	8	0	0	2	0	0	0	0	0	0	10
▨ term	28	0	0	15	3	0	0	0	2	0	48
□ commonplace	9	0	1	0	1	1	0	0	0	1	13
▨ reference	8	0	2	19	6	0	4	0	0	8	47
■ quotation	57	10	9	21	8	1	1	1	3	3	114

Figure 2: The proportions of the different functional types of switching in the *Lampeter Corpus* sermons, with the number of occurrences for each type given in table form

4.2.1 Intertextuality

Through intertextual switches the author positions his sermon in a complex network of earlier texts and engages in discussion and debate with them. The other texts can be used in an authoritative function to support the argumentation or corroborate its orthodoxy (cf. Pahta 2004: 86), but also as examples of erroneous thinking. At the same time, such switches indicate that the author is well-read and able to understand and correctly select passages relevant to the matter at hand. The switches also suggest how the author would like his hearers and readers to be perceived, or what qualities he expects from them. Although the sermons are mostly in English, many of the details are accessible only to those with the ability to understand and identify the foreign-language passages.[7]

Intertextual switches account for 75 percent of the switches in the material, and examples occur in all of the ten sermons. The three subcategories – quotations, commonplaces, and references – are illustrated by examples (4)–(7) below. *Quotation* is used here as a general term for any strings of words that can be traced to an earlier source. Quotations are almost always accompanied by an overt or oblique indication of the derivative relationship, as in (1) above with its reference to Gregory and in (4) below, where the code-switched expression is attributed to "Popish writers". However, quotations also include allusions and paraphrases. For example, the Latin in (5) clearly alludes to the phrase "*Qui* [. . .] *student magis alta quam apta proferre*" ('those who would rather mention lofty than appropriate things') in William of Tournai's preface to his *De instructione puerorum* (Corbett 1955: 11).

(4) [. . .] to this Gregory 7. is by Popish writers that glory given, that he was ***Imperii Pontificii fundator*** ['the founder of the papal empire/sovereignty']; others his Predecessors attempted the reaching at the temporal power, but was not till now attained. (RelA1679: 13)

(5) I study not ***alta*** but ***apta proferre***. (RelA1642: 21)

By contrast, foreign-language strings that appear in at least one earlier text but could not be traced with certainty to a particular source are categorized as *commonplaces*. Example (6) illustrates these typically proverbial or idiomatic expressions. The third subcategory of *references* comprises switches that point the reader to other works. The key foreign-language element here is typically a book title, as in (7). Only cases where the entire switch consists of a reference are counted; when a single code-switched segment includes both a quotation and a reference, it is classified here as a quotation.

---

[7] About 54 percent of the code-switched quotations and a quarter of the commonplaces in the data have an English explanation that is close enough to count as a gloss. The presence of a translation suggests that all hearers or readers were not expected to understand the foreign segment. Since the semantic content is also transmitted in the gloss, retaining the original language alongside the vernacular points to other functions, such as providing authority, proof, clarification or emphasis. Given its long history (see e.g. Wenzel 1994: 109 and Schendl, this volume), the practice may also have a conventional aspect.

(6) And as in Persons and Opinions, so in things too, ***quo antiquius eò melius***, *by how much the more Ancient a thing is, by so much the more is it prized and valued* by those that covet a strict inspection into the Periods that did preceed their own, [. . .] (RelA1682: 2)

(7) The Learned Grotius, as if he had been alive and wrote against the Commentators upon his excellent Book, ***de jure Belli & Pacis***, unanswerably takes off this imputation, [. . .] (RelA1708: 14–15)

The prototypical function of code-switching in the data is clearly quotation (see Figure 2). All ten sermons include code-switched quotations, and overall they make up just under a half of all switches. Also references are fairly common, occurring in six sermons and accounting for a fifth of all switches. Commonplaces play only a minor role. The texts with more frequent switching tend to show a wider variety of functions. Otherwise the distribution of switch types does not suggest clear patterns at this level of analysis, and the very small number of switches in the four sermons RelA1696–RelA1721 distorts the picture. All three foreign languages are attested in quotations, whereas references and commonplaces only occur in Latin. Although code-switching is strongly associated with intertextuality, relationships with other texts are not exclusively indicated by using the foreign languages; instead, the sermons also cite their sources in English. Especially the numerous quotations from and references to the Bible are usually in the vernacular.

The source texts indicated by the code-switched references and quotations form five groups which reflect distinct time periods as well the status of the sources in the Christian tradition: the classical pagan authors of ancient Greece and Rome, the Bible, patristic texts from the early centuries of the Christian church, and medieval and early modern sources. Figure 3 shows the proportions of these categories in the data. The total number of citations, i.e. code-switched quotations and references, is 154.[8] Citations from early modern and patristic sources together make up roughly two thirds of all occurrences; the other source types occur much less often. In three cases the date of the source is too uncertain for allocation into one of these groups.

---

[8] The count used in this part of the analysis is based on the following principles: quotations from and references to different sources within a single code-switched passage are counted separately; references which cite both a medieval author and an early modern edition are counted once for both categories; and cases where a quotation has a separate code-switched reference are counted as one citation rather than two. Switches between Latin and Greek within quotations are ignored, but each quoted segment with a switch back to English is counted separately.

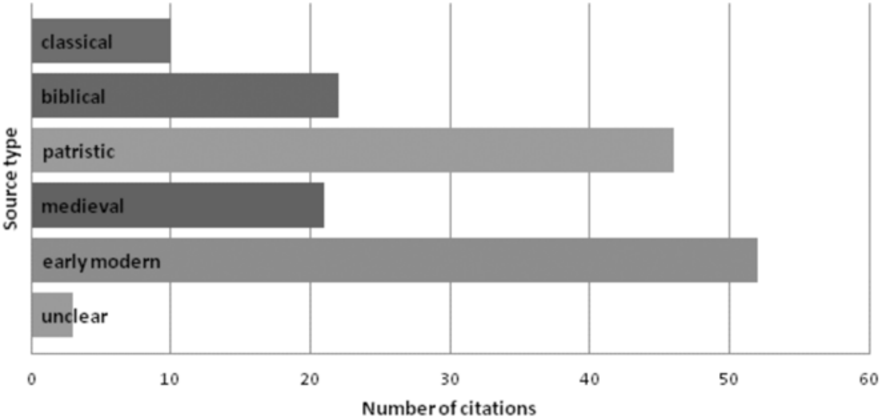


Figure 3: The number of citations for the different source types in the *Lampeter Corpus* sermons

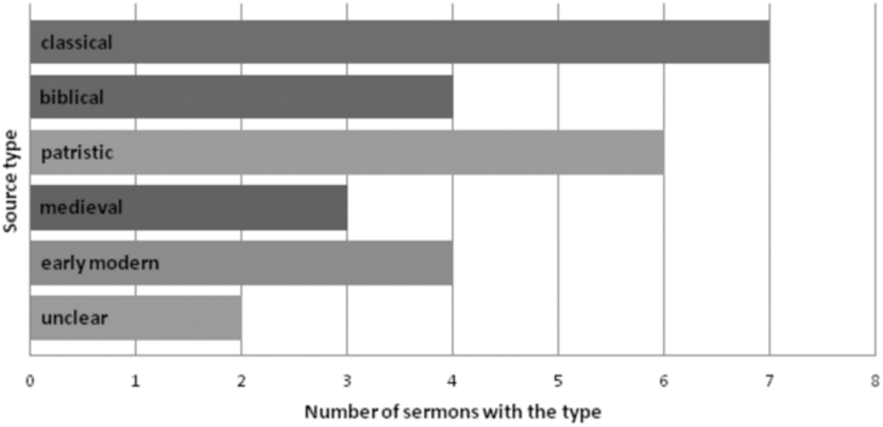


Figure 4: The number of *Lampeter Corpus* sermons citing the different source types

However, as Figure 4 shows, the overall frequency with which the different types of sources are cited does not correspond to their distribution across the sermons. Although classical sources are cited only ten times, this is done in seven of the *Lampeter Corpus* sermons. Code-switched citations of patristic sources are also fairly widely distributed, but biblical, medieval and early modern sources are less typical. RelA1642 and RelA1679, the two sermons with the most frequent switching, are the only ones to show the full range of source types. Medieval sources are only cited in RelA1642, RelA1669 and RelA1679. Other possible patterns in the distribution of the source types are less clear.

The Hebrew switches are both classified here as biblical quotations. About three quarters of the quotations in Greek are also from the Bible, but there are a few from patristic sources as well. It is notable that the Latin Vulgate is not quoted in these sermons at all, although prominent English divines of the early seventeenth century such as Lancelot Andrewes (1555–1626) and John Donne (1572–1631) had still used it extensively (Ferrell 2011: 30).

4.2.2 Terminology

Terms are here understood as code-switched expressions that are characteristic of a specialized field and are used to name or specify a particular designatum, as in examples (8) and (9) below. Unlike quotations, they are not traceable to a single source. Code-switched terms in sermons may reflect a need for precision, or even necessity if no established vernacular equivalent was available.

As with intertextual switches, the use of specialized terminology is also a way of demonstrating one's membership in a restricted professional community, and the two types place similar demands on the audience (cf. Pahta 2004: 81–83). In practice the boundary between them is fluid and their functions overlap. On the one hand, quotations with a source reference can be used in a manner similar to terms, and, on the other, terms have been introduced into the professional discourse at some point in time, and at least implicitly carry a reference to the earlier contexts in which they have been defined and used. Moreover, some expressions which are here classified as terms may have been intended as allusions to a specific source by the author, but were not identified as such in the analysis due to their brevity.

(8) Upon a sleight confession of sinne, they may receive a cheape absolution, and thereby are emboldned to renue the commission of it. That ***Taxa pœnitentiaria*** ['penance charge'], amongst the Papists, that low rate which is put upon the greatest sinnes in the Popes Custome house; [. . .] (RelA1642: 41–42)

(9) [. . .] with them in the act of conversion ***prima causa*** ['the first cause'] depends upon ***secunda*** ['the second (cause)'], the power of Gods grace must wayte upon the concurrence of our good nature. (RelA1642: 6)

Specialized terms make up about a fifth of all switches in the sermons. Unlike intertextual switches, they only occur in four sermons, in which their proportion of the total number of switches ranges between one sixth and two fifths (see Figure 2). The majority are in Latin, and nearly all are single nouns or short noun phrases. Looking at the fields of learning involved, about two thirds of the terms pertain to religion and a quarter to philosophy or logic. Apart from abstract concepts, titles and epithets form a fairly large part of the examples.

5 Discussion

Quantitative analysis of the *Lampeter Corpus* data supports and adds detail to the findings of earlier work based on more limited datasets and a more qualitative approach to the functions of code-switching in sermons. The results indicate that code-switching in sermons published in pamphlet form between 1640–1740 is overwhelmingly prefabricated and conventional. The most typical instances are short Latin quotations from classical or patristic sources, but references and specialized terms can also be expected alongside an occasional one-word switch into Greek.

Set against the results presented above, bishop Burnet's view that the use of foreign languages in English sermons was by 1692 a thing of the past is clearly prescriptive rather than descriptive: code-switching in sermons continues throughout the period covered by the *Lampeter Corpus*. However, there are indications of diachronic change in the code-switching patterns. The frequency of switching in the data shows a decline over time. Moreover, languages other than English and Latin virtually disappear from the data in the later decades. The modern foreign languages such as French and Italian found in other types of writing from this period are entirely absent from the material, as are the Celtic languages and other vernaculars used in the British Isles. The sermons appear to form two groups, with authors born before the 1640s switching more than the authors born in the latter half of the seventeenth century. Although caution is needed in interpreting patterns affected by a number of simultaneous factors, this distribution could be related to stylistic influences picked up early in the clergyman's career (cf. Mitchell 1932: 106). Other characteristics of switching in the data remain fairly stable over time: the average length of switches does not change markedly, and in terms of functions, code-switching is used in quotations in all of the sermons.

The data also suggest differences in the code-switching patterns of different religious communities which have not been apparent in earlier studies. The sermons with the fewest switches are by the two nonconformist authors, and these sermons are among the minority with no Greek. This finding can be interpreted in two interconnected ways: on the one hand, these two ministers lacked the university education available to their counterparts in the established churches of England and Ireland, which may have resulted in a poorer command of the learned languages. On the other, they operated within a different tradition where the use of the learned languages was not expected by hearers and readers. Nevertheless, intertextual switches into Latin occur in their sermons as well, giving a functional profile that is in line with the other sermons in the material.

The occasions and audiences of the sermons are very varied, which makes it difficult to quantify these factors and relate them to the code-switching patterns. In general terms, the sermons with the most switching and the most varied functional types of switches seem to have been given before audiences that included high-ranking and well-educated members of society. The variety of religious affiliations, topics and audiences in the material results from decisions made by the compilers of the corpus, and makes the material more representative of the sermon pamphlets published at the time. At the same time, however, the material is not ideal for answering research questions that require weighing the external variables against each other. Due to these limitations the *Lampeter Corpus* sermons are best considered a “diagnostic” corpus (cf. Rissanen 2008: 59). A larger and more structured set of material, such as the *Corpus of English Religious Prose* being compiled at the University of Cologne (see Kohnen 2010: 534–535), will hopefully provide a means to evaluate the approaches and results of this study and continue the analysis in the future.

Electronic resources

Early English Books Online. 2003–2014. ProQuest LLC. http://eebo.chadwyck.com/home (accessed 16 April 2014).

Eighteenth Century Collections Online. 2008. Detroit: Gale Cengage Learning. http://find.galegroup.com.libproxy.helsinki.fi/ecco/ (accessed 16 April 2014).

Lampeter Corpus = Schmied, Josef J., Claudia Claridge & Rainer Siemund (compilers). 1998. *The Lampeter Corpus of Early Modern English Tracts*. Chemnitz: [Chemnitz University of Technology]. https://www.tu-chemnitz.de/phil/english/sections/linguist/real/independent/lampeter/lampr.htm (accessed 30 November 2014).

Library of Latin Texts (Series A & B). 2012. Turnhout: Brepols. http://clt.brepolis.net.libproxy.helsinki.fi/cds/Default.aspx (accessed 16 April 2014).

Oxford Text Archive. https://ota.ox.ac.uk/ (accessed 30 November 2014).

Patrologia Latina Database. 1996–2014. ProQuest LLC. http://pld.chadwyck.co.uk/ (accessed 16 April 2014).

Thesaurus linguae Graecae. 2009. Irvine: University of California, Irvine. http://stephanus.tlg.uci.edu/ (accessed 16 April 2014).

References

Barrie-Curien, Viviane. 1993. Clerical recruitment and career patterns in the Church of England during the eighteenth century. In W. M. Jacob & Nigel Yates (eds.), *Crown and mitre: Religion and society in northern Europe since the Refomation*, 93–104. Woodbridge: The Boydell Press.

Burke, Peter. 1991. *Heu domine, adsunt Turcae*: A sketch for a social history of post-medieval Latin. In Peter Burke & Roy Porter (eds.), *Language, self, and society: A social history of language*, 23–50. Cambridge: Polity Press.

[Burnet, Gilbert.] 1692. *A discourse of the pastoral care. Written by the Right Reverend Father in God, Gilbert, Lord Bishop of Sarum*. London: Printed by R. R. for Ric. Chiswell. Early English Books Online. http://gateway.proquest.com/openurl?ctx_ver=Z39.88-2003&res_id=xri:eebo&rft_id=xri:eebo:citation:12623030 (accessed 1 December 2011).

Claridge, Claudia. 1999/2003. *"Life is ruled and governed by opinion":* The Lampeter Corpus of Early Modern English Tracts *manual of information*. http://khnt.hit.uib.no/icame/manuals/LAMPETER/LC-manual.pdf (accessed 24 February 2010).

Clarke, M. L. 1959. *Classical education in Britain 1500–1900*. Cambridge: Cambridge University Press.

Corbett, James A. [(ed.)]. 1955. *The* De instructione puerorum *of William of Tournai, O.P.* Notre Dame: The Mediaeval Institute, University of Notre Dame.

Davies, Horton. 1961. *Worship and theology in England*, vol. 3: *From Watts and Wesley to Maurice, 1690–1850*. Princeton, NJ: Princeton University Press.

Davies, Horton. 1975. *Worship and theology in England*, vol. 2: *From Andrewes to Baxter and Fox, 1603–1690*. Princeton, NJ: Princeton University Press.

Ettenhuber, Katrin. 2011. The preacher and patristics. In Peter McCullough, Hugh Adlington & Emma Rhatigan (eds.), *The Oxford handbook of the early modern sermon*, 34–53. Oxford: Oxford University Press.

Ferrell, Lori Anne. 2011. The preacher's Bibles. In Peter McCullough, Hugh Adlington & Emma Rhatigan (eds.), *The Oxford handbook of the early modern sermon*, 21–33. Oxford: Oxford University Press.

Fletcher, Alan J. 2013. Written versus spoken macaronic discourse in late medieval England: The view from a pulpit. In Judith A. Jefferson & Ad Putter (eds.), *Multilingualism in medieval Britain (c. 1066–1520): Sources and analysis*, 137–151. Turnhout: Brepols.

Gardner-Chloros, Penelope. 2009. *Code-switching*. Cambridge: Cambridge University Press.

Genette, Gérard. 1997. *Paratexts: Thresholds of interpretation*. Trans. Jane E. Lewin. Cambridge: Cambridge University Press. (Originally published in French as *Seuils*, 1987.)

Gordon, Alexander. 1897. Sedgwick, Obadiah (1600?–1658). In Sidney Lee (ed.), *Dictionary of national biography*, vol. LI: *Scoffin–Sheares*, 183–184. London: Smith, Elder, & Co.

Green, Ian. 2006. Orality, script and print: the case of the English sermon c. 1530–1700. In Heinz Schilling & István György Tóth (eds.), *Cultural exchange in early modern Europe*, vol. I: *Religion and cultural exchange in Europe, 1400–1700*, 236–255. Cambridge: European Science Foundation & Cambridge University Press.

Green, Ian. 2009. *Continuity and change in Protestant preaching in early modern England*. Friends of Dr Williams's Library, 60th lecture. London: Dr Williams's Trust.

Groeger, Dorothee. 2010. *The pamphlet as a form of publication: A corpus-based study of early modern religious pamphlets*. Aachen: Shaker Verlag.

Halmari, Helena & Timothy Regetz. 2011. Syntactic aspects of code-switching in Oxford, MS Bodley 649. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English*, 115–153. Berlin & Boston: Mouton de Gruyter.

Heller, Monica. 1988. Introduction. In Monica Heller (ed.), *Codeswitching: Anthropological and sociolinguistic perspectives*, 1–24. Berlin, New York & Amsterdam: Mouton de Gruyter.

Iglesias-Rábade, Luis. 1996. The multi-lingual pulpit in England (1100–1500). *Neophilologus* 80(3). 479–492.

Ivimey, Joseph. 1823. *A history of the English Baptists: Comprising the principal events of the history of Protestant dissenters, from the revolution in 1668* [sic] *till 1760; and of the London Baptist churches, during that period*, vol. 3. London: Printed for B. J. Holdsworth. Google Books. http://books.google.fi/books?id=djrSAAAAMAAJ&dq=editions%3AUOM39015012926864&pg=PR1#v=onepage&q&f=false (accessed 31 May 2010).

Kohnen, Thomas. 2010. Religious discourse. In Andreas H. Jucker & Irma Taavitsainen (eds.), *Historical pragmatics*, 523–547. Berlin & New York: Mouton de Gruyter.

Kytö, Merja (compiler). 1996. *Manual to the diachronic part of the* Helsinki Corpus of English Texts*: Coding conventions and lists of source texts*, 3rd edn. Helsinki: Department of English, University of Helsinki. http://khnt.hit.uib.no/icame/manuals/HC/INDEX.HTM (accessed 29 April 2012).

Lessenich, Rolf P. 1972. *Elements of pulpit oratory in eighteenth-century England (1660–1800)*. Köln & Wien: Böhlau Verlag.

Machan, Tim William. 2011. The visual pragmatics of code-switching in late Middle English literature. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English*, 303–333. Berlin & Boston: Mouton de Gruyter.

Mitchell, W. Fraser. 1932. *English pulpit oratory from Andrewes to Tillotson: A study of its literary aspects*. London, New York & Toronto: Society for Promoting Christian Knowledge & Macmillan.

Mullett, Michael. 1994. Radical sects and dissenting churches, 1600–1750. In Sheridan Gilley & W. J. Sheils (eds.), *A history of religion in Britain: Practice and belief from pre-Roman times to the present*, 188–210. Cambridge, MA & Oxford: Blackwell.

Nurmi, Arja & Päivi Pahta. 2004. Social stratification and patterns of code-switching in Early English letters. *Multilingua* 23. 417–456.

Nurmi, Arja & Päivi Pahta. 2010. Preacher, scholar, brother, friend: Social roles and code-switching in the writings of Thomas Twining. In Päivi Pahta, Minna Nevala, Arja Nurmi & Minna Palander-Collin (eds.), *Social roles and language practices in Late Modern English*, 135–162. Amsterdam & Philadelphia: John Benjamins.

O'Day, Rosemary. 1982. *Education and society 1500–1800: The social foundations of education in early modern Britain*. London & New York: Longman.

OED = *Oxford English Dictionary*. 1989 (2nd edn.), 2000– (3rd edn.). Oxford: Oxford University Press. http://www.oed.com/ (accessed 16 April 2014).

Oxford Dictionary of National Biography. 2004–. Oxford: Oxford University Press. Online edn. http://www.oxforddnb.com/ (accessed 22 March 2010).

Pahta, Päivi. 2004. Code-switching in medieval medical writing. In Irma Taavitsainen & Päivi Pahta (eds.), *Medical and scientific writing in late medieval English*, 73–99. Cambridge: Cambridge University Press.

Pahta, Päivi & Arja Nurmi. 2006. Code-switching in the *Helsinki Corpus*: A thousand years of multilingual practices. In Nikolaus Ritt, Herbert Schendl, Christiane Dalton-Puffer & Dieter Kastovsky (eds.), *Medieval English and its heritage: Structure, meaning and mechanisms of change*, 203–220. Frankfurt am Main: Peter Lang.

Raymond, Joad. 2003. *Pamphlets and pamphleteering in early modern Britain*. Cambridge: Cambridge University Press.

Rissanen, Matti. 2008. Corpus linguistics and historical linguistics. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: An international handbook*, vol. 1, 53–68. Berlin & New York: Walter de Gruyter.

Rivers, Isabel. 2009. Religious publishing. In Michael F. Suarez & Michael L. Turner (eds.), *The Cambridge history of the book in Britain*, vol. V: *1695–1830*, 579–600. Cambridge: Cambridge University Press.

Schaffer, Simon & Larry Stewart. 2005. Vigani and after: Chemical enterprise in Cambridge 1680–1780. In Mary D. Archer & Christopher D. Haley (eds.), *The 1702 chair of chemistry at Cambridge: Transformation and change*, 31–56. Cambridge: Cambridge University Press.

Schendl, Herbert. 1996. Text types and code-switching in medieval and Early Modern English. *VIEWS* 5(1–2). 50–62.

Schendl, Herbert. 2000a. Linguistic aspects of codeswitching in medieval English texts. In D. A. Trotter (ed.), *Multilingualism in later medieval Britain*, 77–92. Cambridge: D. S. Brewer.

Schendl, Herbert. 2000b. Syntactic constraints on code-switching in medieval texts. In Irma Taavitsainen, Terttu Nevalainen, Päivi Pahta & Matti Rissanen (eds.), *Placing Middle English in context*, 67–86. Berlin & New York: Mouton de Gruyter.

Schendl, Herbert. 2013. Code-switching in late medieval macaronic sermons. In Judith A. Jefferson & Ad Putter (eds.), *Multilingualism in medieval Britain (c. 1066–1520): Sources and analysis*, 153–169. Turnhout: Brepols.

Smith, Helen & Louise Wilson (eds.). 2011. *Renaissance paratexts*. Cambridge: Cambridge University Press.

Spolsky, Bernard. 2003. Religion as a site of language contact. *Annual Review of Applied Linguistics* 23. 81–94.

Spurr, John. 2006. *The Post-Reformation: Religion, politics and society in Britain 1603–1714*. Harlow: Pearson Education.

Taavitsainen, Irma. 2001. Changing conventions of writing: The dynamics of genres, text types, and text traditions. *European Journal of English Studies* 5(2). 139–150.

Tagliamonte, Sali A. 2012. *Variationist sociolinguistics: Change, observation, interpretation*. Malden, MA, Oxford & Chichester: Wiley-Blackwell.

Trueman, Carl. 2011. Preachers and medieval and Renaissance commentary. In Peter McCullough, Hugh Adlington & Emma Rhatigan (eds.), *The Oxford handbook of the early modern sermon*, 54–71. Oxford: Oxford University Press.

Weinreich, Uriel, William Labov & Marvin I. Herzog. 1968. Empirical foundations for a theory of language change. In W. P. Lehmann & Yakov Malkiel (eds.), *Directions for historical linguistics: A symposium*, 95–188. Austin, TX & London: University of Texas Press.

Wenzel, Siegfried. 1994. *Macaronic sermons: Bilingualism and preaching in late-medieval England*. Ann Arbor, MI: The University of Michigan Press.

Wood, Anthony à. 1691. *Athenæ Oxonienses: An exact history of all the writers and bishops who have had their education in the most ancient and famous University of Oxford, from the fifteenth year of King Henry the Seventh, Dom. 1500, to the end of the year 1690 [...] To which are added, the Fasti or Annals, of the said university, for the same time*, vol. 1. London: Printed for Tho. Bennet. Early English Books Online. http://gateway.proquest.com.libproxy.helsinki.fi/openurl?ctx_ver=Z39.88-2003&res_id=xri:eebo&rft_id=xri:eebo:citation:6 (accessed 19 April 2010).

Arja Nurmi, Jukka Tyrkkö, Anna Petäjäniemi and Päivi Pahta

9 The social and textual embedding of multilingual practices in Late Modern English: A corpus-based analysis

1 Introduction

Recent research has established that multilingual practices, evidenced in the alternating use of two or more languages, are characteristic of language use in various types of English writings from different historical periods. While several single genres and topic domains have received attention in this body of research, most studies are based on relatively small datasets. What we still need is a credible overview of the frequency and type of multilingual practices based on systematic corpus-based study. With the current availability of large masses of electronic text from historical periods of English, work on this scale is finally becoming possible.

This corpus-based study sets out to provide baseline evidence of the frequency and typology of multilingual practices in late modern England. The data comes from the 34-million-word *Corpus of Late Modern English Texts* 3.0 (CLMET3), where the multilingual passages have been identified using a range of complementary automatic and semi-automatic techniques, including a new corpus tool, *Multilingualiser*, developed specifically for processing multilingual data. The corpus has been enhanced with sociolinguistic and text-typological background information to facilitate the analysis of multilingual practices and language-external factors. The enhanced data allows us in this study to present an evidence-based overview of (1) the frequency of foreign-language passages in written English in 1710–1920, (2) the variety of languages used in these texts in addition to English, (3) the connections of multilingual practices and the social variables describing the authors of each text, and (4) the further text-typological features associated with the use of multilingual practices.

2 Multilingual practices in written language in late modern England

2.1 Late Modern English and England

The late modern period, conventionally defined as covering the eighteenth and nineteenth centuries, was characterised by a rising awareness of English as a

DOI 10.1515/9781501504945-009

major European language. The major changes to English lexis, grammar, spelling and orthography were mostly over by the end of the seventeenth century, and both public and learned discourses concerning the need to standardise and maintain the purity of English emerged for the first time. During the eighteenth century in particular, influenced by Enlightenment ideals, both dictionaries and grammars were published for the first time at great volume, and ordinary men and women felt a growing need to speak "proper" English and many took it upon themselves to attend lessons and read books that promised to improve elocution and, thereby, one's standing in polite society.

However, just as a good command of educated English was one of the marks of an educated and cultured person, so was a sufficient knowledge of foreign languages, in particular Latin, Greek and French. Latin and Greek were still the universal languages of learning and formed the backbone of young gentlemen's private education, and members of *le ton*, women in particular, were expected to read and speak French. A smattering of Italian, perhaps German as well, was expected to stick during one's grand tour, the common rite of passage for young men of the upper middle class, and Hobson-Jobson words were popular even in the language of those who had never been to the colonies.

A notable aspect in all this is the fact that much of this multilingualism took place within the community of English speakers. Although there were from time to time contacts with continentals and other foreigners, much of the code-switching was intended to index membership in a certain cultured and worldly society. The common attitude is expressed well in a passage from Arthur Conan Doyle's *Rodney Stone*, in which a boy's uncle imparts wisdom about code-switching to his nephew:

> "You sing, don't you, nephew?" he asked, suddenly.
>
> "Yes, sir, a little."
>
> "A baritone, I should fancy?"
>
> "Yes, sir."
>
> "And your mother tells me that you play the fiddle. These things will be of service to you with the Prince. Music runs in his family. Your education has been what you could get at a village school. Well, you are not examined in Greek roots in polite society, which is lucky for some of us. It is as well just to have a tag or two of Horace or Virgil: '**sub tegmine fagi**,' or '**habet foenum in cornu**,' which gives a flavour to one's conversation like the touch of garlic in a salad. It is not **bon ton** to be learned, but it is a graceful thing to indicate that you have forgotten a good deal. Can you write verse?" (Conan Doyle: *Rodney Stone*, 1896)

The late modern period was also the time when the novel emerged as the dominant form of English prose literature. This was a natural result of expanding literacy among the lower middle and the working classes, but also of fundamental change in how reading was increasingly viewed as a pleasurable and leisurely pursuit, rather than something associated with utility or devotional activity.[1] Novels, along with other popular genres of narrative prose such as travelogues and biographies, were frequently set in foreign lands or featured foreign characters, giving authors reason to use foreign languages as a means of characterisation and adding local flavour to the stories.

2.2 Multilingual practices in writing

The frequency of multilingual passages in texts can be assumed to vary according to genre, audience and writer. The selection of languages available to a writer depends on the writer's educational opportunities and the contemporary trends in language use.

Based on results obtained from the multi-genre *Helsinki Corpus*, the frequency of switched passages on average ranges from 1.8/1,000 words in Old English to 2.2 in Middle English and 1.0 in Early Modern English (Pahta and Nurmi 2006). The vast majority of switched passages were in Latin during all three periods (1.8 for OE, 2.2 for ME and 0.9 for EModE), and the role of other languages was negligible. The second most frequent language of switched passages in Old English and Early Modern English was Greek (0.02 and 0.06 respectively), while during the Middle English period French came second (0.03/1,000 words). Contemporary European languages such as Italian and Spanish only started appearing during the Early Modern English period, and only sparsely then. The topic domains most favouring switching at all ages were religion and science. These results are in agreement with e.g. Voigts (1996), Hunt (2000) and Pahta (2003, 2004a, 2004b) as regards scientific writing, and with e.g. Wenzel (1994), Machan (1994), Iglesias-Râbade (1996), Schendl (2000), Halmari and Adams (2002) and Pahta and Nurmi (2011) in the case of religious texts. For the eighteenth century, the trend with scientific writing containing much multilingual material continues (Nurmi and Pahta 2010), while the domain of religion is increasingly vernacular (Nurmi and Pahta 2010; Tuominen, this volume).

In personal correspondence, similar trends can be observed. In Late Middle English personal letters, the frequency of switching varies, but the average is 0.55/1,000 words for Latin passages and 0.16 for French, other languages only appearing as isolated instances. The writer's social status and profession were decisive for frequency of switching and language choice (Nurmi and Pahta

---

[1] The love of reading was an emergent concept across Europe during the eighteenth century. See Mäkinen (2013).

2004). This can be linked to the daily pursuits of the writers, so that the litigious gentry included legal Latin in their letters, while merchants involved in foreign trade brought elements of French, Italian and Dutch to discussions of business with partners on the continent. The group that stands out most in terms of Latin usage are the clergy, who reach a frequency of 1.85 Latin passages / 1,000 words in the data. The use of foreign languages in the merchants' letters can be related to multilingual practices evident in business documents of the time (see e.g. Wright 1992, 1995, 1998, 2000, 2001; Rothwell 2000), while the use of Latin in the domain of law has been noted in Davidson (2003, 2005), as well as Nurmi and Pahta (2013).

One vital conditioning factor in the case of multilingual practices is education. Education was the means of gaining access to both classical languages of higher learning and the contemporary languages, which were needed for purposes of daily work in some cases, but, particularly in the course of the eighteenth century, became increasingly the markers of membership in the educated elites. From the seventeenth century onwards, French was the lingua franca of the educated European, while Italian was the language of music and arts. This is reflected in the multilingual practices evident in personal correspondence in the course of the eighteenth century. In the case of women, the availability of formal education or the means of self-study was dependent on both social status and the benevolence of men, whether fathers or husbands. It is only in the eighteenth century that women begin to show multilingual elements in their writing to a greater extent, the frequency of switched passages ranging from 0.2 in the fifteenth century to 1.0 in the eighteenth (Nurmi and Pahta 2012).

Another relevant feature, in addition to genre, topic domain and the writer's educational opportunities, is the reader. One writer could vary both the languages used and the frequency of switching according to the intended recipient of a letter, for example. So clergyman and Classical scholar Thomas Twining had an average of 6.8 or 6.7 switched passages / 1,000 words, when writing to Charles Burney the father and the son (respectively), but only 3.9 when writing to Fanny Burney (daughter and sister to the two Charleses). The languages Twining chose for each recipient also match his relationship with them: in letters to Charles Burney sr. Twining switches frequently into French, but also into Italian, which was the language of their shared interest, music. When writing to the younger Charles Burney, a Classical scholar like Twining himself, the most frequent languages in addition to French are Latin and Greek. In the letters to Fanny Burney, almost all foreign-language passages are in French (Nurmi and Pahta 2010).

Thus, based on earlier research, we have arrived at three main hypotheses concerning the social variables related to multilingual practices. They are:

1. Education will play a role in the use of multilingual resources, because individual multilingualism (as opposed to societal multilingualism) is more common when there is access to structured language teaching (e.g. Nurmi and Pahta 2004, 2010, 2012; Pahta and Nurmi 2009, 2011).
2. Texts that (and authors who) have contacts with foreign environments (e.g., a novel set in France, a letter writer living in Italy, a travelogue) are more likely to show multilingual practices (e.g. Nurmi and Pahta 2004, 2012; Pahta and Nurmi 2006, 2009).
3. Gender and social class are related to educational opportunities, but they will not provide significant results in this data due to the scarcity of female writers and the relative uniformity of informants' social background.

3 Material and methods

3.1 Material

As primary data we used the *Corpus of Late Modern English Texts* 3.0 (CLMET3).[2] The corpus comprises 333 full-length texts published in 1710–1920; only texts by contemporary authors are included, so there are no reprints of early modern books. The combined word count is 34.3 million words covering a wide selection of genres ranging from academic treatises and learned essays to travelogues, biographies, novels and plays. Because the main purpose of CLMET3 is to serve as a generic corpus of Late Modern English, the selection of texts was primarily meant to serve the overall representativeness of the period and only secondly to represent the individual genres within it. The texts were harvested from the Project Gutenberg open access archive, edited for corpus use and checked for authenticity. The archive source introduces an opportunistic selection bias by favouring printed and generally prominent texts over a truly random sampling; however, in the present study this works in our favour because the primary research question concerns multilingualism in texts written for a general readership and one of the key objectives is to establish baseline evidence for the frequency of code-switching in the types of books an average reader might have encountered in the late modern period. Some authors, e.g. Jane Austen, Edmund

---

[2] CLMET3 was compiled by Hendrik de Smet, Hans-Jürgen Diller and Jukka Tyrkkö. A CQP-ready version of CLMET3.0 was released in October 2015 by Hendrik De Smet, Susanne Flach and Jukka Tyrkkö. The new version of the corpus, CLMET3.1, also comes with a new cleaned-up version of part-of-speech tagging. Like CLMET3.0, the new corpus is freely available from Hendrik De Smet at https://perswww.kuleuven.be/~u0044428/.

Burke and Lewis Carroll, are represented in the corpus by more than one text, but none has more than three. All authors are native speakers of English and nearly all flourished in Britain. CLMET3 is quite representative of gendered writing, with 71 texts (22 per cent) written by female authors and 259 (78 per cent) by male authors. The mean age of authors is 41.6 years. The corpus also includes three samples from periodicals with numerous authors and some co-authored texts which for the purposes of sociolinguistic analysis were assigned to the perceived main author; for example, the novel *Diary of a Nobody* (1888) was co-authored by brothers George and Weedon Grossmith, but we list it under the elder brother George. The periodicals were left out of the present analysis, as they represent a highly multi-authored text type and the number of samples is too small to allow reasonable generalisations to be made.

The accurate and uncontroversial assigning of genre labels is a notoriously difficult task.[3] Although ambiguity and uncertainty are to some extent an accurate reflection of the difficulty inherent to invariably simplistic genre labels, for the purposes of corpus linguistic analysis such highly detailed descriptors can render genre an effectively moot variable. Thus, for the purposes of the present study, the genre system in CLMET3 was cleaned up by collapsing certain categories into one, resolving double classifications, and assigning a category to texts which were originally categorised as X, or ambiguous. For example, several texts were classified in CLMET3 as both Biographies and Travelogues or as both Histories and Treatises, and these were revisited and the texts were assigned to the genre category that appeared the most appropriate. The pruned genre system comprises twelve genres, which when necessary can be further combined into the three macro-genres of *Drama*, *Fiction* and *Non-fiction* (see Table 1); although CLMET3 includes a small number of periodical articles as well, we leave them aside in the present study. The overall text counts are typically very modest in most individual genres and thus the macro-genre level is most useful when it comes to statistical analysis of frequency data.

Our focus on sociolinguistic metadata rises, firstly, from the research tradition of stratificational sociolinguistics and, secondly, from issues specific to the study of individual multilingualism identified in previous research by Pahta and Nurmi. In the stratificational tradition (e.g. Labov 1994, 2001; Chambers 2003; Nevalainen and Raumolin-Brunberg 2003), we have tracked the author's age (operationalised as year of birth), gender (as a binary variable; overwhelmingly male) and place of birth (in the broad categories of *South England*, *North England*, *Ireland*, *Scotland*, *Wales* and *Abroad*). We also made an initial classification into social classes, but this proved unworkable, since the clear majority

---

[3] See e.g. Diller, De Smet, and Tyrkkö (2011).

Table 1: Macro-genres and genres in CLMET3 with the number of texts in each category

Macro-genre	Genre	1710–1780	1780–1850	1850–1920	Total
Drama					**74**
	Drama	25	19	30	74
Fiction					**131**
	Novels	36	35	56	127
	Children's story	0	1	3	4
Non-fiction					**125**
	Biography	4	8	7	19
	Essay	3	2	1	6
	History	3	2	5	10
	Instruction	1	0	1	2
	Letters	4	6	4	14
	Religious	0	0	2	2
	Travel	1	5	4	10
	Treatise	18	20	24	62

of the writers represent the middle classes. In order to track some differences among the authors, they were given a main occupation following the information found in the *Oxford Dictionary of National Biography*, and these were classified into four main categories: *professional writer* (either of fiction or non-fiction, e.g. Charles Dickens, Henry Fielding), *academic* (e.g. Charles Darwin and Charles Babbage), *cultured* (a catch-all category including people who made their living in the arts, but whose main occupation was not writing, including e.g. Horace Walpole and Edward Gibbon) and *other* (e.g. Charlotte Brontë and Henry Rider Haggard). An individual writer's occupation has been classified varyingly along their lifespan, since many began in one category (particularly *cultured* or *other*) before they found success in writing and were able to become full-time authors.

For the study of multilingual practices, it is important to be familiar with the writer's linguistic profile, and with extralinguistic factors affecting it. In order to track these, we looked at both educational background and travel history. Both depend on sometimes haphazardly preserved information, and may not have been adequately described in our main source, the *Oxford Dictionary of National Biography*. The educational background of an author has been tracked in the form of five binary categories: *grammar school*, *university*, *other formal education*, *private tuition* and *education abroad*. These give us some indication of the languages available for our authors to learn, as, for example, a university education still meant the study of Latin. However, while education provides us with some idea of language teaching, it is difficult to estimate how much any individual would have actually learned during the course of their education. Likewise, it is possible that an individual may have acquired language skills which have not been recorded in biographies.

The other element of a writer's linguistic profile was their travel history. We have tracked this in broad categories of *Britain* (i.e. the writer never travelled abroad), *Europe* (except French-speaking countries and Italy), *Inner Circle* (following Kachru's classification) and *World* (with the possibility of contact with non-European languages). *French-speaking countries* and *Italy* were singled out, since French and Italian were two frequently occurring foreign languages, and we wanted the opportunity to observe the impact of e.g. travelling in Italy on a writer's use of Italian phrases in their writing. All these variables were tracked as binary options, and, with the exception of the category Britain, one writer could have "yes" in more than one category. Finally, we have made an estimate of the writers' overall linguistic profiles, listing the main foreign languages a writer would probably know based on their education and travel. We have binary categories for *Latin*, *Greek*, *French*, *Italian* and *German*.

3.2 Data extraction and the analytical procedure

The first order of business when analysing the frequency of switched passages is to identify them in a corpus. While such a task is doable by manual analysis when the corpus is suitably small, it becomes prohibitively time-consuming and labour-intensive when the corpus comprises 34 million words. On the other hand, a fully automated computational approach would be likely not to reach the required level of analytical prowess, either missing words and passages that a human reader considers code-switches (false negatives) or, conversely, falsely assigning as code-switches words and passages which are not (false positives). To avoid systematic problems of precision and recall, we decided to use a semi-automatic method by designing a multilingualism detection tool which would identify and tag potential multilingual passages at a reasonably high level of precision and then turn the data over for manual pruning. The tool, named *Multilingualiser*,[4] makes use of an iterative, stepwise algorithm which starts with simple dictionary look-up using one or more of the built-in wordlists (Latin, French, Italian and German) or a user-defined wordlist. The tagger then analyses the results of the first pass looking for strings of tags of the same type and their untagged collocates. The number of closely proximate tags increases the likelihood that the tagging is correct and that the untagged items in the same string should also be tagged. For example, take the following sentence from *Red Pottage* (1899) by Mary Cholmondeley:

---

[4] *Multilingualiser* was developed by Jukka Tyrkkö in LiveCode and will run in OS X, PC and Linux. The tool will be available to the research community free of charge once the development reaches the first stable version. A forthcoming development will see the inclusion of first-dating information based on the *Historical Thesaurus of English*. The algorithms and the statistical issues involved are discussed in detail in Tyrkkö, Nurmi, and Tuominen (2017).

(1) Rachel was not by nature *de celles qui se jettent dans l'amour comme dans un précipice*.

In this instance the software should understand that the entire end of the sentence, beginning with *de*, is French. What makes the task difficult is that *nature* is a visual diamorph and could be either French or English, and that *de* and *un* could also be Spanish, Portuguese or Italian (even if the latter would be quite rare).[5] Furthermore, in this particular instance the French dictionary did not include the word *jettent*, the third person plural present and subjunctive form of *jeter* ('throw'). Thus, to work properly the software needs to understand that the unambiguously French function words *celles*, *qui*, *se*, *dans* and *comme*, as well as the equally clear French content words *l'amour* and *précipice*, signal that *de* and *un* have a very high likelihood of also being French, and likewise that the likelihood of *jettent* being French is also very good, especially since it is also not an English word. By contrast, because *nature* is both an English word and a French word and it occurs in a sequence-initial position, it may be assigned as ambiguous with a high likelihood of French – which, in this particular case, it is in fact not. The tool is designed to give the user control over the likelihood weights of a variety of collocate positions within the sequence of items, as well as the freedom to set other related parameters, such as how to treat sentence breaks, capitalised items, known function words, and so on. Additional retrieval methods in *Multilingualiser* include a character n-gram-based method for discovering non-English words using word-initial and word-final character trigram sequences; for example, the sequence *ips-* appears word-initially only in the word *Ipswich* in English. *Multilingualiser* can also ignore proper nouns unless they occur within a longer sequence of foreign items.

The first round of tagging and manual pruning was performed using *Multilingualiser* and its in-built tag editor. The initial results of potential switched passages revealed that one-word instances showed a poor recall rate, since many of them were established loans, such as *auditorium* or *inferno*, or English

---

[5] For an introduction to and further discussion of the term visual diamorph, see ter Horst and Stam, this volume.

words which share a form with a foreign item, most typically French or Latin (See Tyrkkö, Nurmi, and Tuominen 2017 for more on this). This lead to our focusing – at least at this stage of research – on stretches of two or more words in a language besides English. Obviously, this method disregards some genuine code-switches, but our overall view, based on analysis of the initial datasets, is that the number of instances overlooked by this method is negligible. The segments of foreign words were then manually pruned for any remaining items that members of the research team considered to be English despite appearances, the results of the language identification were checked and corrected when necessary, and the rare languages were manually identified.

The switched passages were then manually classified into three main categories, to a large extent based on previous work by Nurmi and Pahta (see e.g. Nurmi and Pahta 2010, 2013; Pahta and Nurmi 2009). The categories present a continuum from more established to less established switches. Conventionalised passages, such as *cara sposa*, *fille de chambre* or *terra firma*, are typically 2–3 words long, appear frequently in English texts and may be familiar to a reader with very poor language skills (see examples 2–3). There are many expressions which could be classified as terms, including *a priori*, *beau monde*, *carte blanche*, *canto fermo* and *ipso facto*. We have chosen to include all such borderline cases at this stage of our study.

(2) The fair *fille de chambre* came close up to the bureau where I was looking for a card (Sterne, *A Sentimental Journey*)

(3) Now I assert, that whoever reasons after this manner, does *ipso facto* believe the actions of the will to arise from necessity (Hume, *A Treatise of Human Nature*)

The second category, requiring somewhat more linguistic skill from both reader and writer is that of prefabricated expressions. Typical examples include (4) and (5). These are usually quotations, proverbs and maxims; we have classified instances in this category only if the quotation could be identified as coming from a recognised source or the expression was listed as a proverb or otherwise appears to have been in general use. Also identification as a quotation within the corpus text itself has been accepted as proof of the prefabricated status, as in (4). Quotations are not necessarily of written data: they can also be reported speech (6).

(4) that it may always apply to itself that celebrated passage in **Lucan**, *Nec quenquam jam ferre potest Caesarve priorem, Pompeiusve parem*. Indeed, [. . .] (Fielding, *Amelia*)

(5) All which, from the words, *De gustibus non est disputandum*, and whatever else [. . .] (Sterne, *The Life and Opinions of Tristram Shandy*)

(6) D'Aubreu, the pert Spanish minister, said the other day at court to poor Alt, the Hessian, '*Monsieur, je vous félicite, Munster est pris*.' (Walpole, *Letters*)

Finally, the third category of switched passages contains all instances not classifiable as conventionalised or prefabricated. These are often longer than the other two, and require a higher level of command of the other language included as well. Reported speech in fiction (example 7) is included in this category, since it is the author's own production, not that of an actual interlocutor of his or hers. Similarly, a writer reporting his/her own words in conversation is regarded as free switching, even if they were first put together at an earlier time (8).

(7) 'Mr Western,' answered the lady, 'you may say what you please, *je vous mesprise de tout mon coeur*. I shall not therefore be angry.' (Fielding, *The History of Tom Jones*)

(8) he asked me if Mr. Pitt was like his sister, I told him, '*Qu'ils se ressembloient comme deux gouttes de feu*.' (Walpole, *Letters*)

The analysis of frequency data will be presented in Section 4, but a brief general comment about the statistical nature of this type of data is in order here. As is well-attested, corpus linguistic data is rarely, if ever, derived from truly random samples and the linguistic features under investigation are even more rarely seen to be normally distributed, that is, linguistic data often does not follow a Gaussian curve. Consequently, both the descriptive and inferential statistical methods used ought to be robust. To take a simple example, the mean (or average) frequency of a small dataset can be easily skewed by one of two texts that show an unusually high frequency. While some linguists choose to interpret this as merely an artefact of the fickle nature of linguistic data, it is often wiser to use methods which deal with outliers in a more organised fashion – or at the very least, it is important to be aware of outliers in the data. The normality of a dataset can be examined in a number of different ways, but one of the most convenient of these is a visual examination of a quantile-quantile plot where the quantiles of a primary dataset are plotted against normal quantiles. In a normally distributed dataset the data points follow a diagonal straight line, while in a skewed dataset they do not. As Figure 1 illustrates, the distribution of switched chunks in CLMET3 is far from normal: there are many texts with very few if any tags and then a smaller number with high frequencies.

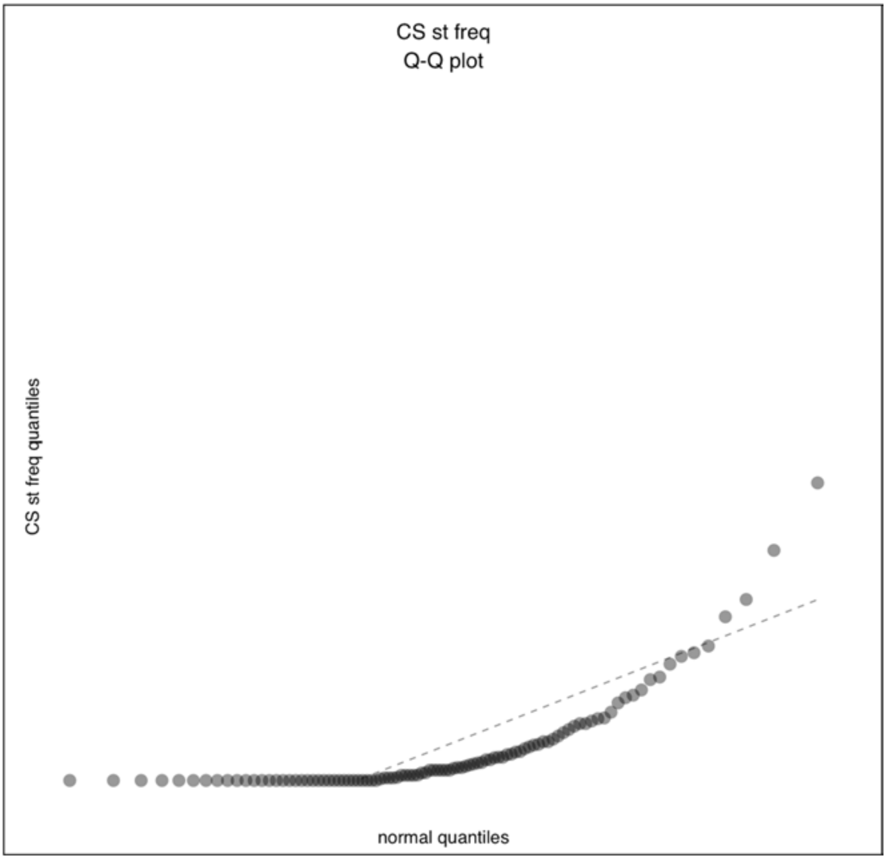


Figure 1: Quantile-quantile plot showing the non-normality of the distribution of code-switched segments (of two words or longer) in CLMET3

To deal with the distributional properties of our data, we will therefore use robust non-parametric methods of inferential analysis without making assumption of normality. Although it may be argued that some parametric tests such as one-way Anova are not particularly sensitive to non-normality, our data is too strongly skewed for us to consider them. In practice then, we use the Wilcoxon ranked sum test (also known as the Mann-Whitney U test) in our monofactorial analysis, comparing each pair of factor levels separately. Similarly to the parametric independent sample Student's t-test, the null hypothesis in the Wilcoxon test is that the two samples tested come from the same population and the alternative hypothesis is that the populations are different. For the Wilcoxon test the values of the continuous variable (here, standardised frequencies) are ranked, that is, all the datapoints are combined into a single dataset, the values are organised according to size and assigned ordinal ranks, after which the ranks are sorted back to the original groupings. We report the means and standard deviations as well as the z-scores used in calculating the approximate p-value. The null hypothesis is rejected at the significance level of 0.05; in the notation, we denote $p < 0.05$ with *, $p < 0.01$ with ** and $p < 0.001$ with ***.

4 Results: Overview of languages and their frequencies, genre findings, sociolinguistic findings

The analysis reveals that the mean frequency of foreign-language segments of two words or longer (hereafter CS segments) is 0.14/1,000 words and the median is 0.04/1,000 words (Figure 2).[6] This is clearly lower than the result for the Early Modern English period of the *Helsinki Corpus*, as well as the frequencies attested in personal correspondence in the eighteenth century. These numbers hide a wide range, however: the highest frequency of CS segments, 2.59/1,000 words, is found in Laurence Sterne's *A Sentimental Journey through France and Italy* (1768). The mean chunk length in the corpus is 13 in words and the median 4 words. The longest chunk in the corpus is 2,944 words in *Stories from the Italian Poets* (1845) by Henry James. There are 104 texts with no CS segments, though it is important to remember that single-word switches are not counted here. There is no correlation across the timeline, which means that in terms of both frequency and switch type CS segments occurred in a steady manner throughout the late modern period.

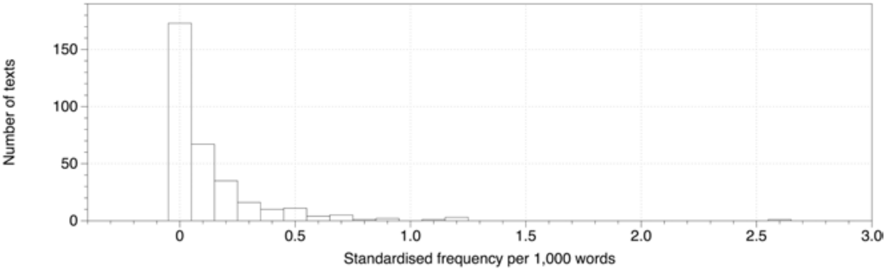


Figure 2: Frequency of code-switched segments (of two words or longer) in CLMET3 by text

---

[6] Although switching back to English from a foreign language segment is technically also a code-switch, we count switch-points exclusively as transitions from English to another language.

The majority of the CS segments in CLMET3 are in French or Latin; the two are essentially equal when it comes to standardised frequencies of chunks (see Table 2). There are twenty-one different languages attested altogether; the identification of five is uncertain.

Table 2: CS segments in CLMET3

Language	CS segments	Total word count in the CS segments
Arabic	18	53
Buginese?	1	6
Dutch	3	9
Egyptian?	2	6
French	1937	21924
Gaelic	1	4
German	110	2630
Greek	270	3018
Hawaian?	1	5
Hindi	1	2
Indonesian	6	12
Irish	6	21
Italian	358	17527
Latin	2156	23544
Lenape?	1	18
Malay	1	2
Malay?	1	2
Portuguese	19	108
Romani / Caló	8	41
Samoan	1	8
Sinhalese	1	2
Spanish	26	154

As Table 2 shows, the foreign languages can be roughly divided into two groups: high-frequency languages and low-frequency languages. Latin and French are by far the most common languages in terms of number of individual foreign-language segments, but Italian comes relatively close in word count, mainly due to several extremely long switches. The rare languages typically occur in a single text in CLMET3, and consequently any differences observed between the low-frequency languages are essentially random artefacts and they cannot be considered indicative of language-specific frequency differences; for example, although CLMET3 happens to feature Malay or (possibly invented) Egyptian in small frequencies, one should not expect to encounter these languages in Late Modern English as a matter of course. The correlation between the number of segments and word count is illustrated in Figures 3 and 4.

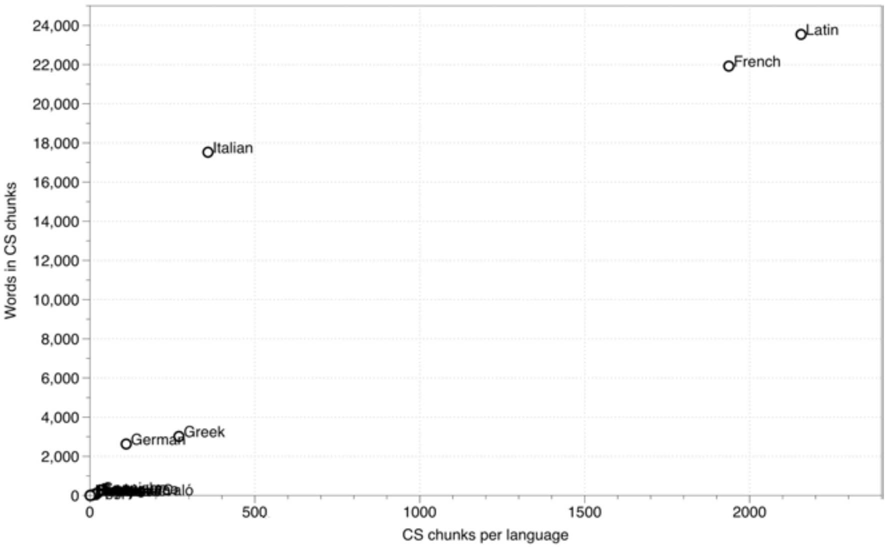


Figure 3: Scatter plot displaying correlation between number of CS segments and their combined word counts

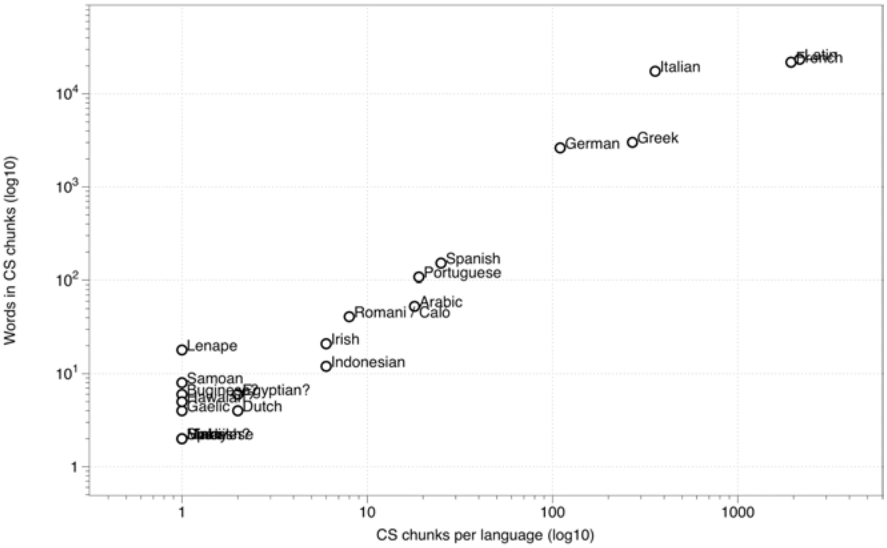


Figure 4: Scatter plot displaying correlation between number of CS segments and their combined word counts using $\log_{10}$ scale

The switch type variable reveals that the vast majority of CS segments are pre-fabricated in nearly all the languages (see Figure 5). With the exception of Hindi and Irish, both of which are extremely infrequent, French is the only language where the majority of CS segments are not pre-fabricated, but rather conventionalised. Perhaps the most striking observations concern Latin which, despite endless years of cramming in school, is hardly ever used in a free and original fashion: only 77 (3.5%) chunks out of 2,154 have been categorised as free switches.

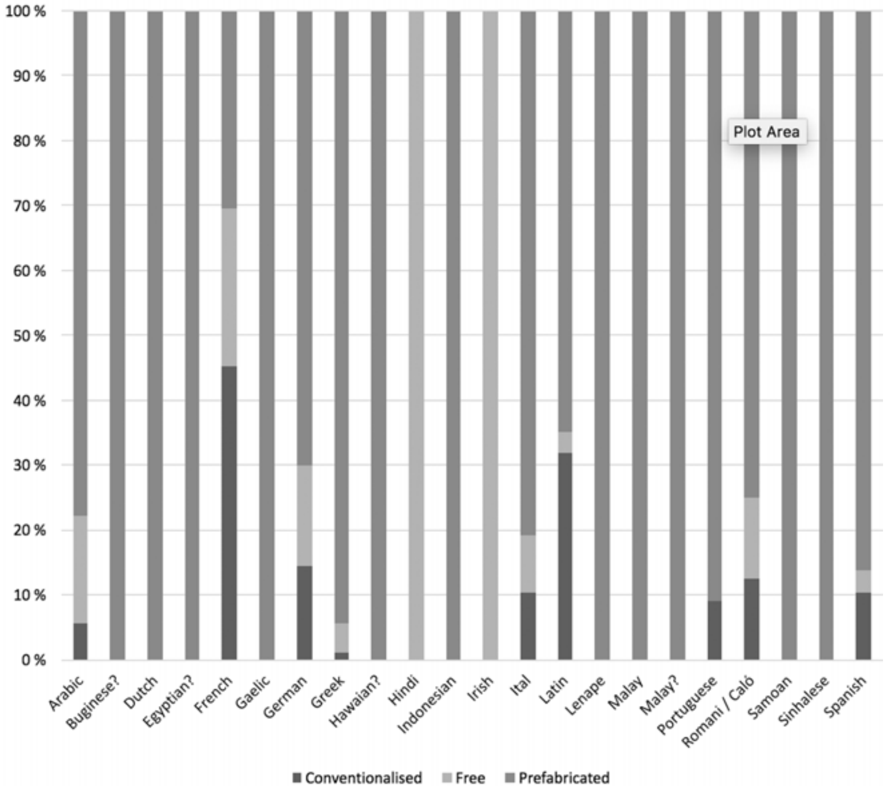


Figure 5: Switch type proportions by language

Moving on to monofactorial analysis and starting with the macro-genres, we see that Non-fiction texts generally contain more switched passages than either Drama or Fiction, though there are notable outliers in both; the differences between Non-fiction and both Drama and Fiction are significant. This finding agrees perfectly with the previous observations concerning the prevalence of pre-fabricated switches: the most typical function of foreign language use in Non-fiction is quoting from original sources.

Table 3: All CS segments in macro-genres

Macro-genre	n	Overall mean (/1,000 words)	Standard deviation	Wilcoxon Each Pair Comparisons:
Drama	74	0.11	0.21	Non-Fiction vs. Fiction Z = 5.44, p = *** Non-Fiction vs. Drama Z = 5.22, p = *** Fiction vs. Drama Z = 1.69, p = 0.20
Fiction	131	0.10	0.25	
Non-fiction	124	0.20	0.25	

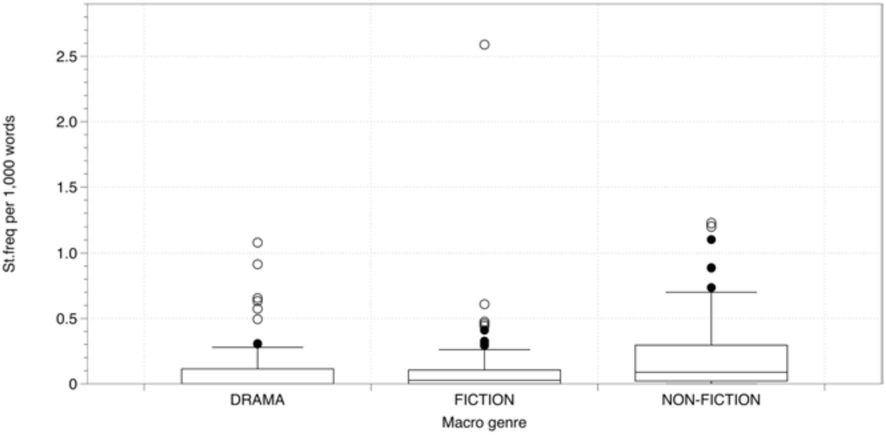


Figure 6: All CS segments in macro-genres

Predictably, Non-fiction writing contains Latin, Greek and German in greater frequencies than Fiction or Drama, while French is used more in Drama and in Fiction. Perhaps against expectations, Italian appears most frequently in Non-fiction. Although Italian was a fashionable language among the upper classes, few possessed real mastery of the language. Knowing this, authors of Fiction or Drama usually limited the use of Italian to conventionalised greetings and exclamations. By contrast, Non-fiction authors who use Italian, such as Edward Gibbon, do so knowing that their readers are likely to understand, or at least appreciate, long quotes in the original language.

Links between sociolinguistic parameters and code-switching confirm many of the hypotheses noted earlier in Section 2.2. For example, it appears clear that while gender or place of birth do not predict the multilingual practices in any way, the author's occupation plays a central role in the use of foreign languages, with Academic authors using multilingual resources most frequently and Professional Writers the least. Statistically significant differences can be observed

Table 4: Appearance of most frequent languages in the three supergenres

LATIN

Macro-genre	n	Overall mean (/1,000 words)	Standard deviation	Wilcoxon Each Pair Comparisons:
Drama	74	0.07	0.19	Non-Fiction vs. Fiction Z = 4.55, p = *** Non-Fiction vs. Drama Z = 4.05, p = *** Fiction vs. Drama Z = 1.28, p = 0.19
Fiction	131	0.02	0.05	
Non-fiction	124	0.16	0.66	

FRENCH

Macro-genre	n	Overall mean (/1,000 words)	Standard deviation	Wilcoxon Each Pair Comparisons:
Drama	74	0.14	0.54	Non-Fiction vs. Fiction Z = 2.22, p = * Non-Fiction vs. Drama Z = 2.07, p = ** Fiction vs. Drama Z = 1.51, p = 0.13
Fiction	131	0.03	0.07	
Non-fiction	124	0.22	1.94	

GERMAN

Macro-genre	n	Overall mean (/1,000 words)	Standard deviation	Wilcoxon Each Pair Comparisons:
Drama	74	0.002	0.013	Non-Fiction vs. Drama Z = 1.93, p = * Non-Fiction vs. Fiction Z = 1.04, p = 0.29 Fiction vs. Drama Z = 1.21, p = 0.22
Fiction	131	0.002	0.007	
Non-fiction	124	0.006	0.03	

ITALIAN

Macro-genre	n	Overall mean (/1,000 words)	Standard deviation	Wilcoxon Each Pair Comparisons:
Drama	74	0.003	0.02	Non-Fiction vs. Drama Z = 3.17, p = *** Non-Fiction vs. Fiction Z = 2.24, p = * Fiction vs. Drama Z = 1.49, p = 0.13
Fiction	131	0.004	0.02	
Non-fiction	124	0.02	0.10	

GREEK

Macro-genre	n	Overall mean (/1,000 words)	Standard deviation	Wilcoxon Each Pair Comparisons:
Drama	74	0.01	0.05	Non-Fiction vs. Drama Z = 2.57, p = *** Non-Fiction vs. Fiction Z = 2.84, p = *** Fiction vs. Drama Z = 0.50, p = 0.61
Fiction	131	0.002	0.02	
Non-fiction	124	0.017	0.07	

between Professional Writers and Academics, and between Professional Writers and Cultured authors. By contrast, Academic and Cultured authors cannot be said to differ, nor Professional Writers and Others. This division can be understood in terms of the authors' professions, but also as a reflection of their respective audience designs.

Table 5: All CS segments by occupation

Occupation	n	Overall Mean (/1,000 words)	Standard deviation	Wilcoxon Each Pair Comparisons:
Academic	30	0.16	0.18	Cultured vs. Academic Z = 0.25, p = 0.79 Professional Writers vs. Others Z = 1.04, p = 0.29 Others vs. Cultured Z = 1.69, p = 0.09 Others vs. Academic Z = 1.75, p = 0.08 Professional Writers vs. Cultured Z = 2.66, p = *** Professional Writers vs. Academic Z = 2.66, p = ***
Cultured	56	0.22	0.30	
Other	73	0.12	0.20	
Professional Writer	170	0.11	0.25	

The situation changes slightly when we turn to specific languages. Focusing on the two most frequently used foreign languages, we see that Academics and Others use Latin the most frequently, and that the differences in Latin use are statistically significant between Academics and Others, Academic and Cultured authors, and Professional Writers and Academics. By contrast, French is used most frequently by Cultured authors and the least by Academics.

The switch types correlate with the author's occupation as well. When it comes to conventionalised CS segments, all occupations use them more or less equally and there are no statistically significant differences. Free switches are

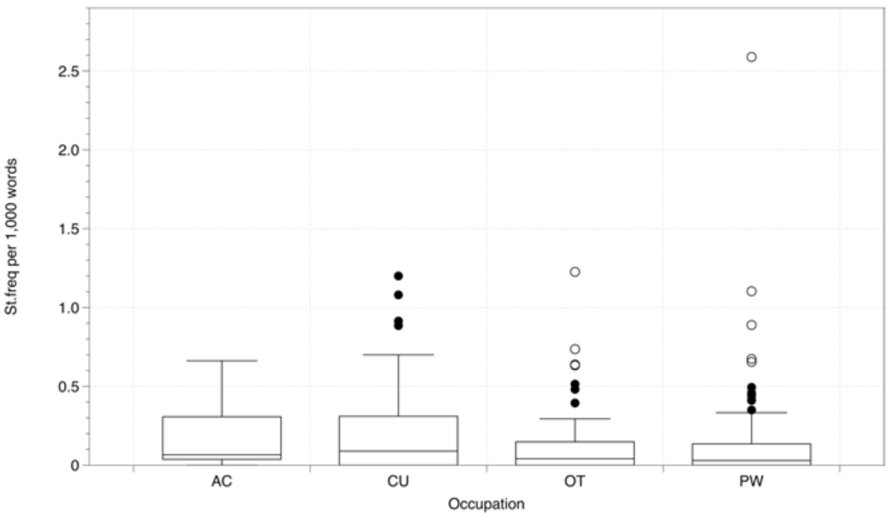


Figure 7: Box plot of multilingual practices and occupation

used the most by Cultured authors and a statistically significant difference is seen between them and Others. Finally, pre-fabricated CS segments are used similarly by Academic and Cultured authors, and Professional Writers and Others, respectively, the two pairs being statistically different.

If we treat education as a single factor, it is immediately clear that it is perhaps the single best predictor when it comes to high rates of multilingual practices. Even grammar school prepares a person for including multilingual elements in their writing – 0.14/1,000 words (n = 60) vs. 0.12/1,000 words (n = 273), p = * – but university education is the real threshold. The mean frequency of switches in texts written by university graduates is 0.17/1,000 words (n = 139) compared to 0.09/1,000 words (n = 194) in texts written by those who did not attend a university; the difference is statistically significant (Figure 8).

When it comes to specific languages, Latin and Greek are strongly associated with university education, but French, Italian and German show no statistically significant differences between the two groups (Table 8). Note that although the frequency differences between the two groups appears quite large when it comes to French, the difference is not statistically significant because there is great variation within each group. Naturally, there are significant covariances here when it comes to the author's education and the types of text they are likely to produce; for example, Academic texts are usually written by university graduates.

Table 6: Most frequent languages by occupation

LATIN

Occupation	n	Overall Mean (/1,000 words)	Standard deviation	Wilcoxon Each Pair Comparisons:
Academic	30	0.11	0.16	Cultured vs. Academic Z = 1.85, p = * Professional Writers vs. Others Z = 0.69, p = 0.48 Others vs. Cultured Z = 0.42, p = 0.67 Others vs. Academic Z = 2.38, p = ** Professional Writers vs. Cultured Z = 1.11, p = 0.26 Professional Writers vs. Academic Z = 3.25, p = ***
Cultured	56	0.08	0.18	
Other	73	0.17	0.85	
Professional Writer	170	0.04	0.12	

FRENCH

Occupation	n	Overall Mean (/1,000 words)	Standard deviation	Wilcoxon Each Pair Comparisons:
Academic	30	0.03	0.04	Cultured vs. Academic Z = 1.25, p = 0.20 Professional Writers vs. Others Z = −1.01, p = 0.82 Others vs. Cultured Z = −1.63, p = 0.3 Others vs. Academic Z = −0.21, p = 0.82 Professional Writers vs. Cultured Z = −2.78, p = *** Professional Writers vs. Academic Z = −0.88, p = 0.37
Cultured	56	0.18	0.59	
Other	73	0.34	2.53	
Professional Writer	170	0.04	0.11	

Education abroad also predicts an apparent tendency for multilingual practices. Those educated abroad have an overall mean frequency of 0.22/1,000 words (n = 33) while those educated in Britain show a frequency of 0.12 (n = 296); z = 2.98, p = ***. As for specific languages, education abroad has a statistically significant effect on the use of Latin, French and Greek, but not on Italian and German. Interestingly, the higher frequency is associated with the use of conventionalised switches but not pre-fabricated or free switches. The data also shows

Table 7: Switch types by occupation

CONVENTIONALISED

Occupation	n	Overall Mean (/1,000 words)	Standard deviation	Wilcoxon Each Pair Comparisons:
Academic	30	0.06	0.19	Cultured vs. Academic Z = −0.19, p = 0.84 Professional Writers vs. Others Z = −0.61, p = 0.54 Others vs. Cultured Z = −0.95, p = 0.33 Others vs. Academic Z = −0.61 p = 0.52 Professional Writers vs. Cultured Z = −1.70, p = 0.26 Professional Writers vs. Academic Z = −1.11, p = 0.26
Cultured	56	0.08	0.20	
Other	73	0.05	0.09	
Professional Writer	170	0.05	0.12	

FREE

Occupation	n	Overall Mean (/1,000 words)	Standard deviation	Wilcoxon Each Pair Comparisons:
Academic	30	0.003	0.01	Cultured vs. Academic Z = 1.81, p = 0.06 Professional Writers vs. Others Z = 1.62, p = 0.10 Others vs. Cultured Z = −2.15, p = * Others vs. Academic Z = 0.24, p = 0.80 Professional Writers vs. Cultured Z = −0.96, p = 0.33 Professional Writers vs. Academic Z = 1.39, p = 0.16
Cultured	56	0.03	0.11	
Other	73	0.008	0.03	
Professional Writer	170	0.02	0.12	

PREFABRICATED

Occupation	n	Overall Mean (/1,000 words)	Standard deviation	Wilcoxon Each Pair Comparisons:
Academic	30	0.09	0.16	Cultured vs. Academic Z = –0.10, p = 0.91 Professional Writers vs. Others Z = –1.17, p = 0.24 Others vs. Cultured Z = –1.85, p = * Others vs. Academic Z = –1.72, p = 0.08 Professional Writers vs. Cultured Z = –3.32, p = *** Professional Writers vs. Academic Z = –2.90, p = ***
Cultured	56	0.10	0.17	
Other	73	0.06	0.15	
Professional Writer	170	0.04	0.12	

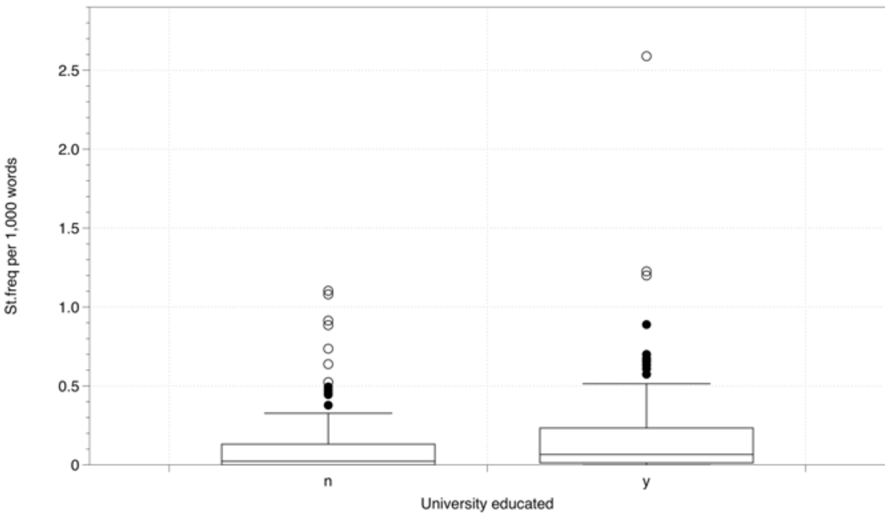


Figure 8: CS segments and university education

Table 8: Most frequent languages and university education

Language	St.freq. among university educated	St. freq, among non-university graduated	Z	p
Latin	0.13	0.05	3.33	***
French	0.21	0.07	0.65	0.51
Italian	0.01	0.009	0.67	0.50
German	0.005	0.002	−0.01	0.99
Greek	0.011	0.009	2.07	*

that spending time in a French-speaking country increased one's likelihood of using French, Latin, Italian and even German, but not Greek, while spending time in Italy increased the likelihood that a person code-switches to any of the five most common foreign languages.

5 Conclusion

To our knowledge, the present study is the first corpus linguistic study of historical multilingual practices to attempt a frequency-based analysis of code-switching using a medium-sized, genre-stratified corpus. One of the main motivations behind this study is the desire to go beyond hand-picked examples and the idiolects of specific authors, and to understand multilingualism as a more general feature of language use. By establishing evidence-based baseline frequencies for foreign content in English texts we provide a backdrop against which the findings of more intensively focused studies can be evaluated, but also to recognise outliers and significant predictors which smaller datasets cannot provide and to test hypotheses formed on the basis of small-scale studies.

Based on earlier studies of historical multilingualism, we knew that the overall frequency of foreign content is generally relatively low in English texts, and consequently a semi-automatic method of discovery was developed for that purpose in the course of the project (see e.g. Tyrkkö, Nurmi, and Tuominen 2017). Our experience with the Multilingualiser suggests that further training of the tool and analysis of the results may help in automatic separation of visual diamorphs, improving precision. Other issues that we will continue to work on involves developing a more data-driven method for disambiguating between conventionalised short phrases that ought to be considered part of the English lexicon and those that remain effectively foreign, and exploring the context-dependent nature of "foreignness" across genres and registers.

It is hardly surprising that the main findings follow what has been established in earlier studies. The prevalence of Latin and French and the link of multilingual practices to the author's university education and profession, as well as to the intended readership of the text all play a role in the big picture of when and how multilingual practices are activated in the conscious or subconscious selection of resources from a writer's repertoire. Consequently, we argue that the three hypotheses set up in Section 2.2 are confirmed: an author's education has been shown to play a major role in the frequency of multilingual practices, texts directly associated with foreign countries likewise features highest frequencies of foreign content, and gender and social class cannot be adequately examined due to the skewed nature of the present dataset. In addition to these findings, a key observation concerns the complexity of code-switching as a phenomenon: specific foreign languages and switch types are used at different frequencies depending on the macro genre and the author's sociolinguistic background variables. However, it is important to note the numerous co-variances within the metadata, particularly between the strongest predictor variables listed above. This does not mean that the individual factors would not predict the use of multilingual resources, but we need to be careful to note that the dynamics are more complex than they may appear. These findings clearly show that a monofactorial analysis needs to be explored further with more sophisticated multifactorial approaches (see e.g. Tyrkkö and Nurmi 2017).

Questions still requiring answers include the language proficiency of the reading public, which is difficult to chart in variables that are easy to describe consistently. Tyrkkö, Nurmi, and Tuominen (2017) describes some attempts in this direction, but further means of connecting the intended and actual reading public to the texts are needed. Another major question is the changing global context. In our data, the expanding number of languages over the timeline and the inclusion of exotic languages echoes of the expansion of the British Empire and the exploration of the world in the spirit of the Enlightenment, but approaching these wider questions would require a more specific dataset, connected to those particular endeavours in more intimate detail. At the same time, it should be noted that the cultural context of even the home environment of the writers and readers of the texts we have studied here keeps changing. Processes such as democratisation, mediatisation and secularisation brought changes not only to languages deemed important to study, but to access to education. What was a correct way of writing kept changing, including the foreign prestige languages, and the increasingly important role of English as a language of significance both at home and abroad should not be overlooked when considering these political, social and philosophical shifts.

References

Chambers, J. K. 2003. *Sociolinguistic theory: Linguistic variation and its social significance*. Malden: Blackwell.

CLMET3 = *Corpus of Late Modern English Texts* 3.0. https://perswww.kuleuven.be/~u0044428/.

Davidson, Mary Catherine. 2003. Code-switching and authority in late medieval England. *Neophilologus* 87(3). 473–486.

Davidson, Mary Catherine. 2005. Discourse features of code-switching in legal reports in late medieval England. In Janne Skaffari, Matti Peikola, Ruth Carroll, Risto Hiltunen & Brita Wårvik (eds.), *Opening windows on texts and discourses of the past*, 343–351. Amsterdam & Philadelphia: John Benjamins.

Diller, Hans-Jürgen, Hendrik De Smet & Jukka Tyrkkö. 2011. A European database of descriptors of English electronic texts. *The European English Messenger* 19(2). 21–35.

Halmari, Helena & Robert Adams. 2002. On the grammar and rhetoric of language mixing in *Piers Plowman*. *Neuphilologische Mitteilungen* 103(1). 33–50.

Helsinki Corpus. http://www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus/.

Historical Thesaurus of English. http://historicalthesaurus.arts.gla.ac.uk/.

Hunt, Tony. 2000. Code-switching in medical texts. In D. A. Trotter (ed.), *Multilingualism in later medieval Britain*, 131–147. Cambridge: D.S. Brewer.

Iglesias-Rábade, Luis. 1996. The multi-lingual pulpit in England (1100–1500). *Neophilologus* 80(3). 479–492.

Labov, William. 1994. *Principles of linguistic change*, vol. 1: *Internal factors*. Oxford: Blackwell.

Labov, William. 2001. *Principles of linguistic change*, vol. 2: *Social factors*. Oxford: Blackwell.

Machan, Tim William. 1994. Language contact in *Piers Plowman*. *Speculum* 69(2). 359–385.

Mäkinen, Ilkka. 2013. Why people read: Jean-Jacques Rousseau on the love of reading. In Dorrit Gustafsson & Kristina Linnovaara (eds.), *Essays on libraries, cultural heritage and freedom of information*, 127–136. Helsinki: National Library of Finland.

Nevalainen, Terttu & Helena Raumolin-Brunberg. 2003. *Historical sociolinguistics: Language change in Tudor and Stuart England*. London: Pearson Education.

Nurmi, Arja & Päivi Pahta. 2004. Social stratification and patterns of code-switching in early English letters. *Multilingua* 23(4). 417–456.

Nurmi, Arja & Päivi Pahta. 2010. Preacher, scholar, brother, friend: Social roles and code-switching in the writings of Thomas Twining. In Päivi Pahta, Minna Nevala, Arja Nurmi & Minna Palander-Collin (eds.), *Social roles and language practices in Late Modern English*, 135–161. Amsterdam & Philadelphia: John Benjamins.

Nurmi, Arja & Päivi Pahta. 2012. Multilingual practices in women's English correspondence 1400–1800. In Mark Sebba, Shahrzad Mahootian & Carla Jonsson (eds.), *Language mixing and code-switching in writing: Approaches to mixed-language written discourse*, 44–67. New York & London: Routledge.

Nurmi, Arja & Päivi Pahta. 2013. Multilingual practices in the language of the law: Evidence from the *Lampeter Corpus*. In Jukka Tyrkkö, Olga Timofeeva & Maria Salenius (eds.), *Ex philologia lux: Essays in honour of Leena Kahlas-Tarkka*, 187–205. Helsinki: Société Néophilologique.

Oxford Dictionary of National Biography. http://www.oxforddnb.com/.

Pahta, Päivi. 2003. On structures of code-switching in medical texts from medieval England. *Neuphilologische Mitteilungen* 104(2). 197–210.

Pahta, Päivi. 2004a. Code-switching in medieval medical writing. In Irma Taavitsainen & Päivi Pahta (eds.), *Medical and scientific writing in Late Medieval English*, 73–99. Cambridge: Cambridge University Press.

Pahta, Päivi. 2004b. 'So seiþ *idem comentator*': Code-switching and organisation of knowledge in John Trevisa's translation of *De proprietatibus rerum*. In Alicia Rodríguez Álvarez & Francisco Alonso Almeida (eds.), *Voices on the past: Studies in Old and Middle English language and literature*, 35–48. A Coruña: NetBiblo.

Pahta, Päivi & Arja Nurmi. 2006. Code-switching in the *Helsinki Corpus*: A thousand years of multilingual practices. In Nikolaus Ritt, Herbert Schendl, Christiane Dalton-Puffer & Dieter Kastovsky (eds.), *Medieval English and its heritage*, 203–220. Frankfurt: Peter Lang.

Pahta, Päivi & Arja Nurmi. 2009. Negotiating interpersonal identities in writing: Code-switching practices in Charles Burney's correspondence. In Arja Nurmi, Minna Nevala & Minna Palander-Collin (eds.), *The language of daily life in England (1400–1800)*, 27–52. Amsterdam & Philadelphia: John Benjamins.

Pahta, Päivi & Arja Nurmi. 2011. Multilingual discourse in the domain of religion in medieval and early modern England: A corpus approach to research on historical code-switching. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English*, 219–251. Berlin & Boston: Mouton de Gruyter.

Rothwell, William. 2000. Aspects of lexical and morphosyntactical mixing in the languages of medieval England. In In D. A. Trotter (ed.), *Multilingualism in later medieval Britain*, 213–232. Cambridge: D. S. Brewer.

Schendl, Herbert. 2000. Syntactic constraints on code-switching in medieval texts. In Irma Taavitsainen, Terttu Nevalainen, Päivi Pahta & Matti Rissanen (eds.), *Placing Middle English in context*, 67–86. Berlin & New York: Mouton de Gruyter.

Tyrkkö, Jukka & Arja Nurmi. 2017. Analysing multilingual practices in Late Modern English: Parameter selection and recursive partitioning in focus. In Sebastian Hoffmann, Andrea Sand & Sabine Arndt-Lappe (eds.), *Exploring recent diachrony: Corpus studies of lexicogrammar and language practices in Late Modern* English (Studies in Variation, Contacts and Change in English 18). Helsinki: Varieng. http://www.helsinki.fi/varieng/series/volumes/18/tyrkko_nurmi (accessed 8 June 2017).

Tyrkkö, Jukka, Arja Nurmi & Jukka Tuominen. 2017. Semi-automatic discovery of multilingual elements in English historical corpora: Methods and challenges. In Arja Nurmi, Tanja Rütten & Päivi Pahta (eds.), *Challenging the myth of monolingual corpora*, 172–199. Leiden & Boston: Brill Rodopi.

Voigts, Linda Ehrsam. 1996. What's the word? Bilingualism in late-medieval England. *Speculum* 71(4). 813–826.

Wenzel, Siegfrid. 1994. *Macaronic sermons: Bilingualism and preaching in late-medieval England*. Ann Arbor, MI: The University of Michigan Press.

Wright, Laura. 1992. Macaronic writing in a London archive, 1380–1480. In Matti Rissanen, Ossi Ihalainen, Terttu Nevalainen & Irma Taavitsainen (eds.), *History of Englishes: New methods and interpretations in historical linguistics*, 762–770. Berlin & New York: Mouton de Gruyter.

Wright, Laura. 1995. A hypothesis on the structure of macaronic business writing. In Jacek Fisiak (ed.), *Medieval dialectology*, 309–321. Berlin & New York: Mouton de Gruyter.

Wright, Laura. 1998. Mixed-language business writing: Five hundred years of codeswitching. In Ernst Håkon Jahr (ed.), *Language change: Advances in historical sociolinguistics*, 99–118. Berlin & New York: Mouton de Gruyter.

Wright, Laura. 2000. Bills, accounts, inventories: everyday trilingual activities in the business world of later medieval England. In D. A. Trotter (ed.), *Multilingualism in later medieval Britain*, 149–156. Cambridge: D. S. Brewer.

Wright, Laura. 2001. Models of language mixing: Code-switching versus semicommunication in medieval Latin and Middle English accounts. In Dieter Kastovsky & Arthur Mettinger (eds.), *Language contact in the history of English*, 363–376. Frankfurt am Main: Peter Lang.

Šime Demo

10 Mining macaronics

1 Introduction

1.1 Macaronic Latin

There is no universally accepted definition of the macaronic style. The term was first employed by Michele "Tifi" degli Odasi in the late 15th century to describe a curious mixture of Latin and a northern Italian dialect in which Latin endings were attached to vernacular stems with excessive frequency, verse was hexametrical, and themes were usually burlesque, lampooning classical epics.[1] It was sometimes also used to describe the language of all multilingual, humorous, or unconventional works (such as tautograms). More recently, with the development of medieval studies in the 20th century, medievalists started to use the term to designate linguistically mixed texts from that period, regardless of the way in which languages combined.[2] Recently, there have been attempts to comprehensively embrace both traditions.[3]

In the present paper I use the original meaning of the term, which designates the language of the literary genre that arose from the humanist culture, quickly spread throughout Europe, lasted for several centuries, and included about a dozen language pairs. This implies that the style was employed intentionally, that it is characterised by densely distributed lexical hybrids with vernacular stems and Latin endings, that the dominant language is Latin, that the content is usually humorous, and that the macaronic poetry itself is an offshoot of the humanist culture (in that, for example, it applies classical metrical form or regularly references ancient classical works). Although such works form a clearly defined tradition, admissible outliers exist in macaronic prose (as found in the introductions to macaronic poems in particular) as well as in non-macaronic works that nonetheless contain macaronic parts (such as Rabelais's *Gargantua and*

---

[1] See e.g. Genthe (1836: 63), Delepierre (1852: 14), Appleton Morgan (1872: 9), Paoli (1959: 20, 57), Dahl (1962: 14–15), Paccagnella (1979: 85), Garavini (1982: 40), Domínguez Leal (2000: 101).

[2] See Sullivan (1932), Wehrle (1933), Oakden (1934), Beatie (1967), Wenzel (1994: 4–6), Zink (1995: 116), Scahill (2003: 22), Bourgain and Hubert (2005: 218–220), Horner (2006), Classen (2010: 478–480).

[3] E.g., Boehme (2011), in which the author covers a span of 1000 years of what she considers English macaronic texts. Concerning the problem of the definition, see Demo (2014).

DOI 10.1515/9781501504945-010

Pantagruel and Moliere's *Imaginary Invalid*). Occasionally, integral vernacular words are likewise used.

One example of the Neo-Latin macaronic style, found in verses written by the greatest macaronic poet, Italian Teofilo Folengo (1491–1544), is given in (1):

(1) An poterit passare maris mea gondola scoios,
quam recomandatam non vester aiuttus habebit? (Cocaius [Folengo] [1552] 1911: v. I.7–8)

'Will my gondola be able to pass the rocks of the sea, if it will not be entrusted to your assistance?'

The Neo-Latin macaronic style has no fixed vocabulary, even within a single language pair; nonetheless it is predictable because of the unequivocal rules of its formation. In this sense, it transcends mere code-switching and instead places itself on the cusp of a completely novel idiom, i.e., an original mixed language. If observed more straightforwardly as characterised by code-switching, however, Neo-Latin macaronics formally comes close to what Muysken (2000: 122–153) terms "congruent lexicalisation".

1.2 Status quaestionis and research problems

Although Neo-Latin macaronic poetry is a case – and an eccentric one, at that – of code-switching at work, it has seldom been analysed with regard to its structural properties in general and to switching in particular. Most of the research focuses on philological and literary problems or historical context, complemented by sporadic analyses of the sociolinguistic framework.[4] Those few studies that undertake close linguistic analysis have focused on just one language pair at a time.[5] Additionally, most of the authors did not have the advantage of using computer technology in their research.

Neo-Latin, the dominant language in macaronics, is a very specific linguistic form. It has many features of an artificial (i.e., not completely natural) language, because educational and literary conventions have always demanded the use of set phrases from classical Roman antiquity. Macaronic Neo-Latin is even more artificial, because its own rules must be superimposed upon the rules of Neo-Latin and consciously followed in order to achieve the desired effect.

---

[4] The body of scholarship in the field has grown considerably; a good list of selected references is found in Sacré (2007).

[5] E.g., Paoli (1959), Chiesa (1972), Paccagnella (1979: 72–106), Paratore (1979), Brusamolino Isella (1993) for Italian; Giraud (1977) for Provençal; Keipert (1988) for Polish.

All Neo-Latin macaronic works, regardless of the Embedded Language (hence: EL),[6] have a common linguistic basis and a unique overall makeup. Because they constitute a finite corpus, it is possible to compare them, to identify their rules of composition, and to observe the flexibility of as well as the occasional exceptions to those rules. The present study focuses primarily on building a foundation for the digitally supported, integrated, and comparative linguistic study of the entire Neo-Latin macaronic tradition. Since digital technology facilitates the automatic processing of large amounts of data, its application may represent a considerable methodological improvement on existing studies in the field. In addition, it can assist in the identification of marginal, theoretically problematic, and ambiguous macaronic phenomena.

A two-stage approach is taken here. Firstly, a sample corpus is built and prepared for analysis, i.e., annotated, exploring along the way the extent to which computers can automate the relevant tasks. In the present study the annotation is made on the level of the word, which is but one possible option.

Secondly, two sample analyses on the corpus are carried out. This involved an observation of how various word types (Latin, hybrid, EL) cluster, by means of calculating the sizes of the continuous monolingual text segments. Special attention is given to possible similarities between texts with closely related ELs. Similarly, a test is run to see whether verse beginnings play an important role in achieving the macaronic effect.

1.3 The corpus and the methodology

There are at least 130 Neo-Latin macaronic authors that appear in widely available literature.[7] The present corpus consists of extracts from 60 macaronic poems. It includes 11 different language pairs, with Latin alone represented throughout. The ELs are all Indo-European, equally representing three major European linguistic subfamilies: Germanic (Dutch: 1[8], English: 7, German: 7), Romance (French: 5, Italian: 13, Portuguese: 5, Provençal/Occitan: 5, Spanish: 5), and

---

[6] Latin gives most of the grammatical morphemes and dominates quantitatively; therefore, it can be conveniently termed the Frame Language (hence: FL) and its partner language in macaronics, the EL, although these do not behave in exactly the same way as naturally occurring FLs and ELs as described in Myers-Scotton (1993).

[7] Counting individual works may be misleading; for example, a single short poem cannot be equaled to a long epic or a series of epigrams, although each of them can be represented by one bibliographical entry.

[8] Figures represent the numbers of poems included in the corpus.

Slavic (Croatian: 1, Czech: 1, Polish: 10). Within individual language pairs, the poems were selected randomly.

Typical Latin macaronic works are scarcely recognised outside the listed linguistic combinations. In terms of authorship time span, the earliest writing is from c. 1484 and the most recent is from 1969, but the majority date from the 16th to 18th centuries.

Our corpus has 3,274 verses in total, with chunks containing an average of 54.57 verses. The total word count is 20,603, which averages out to 343.42 words per poem. The topics are various: political, religious, erotic, intellectual and so on. The narrative framework ranges from playful occasional poetry to long epic-like parodies. The authors are also very diverse – we encounter nationally important poets such as Teofilo Folengo, Jan Kochanowski, Remy Belleau, and Tomás de Iriarte, but also less known names such as Bartolomeo Bolla, and João da Silva Rebello, or even people who made a name in other scholarly fields such as historian and philologist Antonio Muratori, theologian Alexander Geddses, or biologist Edward Forbes, as well as several anonymous works.

Hunston (2009: 160) lists three methodological requirements for a corpus: representativeness, balance, and size. The corpus employed herein is representative (all periods as well as most of the existing language combinations and major language families are covered); it does not contain, e.g., Basque, Finnish, Estonian, or Hungarian, however, because Neo-Latin macaronic works in those languages, if they exist, have not been widely accessible to the Neo-Latin research community. Meanwhile, as some languages (namely Croatian, Czech, and Dutch) have relatively modest macaronic traditions, their parts of the corpus are considerably smaller and the sizes of their chunks are relatively large, while the extracts from other texts are of the approximately same size. The number of poems selected from most language pairs is roughly equal, with more space given to larger traditions, especially to the Italian, which is by far the most comprehensive one. The results obtained by the statistical analysis have been normalised when necessary. Furthermore, although as mentioned there do exist some instances of macaronic prose, only macaronic poetry is included in the present analysis, because it presents the basic form of macaronic language and ensures stylistic consistency across the data. Any comparison of macaronic prose and poetry should take into consideration the differences that generally exist between the language of (Neo-)Latin poetry and its prose counterpart in terms of word choice and order.

The size of the corpus is an especially troublesome matter. The texts are not very numerous because they are mostly unavailable in digital form, and this is for several reasons. Firstly, the texts predate digital publication; secondly, Neo-Latin works are not nearly as often digitised as ancient ones or those written in major modern languages; finally, even within the Neo-Latin tradition, macaronic works are relatively marginal and have thus been only sparingly digitised.

Nevertheless, even a corpus of this size is sufficient to demonstrate the current methodology, as the conclusions of our analysis are not based on low-frequency items. As the case is with any corpus, the larger it becomes, the wider its applicability gets.

Although the corpus is comprised of texts from the past, given the very limited range of diachronic linguistic change in Neo-Latin, the corpus can be said to be historical only regarding the ELs. Thus it can be viewed as a group of historical texts that are comparable on a certain basis.[9]

If a computer is to be used for the processing of text, all characters of the text have to be encoded according to a character encoding scheme. The most basic scheme, the so-called ACSII, can encode all characters of the English (and Latin) alphabet, but is unable to deal with many characters that appear in, for example, French (*ç*), German (*ä*, *ü*, *ö*), Polish (*ł*, *ś*), or Croatian (*č*, *ć*, *š*). A more recent scheme, UTF-8, which can handle more than a million characters, is used in the present corpus. If, e.g., Russian, Greek, or Georgian are to be added in the future, UTF-8 can handle their respective scripts, too.

For text tagging XML encoding is used, as it is a widely accepted standard.[10] Shorthand titles were added in front of each poem. The lines containing the titles were tagged with `<title>Some title</title>`, which serves as a unique delimiter of corpus poems.

Annotated in this way, the corpus becomes a quantitative and qualitative data mine, and as such is able to reveal some high-level properties of macaronic language. The programming language used in text encoding, Perl, also proves convenient in analysis.[11] It supports very complex pattern searching (via the so-called regular expressions), which is used in retrieving data.

---

[9] See Aijmer (2009: 276).

[10] XML (Extended Markup Language) is a standard markup language that enables both humans and computers read a tagged text. All metadata are placed in a tag, called element, that encloses the part of the text it refers to. Elements can contain further details, which are stored in the so-called attributes. For example, a description of the Italian word *verso* can be encoded like this: `<word lang="ita", meaning="verse", pos="noun", gender="masc", number="sing">verso</word>`, where the element `<word>`, giving information about the word *verso*, contains attribute `lang` with the value `"ita"`, attribute meaning with the value `"verse"`, and so on. The main advantages of the XML are its flexibility and platform-independence.

[11] See Hammond (2003) and Bilisoly (2008), books specialising in linguistic research with Perl, which offer some good ideas for the analysis. I would like to thank my colleague Jan Šipoš for having drawn my attention to the usefulness of Perl in linguistic analysis and to the existence of the Morpheus lemmatiser (see below).

2 Building the corpus

2.1 Binary tagging

As already mentioned, the word is the operative level of analysis in the present work. The TEI-recommended[12] <w></w> element for each word is used, to which two attributes are then assigned: lang for the languages involved, and type for the classes of the word tailored to the corpus. Words are identified with the help of regular expressions, and after some minor manual cleanup, every single word in the corpus is tagged with the element of the following type:

```
<w lang="xx" type="xx">some_word</w>
```

where `"xx"` is added as a placeholder for the attribute values, in order to anticipate the final structure of the element.

The rest of the corpus preparation consists mostly of filling in and adjusting the attribute values within the word elements (see Figure 1 for the overview of the process). The first step here is distinguishing between the two major word types that appear in macaronic poetry: Latin words and non-Latin (macaronic) words. Of course, we want, whenever possible, a computer program to do the tedious work of telling them apart. This is a matter of a straightforward comparison between a list of Latin words and the words in the corpus texts. As Latin is a highly inflected language, and it is desirable to account for every possible word, a list of not only lemmas but of all word forms is needed. In order to obtain this one can use a lemmatiser, which is a program that checks an input against a database of word forms and, if the word from the text is found, returns its lemma and possibly other information.[13] Lemmatisation per se is unnecessary here, as the objective is simply identifying words as Latin, but since lemmatisers are connected with databases comprising Latin word forms, they provide the necessary information.

---

[12] TEI (The Text Encoding Initiative), as its web site states, “is a consortium which collectively develops and maintains a standard for the representation of texts in digital form” (see *TEI: Text Encoding Initiative* 2013). Its standards are described in TEI Guidelines and use XML as the markup language.

[13] There are several lemmatisers for Latin available on the Internet, e.g., Collatinus (Ouvrard and Verkerk 2014), or Morpheus, made by the the Perseus Project (Crane 2015); the latter is used in the present study.

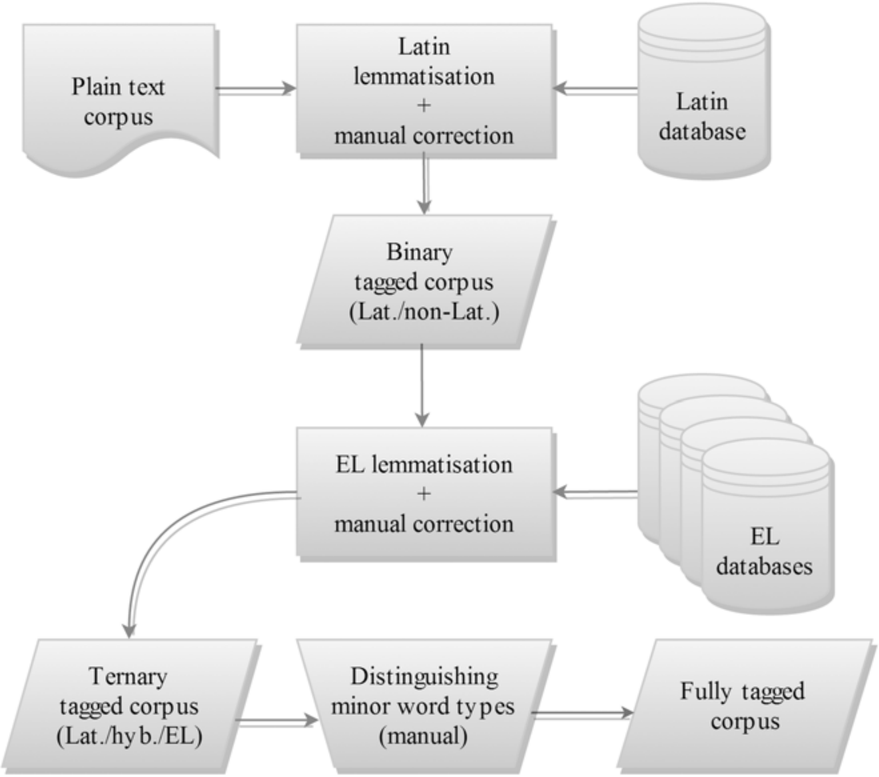

Plain text corpus
Latin lemmatisation + manual correction
Latin database
Binary tagged corpus (Lat./non-Lat.)
EL lemmatisation + manual correction
EL databases
Ternary tagged corpus (Lat./hyb./EL)
Distinguishing minor word types (manual)
Fully tagged corpus

Figure 1: A flowchart diagram showing the procedure of the corpus tagging

Another method would be building from scratch a comprehensive form database, which can be generated from a list of dictionary headwords with unambiguous information about the class of their inflection. Such lists are available on the Internet, and it is possible to produce the remaining forms. If the database and the text to be processed are stored on the same computer, transmission of data across the Internet is unnecessary, and the program executes more quickly.

Of course, the output in such a case would depend on the varieties of Latin covered with the form database. As will be demonstrated further down, ready-made databases of Latin often contain forms that are superfluous to a given search or, conversely, lack forms that are needed. One of the advantages of using a self-built database is that it is possible to predetermine which words and forms are included in a given search. As this process could be extremely time-consuming, an alternative would be to generate the database from an appropriate collection of existing Latin texts.

The lemmatiser gives each word a binary value: it is either Latin or non-Latin. If the word is [+Latin], `"la"` replaces the dummy `lang` attribute value;[14] if it is [-Latin], `″xx″` is left unchanged as an EL code placeholder (more on this below). Furthermore, since the program detects pure Latin words, the attribute value `"p"` (< lat. *purum*) is set as their `type` attribute value. Thus, the Latin word *ergo* 'therefore' now looks like this:

`<w lang="la" type="p">ergo</w>`,

and the hybrid word *dansas* 'dances' (*n.*), like this:

`<w lang="xx" type="xx">dansas</w>`.

As XML code is relatively hard to read, the texts can be transformed into HTML and, thanks to the newly added tags, automatically formatted according to the word types. In the verses given in (2) non-Latin words are boldfaced:

(2) Qui volet ergo bassas apprendere **dansas**
Et cito **dansandi mestrus** in arte fore,
Hunc bene de testa se **forcet** discere librum,
Qui bene **dansandi** monstrat habere modum. (Arena [1529] 1758: v. 1–4)

'Therefore, he who wants to learn low dances and to quickly become master in the art of dancing, let him make himself learn by heart this book, which shows the method of good dancing.'

Because the markup is converted into text formatting, in the above verses anyone familiar with Latin can immediately recognise what is or is not Latin. Such formatting will prove useful further down, as tagging gets increasingly precise and as repeated manual checking of the corpus becomes necessary.

Researchers may wish to distinguish between macaronic words in different languages, as a prospective analysis could hinge on whether the first part of the hybrid is from, for example, Polish or Portuguese. This task can also be automated to a certain degree. Here this is done by relating each poem title to a language[15] (in Perl, using its numerical index in the series of texts) and then generating the two letters of the lang attribute value, each one representing a language.

Up to this point the process has been almost completely automated. As for the size of the corpus, the only restraining factor is the computer memory. However, much work remains to be done if the objective is tagging various types of words in the macaronic realm. This is the topic of the following subsections.

---

[14] Attribute values for individual languages correspond to the codes given in the ISO-639 language names coding standard.

[15] The cases of using more than one non-Latin language in a single poem, being extremely rare, are dealt with manually.

2.2 Correction of the binary output

The corpus – now tagged in a binary fashion – has to be manually corrected. First of all, some Latin words are not recognised as such by the program; these must be moved to the Latin section manually. Their number and kind depend on the quality of the Latin database at hand and, perhaps contentiously in the present case, on our own decision of what counts as a Latin word. Some common problems in this regard include non-standard spellings (e.g., *sydera*, *lachrymas*, *michi*, *terre*, *foemineus*…) and alternate forms (e.g., *miraclum*, *dominabus*) which, despite being considered unequivocally Latin by contemporary readers, are often not present in Classical Latin form lists because they only became common in post-classical periods.[16] Secondly, the database used by the Morpheus lemmatiser does not include proper names. If one were to generate a form list from an existing word list of Classical Latin that included such names (for example, from a digitisation of Lewis and Short 1891), unless Neo-Latin texts are included as sources for the initial word list – which was not performed in the course of the present analysis, but remains as an option for the future – some non-classical examples which are not recognisable as macaronic, but which are nonetheless familiar to Neo-Latin readers, would not be identified by the program (e.g., *Carolus*, *Franciscus*, *Avicenna*, or *Polonia*). The same holds for non-classical common nouns that appear so frequently in Neo-Latin that they even find their way into high-style poetry (e.g., *taba(c)cum* 'tobacco', *paradisus* 'paradise', *Manichaeus* 'Manichean'). In a subsequent manual check, these words were tagged as Latin. Less-common names or *ad hoc* macaronic creations (e.g., *Fransa*, *Espagnolus*, *Wrightus*), meanwhile, were left among non-Latin words.[17]

The examples listed above show that Neo-Latin, like every language that has communicated with other linguistic systems, has admitted words from other languages and simply attached Latin endings to them. This process of borrowing is analogous to macaronic vocabulary building, and in fact takes place with any adoption of a foreign word into an inflectional system of a language; however, in macaronic poetry it is uniquely abrupt and excessive, and readers usually encounter newly-forged words that do not appear outside the macaronic tradition. Such words cannot be considered normal Neo-Latin neologisms, not least because they lack "predictability", which has been recognised as a reliable criterion for distinguishing code-switching from borrowing.[18]

---

[16] I also encountered some unexpected cases such as the absence of *forcipe* 'pincers' (ablative singular), which is certainly a regular Classical Latin word.

[17] In deciding the (not unproblematic) status of names in Neo-Latin, I followed handbooks such as Grässe [1861] (1972) and Hofmann (1698).

[18] See Myers-Scotton (1992).

There is also some cleanup required on the Latin "side" of the corpus, because the program marks as Latin some macaronic formations that are homonymous with esoteric words from the Classical Latin period. One example is *batone*, a word from a French-Latin macaronic, made up of the French *bâton* 'a rod', and an imitation of the Latin ablative ending *-e*, which fits grammatically and semantically into the sentence structure. However, the program recognises this word as coming from the extremely rare Latin word *batus* 'the blackberry-bush' or 'a Hebrew measure for liquids,' in the ablative case (*bato*), joined with the enclitic *-ne*. It is beyond doubt that the Latin *batus* has never come to macaronic readers' minds when encountering *batone*. This is why the existence of such words in Classical Latin is discounted from lists.

Given the currently available databases of Latin word forms, both of the above-noted cleanups must be performed manually, i.e., by reading line-by-line and keying in the changes in the attribute values.

2.3 Ternary tagging

In the macaronic tradition, not all non-Latin words are hybrids. Some of them belong completely to the EL. Therefore it is useful to distinguish between hybrid words and pure non-Latin words. Certainly, there is no database of hybrid macaronic words, but for languages that have reliable form lists (which ideally reflect the state of a language at the time a given poem was written), researchers can find pure EL words by way of yet another automated process. On the Internet there are word lists for (the modern variants of) many languages and, in fact, most of the languages appearing in the present corpus are covered.[19] These lists come in handy at this stage. Here a self-written program is used that first checks for the existence of each word in the applicable word list at appropriate places in the corpus, and then sets the `type` attribute to the value "`p`" if the word is found or "`h`" if it is not. Of course, the fact that the available word lists are not adjusted to the corresponding historical periods and dialects makes them considerably less appropriate for the present purpose, but the subsequent manual check has shown that they do the job pretty well. This might be partly due to the fact that the pure EL words make up only 6% of the total word count in the corpus.

The program used in this stage is slightly more complicated (and slower) than the previous one, because it must perform multiple checks across eleven bulky form databases. After another manual cleanup, the corpus reveals an additional set of word groups: pure words in languages other than Latin. Of course, the new situation that this creates can also be visually formatted to represent the three distinct word types.

---

[19] I used lists available at *WinEdt Dictionaries* (2014).

2.4 Minor word types

After the above described steps the words of the corpus are divided into three groups: Latin, hybrid, and pure EL words, the latter two having been connected to their respective languages. However, the picture is more complicated. All of these groups must be classified into various sub-groups in order for the analysis to be able to provide precise results. This process is very language-specific and, consequently, the most engaging one. Some parts of it can be automated by programming. Others must rely on manual checking because they depend on complex cultural factors that cannot always be precisely predicted or require a rich linguistic documentation – which is still insufficiently available for the historical texts – as to be accomplished computationally. Although the hybrids group contains more entries than any other except for the pure Latin word group, it will be demonstrated below that there are some additional distinctions of word types that can be identified in macaronic language. This is why the `type` attribute is necessary.

At the beginning of an English macaronic the poet states: "*Pandere mens est* | *Skippantes hinc inde fleas*" ('I have the intention to sing about the fleas that hop to and fro', Anonymous [1788] 1852: v. 2–3). In this example, the lemmatiser recognised *fleas* as a Latin word (2nd person singular present subjunctive of *fleo* 'cry'), although it is clear from the context that it is either English or hybrid.[20] An additional example is *patri* in a French macaronic, which is not the dative form of the Latin *pater* 'father', but rather an onomatopoeia representing broken glass.

Such words belong to a group that can be termed *cryptomacaronic words*, which are recognised as Latin by the program, either because of translingual homonymy or due to a glitch in the lemmatiser, but which clearly do not fit that classification when read in context. These words constitute a special class and must be tagged by hand in the corpus (which is made annotator-friendly through graphical formatting). With a sophisticated and syntactic algorithm for Latin, adjusted to the macaronic situation, this stage may in the future be at least partly automated.

---

[20] The structure of the verse suggests that it is hybrid; however, this is not relevant in the present discussion.

Some cryptomacaronic words are pure EL words (e.g., the English 1st person pronoun *I*, which the program used here recognises as the Latin imperative of *ire* ‘go’); others are hybrid (e.g., the Provençal *balando* ‘dancing’, rather than that word’s meaning in Latin, ‘bleating’). Here these words’ values are set to “`pc`” (lat. *pura cryptomacaronica*) and “`hc`” (lat. *hybrida cryptomacaronica*), respectively.

The precise category of cryptomacaronic words is the direct consequence of the corpus processing method in the course of marking it up. If the ELs were subjected to automatic recognition first, lexical overlaps with Latin would be conviniently termed “crypto-Latin”. However, an analysis of a sample has shown that moving the other way round (namely, starting from the ELs), or parallel tagging of Latin and ELs, would require much more manual cleanup. Therefore, starting from Latin, as the most frequent language in the macaronics, is the most economical strategy. In the analyses, cryptomacaronic words will be mostly treated as EL or hybrid words, because they used to be immediately recognised as such by the readers, but a special tag makes the future research of homonymy and its poetic effects possible.

In addition, macaronic authors sometimes played with Latin words without including another language. They mixed stems and endings that do not belong together, but the resulting words should not be considered interlingual (e.g., the future *dicebo* instead of *dicam* ‘I shall say’, or the comparative *vulpior* ‘more fox-like’, derived from the noun *vulpes* ‘fox’). Such forms, going unrecognised by the program but still macaronic, are here given the type attribute value “`h`” while their lang value remains “`la`”.

Hybrid and EL words, detected as such by the computer program or by a human annotator, are certainly distinguishable from basic Latin text, and their salience, producing an effect of “linguistic shock”, is what makes macaronic language special. However, there are word forms that can in a given context belong to more than one of the three types.[21] In order to give more fine-grained distinctions and to enable a research of the role lexical ambiguity plays in the macaronic tradition, it is convenient to include a special set of tags for such words. The fuzziness of interlingual boundaries may have various degrees of representation and importance, depending on the language combined with Latin. For example, Romance languages can be expected to have much more overlap with Latin than Germanic or Slavic languages.

---

[21] The phenomenon of interlingual lexical overlap was used in early modern Spain as an ideological tool; texts composed entirely of words that can be either Latin or Spanish were written. See Woolard and Genovese (2007), who use the term *strategic bivalency*, but also distinguish such practices from the macaronic style (2007: 500), where ambiguity is a byproduct, not a strategy. I learned about their work from a remark by an anonymous reviewer of the present article. Wright (2011: 230) calls such words, as appearing in written texts, *visual diamorphs*.

In the present work such words are distributed into the following categories:

- EL/Hybrid: words that can be EL or hybrid, but not Latin (e.g., the Spanish *segunda*, 'second,' which fits into Latin sentence structure). The `type` attribute value is "`dh`". These words can be cryptomacaronic as well, i.e., they can have some other meaning in Latin (e.g., the Italian *testa* 'head,' which is also Latin for 'brick', 'shell' and in later usage 'skull'). The value for these is "`dc`".
- Latin/EL/Hybrid: words that can fall into any of the three groups (e.g., the Spanish *fortuna* 'luck' or 'destiny', can have the same meaning in the Latin nominative). The value is "`d3`".
- Latin/EL: words that normally do not receive any ending and consequently cannot be macaronically hybridised (e.g., the French/Latin preposition *de*). These are generally function words that are found in fixed phrases, and very often they are also stop words, which can sometimes be conveniently omitted from searches. The value is "`dp`".
- Latin/Hybrid: this group consists of words like *celebrare* (in Latin, 'celebrate', or a hybrid from the Portuguese *celebrar* 'celebrate'). Although this group theoretically exists, it does not seem plausible to set it apart from ordinary Latin. Such words are not salient enough to break the in-sentence Latin flow, and if they are included as a separate category, every Latin root that has been carried over into another language (Romance or otherwise) would appear in this group. In contrast to the previous two categories, these words are not liable to be perceived as non-Latin. Distinguishing such words from Latin words would only show the number of EL stems derived from Latin, without revealing much about the macaronic nature of the texts.

In the present work, the tagging of cryptomacaronic and ambiguous words was a manual undertaking.

2.5 Word types scheme

As a result of the tasks described above, each word in the corpus is now tagged with one of the 12 different `lang` attribute values and then again with one of the 8 complex `type` attribute values (see Table 1 and Table 2).

Table 1: `lang` attribute values

Language	Value	Language	Value	Language	Value	Language	Value
Latin	`la`	Portuguese	`pt`	Dutch	`nl`	Polish	`pl`
Italian	`it`	Occitan	`oc`	English	`en`	Czech	`cs`
Spanish	`es`	French	`fr`	German	`ge`	Croatian	`hr`

Table 2: `type` attribute values

type		**value**	
		overt	**cryptomacaronic (*cryptomacaronica*)**
monoglossic (*pura*)		`p`	`pc`
hybrid (*hybrida*)		`h`	`hc`
ambiguous (*dubia*)	EL/hybrid	`dh`	`dc`
	Latin/EL	`dp`	
	Latin/EL/hybrid	`d3`	

This returns 90 potential attribute combinations (Latin with only “`p`” and “`h`”, and each of the 11 ELs with each of the 8 type attribute values), together representing a powerful categorising tool for future analyses. Visually formatted after the final annotation stage, the text appears as follows in (3):[22]

(3) Qui$_{\mathrm{dp}}$ volet ergo **bassas**$_{\mathrm{dh}}$ apprendere **dansas**$_{\mathrm{dh}}$
Et$_{\mathrm{dp}}$ cito **dansandi**$_{\mathrm{h}}$ **mestrus**$_{\mathrm{h}}$ in arte fore,
Hunc bene de$_{\mathrm{dp}}$ **testa**$_{\mathrm{dc}}$ se$_{\mathrm{dp}}$ **forcet**$_{\mathrm{h}}$ discere librum,
Qui$_{\mathrm{dp}}$ bene **dansandi**$_{\mathrm{h}}$ monstrat habere modum. (Arena [1529] 1758: v. 1–4)[23]

3 Sample analyses

The following section serves to demonstrate the potential of the corpus, through analysis, to reveal trends in the positioning of various groups of words in macaronic poems. First, the sizes of the clusters of individual word types are analysed. Then, a hypothesis is reviewed regarding whether verse beginnings have a special role in displaying the macaronic nature of the poems.

In the cluster size calculation, the words are divided into three main types: Latin (consisting of [`la`]`p`, `dp`, and `d3` types), hybrid (encompassing `h`, `hc`, `dh`, and `dc` types), and pure EL words (comprising [non-`la`]`p` and `pc`). Minor types have been merged into the dominant types because what is relevant is the reader’s linguistic impression. As Latin is the frame language of macaronic poetry and as everything that is not Latin indicates a peculiar text type to the reader, the objective here is to identify a pattern by which macaronic lexical content interrupts Latin. The most salient distinction exists between the main Latin flow of the text and non-Latin interruptions, which are the points of "linguistic shock" that produce the macaronic effect. Within these interruptions, the readers easily distinguish the hybrids from the pure EL words. Words that are linguistically ambiguous do not contribute to the macaronic effect in the same way as the pure non-Latin lexical units because they are processed as Latin words in the course of cognitive grammatical parsing of the text. Cryptomacaronic words are categorised as the words of the language they really belong to, homonymy not being an obstacle to the understanding of the text. Finally, most of these word types are so infrequent that even their separate inclusion into the analysis (which would make it considerably complicated) would not significantly alter the results.

---

[22] Latin words are again in the regular font, while non-Latin words are boldfaced. Marks for all word types, except for pure Latin words, are added in subscript. This example contains five word types out of eight.

[23] See the translation in example (2).

In the verse beginning analysis, only interruptions of Latin text are what interests us; therefore, the division is binary – Latin words are distinguished from the non-Latin (hybrid and pure EL) lexical material.

3.1 Cluster sizes

Various types of linguistically mixed texts can represent different code-switching patterns. For example, longer stretches of the dominant language discourse can be interrupted by extremely short chunks of heteroglossic elements, or a poem may consist of alternating verses in two languages. The annotated corpus of the macaronic texts enables the calculation of the individual same-type cluster sizes and their frequency in each text. In this way, the analyst can obtain an approximate picture of the "granularity" of the mixture, giving information not available by mere word type frequency counts.

Panel A of Figure 2 displays a density plot of the longest clusters in each poem as an indication of the maximum length of a single-language discourse considered acceptable for a macaronic poem. Three curves are generated, representing three linguistic groups of the ELs. The height (i.e., y-axis value) of a particular point of the curve increases as the probability increases that the cluster length at the x-axis under the point is the longest one in a text. Interestingly, in all texts, the longest clusters always consist of Latin words, underscoring the status of Latin as the FL. The figure shows that clusters of over 25 words are very rare, although they appear in all three groups. Apart from that, Germanic macaronic texts seem to be the least prone to larger clusters (dropping to very low densities after the size of 20), while much longer clusters are allowed in the Romance group. For example, at the maximum size of 20 to 25 words, the height of the Romance curve is more than double the height of the Germanic one.

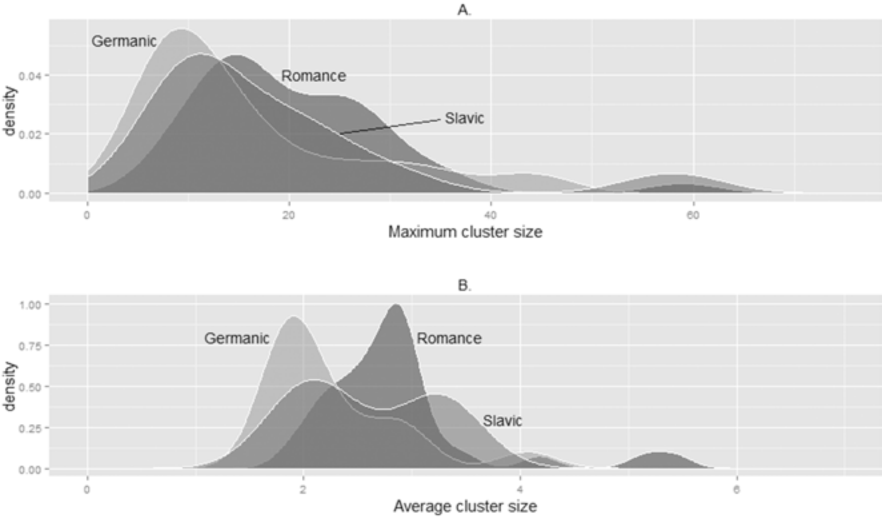


Figure 2: Density plots of maximum (A.) and average (B.) cluster sizes in each text, distributed according to the linguistic group of the EL

As the longest clusters can be isolated outliers, it is useful to look at the average cluster sizes in individual texts, which can be easily calculated with our corpus.[24] Panel B of Figure 2 shows the density plots for average cluster sizes, again distributed into the three linguistic groups. The picture appears to show some typologically interesting trends. First, it is obvious that the Slavic group has a more dispersed, bimodal distribution. This might be accounted for by the fact that some Slavic authors were educated in Italy, while others remained under the strong influence of German poets. A more detailed analysis could give more precise answers. On the other hand, for the Romance texts, the average lengths are grouped around a little less than three, while for the Germanic ones the central value is lower than two. The considerable difference in the distribution between the three groups suggests that the typological properties of the ELs can be suspected to have influence on the average size of the clusters. An analysis of a larger corpus, with divisions on the level of individual languages, periods, and texts, along with appropriate statistical tests, could reveal precise trends.

---

[24] Isolated words are also included as “clusters” of length 1.

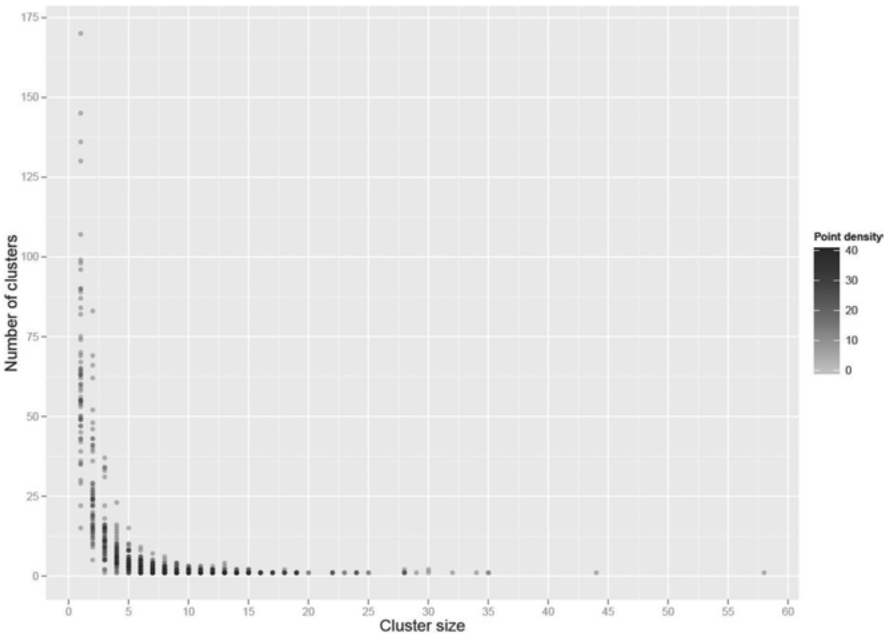


Figure 3: Frequency of clusters of each size in all texts

A considerable discrepancy between the maximal and average cluster lengths suggests that the distribution is not symmetrical. A scatterplot graph representing all cluster lengths in the entire corpus correlated with the frequency of each size (see Figure 3) shows their distribution in the most straightforward fashion. A total of 3,422 clusters is placed on a single graph, in such a way that the darker dots indicate a greater density of values. After relatively high frequencies for one-word "clusters" (i.e., isolated words) the plot exhibits an exponential decay, dropping in number of clusters to fewer than 10 very early on, and then soon down to one or zero. This means that, although Latin has priority in word distribution, the texts tend to be divided into relatively small lexical segments, ensuring frequent switching between Latin and macaronic material.

It would be interesting to compare the macaronic patterns with the rhythm of language switching in various historical bilingual texts, such as those presented in other papers of the present volume. Of course, mere numbers do not suffice. Such comparisons should take into account various levels and types of switching.

3.2 Verse beginnings

Verse is the basic rhythmical unit in poetry. Every verse opening starts a new whole that structures the poem. The beginning of the verse gives flavour to that which follows, and the question arises as to whether macaronic poets sought to “shock” readers with a non-Latin word upfront, or if they began more calmly with a Latin flow to be interrupted later. Therefore, the present analysis compares the percentage of non-Latin words in individual poems with the percentage of verses beginning with a non-Latin word. The corpus markup, as described further above and enriched with XML elements labeling verses, enable the calculation to be performed automatically.

The scatterplot graph comparing these two variables (see Figure 4) shows that the percentage of macaronic verse openings is generally positively correlated to the total percentage of macaronic words. The regression line is added to show the direction of the correlation. The correlation coefficient between the

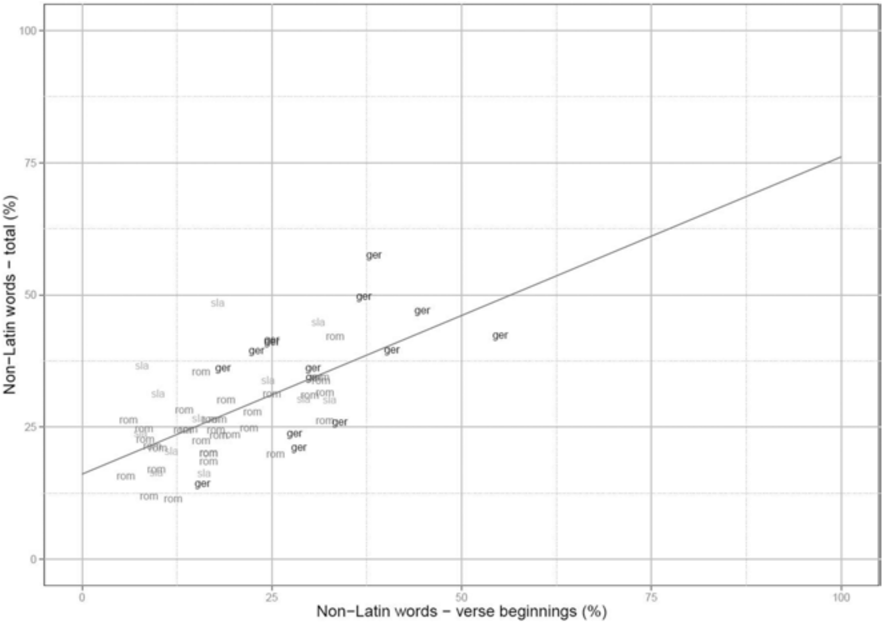


Figure 4: Percentage of non-Latin words at verse beginnings compared to their total percentage in individual works

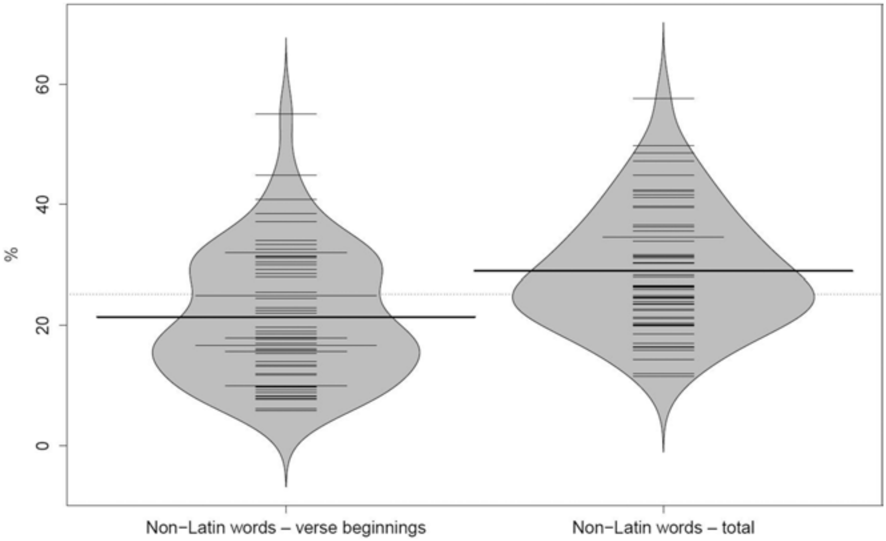


Figure 5: Beanplot graphs of the distribution of non-Latin words at verse beginnings and in entire poems. The height of the shorter horizontal bars represents the percentages in individual poems

two variables ($\rho = 0.62$)[25] points to a strong positive correlation, which means that an increase in one variable corresponds to an increase in the other. Moreover, as shown by Figure 5 – in which the horizontal width of the shape reflects the frequency of the poems grouping around that particular percentage – non-Latin words are even less frequent at the verse beginning than in the poems taken as a whole. This indicates that not only can no special role be attributed to verse beginnings in achieving the macaronic effect, but also the poets opened their verses more readily in Latin than macaronically.

To make Figure 4 richer and to stimulate thoughts of further investigation, the values are marked with the abbreviated names of the linguistic groups of the ELs (Germanic, Romance, and Slavic) rather than dots. Some overall trends emerge. Germanic texts tend to have a higher percentage of non-Latin words, both at the verse beginnings and in total. Furthermore, most of them are placed marginally with respect to the entire corpus. On the other hand, Romance poems flock relatively close to the regression line and seem to be less dispersed than the other groups.

---

[25] The Spearman correlation coefficient, which is used here for the calculation of the correlation, can take values from −1 to 1; −1 indicates a perfect negative association of ranks (as the variable x grows, the variable y decreases), and 0 means the absence of the correlation, while 1 indicates a perfect positive association of ranks. The strength of the correlation is interpreted in the following way: $0 < \rho < 0.19$: very weak, $0.2 < \rho < 0.39$: weak, $0.40 < \rho < 0.59$: moderate, $0.6 < \rho < 0.79$: strong, and $0.8 < \rho < 1$: very strong.

4 Conclusion

The present paper has had a twofold purpose: to describe the compilation of an annotated digital corpus of macaronic poems, and to demonstrate the usefulness of the corpus in the context of comparative analyses. The increased initial effort and challenging methodological choices required in the first task pay off in the analyses, which are quick and accurate regardless of the corpus size.

The computing tools and procedures described in the present paper can be modified and employed in tagging corpora of other kinds of linguistically mixed texts. Apart from that, some of the methodological challenges would probably at least partly overlap with those dealt with here. Within the research of macaronics, future development should involve the increase in the size of the corpus as well as the tagging at other levels such as morphologic, syntactic, semantic, and extralinguistic.

The analysis of the cluster sizes has pointed to at least two tendencies. First, genetic linguistic groups of the ELs seem to play a role in the relative frequency of choices between Latin, hybrid, and pure EL lexical material (Figure 2). Second, although longer clusters of monolingual discourse occasionally appear, the increase in the size of clusters corresponds to an exponential decrease in their frequency, which emphasises the high level of granularity of the macaronics in this respect (Figure 3).

The verse beginnings are not more frequently macaronic than would be expected from the overall frequency of macaronic words in individual texts; on the contrary, the two variables are highly correlated (Figure 4). Moreover, the total frequencies of the macaronic words tend to be slightly higher than their frequencies at the verse opening (Figure 5). Apart from that, Figure 4 suggests that at least the texts with Germanic ELs have different distributions than the rest.

One direction of further research would be to compare the frequencies of various word types in different languages, language groups, periods, and works. Precise statistical calculations could reveal whether the differences are related to certain external conditions or whether they may be produced by chance. If significant differences exist, the explanations for them should be researched. The same data can reveal the level of uniformity of the linguistic choices in the entire macaronic tradition. For example, if the frequencies of Latin words in macaronic works are grouped around a central value, this would mean that the authors respected the implicit rule declaring the desirable ratio of Latin versus non-Latin lexical material. A further step would be the comparison of the macaronics with other kinds of linguistically mixed works.

The quantity of multilingual texts from the past that is coming to the attention of the research community will only increase in the future, and the variety of interesting findings will multiply. Such complexity can only be adequately handled by the use of information technology, not only in quantitative research, but also in other types of comparisons. This is the reason it is sound to regard digitisation as a necessary ingredient of every major "next step" in code-switching research.

References

Aijmer, Karin. 2009. Parallel and comparable corpora. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: An international handbook*, vol. 1, 176–192. Berlin & New York: Walter de Gruyter.

Anonymous. [1788] 1852. Buggiados liber unicus. Carmen maccheronicum [The only book of the Buggias. A macaronic poem]. In Octave Delepierre (ed.), *Macaronéana, ou mèlanges de littérature macaronique*, 348–349. Brighton: Gancia.

Appleton Morgan, James. 1872. *Macaronic poetry*. New York: Hurd & Houghton and Cambridge: Riverside Press.

Arena, Antonius. [1529] 1758. Introductio ad bassas dansas [Introduction to low dances]. In *Ad suos compagnones, qui sunt de persona friantes* [. . .], 29–30. London [i.e. Paris]: [s.n.].

Beatie, Bruce A. 1967. Macaronic poetry in the Carmina Burana. *Vivarium* 5(1). 16–24.

Bilisoly, Roger. 2008. *Practical text mining with Perl*. Hoboken: John Wiley & Sons.

Boehme, Julia. 2011. *The macaronic technique in the English language in texts from the Old English, medieval and early modern periods (9th to 18th centuries): A collection and discussion*. Glasgow: University of Glasgow MRes thesis.

Bourgain, Marie-Clotilde & Pascale Hubert. 2005. *Le latin médiéval*. Turnhout: Brepols.

Brusamolino Isella, Silvia. 1993. Superficie grafica e strati linguistici nel Baldus del Folengo: un esempio. In Giorgio Bernardi Perini & Claudio Marangoni (eds.), *Teofilo Folengo nel quinto centenario della nascita (1491–1991)*, 185–203. Florence: Olschki.

Chiesa, Mario. 1972. La tradizione linguistica e letteraria cristiano-medievale nelle "macaronee". *Giornale storico della letteratura italiana* 149. 48–86.

Classen, Albrecht. 2010. The *Carmina Burana*: A mirror of Latin and vernacular literary traditions from a cultural-historical perspective: Transgression is the name of the game. *Neophilologus* 94(3). 477–497.

Cocaius, Merlinus [Teofilo Folengo]. [1552] 1911. Baldus. In Alessandro Luzio (ed.), *Le maccheronee*, vols. 1–2, (1) 47–349; (2) 3–139. Bari: Laterza.

Crane, Gregory (ed.). 2015. *Perseus digital library*. http://www.perseus.tufts.edu/hopper/ (accessed 21 January 2016).

Dahl, Jürgen. 1962. *Maccaronisches Poetikum oder Nachtwächteri veniunt cum Spiessibus atque Laternis, mit Illustrationen aus alter und neuer Zeit*. Ebenhausen bei München: Langewiesche-Brandt Verlag.

Delepierre, Octave (ed.). 1852. *Macaronéana, ou mèlanges de littérature macaronique*. Brighton: Gancia.

Demo, Šime. 2014. Towards a unified definition of macaronics. *Humanistica Lovaniensia* 63. 83–106.

Domínguez Leal, José Miguel. 2000. Concepto de poesía macarrónica. *Calamus Renascens* 1. 101–110.

Garavini, Fausta. 1982. Ecriture critique et genre macaronique. In Henri Weber, Claude Longeon & Claude Mont (eds.), *Les rapports entre les langues au XVIème siecle. Actes du colloque de Sommières, 14–17 septembre 1981*. [Special issue]. *Bulletin de l'Association d'étude sur l'humanisme, la réforme et la renaissance* 15(2). 40–46.

Genthe, Friedrich Wilhelm. 1836. *Geschichte der macaronischen Poesie und Sammlung ihrer vorzüglichsten Denkmale*, 2nd edn. Leipzig: Eduard Meissner.

Giraud, Yves. 1977. La Babel facétieuse du macaronique: A. d'Arena décerveleur de langage. Discussion. *Bulletin de l'Association d'étude sur l'humanisme, la réforme et la renaissance* 7. 60–65.

Grässe, Johann Georg Theodor. [1861] 1972. *Orbis Latinus: Lexikon lateinischer geographischer Namen des Mittelalters und der Neuzeit*. Helmut Plechl with Sophie-Charlotte Plechl (eds.). Braunschweig: Klinkhardt & Biermann.

Hammond, Michael. 2003. *Programming for linguists: Perl for language researchers*. Malden etc.: Blackwell.

Hofmann, Johann Jacob. 1698. *Lexicon universale, historiam sacram et profanam* [. . .] *explanans* [Universal lexicon, explaining sacred and worldly history]. Leiden: Jacob. Hackius, Cornel. Boutesteyn, Petr. Vander Aa & Jord. Luchtmans.

Horner, Patrick J. (ed.). 2006. *A macaronic sermon collection from late medieval England: Oxford, MS Bodley 649*. Toronto: Pontifical Institute of Mediaeval Studies.

Hunston, Susan. 2009. Collection strategies and design decisions. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: An international handbook*, vol. 1, 154–168. Berlin & New York: Walter de Gruyter.

Keipert, Helmut. 1988. Sprachprobleme der makkaronischen Dichtung in Polen. *Die Welt der Slaven* 33(2). 354–388.

Lewis, Charles & Charlton T. Short. 1891. *A Latin dictionary; Founded on the translation of Freund's Latin-German lexicon, edited by E. A. Andrews, LL.D.* New York: Harper & Brothers and Oxford: Clarendon Press.

Myers-Scotton, Carol. 1992. Comparing codeswitching and borrowing. *Journal of Multilingual and Multicultural Development* 13(1–2). 19–39.

Myers-Scotton, Carol. 1993. *Duelling languages: Grammatical structures in codeswitching*. Oxford: Oxford University Press.

Muysken, Pieter. 2000. *Bilingual speech: A typology of code-switching*. Cambridge: Cambridge University Press.

Oakden, James P. 1934. William O. Wehrle. *The macaronic hymn tradition in medieval English literature* (book review). *Modern Language Review* 29(1). 84–86.

Ouvrard, Yves & Philippe Verkerk (eds.). 2014. *Collatinus-web: lemmatiseur de textes latins*. http://collatinus.fltr.ucl.ac.be/ (accessed 21 January 2016).

Paccagnella, Ivano. 1979. *Le maccaronee padovane: tradizione e lingua*. Padua: Editrice Antenore.

Paoli, Ugo Enrico. 1959. *Il latino maccheronico*. Florence: Le Monnier.

Paratore, Ettore. 1979. Il maccheroneo folenghiano. In Mario Chiesa & Ettore Bonora (eds.), *Cultura letteraria e tradizione popolare in Teofilo Folengo*, 37–61. Milan: Feltrinelli.

Sacré, Dirk. 2007. Makkaronische Dichtung. In Hubert Cancik & Helmuth Schneider (eds.), *Der neue Pauly: Enzyklopädie der Antike, Rezeptions- und Wissenschaftsgeschichte*, vol. 15/1, 281–285. Leiden: Brill.

Scahill, John. 2003. Trilingualism in early Middle English miscellanies: Languages and literature. *The Yearbook of English Studies* 33. 18–32.

Sullivan, Carmeline. 1932. *The Latin insertions and the macaronic verse in Piers Plowman*. Washington, DC: The Catholic University of America.

TEI: Text Encoding Initiative. 2013. http://www.tei-c.org/index.xml (accessed 21 January 2016).

Wehrle, William O. 1933. *The macaronic hymn tradition in medieval English literature*. Washington, DC: The Catholic University of America.

Wenzel, Siegfrid. 1994. *Macaronic sermons: Bilingualism and preaching in late-medieval England*. Ann Arbor, MI: University of Michigan Press.

WinEdt Dictionaries. 2014. http://www.winedt.org/Dict/ (accessed 21 January 2016).

Woolard, Kathryn A. & E. Nicholas Genovese. 2007. Strategic bivalency in Latin and Spanish in early modern Spain. *Language in Society* 36(4). 487–509.

Wright, Laura. 2011. On variation in medieval mixed-language business writing. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English*, 191–218. Berlin & Boston: Mouton de Gruyter.

Zink, Michel. 1995. Plurilingualism, hermeticism, and love in medieval poetics. *Comparative Literature Studies* 32(2). 112–130.

Tom ter Horst and Nike Stam

11 Visual diamorphs: The importance of language neutrality in code-switching from medieval Ireland

1 Introduction

Even before its early conversion to Christianity, Ireland – on the edge of Europe – came into contact with Latin through trade with the Romans and through settlements in the Latin-speaking parts of Britain. When Christianity became established in Ireland in the course of the fifth century, the church introduced books and writing in Latin.[1] This means that, from an early stage in Irish history Latin and Irish existed side by side among the intellectual elite. Evidence for this theory emerges in the earliest written vernacular texts from the eighth century, which display much interaction between Irish and Latin.[2] Between the ninth and eleventh centuries, this use of a single bilingual code of unmarked code-switching seems to become increasingly acceptable in formal scholarly texts.[3]

One of the most notable characteristics of this mixing of languages is the occurrence of what Laura Wright (2011: 191–218) has called “visual diamorphs”: words that by their form can be assigned to both languages involved in a bilingual situation. In this article, we will take a closer look at, and provide a further analysis of, visual diamorphs as we find them in the two texts that constitute the corpus of our respective research projects: a group of homilies from the manuscript called *An Leabhar Breac* (*LB*; ‘The Speckled Book’), and a commentary written for a martyrology called *Félire Óengusso* (‘The Martyrology of Óengus’).[4]

Acknowledgment: We are very grateful to our anonymous reviewers, who have supplied us with many helpful comments. Our gratitude also extends toward the organisers of the Historical Code-switching: The Next Step symposium in 2014 for a very inspiring and energising conference. We are indebted to the Dutch Scientific Organisation (NWO) for their funding of the research project ‘Bilingualism in Medieval Ireland’ in which we are employed.

---

[1] For the development of literacy in Irish and Latin, see Johnston (2013), especially (2013: 15): “Within a century of Patrick’s mission [431 AD] there is evidence for serious literacy in Latin and within two centuries he himself was the hero of some of the first major productions of Hiberno-Latin hagiography” and (2013: 27–58).

[2] Cf. the bilingual glosses described by Bisagni (2013–2014), especially (2013–2014: 56–58).

[3] Cf. Bronner (2005); Müller (1999); Ó hAodha (1978).

[4] These corpora are the focal points of our respective Ph.D. projects, which were published in 2017.

DOI 10.1515/9781501504945-011

2 Corpora and methods

This paper presents the results of an analysis based on a representative sample from the two corpora. The corpora have been fully encoded in the computerised mark-up language XML. In order to analyse code-switching, computerised resources have considerable added value. They may eliminate many of the errors of manual labour and prevent biases based on preconceived notions. For computerised analysis, however, visual diamorphs present a problem. As stated above, these words could be assigned to two languages at the same time. XML-based editors, however, do not handle this overlap well, as they usually require a single identification for every element encoded. This issue is ameliorated by categorising syntactic constituents as either of the two languages, restricting visual diamorphs to the level of the word.

Problems of both a lexical and a syntactic nature are also encountered. For example, lexical overlap occurs in prepositions such as *in* and *de*, since they can be labelled either Irish or Latin. In addition, syntactic overlap occurs in the clause *is locusta quasi longa [h]asta atberar fria* 'locust, like a long lance, is said to it'. The Irish verb phrase *is … atberar fria* completely surrounds the Latin predicate *locusta quasi longa [h]asta*. In such cases, it is perhaps more profitable to denote the overall occurrence of code-switching and analyse the specifics manually. To attempt to cover such details in coding obstructs the retrieval process of both researcher and user.

The preferred method for signalling code-switches in XML is the language-tag (@lang=""). Apart from the standard tag for Latin ("la") and Irish ("ga", for *Gaelic*) we added the custom tag "ga-la" for visual diamorphs. Although this is not standard TEI encoding, it is a convenient way to search for ambiguous elements in our query. As stated, we have opted to assign this tag to the highest part of speech where the relevant switch occurred. Deciding whether a particular ambiguous preposition is Latin or Irish undermines the structural presence of a mixed code. For instance, whether the phrase *in celum* 'in heaven, the heaven', contains the Latin preposition *in* or the Irish article *in* is a matter for humans, not computers.[5]

---

[5] Specific categories of code-switching, as per the gradual division advanced in this paper, can be indicated by @type tags. Additionally, we found it fruitful to tag a host of metadata including proper names of persons or places or even foreign and formulaic expressions. In terms of abbreviations, other problems arise. In keeping with the distinction by Driscoll (2006: 259) between lexical and graphemic abbreviations, we think it useful to differentiate between the surface form in the manuscript and the editorial expansion. For this purpose the element @choice is deemed a suitable choice; the string *imper-*, for instance, could be rendered *impeir* in editorial mode and i~p- in diplomatic mode. In this way both original data and editorial information are available to the user.

The two corpora used for this research differ both in content and form. *An Leabhar Breac* is a fifteenth-century compilation of mainly religious texts, most notably bilingual sermons, almost exclusively written by a single scribe (Mulchrone et al. 1926–1970, 27: 1230).[6] These sermons are not contemporaneous with the manuscript and have dates of origin ranging from the eleventh century to the fourteenth century, relying on religious traditions from the ninth century (Atkinson 1887; Mac Eoin 1996: 195; Mac Donncha 1976: 66–67).[7] The martyrology called *Félire Óengusso* has been dated to the turn of the eighth century, and is accompanied by a bilingual commentary which has been dated to the twelfth century.[8] The martyrology has come down to us in nine manuscripts, the oldest of which dates to the fourteenth century and the youngest to the seventeenth century (Ó Riain 2006: 174–175).

Because of the complex transmission of both texts, they both pose problems when considering the intentions behind their bilingual composition (O'Sullivan 2005: 511). The homilies might be rewritings resulting from patchwork, while the commentary might be the result of compilation by various copyists, adding their own knowledge to older existing glosses. Therefore, it is not always clear whether the code-switches found in these texts are the result of the bilingualism of one individual or the outcome of a process of text compilation that took place during several decades, or even centuries. However, the survival of these texts in their bilingual shape throughout the centuries at least suggests that they continued to be read bilingually.

Similar bilingual intellectual environments have been described for medieval England by Herbert Schendl. Although the sociolinguistic situation of England changed considerably during the Middle Ages, the later period can be characterised by a sanctioning of the vernacular in interchange with Latin. This bilingual equilibrium, strongly similar in England and Ireland at the time of our corpora, produced what Siegfried Wenzel has called "a mixed language for bilingual audiences" (Wenzel 1994: 127; Schendl 2011: 47–94). This in-group use is consistent with the sociolinguistic status of modern code-switching. Although the recognition of the bilingual character and backdrop of our texts enables us to analyse them as mixed code, we will have to remain wary of switches that are the result of the complex transmission of the texts or of compilation rather than the result of the individual bilingualism of a scribe.[9]

---

[6] Described by O'Rahilly et al. (1926–1970, 27: 1230); partially edited by Atkinson (1887). In research into Old English homiletic texts, bilingual sermons are often referred to with the term *macaronic*. However, this word originally refers to "a burlesque form of verse in which vernacular words are introduced into the context of another language (originally and chiefly Latin), often with corresponding inflections and constructions; gen. of or designating any form of verse in which two or more languages are mingled together." (macaronic, adj. and n., *OED Online*). Therefore, this study refers to mixed-language text as *bilingual*.

[7] On the dating of Irish manuscripts and the language in them see for example Koch (2006: 985, 993–994).

[8] For a discussion on the date of the calendar, see Dumville (2002: 19–34), Ó Riain (1990: 21–38) and Thurneysen (1908: 6–8). For editions of the text, see Stokes (1880, 1905).

3 Visual diamorphs

In a mixed code, an important role is laid out for elements functioning in both languages: the visual diamorph or, for spoken language, the homophonous diamorph. The concept of homophonous diamorphs has been described by Pieter Muysken (2000: 133), who stated that when languages are similar or are perceived to be similar by bilingual speakers, switching is facilitated by specific words that sound alike in both languages.[10] Muysken refers to Michael Clyne (1967), who has suggested that "the distinction between two codes may be neutralised at the point where they share a pair of homophonous diamorphs" (Muysken 2000: 133). An example of this would be the form of the copula *is* in English and Dutch; if *is* were preceded by English and followed by Dutch, it can itself be assigned to either language. Even in the mind of the speaker these words might not be assigned to a particular language. Instead, these words serve as neutral elements that leave the speakers a choice in which language to continue their sentence. By providing this neutral zone, or language vacuum, diamorphs can facilitate or even trigger a code-switch. Evidence from linguistics thus suggests that homophonous diamorphs play an important role in bilingual communication. Our question, then, is whether the role of visual diamorphs in a written, historical corpus is just as important. We would like to know how often code-switches are triggered by these language-neutral elements, and whether there is a way to classify and categorise different diamorphic switches. In the end, we hope that this might lead to a more refined and layered analysis of historical code-switching.

According to Laura Wright's research into Middle English business writing, there is a fivefold division in potential diamorphs (Wright 2011: 194–195):

---

[9] See especially Bisagni (2013–2014: 51). Intersentential switches might be less likely to have been written by a single scribe. However, we cannot entirely neglect them, as they are still the products of a bilingual environment and still constitute a switch in written language.

[10] Muysken considers homophonous diamorphs inextricably linked to his third type of code-switching: "congruent lexicalisation" (Muysken 2000: 122–153).

- **function words**, for example prepositions such as *de* and *in* or conjunctions such as *et*;
- **bound morphemes**, inflected endings such as noun plural marker *-is*;
- **borrowings** from another language, such as *spitell* ‘guest-room’;
- **bare roots** of words with their endings abbreviated in the manuscript such as *turtull-*; and
- **bound morphemes** that are again suppressed by abbreviation such as *cariandꝭ*.

These last two categories, “bare roots with abbreviated suffixes” and “bound morphemes that are suppressed”, could be clustered together due to the fact that their ambiguity derives from abbreviation. The category of shared bound morphemes is not relevant for our sources; Irish and Latin share few bound morphemes, and no examples in which they acted as diamorphs were encountered during the research for this paper. While this particular category does not seem to be applicable to our corpora, they do present us with an additional class of potential diamorphs. We would like to integrate this additional class into the fivefold division presented by Laura Wright (2011) in a slightly rearranged order. By rearranging the order of the potential diamorphs, we hope to arrive at a gradual scale running from diamorphs that function in both languages at a lexical level, via function words, through abbreviated words, to diamorphs that function in both languages on a pictorial level as symbols (Driscoll 2006, 2009),[11] which we will call emblems. We will also discuss an alternative strategy to connect two languages, and contribute examples for each category from our corpora as we climb along the scales.

The cline from lexical loans to integrated items shares some similarities with the linguistic process of grammaticalisation (Hopper and Traugott 1993; Traugott and Dasher 2002). According to this theory a lexeme loses its referential meaning and over time acquires a grammatical meaning. An example is the English word *while*, which was originally a (lexical) noun but came to be regarded as a (grammatical) conjunction (Brown and Miller 2014). One difference with our conceptualisation of diamorphs is that these of course do not generally move between categories. Unlike grammaticalisation, the scale of diamorphs does not primarily indicate a diachronic change. Instead, the diachronic dimension of diamorphs might lie in the degree to which the two languages involved can be integrated. The preliminary stages of code-switching are likely to have conveyed stricter distinctions between the two codes, illustrated by the use of borrowings and the sharing of abbreviations. At the stage in which two languages are more intimately intertwined, such as Latin and Irish in the later Middle Ages, switching in function words and emblems can be expected to be more acceptable. In this respect it may be hypothesised that there are indeed structural similarities between the occurrence of diamorphs and the diachronic process of grammaticalisation. Nonetheless, the directionality of grammaticalisation would suggest a different ordering of categories in which abbreviations form an earlier step and function words a later process. In the case of diamorphs, by contrast, the categories are grouped according to their visual representation. This procedure results in an order running from words written in full to words rendered in highly abstract form, as Figure 1 displays.

---

[11] See Nurmi et al., in this volume.

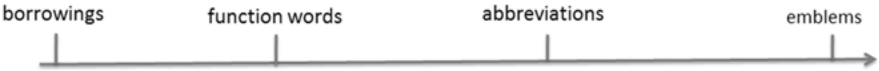


Figure 1: Diamorphs on scale

4 Irish-Latin diamorphs

4.1 Diamorphs in historical corpora

The fact that diamorphs occupy an important place in code-switching is confirmed by the number of occurrences in our corpora. Out of 271 switchpoints analysed so far we counted 202 instances where the switches are connected to diamorphs. In other words, three-quarters, or 75%, of the switches in our corpora occur in conjunction with a diamorph. It has to be noted that this figure refers to switches below the sentential level. Intersentential code-switching is expected to contain few diamorphs as these are defined at the level of the word. Furthermore, the boundaries between sentences are the least informative of a grammatical phenomenon like code-switching and the most susceptible of an unconnected addition to a previous utterance (Bisagni 2013–2014: 25). Therefore, at this point, it is important to split these results up according to two categories: diamorphs that were most likely written down by the same author and diamorphs that can signal a break in authorship. Especially for the *Félire Óengusso* – which mainly consists of short glosses – some diamorphs might not be markers of intrasentential switching. In fact, they may not even have a syntactic relationship to the text preceding them. In a way, these medieval texts functioned like open source does nowadays; they were scholarship in progress, prone to additions by their audience. This is especially true for a commentary like that on the *Félire Óengusso*. It would be easy for a scribe to add onto existing glosses by using the diamorphic abbreviations *.i.* 'that is to say' and *ł* 'or'. These brevigraphs basically function as tools for this specific purpose: to add extra information to a sentence or an already existing gloss. This means that these diamorphs are not only often intersentential but also possibly additional. However, they could also still be the product of one author who was keen on synonyms or on providing more than one explanation in his commentary, as often happens. The only thing we can say with certainty is that with such diamorphs, the chance that they were added at a different stage is greater than with diamorphic function words or object nouns. An example of such a potentially composite gloss is the gloss on the Irish word *caín* 'fair, beautiful':

(1) *cai*n *.i.* *alai*nd *.i.* *lucia* [leg, *lucida*]
'Fair, that is to say beautiful that is to say *radiant*' (Jan. 3, Oxford, Bodleian Rawlinson B505 (R1), f. 211v)[12]

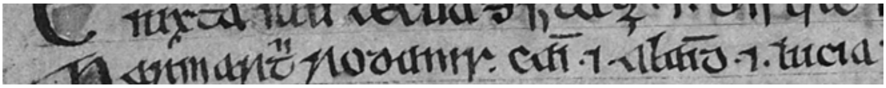


Figure 2: R1 f.211v containing the gloss from example 1. By permission of the Bodleian Library, Oxford

Each of these terms could have been added by a different scribe (Bisagni 2013–2014: 26). However, each of them could also have been gathered by a single scribe trying to find synonyms. As explained above, while we cannot tell definitively whether a particular code-switch is the result of one or more scribes, ultimately, it functions as a code-switch to the reader of the text, and therefore is a code-switch in use, if not in production. Even if a diamorph such as *.i.* introduces a switch that has been added by a different scribe, it still demonstrates that the diamorph facilitated a language switch to, for example, Latin while the copyist was reading Irish.

Returning to the results of our study, the first two months of the martyrology contain 53 switches, 41 of which are diamorphic. Of these 41, 15 involve the break diamorphs *.i.* and *ł*. Break diamorphs, the diamorphs that could indicate a break in authorship, thus make up 37% of the total amount of diamorphs. If 7 'and' would be included, this would be 46%. For the two homiletic quires of the *Leabhar Breac* that are under investigation there are 161 diamorphs in 218 switches. In 17 cases these are break diamorphs containing *.i.* or *ł* (11% of the diamorphs), or 57 including 7. (35% of the diamorphs). The low percentages of

---

12 In the text, abbreviations have been expanded in normal font. Editorial additions are signalled by square brackets. In the translation, underscore means diamorph; *italics* signify *Latin.*

break diamorphs in the *Leabhar Breac* as compared to the *Félire* are the result of the ample glossing in the latter.

Table 1: Diamorphs in code-switching

	Number of switches	Percentage of total	Percentage of diamorphs
Total number of switches	271	100%	
Diamorphs	202	75%	100%
Break diamorphs	32	12%	16%
Break diamorphs incl ⁊	76	28%	38%

4.2 Names and borrowings

As in the corpus analysed by Laura Wright (2011), borrowings form a considerable part of the total count of diamorphs. To analyse borrowings as diamorphs, one has to be aware of the thin line between code-switches and borrowings or loans. A common way to differentiate between the two is to see code-switches as foreign words not yet integrated into the recipient language, and loans or borrowings as “domesticated” words showing signs of adaptation into a native system. As such they can be divided up along the notion of “listedness” as used by Muysken (2000): “the degree to which a particular element or structure is part of a memorised list which has gained acceptance within a particular speech community”.[13] In other words, borrowings are mainly reproductive (they are already part of the accepted vocabulary of speakers) whilst switches are creative (they are created by the speakers on the spot). This means that the dividing line between the processes is the degree to which foreign elements are accepted in the receiving language, and it is exactly at this point that diamorphs come into play. By using a language-neutral element, the edges of both languages are smoothed and the two become more integrated.

An example of grammatical integration is found for the phrase *Na* Iudei *vero* ‘the Jews, however’, where the originally Latin *Iudei* is modified by the Irish article *na*. Further, semantic integration may be seen with the phrases *.i. o* lex ⁊ *o chánoin* ‘that is, from secular law and from canon law’, and ⁊ *amal dosbeir* pax *don lebar* ‘and as he gives a kiss upon the book’, where the Latin meanings of *lex* ‘law’ and *pax* ‘peace’ have narrowed in Irish to ‘secular law’ and ‘kiss of peace' respectively.[14] Interestingly, the last two examples show a tendency to leave their case endings, dative and accusative respectively, unexpressed. This creates so-called bare forms that are not uncommon in code-switching.[15] We will return to these bare forms below.

---

[13] Cf. Muysken (2000: 15, 72), who differentiates between code-mixing, established code-mixing, nonce loans and established loans. His insertional code-mixing functions in much the same way as borrowing. See also the article by Herbert Schendl in this volume.

Related to the diamorphic class of borrowings is a group of words consisting of place-names and personal names. According to Durkin (2014: 11), the often neglected or dismissed proper names convey exactly the notions of identity and status associated with conscious code-switching. James Adams was one of the first, and few, to argue this for historical code-switching; he states that "names by their very nature express identity and code-switching has the power to [make] more explicit nuances implicit in a name" (Adams 2003: 145, 245). Foreign names in our corpora, too, are sometimes left uninflected, as can be seen in the following example from the commentary:

(2) *.i. orned andreas* in*achaia*
'that is, the ordaining of Andreas in Achaia' (Feb. 6 R1, f. 211v)

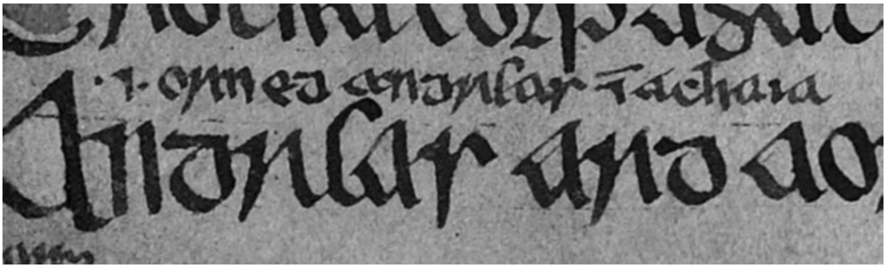


Figure 3: R1 f211v, containing the gloss from example 2. By permission of the Bodleian Library, Oxford

Here, the foreign name *Andreas* is a bare form, since it does not show a genitive marking after *orned*, an old loan of Latin *ordinatio*. Even old loans like *orned* may play a role in triggering a code-switch or at least in activating other language registers. Ad Putter (2011: 299) has found examples of this for medieval England, where a code-switch into French was often preceded by an English word of French origin. This again leads to a reconsideration of the boundary between borrowing and code-switching; even an integrated loan can still trigger language shift to its former donor language. In total, borrowings of all kinds make up 12% of the diamorphs in the commentary and 24% in the homilies.

---

[14] See the *Dictionary of the Irish Language* (eDIL s.v. "pács, (páx)" or dil.ie/34124 and eDIL s.v. "léx" or dil.ie/30088).

[15] Appel and Muysken (1987: 124–125); Muysken (2000: 30, 62–63); See also Bisagni (2013–2014: 46), including n126, n127 and n128, where he refers to, amongst others, Myers-Scotton 2009: 34; 2002: 113, 127–129, 227–228 and Gardner-Chloros 2009: 108.

4.3 Function words

In this same gloss on the ordaining of Andreas, we find an example of our next category of diamorphs: function words. In this gloss, names are not the only neutral signs: apart from the sign *.i.*, to which we will return later, we find the Latin and Irish preposition *in* as in the phrase *in Achaia*. Along with the preposition *de* ‘of, from’ these two function words are facilitators of code-switching. The preposition *de* is used in the gloss on a female saint, whose provenance is explained as follows:

(3) *.i. uirgo de sil chonaire*
‘that is, *a virgin* from the Sil Chonaire’ (Feb. 11 R1, f. 212r)

Uirgo is clearly Latin, but *de* can be used as a preposition in both languages. The proper name is Irish and might have been hard to translate, if this was desired at all. Here, the ambiguous preposition *de* succeeds in connecting the Latin *uirgo* to the Irish name of her people.[16] A less ambiguous example of a function word can be found in a gloss on the first of January:

(4) *huasal* indail .i. in *cloch de [quo] f*ac*ta* est *circumcisio xpí*sti. *q*uia *p*er*lapidem agebat*ur *prius circ*um*cisio.*
‘Noble the rock, that is, the stone with which *was done the circumcision of Christ because the circumcision was carried out with a stone before.*’ (Jan. 1 R1, f. 211r.)

In cloch ‘the stone’, is Irish but the relative sentence following it is Latin. If the preposition were to be realised as though it were Irish, it would have been difficult for an Irish relative clause to follow due to the requirements of Irish syntax.[17] However, if we follow Myers-Scotton (2002: 69–86), content words and function words are activated in the mind at the level of the mental lexicon before the syntactic structure of the sentence is made up at the level of the formulator, which would allow the preposition to remain diamorphic.

---

[16] The possibility of switching between prepositions and complements has been noted by Halmari and Regetz (2011) as well as Ingham (2011).

[17] Old and Middle Irish requires the use of a relative particle, to be combined with the preposition. The resulting shape of the preposition would then be *dia* (Thurneysen [1947] 1970: §492).

Another functional type of diamorph is a word such as *in*, which, apart from being an Irish article, as evinced by *in cloch*, ‘the stone’, above, can function as a preposition both in Latin and in Irish. It is often used before books of the Bible. While these are Latinate, the preposition, and therefore the phrase, seems ambiguous. In this event, the Irish conventions of spelling may influence the determination of language. Because Irish has strong initial stress, non-initial vowels tend to become interchangeable, and sometimes the Latin nouns show signs of this accommodation of reduced vowels (*in apocolipsi* ‘in the Book of Revelations’, *in dialago* ‘in Gregory the Great’s Dialogues’). Furthermore, the adaptation of the Irish sound system to the Latin alphabet created confusion over the phonetic value of some consonants (*in eclestiasico* ‘in Ecclesiasticus’, *in ezetsiele* ‘in Ezekiel’).[18] This increases the vernacular appearance of the phrases, even if they are idiomatic expressions strongly sanctioned by the Latin liturgy of the Church (Bisagni 2013–2014: 48–50).

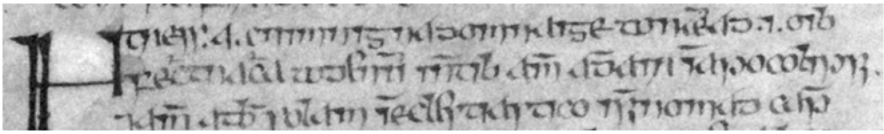


Figure 4: *LB* p. 243a: i.e. line 2 *inapocolipsi*; l[in]e 3 *ineclestiastico*. By permission of the Royal Irish Academy © RIA

By contrast, the Irish preposition *in* does not always look the same as the Latin preposition. Its usual shape is simply the letter *i*, but an *n*- will be attached to the following word if this starts with a vowel, *d* or *g*, and the two words will usually be written together. This means that only in these cases it can function as a diamorph. Conversely, in the case of *i mmalicia* ‘in Malachias’, the doubling of the consonant makes the prepositional phrase linguistically Irish, although the outcome looks a lot like Latin. We even have contrastive pairs *in libro* and *illibro exódi* ‘in the Book of Exodus’ (note the Irish diacritic), both unambiguous linguistically. However, before a vowel, *d* or *g*, Latin *in* and Irish *i* + *n*- are often indistinguishable and thus diamorphic. This overlap even holds for instances where the nasal is abbreviated by a suspension stroke, a procedure that is acceptable in both languages.

---

[18] The Ogham symbol *straif* (also *straiph*, *zraif*, *sraif*, *sraiph*, *sraib*), with phonetic value /st/, /ts/ or /sw/, was written <z> in the Latin alphabet, as witnessed in the *Leabhar Breac* by such spellings as *evangelizis* /st/ and *baptistare* /z/; for the inclusion of orthographical interference in the overarching category of code-switching, see Kopaczyk (this volume).

A complex case is *in oin uocabulo* ‘in one word’, where the Irish numeral modifying the Latin noun increases the chances for an overall Irish interpretation of the phrase. Though the preposition could well be interpretated as Latin *in*, this would leave the Irish numeral stranded between the noun it modifies and the preposition that modifies the noun. On the other hand, Irish *i n-* before a vowel would result in a simpler analysis: an Irish preposition plus numeral and a Latin noun to be considered an insertion (Muysken 2000: 60). Function words occur as diamorphs in *Félire Óengusso* 5 times (12%) and in the *Leabhar Breac* 21 times (13%).

4.4 Abbreviations

Comparable to Wright’s (2011) category of abbreviated roots are diamorphs that consist of abbreviated words. Some words are abbreviated ambiguously, and they can consequently be resolved as both Latin and Irish. Three examples are the titles *aps*. ‘apostle’, *eps*. ‘bishop’ and *imp*. ‘emperor’, as found in the following examples. Apostle, Latin *apostolus* or Irish *apstal*, appears in

(5) *Atbert* vero *eoin* aps~
‘Then John the Apostle said’ (*LB* 187a50)

Bishop, either *episcopus* or *epscop*, is seen in

(6) .i. eps~ *mel* ⁊ *melchu* <nomina eorum>
‘that is, bishop Mel and Melchu [were] their names’ (*LB* 62a33)

Emperor, *imperator* or *impeir*, is found in

(7) *Corochuindig ar foicc cesar* .i. imp~ *naroma*
‘It failed through evil of Caesar, that is, the emperor of Rome’ (*LB* 187a21)

The category consisting of abbreviations is only marginally attested in both *Félire Óengusso* (2%) and in the *Leabhar Breac* (9%). These instances, however, do illustrate the convenience of using diamorphs for abbreviations. The above brevigraphs are readily understood to stand for a specific title or designation that is basically irrespective of language. The established title can thus be retained while allowing its solution to be either Irish or Latin. Though this proved no obstacle to the medieval audience, it is slightly more cumbersome to the modern computer. As discussed, each example would ideally be assigned one language and one analysis; however, this does not correspond to the diamorphic nature of such abbreviations. Closely related to these examples are abbreviated expressions such as *ut d*ixit ‘as he said’ and *no*men *e*ius ‘his name’. Because these are so highly formulaic, rather than to integrate them in the category of abbreviated words or phrases we would like to add a new category to the scale of diamorphs.

5 Emblems

At the end of the scale, we propose a category of diamorphs operating on a more pictorial level than abbreviations, even though the examples within this new category can often be seen as abbreviations. These symbols could be called “emblems”, since they are visual representations of an underlying concept, expressed not literally but rather graphically; not unlike an emblem.[19] Firstly, this category consists of Tironian notes and abbreviations that originated in Latin palaeography, but that have come to function as visual symbols for concepts rather than abbreviations of a word, a bit like a hieroglyph or a Chinese character. Many vernacular abbreviations find their origin in the extensive Latin shorthand system. These corresponds to Driscoll’s (2006: 259) category of “graphemic abbreviations”.

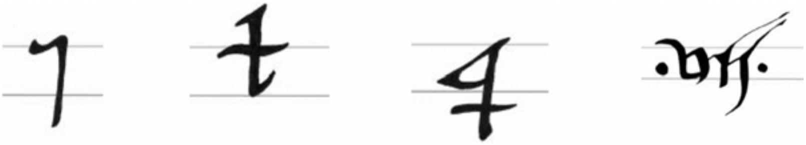


Figure 5: Examples of emblematic diamorphs: *et*, *uel*, *quia*, numerals. © Stichting A. G. van Hamel

As such, these signs can be used regardless of language. For instance, the previously mentioned Latin abbreviation for *id est*, .i., is just as easily expanded as Irish *éd on* ‘that is’ and it is employed in all language contexts. Another example of this is the Tironian note for the word ‘and’, represented by a sign similar to a seven (7). This sign may occur in a Latin context to mark *et*, but also in Irish contexts to mark the Irish word for ‘and’, *ocus*. The convergence of languages even enable it to be inserted in a Latin or Irish word to carry the phonetic value of *et*, for example in the Irish word for hundred: *cét* c7. Although of course this is by itself not a diamorph, it is a sign of the interweaving of codes. The same goes for the sign for 'or', Latin *uel*, Irish *nó*, represented by an *ɫ* with a stroke through it, and s̄ for the *s* with a macron over it, the sign for 'but', Latin *sed*, Irish *acht*.[20] A final example of this category is the abbreviation that stands for Latin *quia* and Irish *ar* meaning 'because', which eventually could even stand for Latin *-ar-*. In the same passage, this symbol could be used for Latin *quia*, for *-ar-* in Latin *carnem*, as well as for *-ar-* in Irish *ní tharmnaig* (*LB* p. 61b).

---

[19] Appel and Muysken (1987: 118) call tag-switches “emblematic”; however, this terminology is outdated. See also the article by Voigts (1989: 92) on pictorial signs in mago-medical texts, which she calls “sigils” or “carectères”.

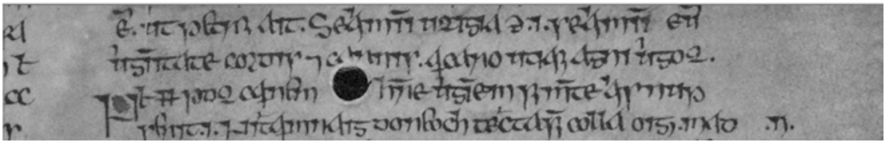


Figure 6: Leabhar Breac, p. 61b, line 2 *quiacaro utiq*ue *agni ui*rgo est. / *N*ihi*l* e*n*im *p*[ro]*d*est *carnem*. By permission of the Royal Irish Academy © RIA

This process also applies to the superscript *a* over a consonant, such as Latin t^a for *tua*; this method could then be applied to a similar circumstance in Irish t^a for *tra*, then reused in Latin *sinist*^{a} for *sinistra*. (*LB* p. 70b.)

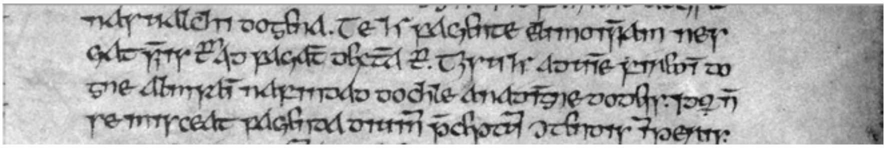


Figure 7: Leabhar Breac, p. 70b, line 2 [*nes*]*ciat s*in*is*tra *q*uid *facia*nt [sic] *dext*era *tua*. By permission of the Royal Irish Academy © RIA

Originally, each of these symbols was only part of one (visual) code, i.e. Latin. Over time, however, as Irish assumed a prominent place in liturgy, their use expanded to include the equivalent Irish meaning as well (O'Sullivan 2005: 539). As a result, these ambiguous signs could now operate freely in both an Irish and a Latin context, whether or not the resulting word is still a diamorph. In the commentary on the *Félire Óengusso*, a switch following such an emblematic diamorph often means a total switch of language systems, that is, a full alternation from one language to the other (Muysken 2000: 5). For example, the gloss on the sea creature Muirgein states that she was baptised by Comgall and afterwards told him all about her adventures. The fact that she related things to Comgall is stated in a well-constructed Latin sentence:

---

[20] Bisagni (2013–2014: 28–29) notes a gloss containing Latin *contra* 'against' solved as Irish *fri* 'against' in another codex. The extent of such copying of glosses appears to be limited.

(8) ⁊*coros baist comgall iars*in ⁊*narrauit* omn*ia q*uae *accederant sibi* in*aquis*.
‘And Comgall baptised her after that, and *she related everything that had happened to her in the water.*’ (Jan. 27 R1, f. 211v)

In this case, the symbol is clearly functioning as a language-neutral bridge between two languages (Gardner-Chloros 2009: 1). We also include Roman numerals as in

(9) *Atbert in duine fris .xui. demones uenerunt nunc in ciuitatem*
‘The man said to him: Sixteen *demons have now come into the city*’
(*LB* 60b50)

where the visual representation of the number facilitates a switch.

Comparable cases are adverbial abbreviations such as those for *v*ero and *h*autem ‘however’. Albeit Latin in form, these can function in contexts that are either partly or fully Irish; alongside *Abram v*ero a ainm ‘Abraham then [is] his name’, and *ministrantes h*autem (or h*immorro*) *ind a aingil* ‘*Those who serve* then [are] the angels’, where the diamorph can be said to stand at the language boundary, appears *atbert v*ero *eoin* aps*tal* ‘Then says John the Apostle’. The Latin word form *vero*, in this last case, may well have functioned merely as the conventional abbreviation of the written language, to have been rendered by the vernacular equivalent (*h*)*immorro* in spoken form. Hence, even words that are not diamorphous in *writing* could have been realised as diamorphs when *speaking*, because they may function rather as visual signs.[21]

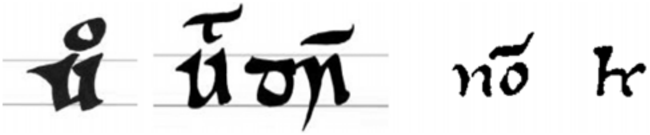


Figure 8: Other diamorphs: *uero, ut dicitur, nomen, hautem* (or *tra, amal asbert, ainm, immorro*). © Stichting A. G. van Hamel, and A. Capelli, *Dizionario di Abbreviature Latini ed Italiani* (1912)

We briefly touched on the fact that there is also a category of switches that can be considered formulaic phrases. They are in a way related to the pictorial diamorphs since they are often abbreviated and since they are of such standard stock that they may have functioned in both the Latin and the Irish repertoire. These Latin formulae even seemed to have had Irish equivalents, which might have been uttered when the Latin written formula was encountered. An example of such a phrase is *ut d*ixit, 'as he said'. *Ut dixit* is often used to introduce direct speech, and may function more as a text mark-up for a reader than as a code-switch proper.[22] Because of this specific textual function, the reader may have automatically interpreted the Latin phrase with its Irish equivalent *ama(i)l asbert*, whether he was speaking aloud or reading silently.

---

[21] See Voigts (1989: 102); *R* stands for Latin *recipe* or English *take*, *m* for Latin *manipulus* or English *handful*.

To underline the ambiguity of these phrases, and therefore their rightful position in the scale of diamorphs, these phrases can be continued as if they were Irish. For example, *ut dixit* 'as he said' could be used with the Irish *fria* meaning 'to him', resulting in the mixed phrase 'as he told him'. A native counterpart is also found for the religious phrase *am*en 'may it be so', rendered *ro-p fír* in Irish. Other examples of formulaic language are sentences where a name is explained by using phrases such as *a quo nominatur* '*from which* is named', and *no*men *eius*, *urbis* or *amnis*, 'the name *of a person, a city or a river*'. *Nomen* is often followed by a Latin noun, but it is also occasionally followed by Irish, as the gloss *nomen muccada* 'name of a swineheard' demonstrates; note that the Irish genitive case is here demanded by the Latin noun.[23] The opposite also happens, where Latin nouns are used to specify *a h-ainmm* 'its name (referring to *locusta*, a grammatically feminine word)', the Irish equivalent of the formulaic construction *nomen eius*, as happens with *locusta*. 7 *brucus a h-ainmm* 'a *locusta*, and a *locust* [is] its name'.[24] What may have been the spoken version here made it to the written stage. The "emblems" category is the largest category in both corpora: the commentary and the homilies have 73% and 54% of their respective switches here.

The congruence of Irish and Latin in our texts means that assigning words to one language or the other is rather complicated. As stated by Clyne (2003: 163), homophonous – or in our case visual – diamorphs can be seen as a method to neutralise the codes involved in a code-switch. Names of people and places are a prominent problem as these could be considered diamorphs, their original form functioning in both Irish and Latin contexts. However, such names may also employ another neutralisation strategy (Durkin 2014: 11), in that their form is often left uninflected. Forfeiting the cases on names, by using a bare form for oblique functions (as above in *orned Andreas* and *o lex*), is one way of neutralising. By contrast, Irish scribes had a choice between using a Latinate version of a name or an Irish version of a name (for example, *Petrus* or *Petar*). These two codes could even be used in the same context e.g.

---

[22] Müller (1999: 85n19) calls this metatextual use "Hervorhebung".

[23] The full quote: duin blesce. .i. blesc no*men* meret*ri*cis q*ui* h*ab*itabat ibi an*te* *sancto*s .i. fi*n*tan 7fi*n*dlug ł blesc no*men* muccuda rig huacuanach q*ui* i*n*illo loco sepe morabat*ur* 7 sic id*eo* loc*us* abeo no*min*atur.[..] 'Duin Blesc .i. blesc [is] the name of a prostitute who lived there before the saints .i. Fintan and Findlug or Blesc, the name of the swineherd of the king of the Huí Cuanach who used to abide in that place often and it is thus then the place was named from it.' (Jan. 3 R1, f. 211v)

[24] For *brucus*, see Borsje (1994: 91).

(10) *co tarla Bonifatius comorba Petair*
‘When it fell upon Boniface, the successor of Peter’

The fact that Irish equivalents of names were available supports the notion that the choice for the neutral Latinate form in an Irish context could be conscious. Sometimes the Latin name is modified by Irish determiners, as in *Bonifatius sin*, ‘that Boniface’, ‘the abovementioned Boniface’. At other times, it is left to look like Latin, as in *ar foicc Cesar .i. imper na Roma* ‘through the evil of Caesar, that is, the emperor of Rome’ (Dumville 1977–1978: 461–467). The two names *Cesar* and *Roma* could be Latin or Irish as they are left undeclined, and they are thus as ambiguous as the conjunction *.i.* and the abbreviation *imper*. Conversely, a Latin name can be declined as if it were Irish, as in *triar fer o Galilee co Ierusalem* ‘three men [went] from Galilea to Jerusalem’, where *Galilee* has the dative case demanded by the Irish preposition. Were we to read an underlying Latin preposition *de*, the case should have been ablative.[25]

An additional example is *fri himnum dicat* 7 *imnum míchíl* ‘with a *Hymnum Dicat* and a Hymn of Michael’, where the noun has a Latinate ending with an accusative case prompted by the Irish preposition. By contrast, the corresponding Latin preposition *cum* would have yielded ablative case.[26] Here, there appears to be a government relation between the Irish prepositions and the accompanying Latin nouns that display the case demanded in Irish rather than Latin. This governing of inflections, as an alternative to the aforementioned neutralisation of endings, acts as a bridge between the two languages Bisagni (2013–2014: 34). Such names are therefore both Latin and Irish at the same time, confirming their right to be considered diamorphs in this discussion.

---

[25] That the writer knew the correct Latin inflections is evinced by the phrase *A galilea quia ibi educatus est*, *LB* p.41b. Still he chose not to use the Irish equivalent *Galail*. See Bisagni (2013–2014: 46) for similar bridges between codes.

[26] The Latin *(h)imnum* is not to be confused with the Irish *(h)immun(d)*, which is found in the previous phrase: *frihimmund patraic* 7 *immund nanapstal* ‘with a Hymn of Patrick and a Hymn of the Apostles’.

6 Conclusion

As it appears, diamorphs are only one of several phenomena that neutralise the divergences between codes, and it appears that these strategies can be combined if necessary. We have seen the use of former loans such as *pax* and *lex* without grammatical modification and can now also include the integration of names into this spectrum. The use of ambiguous abbreviations and function words is also part of the same continuum. From our examples, it has become clear that there are many similarities between the types of diamorphs that occur in both corpora and the frequency in which they do so: abbreviations constitute by far the smallest group, followed by function words. The largest group in both corpora is that of emblems with 73% in the commentary and 54% in the homilies. For borrowings, however, there is substantial difference in the percentages in which they occur: 12% in the commentary versus 24% in the homilies. This is mainly due to the fact that the latter corpus contains more diamorphic names. These numbers, together with the overall percentage of diamorphs (75% of the switches), demonstrate the prevalence of ambiguous language assignment in bilingual texts.

Some instances above suggest that the overall language of a segment dictates the way in which diamorphs should be rendered, leading, for example, to the Irish reading of Latin emblems such as *vero*. It is important, however, to postpone such a linguistic verdict on diamorphs in general, as the very essence of diamorphs is that they cannot be – and do not have to be – indexed according to language. Whatever their realisation, we can conclude that diamorphs play an important role in triggering or facilitating switches. This is indicated both quantitatively, for the majority of the switches in our corpora are introduced by diamorphs, and qualitatively, as these signs come in all shapes and sizes. Expanding on the partition made by Wright (2011), we propose a new class of pictorial diamorphs, to be labelled emblems, and with this, we wish to transform the existing view on diamorphs into a gradual scale that moves from lexical borrowings to emblematic symbols. This scale does justice to the intermediate linguistic status of such diamorphs, and will allow for a sensitive and subtle analysis of historical code-switching.

References

Adams, J. N. 2003. *Bilingualism and the Latin language*. Cambridge: Cambridge University Press.

Appel, René & Pieter Muysken. 1987. *Language contact and bilingualism*. Amsterdam: Amsterdam University Press.

Atkinson, Robert. 1887. *The passions and the homilies from Leabhar Breac: Text, translation, and glossary.* Dublin: The Academy.

Bisagni, Jacopo. 2013–2014. Prolegomena to the study of code-switching in the Old Irish glosses. *Peritia* 24–25. 1–58.

Borsje, Jacqueline. 1994. The *bruch* in the Irish version of the Sunday Letter. *Ériu* 45. 83–98.

Bronner, Dagmar. 2005. Codeswitching in medieval Ireland: The case of the *Vita Tripartita Sancti Patricii*. *Journal of Celtic Linguistics* 9. 1–12.

Brown, Keith & Jim Miller. 2014. *The Cambridge dictionary of linguistics*. Cambridge: Cambridge University Press.

Clyne, Michael. 1967. *Transference and triggering: Observations on the language assimilation of postwar German-speaking migrants in Australia*. The Hague: Martinus Nijhoff.

Clyne, Michael. 2003. *Dynamics of language contact: English and immigrant languages*. Cambridge: Cambridge University Press.

Dictionary of the Irish Language. Cambridge University and Queen's University Belfast. dil.ie/34124 and dil.ie/30088 (accessed 2 February 2016).

Driscoll, M. J. 2006. Levels of transcription. In Lou Burnard, Katherine O'Brien O'Keeffe & John Unsworth (eds.), *Electronic textual editing*, 254–261. New York: The Modern Language Association of America.

Driscoll, M. J. 2009. Marking up abbreviations in Old Norse-Icelandic manuscripts. In M. G. Saibene & M. Buzzoni (eds.), *Medieval texts – contemporary media: The art and science of editing in the digital age*, 13–34. Como: Ibis.

Dumville, David N. 1977–1978. The Welsh Latin annals. *Studia Celtica* 12–13. 461–467.

Dumville, David N. 2002. *Félire Óengusso:* Problems of dating a monument of Old Irish. *Éigse* 33. 19–34.

Durkin, Philip. 2014. *A History of loanwords in English*. Oxford: Oxford University Press.

Gardner-Chloros, Penelope. 2009. *Code-switching*. Cambridge: Cambridge University Press.

Halmari, Helena & Timothy Regetz. 2011. Syntactic aspects of code-switching in Oxford, MS Bodley 649. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English* (Topics in English Linguistics 76), 115–153. Berlin & Boston: Mouton de Gruyter.

Hopper, Paul & Elizabeth Traugott. 1993. *Grammaticalization* (Cambridge Textbooks in Linguistics). Cambridge & New York: Cambridge University Press.

Ingham, Richard. 2011. Code-switching in the later medieval English lay subsidy rolls. In Herbert Schendl & Laura Wright (eds), *Code-switching in early English* (Topics in English Linguistics 76), 95–114. Berlin & Boston: Mouton de Gruyter.

Johnston, Elva. 2013. *Literacy and identity in early medieval Ireland* (Studies in Celtic History 33). Woodbridge: The Boydell Press.

Koch, John. 2006. *Celtic culture: A historical encyclopedia*, vol. 2. Santa Barbara: ABC-CLIO.

Mac Donncha, Frederic. 1976. Medieval Irish homilies. In Martin McNamara (ed.), *Biblical studies: The medieval Irish contribution*, 59–71. Dublin: Dominican Publications.

Mac Eoin, Gearóid. 1996. Observations on some Middle-Irish homilies. In Próinséas Ní Chatháin & Michael Richter (eds.), *Irland und Europa im früheren Mittelalter: Bildung und Literatur / Ireland and Europe in the early Middle Ages: Learning and literature*, 195–211. Stuttgart: Klett-Cotta.

Mulchrone, Kathleen, Thomas F. O'Rahilly, Elizabeth FitzPatrick & A. I. Pearson (eds.). 1926–1970. *Catalogue of Irish manuscripts in the Royal Irish Academy*, vol. 27. Dublin: Royal Irish Academy.

Müller, Nicole. 1999. Kodewechsel in der irischen Übersetzungsliteratur: Exempla et Desiderata. In Hildegard Tristram & Erich Poppe (eds.), *Übersetzung, Adaptation und Akkulturation im insularen Mittelalter* (Studien und Texte zur Keltologie 4), 73–86. Münster: Nodus Publikationen.

Muysken, Pieter. 2000. *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.

Myers-Scotton, Carol. 2002. *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.

Myers-Scotton, Carole & J. Jake. 2009. A universal model of code-switching and bilingual language processing and production. In Barbara E. Bullock & Almeida J. Toribio (eds.), *Cambridge handbook of linguistic code-switching*, 336–357. Cambridge: Cambridge University Press.

OED Online = Oxford English Dicitonary Online. http://dictionary.oed.com (accessed 1 February 2016).

Ó hAodha, Donncha (ed.). 1978. *Bethu Brigte*. Dublin: Dublin Institute for Advanced Studies.

Ó Riain, Pádraig. 1990. The Tallaght martyrologies redated. *Cambridge Medieval Celtic Studies* 20. 21–38.

Ó Riain, Pádraig. 2006. *Feastdays of the saints: A history of Irish martyrologies* (Subsidia Hagiographica 86). Bruxelles: Société des Bollandistes.

O'Sullivan, William. 2005. Manuscripts and palaeography. In Dáibhí Ó Cróinín (ed.), *A new history of Ireland*, vol. 1, 511–548. Oxford: Oxford University Press.

Putter, Ad. 2011. Code-switching in Langland, Chaucer and the Gawain Poet: Diglossia and footing. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English* (Topics in English Linguistics 76), 281–302. Berlin & Boston: Mouton de Gruyter.

Schendl, Herbert. 2011. Beyond boundaries: Code-switching in the leases of Oswald of Worcester. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English* (Topics in English Linguistics 76), 47–94. Berlin & Boston: Mouton de Gruyter.

Stichting A. G. van Hamel voor Keltische Studies (A. G. van Hamel Foundation for Celtic Studies) – Tionscadal na Nod. http://www.vanhamel.nl/codecs/Project:TnN_introduction (accessed 2 February 2016).

Stokes, Whitley. 1880. On the Calendar of Oengus. *The Transactions of the Royal Irish Academy: Irish Manuscript Series* 1. 1–352.

Stokes, Whitley. 1905. *Félire Óengusso Céli Dé: The martyrology of Oengus the Culdee*. Dublin: Dublin Institute for Advanced Studies.

Thurneysen, Rudolf. 1908. Die Abfassung des *Félire* von Oengus. *Zeitschrift für Celtische Philologie* 6. 6–8.

Thurneysen, Rudolf. [1947] 1970. *A grammar of Old Irish*. Dublin: Dublin Institute for Advanced Studies.

Traugott, Elizabeth & Richard Dasher. 2002. *Regularity in semantic change* (Cambridge Studies in Linguistics 96). Cambridge & New York: Cambridge University Press.

Voigts, Linda Ehrsam. 1989. The character of the *carecter*: Ambiguous sigils in scientific and medical texts. In A. J. Minnis (ed.), *Latin and vernacular: Studies in late-medieval texts and manuscripts* (Manuscripts Conferences: Proceedings Series 1), 91–109. York: D. S. Brewer.

Wenzel, Siegfried. 1994. *Macaronic sermons*: *Bilingualism and preaching in late-medieval England* (Recentiores: Later Latin Texts and Contexts). Ann Arbor, MI: University of Michigan Press.

Wright, Laura. 2011. On variation in medieval mixed-language business writing. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English* (Topics in English Linguistics 76), 191–218. Berlin & Boston: Mouton de Gruyter.

Alpo Honkapohja

12 “Latin in recipes?” A corpus approach to scribal abbreviations in 15th-century medical manuscripts

1 Introduction

As anyone who has worked with medieval manuscripts will know, a considerable number of words in them are written using an abbreviation system in which syllables, letters and word endings are substituted with graphs, which have more in common with syllabic scripts and logograms than the kind of Latin alphabetic script used in modern printed text (see e.g. Honkapohja 2013: 5.1). This medieval abbreviation and suspension system had the obvious use of saving space in a handwritten literary culture, but since many of the abbreviations can be expanded in Middle English as well as Latin, it has also been linked to language-independent communication in registers such as business accounts (e.g. Wright 2000, 2012) and scientific writing (Voigts 1989). This language independence makes abbreviations an important part of multilingual practices in late medieval England, and one that has hitherto received only limited attention.

The aim of the current paper is to examine abbreviations in the text type of recipes in multilingual medical manuscripts – using an approach which contrasts their use in Latin and Middle English. Despite the possibilities of XML markup for this type of approach, the abbreviations are not encoded in most corpora, and the possible information contained in them is not made use of in historical corpus linguistics. The current paper is a pilot study which tests the applicability of this kind of methodology to abbreviations. The data comes from five manuscripts, which are part of a group originally described by Voigts (1990). I use a TEI P5 XML – based encoding, which makes it possible to examine the abbreviations and their expansions quantitatively, applying methodologies developed for corpus linguistics.

The article is divided into the following sections. Section 2.1 presents the source for the data, the Voigts–Sloane Sibling Group. Section 2.2 contains definitions and encoding for abbreviations. Section 2.3 briefly discusses recipes as a text type and gives the criteria used for selecting them. Section 3.1 presents the abbreviation frequencies. Section 3.2 looks in more detail into what gets abbreviated in Latin and Middle English. Section 3.3 contains an analysis of lexical

DOI 10.1515/9781501504945-012

words and function words. Section 3.4 looks at medical vocabulary and measurements. Section 4 contains conclusions and discussion of future directions, of which there are many.

2 The data

2.1 What is the Voigts–Sloane Sibling Group?

Originally described by Voigts (1990), The Sloane Sibling Group consists of six manuscripts, copied between 1450 to 1490 in London or its metropolitan area, which contain an anthology of twelve medical and astrological texts in a standard order.[1] The group has received interest for two reasons. The first is that the manuscripts may be evidence of standardisation in the production of medical codices slightly before the printing press (see e.g. Green 2000 and Jones 2004). The second is that they are highly bilingual manuscripts, containing Latin texts copied for the commercial book market rather than in a university setting, fairly late in the 15th century, when all medical genres can be found as English translations (see e.g. Voigts 1996; Pahta 2004 and Honkapohja 2011).

The defining characteristic of the Voigts–Sloane Sibling Group manuscripts is the Sibling Set Text: an anthology of twelve medical and astrological texts, which appear in the same order[2] in all of the codices. The anthology was copied by a single hand, a different one for each manuscript, who was responsible for both the Latin and Middle English texts. A list of the texts is given in Table 1 below, with Middle English texts marked in bold.[3]

---

[1] The Sloane Group as originally described by Voigts (1990) consists of eleven manuscripts, including the "core" group defined by codicological similarities and the "sibling" group defined by the recurring anthology of texts. One manuscript, Sloane 2320, belongs to both groups, since the first three quires in the codex, which form an independent booklet, contain the sibling anthology and the manuscript is in the layout characteristic of the core group. I argue in my PhD dissertation that these groups should be considered separately (Honkapohja 2017). In this paper, I concentrate on the "sibling" group.

[2] One manuscript, Trinity O.1.77, switches the order of two distillatory recipes (5. and 7.). Two manuscripts, Gonville & Caius 336/725 and Tokyo Takamiya 33, omit the astrological texts, 11 and 12.

[3] I use titles taken from the eTK (Latin) or eVK2 (Middle English), except for the *Twenty-jordan series* uroscopical treatise and the plague treatises by John of Burgundy, which have been subject to thorough studies by Tavormina (2005) and Matheson (2005, 2006) respectively. The eleventh treatise, *De conditionibus septem planetarum*, is not listed as a separate item in the eVK and I use a title from manuscript incipits.

Table 1: The bilingual medical anthology that defines the Voigts–Sloane Sibling Group

1. *Manipulus medicine de digestivis et laxativis*, eTK: 0982F.
2. ***Practica urinarum***, eVK: 3229.00.
3. ***The twenty-jordan series***, eVK: 4401.00.
4. *Expositiones colorum urinarum in ordine*, eTK: 235B.
5. *Aqua mirabilis et preciosa*, eTK: 1325B.
6. *De regimine sanitatis*, eTK: 135M.
7. *De mirabilibus aquae ardentis rectificate*, eTK: 0007K.
8. John of Burgundy: plague treatise (long version), eTK: 0488P.
9. John of Burgundy: plague treatise (English), eVK: 2177.00.
10. John of Burgundy: plague tratise (the epistolary version), eTK: 431K.
11. ***De conditionibus septem planetarum***.
12. *De signis sumptis per lunam in quo signo zodiaci sit*, eTK: 80Q.

The anthology contains both Latin and Middle English texts organised into thematic groups. The first text (1) gives instructions for making laxative and purgative remedies as well as administering them correctly based on the patient's condition and complexion. It is followed by a uroscopical section, consisting of two Middle English texts (2 & 3) and one Latin text (4). Two long recipes (5 & 7) give instructions for making alcohol-based medicaments. They are accompanied by a regimen of health (6), and followed by a section on the plague, consisting of three different versions of a treatise attributed to John of Burgundy (8–10). The anthology concludes with two astrological texts (11–12).

All English texts of this bilingual anthology appear next to Latin ones on the same subject matter, whereas a number of subjects do not get a corresponding English equivalent. There is also a functional distinction. Specialised and technical content, namely, astrological tables and recipes are in Latin, while the Middle English texts recount the same information, but in a condensed form. The Middle English translation of John of Burgundy (9) only lists the ingredients without giving specific amounts (cf. Honkapohja 2017: 127–129), while the Latin versions (8 and 10) contain recipes with exact measurements for ingredients. Actual code-switching in the Sloane Sibling Group is rare; multilingualism in the Group consists mainly of what Schendl and Wright (2011: 23–24) call *intertextual code-switching*: the "occurrence of different monolingual texts collected in a single manuscript" (see also Skaffari, this volume).

The bilingual anthology is found in six manuscripts, five of which are used as material in this study.[4] This includes one quarto-sized manuscript, Sloane

---

[4] I exclude Tokyo Takamiya, since I only have access to it in a rather bad quality microfilm located in the British Library.

2320, where the texts take up the first three quires, and are bound together with booklets containing exclusively Latin treatises. It also includes three pocket-sized manuscripts, Boston MS 19, Sloane 3566 and Trinity O.1.77, and two larger *de luxe* codices, Gonville & Caius 336/725 and Tokyo Takamiya 33, where the main contents are Middle English and the Sibling anthology appears at the end.

Table 2 presents the manuscripts as well as the word count. The Middle English (ME) corpus comprises all texts that are part of the Sibling anthology, but excludes additional texts and marginal comments. The Latin corpus includes only recipes. The reason for the focus on Latin recipes is that abbreviations are very common in the text-type, which makes them a good candidate for finding regularity in a pilot study such as this. The corpus size is small, but it is offset by the fact that it is a parallel corpus of manuscripts which are very close to each other, which enables examining how the different scribes have treated each word and its abbreviations in texts whose contents are very close.

Table 2: The five manuscripts included in the study

Manuscript	**ME**	**Latin recipes**
Sloane 2320, British Library: a quarto-sized paper manuscript.	2421	1672
Sloane 3566, British Library: a small parchment manuscript.	2469	1684
Boston Ballard 19, Medical Library in the Francis A. Countway Library of Medicine: a small paper manuscript.	2434	1673
Trinity College, Cambridge, O.1.77: a small paper manuscript.	2518	1702
Gonville & Caius, Cambridge, 336/725: a *de luxe* parchment manuscript from ca. 1480s–90s.	2252	1667

2.2 Abbreviations and encoding

The contribution of the current paper is to apply corpus linguistic methodology to encode the abbreviations and their expansions. The linguistic problems related to using data where abbreviations are silently expanded are mentioned by a number of scholars, including Wright (2012: 101), Lass (2004), and Grund (2006). Theoretical issues related to their encoding are discussed by Robinson and Solopova (2006), Driscoll (2006, 2009), Stenroos (2007), Rogos (2011, 2012) and Honkapohja (2013). However, their corpus linguistic potential still remains largely unexplored. The only corpus study to get interesting results that I am aware of is Rogos (2012), in which she compared the use of abbreviations in a parallel corpus of 10 *Canterbury Tales* manuscripts in the *Man of Law's Tale* corpus and was able to uncover a number of consistencies in the lexical and orthographic contexts which trigger abbreviation (2012: 7).

The current paper resembles Rogos' study in making use of a parallel corpus, but differs from it in its bilingual data and the XML encoding. The encoding makes use of a feature which was introduced in the P5 version of the guidelines, the <choice> element, which makes it possible to encode both the abbreviation symbol and the abbreviated content for each abbreviation and abbreviated word. Example 1 shows how the word *ī* 'i*n*' is encoded.[5]

(1)

```
<w>
<choice>
<abbr>i<am><g ref="#macron"></am></abbr>
<expan>i<ex>n</ex></expan>
</choice>
</w>
```

I use a different encoding for superscript abbreviations, which are encoded with <hi rend="superscript"> as shown in example 2 below, which shows the encoding for *ponat*^{r} 'ponat*ur*' [is put]. The reason for this is that they are an open category, particularly with Latin case endings, and the @ref for the <g>-element requires a closed set of values. In Tables 3 and 4 both <am> and <hi rend="superscript"> are included in the abbreviation count.

(2)

```
<w>
<choice>
<abbr>ponat<hi rend="superscript">r</hi></abbr>
<expan>ponat<ex>ur</ex></expan>
</choice>
</w>
```

The transcription used for abbreviations is graphemic (cf. Robinson and Solopova 2006: 2; Varila 2014: 158). Thus I make a distinction between what are deemed to be different types of abbreviations, but group together what are deemed to be different forms of the same abbreviation. The practice is analogous to treating, for example, the different s-variants, the tall s, the sigma-shaped s and the kidney-shaped s, as allographs.

This entails two types of difficulty. First, placing something as fluid and variable as the strokes produced by a scribal quill or stylus inside the exact lockers of XML markup, necessarily requires editorial decisions in which several encodings could be possible. I initially attempted to distinguish between a hook ᵓ (e.g., *euᵓy* ‘every’) and a tail-hook a more horizontal variant which occurs at the ends of letters, typically in connection to *t* or *r* (e.g., *watᵓ* ‘water’). However, it proved impossible to have reliable criteria for distinguishing the two. Consequently, both are encoded as hooks. Likewise, it is sometimes difficult to make a distinction between a strikethrough next to a letter with a tall ascender, i.e. *b*, *l* or *h* (e.g., *passetħ* ‘passeth(e?)’) and a macron (typically representing a nasal, e.g., *īto* ‘into’).

---

[5] Principles and difficulties used for representing the different types of abbreviations are discussed in greater detail in Honkapohja 2013.

Second, there are a number of cases in which the expansion is not certain. A notorious example is determining whether a strikethrough or a hook at the end of a Middle English word represents a word-final -*e* or is otiose (cf. Rogos 2012). For the purposes of this article, I represent uncertain expansions with a question mark. Example 3 displays the encoding for *colourᵓ* ‘colour(e)’, where the hook may or may not indicate a final -*e*.

(3)
```
<w>
<choice>
<abbr>colour<am><g ref="#hook"></am></abbr>
<expan>colour<ex>?</ex></expan>
</choice>
</w>
```

2.3 Recipes

The Middle English texts were included in their entirety. Due to the small scale pilot nature of the study, I decided to focus on Latin recipes. The assumption was that they would be a suitable text-type for finding regularities within the abbreviation and suspension system, because they are among the most heavily abbreviated material in the Latin sections, and since they contain a number of language-independent qualities, including apothecaries’ weights as well as Roman and Arabic numerals (cf. Voigts 1989).

The recipe as a text-type can be defined focusing on its communicative function or linguistic features (cf. Taavitsainen 2001: 88), and has been the focus of a number of studies, including (Stannard 1982; Görlach 1992; Carroll 1999, 2004; Taavitsainen 2001; Mäkinen 2004, 2006; Grund 2003; Jones 1998; Alonso-Almeida 2008), concentrating variously on the “social, linguistic and technical features” of the genre/text type, and identifying “a high degree of standardisation of the structure of the recipes” (Alonso-Almeida and Carroll 2004: 26).

In medical texts, recipes can be found independently or embedded in other texts (Taavitsainen 2001: 86; Grund 2003: 462). Both types can be found in the Voigts–Sloane Sibling anthology. The anthology includes longer recipes on the distillation of alcohol-based medicaments (see Voigts 2003), which are given individual incipits and explicits,[6] as well as shorter recipes on a number of topics, which are integrated into the treatise. The latter is characteristic of specialised and academic treatises (cf. Taavitsainen 2001; Pahta and Taavitsainen 2004; Alonso-Almeida and Carroll 2004). These include recipes for emptying the patients' stomach by either route in 1. *Manipulus medicine de digestiuis et laxatiuis* and recipes against the plague found in two versions of the plague treatise attributed to John of Burgundy (Matheson 2005, 2006), texts 8 and 10. Out of the English texts in the manuscripts only the Middle English version of John of Burgundy's plague treatise (9) contains recipes and these lack exact measurements.

In order to accommodate both the longer and shorter recipes in the Latin corpus, I ended up using a simple and inclusive definition. Recipes were identified by the presence of the imperative *recipe* or the graph ℞, which, according to Grund (2003: 461), "may be a conventionalised initial marker", performing a "similar function to that of a heading [. . .] in Latin, bringing the readers' attention to the text as a recipe or signalling that the recipe is beginning". The end of the recipe was identified following scribal punctuation.

3 Analysis

3.1 Abbreviation frequency

Calculating the simple frequencies of abbreviated words in proportion to the word count revealed considerable differences between Latin and English. Tables 3 and 4 give word counts in the Latin and Middle English corpus. Figures 1 and 2 display the same information as bar diagrams.

To begin with, the relative frequency of abbreviated words is considerably higher in the Latin than in Middle English. The number of abbreviated words is particularly low in Middle English texts in the two larger manuscripts, Sloane 2320 and Gonville & Caius 336/725, in which only 3.7% and 3.5% of the words are abbreviated, respectively.

Perhaps expectedly, abbreviated words are most frequent in the three pocket-sized manuscripts, Sloane 3566, Boston MS 19 and Trinity O.1.77. The manuscript that contains the highest number of abbreviations is Trinity O.1.77, and this is the case with both Latin and Middle English. The number of abbreviated Latin words in Latin recipes is as high as 45.5%, in other words, almost every other word contains an abbreviation of some kind. The contrast to the 3–4% of the Middle English texts in the two larger manuscripts is considerable.

---

[6] These are 4. *Aqua mirabilis et preciosa* and 7. *De mirabilibus aquae ardentis rectificate*.

Table 3: The Latin corpus

Manuscript	Word count	The number of words containing abbreviations	The percentage of abbreviated words %
Sloane 2320	1672	434	25.9%
Sloane 3566	1684	654	38.9%
Boston MS 19	1673	515	30.8%
Trinity O.1.77	1702	775	45.5%
Gonville & Caius 336/725	1667	556	33.4%
Overall:	**8423**	**2934**	**34.8%**

Table 4: The Middle English corpus

Manuscript	Word count	The number of words containing abbreviations	The percentage of abbreviated words %
Sloane 2320	2421	89	3.7%
Sloane 3566	2469	235	9.5%
Boston MS 19	2434	241	9.9%
Trinity O.1.77	2518	465	18.5%
Gonville & Caius 336/725	2252	78	3.5%
Overall:	**12094**	**1108**	**9.2%**

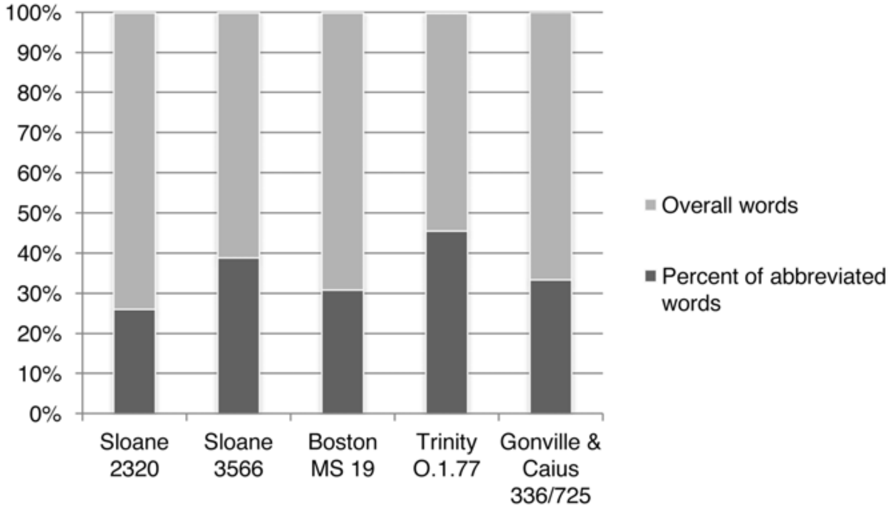


Figure 1: Latin recipes: percentage of abbreviated words of the overall word count.

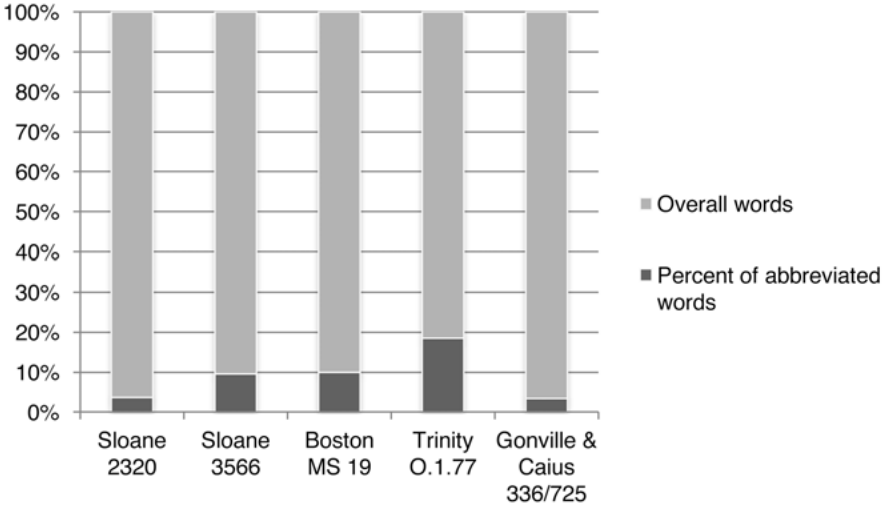


Figure 2: Middle English: percentage of abbreviated words of the overall word count.

3.2 What gets abbreviated?

We now move on to considering the most frequent abbreviated words used in the manuscripts. These are given in Tables 5 and 7. The count is a token count, but I conflate grammatical and derivational endings as well as common spelling variants, such as the variation between *i* and *y* as well as *i* and *j* into the same count. Thus ME *pimꝑnelł* and *pymꝑnelł* ‘pim*per*nel’ and Latin *pimꝑnella* ‘pim*per*nella’ (NOM) and *pimꝑnellā* ‘pim*per*nella*m*’ (ACC) are counted as a single item as are *om̄ia* ‘om*n*ia’ [all, every], *om̄ibus* ‘om*n*ibus’ (GEN) and *om̄j* ‘om*n*j’ (DAT or ABL) as well as derivational endings, which use same abbreviation for the lexical item such as puluᵓe ‘pulv*ere*’ [powder] and puluᵓizata ‘pulv*er*isata’ [made into powder]. On the other hand, the following forms are all considered to be separate: ʒ ‘drachm’ (found 33 times in Sloane 2320), *dragmā* ‘drachma*m*’ (ACC, found once in Sloane 2320) and *dragmatᵓ* ‘drachmat*e*’ (ABL found once in Sloane 2320), because they use a different abbreviation for the same content. Medieval Latin compounds such as *suꝑdominari* ‘su*per*dominari’ (PASS INF) [to be dominated] and *suꝑexaltari* ‘su*per*exaltari’ [to be praised] are counted separately.

Table 5: The 25 most frequent abbreviated Latin words

2320		3566		Boston		Trinity		Gonville	
ʒ 'drachm'	33	ʒ 'drachm'	30	ʒ 'drachm'	28	cū 'cu*m*'	32	ʒ 'drachm'	32
℞ 'recipe'	28	℞ 'recipe'	30	℞ 'recipe'	27	ī 'i*n*'	29	cū 'cu*m*'	29
an^{a} 'of each'	25	℥ 'ounce'	25	℥ 'ounce'	25	ʒ 'drachm'	27	℞ 'recipe'	29
℥ 'ounce'	21	an^{a} 'of each'	23	an^{a} 'of each'	24	℞ 'recipe'	26	an^{a} 'of each'	25
ß 'semis'	16	cū 'cu*m*'	23	ß 'semis'	15	℥ 'ounce'	24	℥ 'ounce'	23
pon(a/e)(n)t^{r} 'ponat*ur*'	10	ß 'semis'	17	ꝑ '*per*'	12	an^{a} 'of each'	21	ß 'semis'	17
ꝑ '*per*'	7	pon(a/e)(n)t^{r} 'ponat*ur*'	11	cū 'cu*m*'	11	ß 'semis'	16	ꝑ '*per*'	11
cū 'cu*m*'	6	ꝑ '*per*'	9	pona(n)t^{r} 'ponat*ur*'	10	bn̄ 'b*ene*'	11	pona(n)t^{r} 'ponat*ur*'	11
pomū 'pomu*m*'	6	fu^{ɔ}rint	7	vn(ā/ū) 'una*m*'	6	ꝑ '*per*'	11	itm̄ 'it*em*'	9
zucar^{ɔ} 'zucar*e*'	6	om̄i(a/ibus /es) 'om*n*ia'	7	zucar^{ɔ} 'zucar*e*'	6	pona(n)t^{r} 'ponat*ur*'	11	col(a/e)(n)t^{r} 'colat*ur*'	7
om̄i(a/ibus) 'om*n*ia'	5	bn̄ 'b*ene*'	6	empla(u)strū 'emplastru*m*'	5	col(a/e)(n)t^{r} 'colat*ur*'	8	zuc(/c)ar^{ɔ} 'zucar*e*'	7
lb 'libra'	5	c^{a} 'c*etera*'	6	pomū 'pomu*m*'	5	itē 'ite*m*'	8	bn̄ 'b*ene*'	6
pimꝑnella(m) 'pim*per*nella'	4	ꝑt(e/is) '*par*t'	6	itm̄ 'it*em*'	4	vn(ā/ū) 'una*m*'	7	pomū 'pomu*m*'	6
pulu^{ɔ}e 'pulv*er*e(/m/isata)'	4	pomū 'pomu*m*'	6	miscea(n)t^{r} 'misceat*ur*'	4	e^{ɔ} 'e*st*'	6	vn(ā/ū) 'una*m*'	7
suma(n)t^{r} 'sumat*ur*'	4	zucar^{ɔ} 'zucar*e*'	6	pimꝑnella(m) 'pim*per*nella'	4	fu^{ɔ}r(a/i)(n)t 'fu*er*it'	6	lb 'libra'	5
conficeant^{r} 'conficeant*ur*'	3	adda(n)t^{r} 'addat*ur*'	5	qd 'q*uo*d'	4	pomū 'pomu*m*'	6	addat^{r} 'addat*ur*'	4
fu^{ɔ}it 'fu*er*it'	3	aq^{a} 'aq*ua*'	5	suma(n)t^{r} 'sumat*ur*'	4	zucar^{ɔ} 'zucar*e*'	6	emplaustrū 'emplastru*m*'	4
m 'manipulus'	3	col(a/e)(n)t^{r} 'colat*ur*'	5	suꝑ 'su*per*'	4	addat^{r} 'addat*ur*'	5	om̄i(a/ibus) 'om*n*ia'	4
quousqz 'quousq*ue*'	3	electuariū'electuariu*m*'	5	aq^{a} 'aq*ua*'	3	ōi 'o*mni*(a/ibus)'	5	pimꝑnella(m) 'pim*per*nella'	4
Э 'scruple'	3	l 'libra'	5	bn̄ 'b*ene*'	3	lb 'libra'	5	suma(n)t^{r} 'sumat*ur*'	4
		pimꝑnella(m) 'pim*per*nella'	4	c^{a} 'c*etera*'	3	pulu^{ɔ}e(m/isata) 'pulv*er*e'	5	suꝑ 'su*per*'	4
		pulu^{ɔ}e(m/isata) 'pulv*er*e'	4	colent^{r} 'colent*ur*'	3	tūc 'tu*n*c'	5	c^{a} 'c*etera*'	3
		tormentilł 'tormentill*a*'	4	conßuat 'con*ser*vat'	3	c^{a} 'c*etera*'	4	dicit^{r} 'dicit*ur*'	3
		usqz 'usq*ue*'	4	dicit^{r} 'dicit*ur*'	3	pimꝑnella(m) 'pim*per*nella'	4	digestiuū 'digestiuu*m*'	3

Table 6: The 25 most frequent abbreviated English words. Spelling variation as well as variation related to the grammatical/derivational ending is indicated in brackets.[7]

2320		3566		Boston		Trinity		Gonville	
matierᵓ ‘matter(e?)’	4	þᵗ ‘th*at*’	36	þᵉ ‘*the*’	18	ī ‘i*n*’	53	þᵉ ‘th*e*’	11
þᵓ ‘th*ere*’	3	þˢ ‘th*is*’	16	m(a/e)n̄ ‘man(n)’	16	þᵗ ‘th*at*’	37	ī ‘i*n*’	5
blac(k)qᵓ ‘black(e?)’	3	ī ‘i*n*’	13	than̄ ‘than(n)’	11	hī ‘hi*m*’	10	watᵓ ‘wat*er*’	5
telletħ ‘telleth(e?)’	3	colourᵓ ‘colour(e?)’	11	þᵗ ‘th*at*’	10	oþᵓ(e) ‘oth*er*’	10	pⁱncipalł ‘p*ri*ncipal’	4
cometħ ‘cometh(e?)’	2	wᵗ ‘w*ith*’	10	wᵗ ‘w*ith*’	9	watᵓ ‘wat*er*’	10	þᵗ ‘th*at*’	4
considᵓration ‘consid*er*ation’	2	þᵘ ‘th*ou*’	8	vndirᵓ ‘under(e?)’	6	wᵗ ‘w*ith*’	10	coloʳ ‘colo*ur*’	3
domina(c/t)ōn ‘dominat*i*on’	2	þᵉ ‘th*e*’	7	thyn̄ ‘thyn(n)’	5	þā(ne) ‘the*n*/tha*n*’	8	sū ‘so*m*e’	3
iijᵈ ‘third’	2	ben̄ ‘be(e)n(n)’	5	whann̄ ‘when(n)’	5	feuᵓe ‘fev*er*(e?)’	8	complexioū ‘complexio*n*’	2
ꝑf(ect/it)ly ‘*p(e/a)r*fectly’	2	circulꝰ ‘circul*us*’	5	feirᵓ ‘fair(e?)’	4	þᵘ ‘th*ou*’	7	dn̄acōn ‘d*omi*nat*i*on’	2
passetħ ‘passeth(e?)’	2	oþᵓ ‘other’	5	feuerᵓ ‘fever(e?)’	4	clens(i/y)ngᵓ ‘cleansing(e?)’	7	oþᵓ ‘oth*er*’	2
pimꝑnelł(e) ‘pim*per*nel’	2	man̄ ‘man(n)’	3	ꝑt(y/ies) ‘*par*ty’	4	lyuᵓ(e) ‘liv*er*’	7	sedimē ‘sedime*n*’	2
vndrᵓ ‘und*er*’	2	coloʳ(e) ‘colo*ur*(e)’	3	watᵓ ‘wat*er*’	4	vndᵓ ‘und*er*’	7		
wᵗ ‘w*ith*’	2	gᵃuel ‘g*ra*uel’	3	airᵓ ‘air(e?)’	2	pⁱncipal(le/y) ‘p*ri*ncipal(ly)’	6		
		lyuᵓ ‘liv*er*’	3	clensyngᵓ ‘cleansing(e?)’	3	sƀstaūnce ‘s*u*bsta*n*ce’	6		
		þᵓof ‘th*er*eof’	2	coloʳ ‘colo*ur*’	3	wheþᵓ ‘wheth*er*’	6		
		born̄ ‘born(n/e)’	2	colourᵓ ‘colour(e?)’	3	matᵓ ‘matt*er*’	6		
		cōnsidre ‘co*n*sider’	2	longᵓ ‘long(e?)’	3	sū(me) ‘so*m*e’	5		
		dominacōn ‘dominat*i*on’	2	othirᵓ ‘other(e?)’	3	werᵓ ‘wer*e*’	5		
		euᵓy ‘ev*er*y’	2	thyngᵓ(s) ‘thing(e?)s’	3	þᵓ(e) ‘th*er*e’	4		
		ouᵓall ‘ov*er*all’	2	euᵓy ‘ev*er*y’	2	coloʳ ‘colo*ur*’	4		
		pⁱncipalle ‘p*ri*ncipal’	2	gᵃuell ‘g*ra*vel’	2	elł ‘else’	4		
		pymꝑnelł(e) ‘pim*per*nel’	2	lyuᵓ ‘liv*er*’	2	īto ‘i*n*to’	4		
		watᵓ ‘wat*er*’	2	pimꝑnelł ‘pim*per*nel’	2	þˢ ‘th*is*’	3		
		wheithᵓ ‘wheth*er*’	2	wᵗin ‘w*ith*in’	2	þᵓof ‘th*er*eof’	3		

---

[7] I include the word *þᵉ* ‘the’ as an abbreviation, even though it does not omit a letter or a syllable. The reason for this is that it is very similar to thorn-based abbreviations like *þˢ* ‘th*is*’ and *þᵗ* ‘th*at*’. Moreover, it does serve the function of saving space by making the word slightly shorter.

These tables reveal striking differences in the most frequent abbreviated words in Latin and Middle English. In Latin, the sequence of most frequent words is fairly close to each other with measurements, such as *ʒ* ‘drachm’, ℥ ‘ounce’ and *ß* ‘semis’ [half]; names of ingredients such as *pomū* ‘pomu*m*’ [fruit], *zucar*ꝰ ‘zucar*e*’ [sugar] and *pimꝑnella*; as well as some verbs related to preparation such as *pona(n)t*^{r} ‘pona(n)t*ur*’ [is/are put].

On the other hand, the frequencies of small function words such as *cū* ‘cu*m*’ [with] and *ī* ‘i*n*’ are more variable in Latin. For example, in the most heavily abbreviated manuscript, Trinity O.1.77, they surpass *ʒ* ‘drachm’ and ℞ ‘recipe’ in frequency, whereas in another pocket-sized manuscript, Boston MS 19, they are considerably less frequent. The abbreviation *cū* is attested 32 times in Trinity and 11 times in Boston; *ī* is attested 29 times in Trinity, but is not abbreviated in the other manuscripts. This kind of difference is in stark contrast with how close the frequencies for names of ingredients such as *zucar*ꝰ, *pimꝑnella* and *pomū*. What it suggests is that the scribes followed their exemplar very closely when it came to names of ingredients and measurements in recipes, but took more liberties with function words. This is examined further in sections 3.3 and 3.4 below.

Meanwhile, Middle English displays considerably more variation in the most frequent abbreviated words. For example, the second most frequent abbreviation used by the Sloane 3566 scribe *þ*^{s} ‘th*is*’ (16) is only used three times by the Trinity scribe. The most frequently abbreviated words in the Trinity manuscript, ī ‘i*n*’ (53) and þ^{t} ‘t*hat*’ (37) do not appear at all in Sloane 2320 and are attested in Gonville & Caius only 5 and 4 times respectively.

3.3 Content words and function words

For further analysis, I grouped the words into content words and function words. In the Latin material, I also separated measurements into their own group, which is discussed in section 3.4. Since the English texts lack exact measurements, I only use the categories “lexical words” and “function words”. Lexical words are “members of open classes” and “the main carriers of meaning in a text” (Biber et al. 2000: 2.2.3.1). Words in this category include: ME *colour(e)* ‘colour’, *matier(e)* ‘matter’, to drink, or Latin *ponatur* ‘is poured/put/poured/placed’. Function words, in contrast, “provide the mortar which binds the text together” and “indicate relationships between lexical words or larger units”. They are members of closed classes (Biber et al. 2000: 2.2.3.2). Since the aim of the study is to analyse scribal abbreviation patterns, I made the decision to group together words that share a form and are abbreviated in the same way, sometimes ignoring different morpho-syntactic functions. For example, the English

Table 7: The most frequent Latin lexical words in all manuscripts

2320		**3566**		**Boston**		**Trinity**		**Gonville**	
pon(a/e)(n)tʳ ‘ponat*ur*’	10	pon(a/e)(n)tʳ ‘ponat*ur*’	11	pona(n)tʳ ‘pona(n)t*ur*’	10	pona(n)tʳ ‘ponat*ur*’	11	pona(n)tʳ ‘ponat*ur*’	11
zucarᵓ ‘zucar*e*’	6	fuᵓint ‘fu*er*int’	7	vn(ā/ū) ‘una*m*’	6	col(a/e)(n)tʳ ‘colat*ur*’	8	col(a/e)(n)tʳ ‘colat*ur*’	7
pomū ‘pomu*m*’	6	om̄es ‘omnes’	7	zucarᵓ ‘zucar*e*’	6	vn(ā/ū) ‘una*m*’	7	zuc(/c)arᵓ ‘zucar*e*’	7
om̄i(a/ibus) ‘om*n*ia’	5	zucarᵓ ‘zucar*e*’	6	empla(u)strū ‘emplastru*m*’	5	eᵓ ‘e*st*’	6	pomū ‘pomu*m*’	6
suma(n)tʳ ‘sumat*ur*’	4	pomū ‘pomu*m*’	6	pomū ‘pomu*m*’	5	fuᵓr(a/i)(n)t ‘fu*er*it’	6	vn(ā/ū) ‘una*m*’	7
pimꝑnella(m) ‘pim*per*nella’	4	cᵃ ‘c*etera*’	6	miscea(n)tʳ ‘misceat*ur*’	4	addatʳ ‘addat*ur*’	5	addatʳ ‘addat*ur*’	4
puluᵓe(m/isata) ‘pulv*ere*’	4	ꝑte ‘*par*te’	6	suma(n)tʳ ‘sumat*ur*’	4	zucarᵓ ‘zucar*e*’	6	emplaustrū ‘emplastru*m*’	4
fuᵓit ‘fu*er*it’	3	addantʳ ‘addant*ur*’	5	cᵃ ‘c*etera*’	3	pomū ‘pomu*m*’	6	om̄i(a/ibus) ‘om*n*ia’	4
conficeantʳ ‘conficeant*ur*’	3	colentʳ ‘colent*ur*’	5	colentʳ ‘colent*ur*’	3	ōi ‘o*mn*i(a/ibus)’	5	pimꝑnella(m) ‘pim*per*nella’	4
emplastrū ‘emplastru*m*’	2	pimꝑnella(m) ‘pim*per*nella’	4	conßuat ‘con*ser*vat’	3	puluᵓe(m/isata) ‘pulv*er*e’	5	suma(n)tʳ ‘sumat*ur*’	4

word *other* can function as an adjective, a pronoun, a noun and an adverb, depending on the context (*Oxford Engish Dictionary* [*OED*]). *This betokenyt as doot þat oþer* 'This means as does the other' (a noun, referring back to the previous colour of urine, Text 2) and *blede on þat oþer side* 'blede on the other side' (an adjective, modifying *side*, Text 9). Both are considered as function words. I also made the decision to analyse some adverbs as lexical words ('principally'), others as function words (closed class adverbs, e.g. *bene* 'well', *item* 'likewise, besides, also'). The ten most frequent words in each category are presented in Tables 7 to 11.

Table 7 presents the ten most frequent Latin lexical words. This includes medical vocabulary, such as the names of ingredients like *zucare*, *pimpernella*, *pulver* and *emplastrum* 'plaster'. It also includes more general vocabulary, such as forms of the verb 'to be' or adjectives such as *omnia* or *una* 'one' as well as *cetera* 'for the rest, the other, remainder' usually belonging to the formulaic phrase *et cetera*. It also includes a number of verbs in their passive form. Figure 3 shows the frequencies of three passive verbs related to preparation *ponat*^{r} 'ponat*ur*', *sumat*^{r} 'sumat*ur*' [is consumed] and *addat*^{r} 'addat*ur*' [is added].

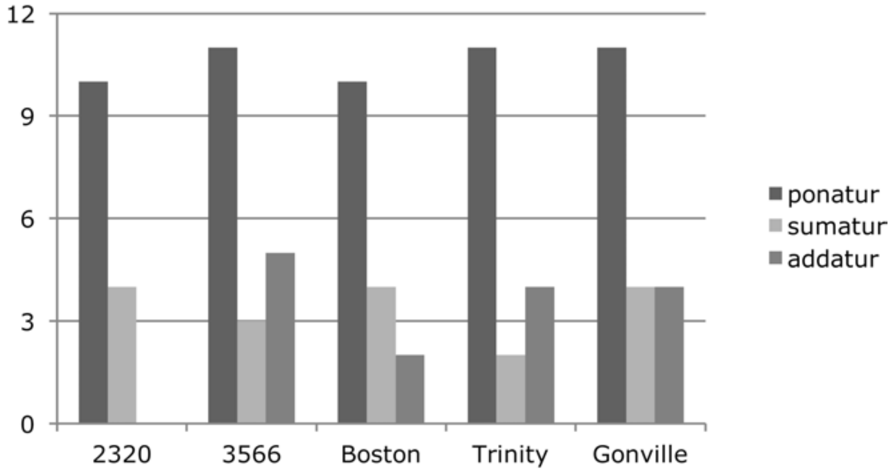


Figure 3: The frequencies of three Latin verbs

The abbreviation for the passive verbs is the superscript r, representing the syllable *ur*. The distribution of the most frequent verb *ponatur* is nearly identical in all of the manuscripts (10, 11, 10, 11, 11). The two others show more variation. The scribes of Sloane 3566 and Trinity expand *sumatur* (4, 3, 4, 2, 4), where the others abbreviate it. This is somewhat surprising, as these are the two smallest manuscripts and normally have the highest abbreviation frequencies. However, the number is the same for the other three. *Addatur*, in contrast, shows more variation (0, 5, 2, 4, 4). The scribe of 2320 always expands it, but both the scribe of the Trinity manuscript and Gonville abbreviate it four times, giving identical distributions. A fourth verb, not displayed in the table, *colat*ʳ ‘colat*ur*’ [is cared for/is cultivated] also shows variation (0, 5, 3, 8, 7).

Looking at the frequencies of Latin lexical words, it thus appears that the frequencies of medical words such as the names of ingredients are most similar across manuscripts, whereas less specialised lexical words display more variation. However, in comparison to function words, the scribes do treat them similarly. The most frequent word in this category, *ponatur*, is also treated nearly identically by the scribes in different manuscripts. This is possibly due to a formulaic function in recipes.

Table 8 shows the abbreviation practices for English lexical words. In contrast to Latin, they have considerably more variation in frequency. For example, the most frequent word is different in all manuscripts, except in Trinity and Gonville, in which the most frequent word is *wat*ᵓ ‘wat*er*’. Its frequencies are a good illustration of the extent of variation found with these words (0, 2, 4, 10, 5). The scribes appear to take more liberties with English words. A possible reason for this is that even medical words on the list are fairly general, including *liver*, *colour*, *substance* and *fever* and would thus have been familiar to the scribes. However, even words which have a specific meaning in medical theory and practice, such as *cleansing* (referring to cleansing places of the body, from which it attempts to expel the “poisonous” matter of pestilence) or *complexion* are treated differently by different scribes.

Table 9 presents the most frequent Latin function words in all the manuscripts. The preposition *cū* ‘cu*m*’ shows much variation in frequencies (6, 23, 11, 32, 29). The frequencies for the preposition *ꝑ* ‘*per*’ are closer to each other, and the same in Trinity and Gonville (7, 9, 12, 11, 11). This is perhaps due to the abbreviation for *per* being one of the most consistently used abbreviations in the period. The closed-class adverb *bn̄* ‘b*e*n*e*’ [well] is expanded varyingly by each scribe, although the scribes of Sloane 3566 and Gonville abbreviate it both six times (0, 6, 3, 11, 6). However, the Trinity scribe abbreviates it as often as eleven times, the Boston scribe only three times, and the scribe of Sloane 2320 always expands the word. The fourth word is the indeclinable adverb *item* (0, 2, 4, 8, 9) – this too shows considerable variation, considering it is a word that is frequently found as a transitional element between recipes, and could be expected to behave in a formulaic way.

Table 8: The most frequent ME lexical words in all manuscripts

2320		**3566**		**Boston**		**Trinity**		**Gonville**	
matier^{ɔ}	4	colour^{ɔ} ‘colour(e?)’	11	m(a/e)n̄ ‘man(n)’	16	wat^{ɔ} ‘wat*er*’	10	wat^{ɔ} ‘wat*er*’	5
Blacq^{ɔ}	3	ben̄ ‘be(e)n(n)’	5	thyn̄ ‘thin’	5	feu^{ɔ}e ‘fev*er*(e?)’	8	p^{i}ncipal(le/y) ‘p*ri*ncipal(ly)’	4
Stomacq^{ɔ}	3	m(a/e)n̄ ‘man(n)’	4	feir^{ɔ} ‘fair(e?)’	4	clens(i/y)ng^{ɔ} ‘cleansing(e?)’	7	colo^{r} ‘colo*ur*’	3
telleħ	3	lyu^{ɔ} ‘liv*er*’	3	feuer^{ɔ} ‘fever(e?)’	4	lyu^{ɔ}(e) ‘liv*er*’	7	complexioū ‘complexio*n*’	2
comeħ	2	colo^{r} ‘colo*ur*’	3	ꝑt(/y/ies) ‘*par*ty’	4	p^{i}ncipal(le/y) ‘p*ri*ncipal(ly)’	6	dn̄acōn ‘d*omi*nat*i*on’	2
consid^{ɔ}ation	2	g^{a}uel ‘g*ra*vel’	3	wat^{ɔ} ‘wat*er*’	4	sƀstaūnce ‘s*u*bsta*n*ce’	6	sedimē ‘sedime*n*’	2
domina(c/t)ōn ‘dominat*i*on’	2	‘cōsidre co*n*sider’	2	air^{ɔ} ‘air(/e?)’	2	mat^{ɔ} ‘matt*er*’	6		
ꝑfectly ‘*pe*rfectly’	2	dominacōn ‘dominat*i*on’	2	clensyng^{ɔ} ‘cleansing(/e?)’	3	wer^{ɔ} ‘wer*e*’	5		
passeħ	2	sedimen̄	2	colo^{r} ‘colo*ur*’	3	colo^{r} ‘colo*ur*’	4		
colo^{r} ‘colo*ur*’	1	born̄	2	colour^{ɔ} ‘colour(e?)’	3	apper^{ɔ} ‘appear’	3		

Table 9: The most frequent Latin function words in all manuscripts

2320		3566		Boston		Trinity		Gonville	
ꝑ ‘*per*’	7	cū ‘cu*m*’	23	ꝑ ‘*per*’	12	cū ‘cu*m*’	32	cū ‘cu*m*’	29
cū ‘cu*m*’	6	ꝑ ‘*per*’	9	cū ‘cu*m*’	11	ī ‘i*n*’	29	ꝑ ‘*per*’	11
quousqz ‘quousq*ue*’	3	bn̄ ‘b*e*n*e*’	6	itm̄ ‘it*e*m’	4	bn̄ ‘b*e*n*e*’	11	itm̄ ‘it*e*m’	9
		vƚ ‘v*el*’	4	qd ‘q*uo*d’	4	ꝑ ‘*per*’	11	bn̄ ‘b*e*n*e*’	6
		usqz ‘usq*ue*’	4	suꝑ ‘su*per*’	4	itē ‘ite*m*’	8	ipm̄ ‘ip*su*m’	3
		ipm̄ ‘ip*su*m’	3	bn̄ ‘b*e*n*e*’	3	tūc ‘tu*n*c’	5	quosqz ‘quousq*ue*’	3
		ꝗ’*quam*‘	3	ipm̄ ‘ip*su*m’	3	q2 ‘q*uia*’	4	vsqz ‘usq*ue*’	3
		quosqz ‘quousq*ue*’	3	sup^{a} ‘sup*ra*’	3	q^{o}d ‘q*uo*d’	4	cont^{a} ‘cont*ra*’	2
		tūc ‘tu*n*c’	3	cont^{a} ‘cont*ra*’	2	cont^{a} ‘cont*ra*’	4	iterū ‘iteru*m*’	2
		cui9 ‘cui*us*’	2	ħ ‘*hoc*’	2	usqz ‘usq*ue*’	4	nūquā ‘nu*n*qua*m*’	2

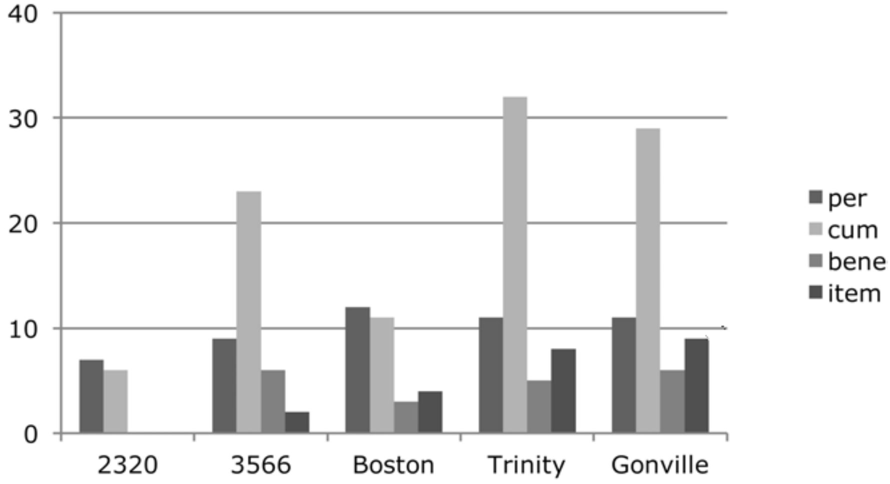


Figure 4: The frequencies of four Latin function words

Table 10 shows the frequencies of the ten most common Middle English function words. The results reveal a wide gap in the frequencies of the three pocket-sized manuscripts and the two larger manuscripts, Sloane 2320 and Gonville, which contain very few abbreviated words in this category. Figure 5 illustrates the frequencies of three frequently abbreviated words *þ*^{t} ‘t*hat*’ (0, 36, 10, 37, 2), *w*^{t} ‘w*ith*’ (2, 10, 9, 10, 0) and *ī* ‘i*n*’ (0, 13, 0, 53, 5). The major differences in frequencies are likely due to the different sizes of the manuscripts, and different requirements for abbreviating words in them. Small function words account for a major part of the higher frequencies in the small manuscripts, whereas the scribes of Sloane 2320 and Gonville, in which the relative frequency of abbreviated Middle English words is less than 4%, expand these words. Notably, the frequencies of Sloane 3566 and Trinity are very close to each other for the words *that* and *with*, but very different for *in*, which the Trinity scribe uses in both Latin (29) and English (53). This, in combination with the fact that these words are rarely abbreviated in the two larger manuscripts, supports the assumption that the scribes had individual repertoires for function words. Frequencies that are close to each other in Sloane 3566, Boston and Trinity most likely result from the scribes being consistent in abbreviating these prepositions.

Table 10: The most frequent ME function words in all manuscripts

2320		**3566**		**Boston**		**Trinity**		**Gonville**	
þ*er* ‘th*ere*’	3	þ^{t} ‘th*at*’	36	þ^{e} ‘th*e*’	18	ī ‘i*n*’	53	þ^{e} ‘th*e*’	11
vndr$^{\text{ɔ}}$ ‘und*er*’	2	þ^{s} ‘th*is*’	16	thañ ‘than(n)’	11	þ^{t} ‘th*at*’	37	ī ‘i*n*’	5
w^{t} ‘w*ith*’	2	ī ‘i*n*’	13	þ^{t} ‘th*at*’	10	hī ‘hi*m*’	10	sū ‘so*m*e’	3
		w^{t} ‘w*ith*’	10	w^{t} ‘w*ith*’	9	oþ$^{\text{ɔ}}$(e) ‘oth*er*’	10	oþ$^{\text{ɔ}}$ ‘oth*er*’	2
		þ^{u} ‘th*ou*’	8	vndir$^{\text{ɔ}}$ ‘under(e?)’	6	w^{t} ‘w*ith*’	10	þ^{t} ‘th*at*’	2
		þ^{e} ‘th*e*’	7	whanñ ‘when’	5	þā(ne) ‘the*n*/tha*n*’	8		
		oþ$^{\text{ɔ}}$ ‘oth*er*’	5	othir$^{\text{ɔ}}$ ‘other(/e?)’	3	þ^{u} ‘th*ou*’	7		
		wheith$^{\text{ɔ}}$ ‘wheth*er*’	2	eu$^{\text{ɔ}}$y ‘ev*er*y’	2	vnd$^{\text{ɔ}}$ ‘und*er*’	7		
		þ$^{\text{ɔ}}$of ‘th*er*eof’	2	w^{t}in ‘w*ith*in’	2	wheþ$^{\text{ɔ}}$ ‘wheth*er*’	6		
		ev$^{\text{ɔ}}$y ‘ev*er*y’	2	sū ‘so*m*e’	2	sū(me) ‘so*m*e’	5		

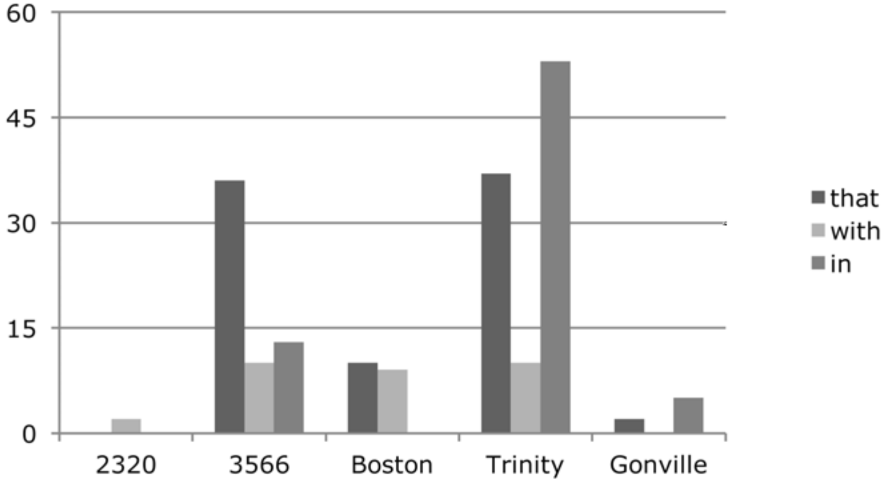


Figure 5: The frequencies of three Middle English function words

3.4 Medical vocabulary and measurements

Looking at Latin and Middle English lexical words and content words revealed variation in the abbreviation practices of the different scribes. However, there are some words which the scribes treated nearly identically from manuscript to manuscript. Table 11 illustrates the abbreviation frequencies of Latin medical content words.

The frequencies for names of ingredients such as *zucarᵓ* ‘sugar’, *pimꝑnella* and *pomū*, of medicament *emplastrum* and the verb *ponatur* have very similar frequencies across the different manuscripts. In these cases, the scribes appear to have followed the exemplar very closely.

In order to compare the scribal treatment of medical vocabulary in Middle English and Latin, I performed a qualitative analysis of the names of various ingredients and medicaments found in the manuscripts, comparing how the scribes have treated them in Latin and Middle English.[8] Some results are illustrated by Tables 12 to 14.

---

[8] The names of the ingredients and medicaments included in the qualitative analysis were vinegar (Lat. *acetum*, ME *vynegre*), adragant (Lat. *diagragante*), agarikon (Lat. *agarici* GEN), various healing waters (*aqua mirabilis*, *aqua ardens*, *aqua camphorae*), ‘aloes succotrini’ (*aloes cicotrini*), Arabian pills (*pillulae arabice*), Armenian Bole (Lat. *bol armeniac*, ME *bole armoniak*), cardamom (*cardamoni*), “berberis”, borage (*borago*), buglossa (*buglossis*), Ellen-wyrt? (Lat. *cameacte*), calamint (*calamente*), chamomile (*camomillis*), Cyclamen (*ciclaminis*), “diarodon

Table 12 shows how the Latin abbreviations for *pimpernel* are distributed across the different manuscripts. There are both similarities and differences. First of all, the abbreviation *ꝑ* ‘*per*’ is used with complete consistency for the middle syllable. Moreover, the scribal treatment of the ingredient in the first recipe, *Si fuerit pauper* ‘if he should be poor’ (text 8),[9] is the same. In the second recipe, some scribes favour the strikethough-l and some the macron for abbreviating the accusative or nominative ending. Four of the scribes have used the same abbreviations also in the third and fourth recipes. The scribe of Sloane 3566 seems to be the odd one out, using a different abbreviation to the others in the latter three recipes. In the final one *Bethazaer*, he is joined by the Trinity scribe in his deviations.

Table 13, in contrast, illustrates scribal treatment of Middle English abbreviations for *pimpernel*. They show increased variation. The scribe of 3566 writes the word with a final *-e* in the first recipe and three scribes write *ł* ‘a strikethrough-l’. In the second recipe, two of the scribes, 2320 and Trinity, write a final *-e*, and the other three a strikethrough-l. The scribes of 2320 and Boston MS 19 use *i* consistently and the scribes of 3566 and Trinity use *y*. The Gonville scribe mixes the two. The difference compared to Latin recipes is that it seems difficult to find patterns in the scribal behaviour, and the likely interpretation is that A) Middle English allows for greater variation in orthography, and B) the scribes are more comfortable with taking liberties with Middle English than with Latin, where they reproduce the exemplar.

However, there are some Latin examples in which the scribal behaviour is very similar across the manuscripts. Table 14 illustrates how *Bole Armeniac*, referring to ‘Armenian bole (a red earth)’ (*Middle English Dictionary*), ‘a soft friable fatty earth, usually of a pale red colour’ (*OED*), is treated by each scribe. The scribes use the same forms for ‘Armenian bole’ in the same location in all

---

abbatis”, dittany (*diptanum*), liverwort (*epaticum*), cuscuta epithimum? (*epithimi*), fumitory (*fumiterre*), galangal (*galange*), clove (*gariofilis*), myrrh (*mirre gummi*), Iris tuberosa (*hermodactili*), mallows (*malvarum*), manna? (*manne*), pearls (*margaritarum*), mastic (*masticis*), melilot (*mellilotis*), muscat nut (*nucis muscatis*), conium oil? (*oleo coni*), eggs (*ovum*, *vitellum*), garden parsley (*petrocilij*), mockorange? (*philadelphis*), pimpernel, portulaca (*portulace*), various roots (*radice celsi*, *radice sambuci*), Pills of Rhazes (*pillulae rasis*), sirup of roses (*sirupo rosaico*), rhubarb (*rhubarbe*), sandalwood (*sandalorum*), Hart’s-tongue? (*scolopendre*), terra sigillata, asparagus (*sporagi*), scilla (*squillae*), camel’s hay? (*squinanti*), gum of storax tree (*storacis*), sugar, tamarind (*tamarindorum*), turbit (*turbith albi*, *turbith agarici*), turpentine (*terebinthe*), tamarisk (*thamarisce*), thyme (*thimi*), tormentil (*tormentille*), furfur triticae (*tritice*), St. John’s Wort (*ypericon*), ginger (*zinziberis*), violet/gillyflower? (*violae*) and scabious (*scabiose*). The Latin words are quoted in the case found in the recipes.

[9] John of Burgundy gives different recipes for rich and poor people. This recipe is directed to the poor.

Table 11: Frequencies of medical vocabulary

2320		3566		Boston		Trinity		Gonville	
pon(a/e)(n)tʳ ‘ponat*ur*’	10	pon(a/e)(n)tʳ ‘ponat*ur*’	11	pona(n)tʳ ‘pona(n)t*ur*’	10	pona(n)tʳ ‘ponat*ur*’	11	pona(n)tʳ ‘ponat*ur*’	11
zucarꝰ ‘zucar*e*’	6	zucarꝰ ‘zucar*e*’	6	zucarꝰ ‘zucar*e*’	6	zucarꝰ ‘zucar*e*’	6	zuc(/c)arꝰ ‘zucar*e*’	7
pimꝑnella(m) ‘pim*per*nella’	4	pimꝑnella(m) ‘pim*per*nella’	4	pimꝑnella(m) ‘pim*per*nella’	4	pimꝑnella(m) ‘pim*per*nella’	4	pimꝑnella(m) ‘pim*per*nella’	4
pomū ‘pomu*m*’	6	pomū ‘pomu*m*’	6	pomū ‘pomu*m*’	5	pomū ‘pomu*m*’	6	pomū ‘pomu*m*’	6
emplastrū ‘emplastru*m*’	2	emplastrū ‘emplastru*m*’	4	empla(u)strū ‘emplastru*m*’	5	emplaustrū ‘emplaustru*m*’	4	emplaustrū ‘emplastru*m*’	4

Table 12: Pimpernel in Latin recipes

Manuscript	2320	3566	Boston	Trinity	Gonville
John of Burgundy (8), Si fuerit pauper	pimꝑnella	pimꝑnella	pimꝑnella	pimꝑnella	pimꝑnella
John of Burgundy (8), Bethazaer	pimꝑneƚƚ ‘pim*per*nell*am*’	pimꝑnellā ‘pim*per*nella*m*’	pimꝑneƚƚ ‘pim*per*nell*am*’	pimꝑneƚƚ ‘pim*per*nell*am*’	pimꝑneƚƚ ‘pim*per*nell*am*’
Epistolary Burgundy (10), Item facias tibi pulverem de istis herbis	pimꝑnella	pimꝑneƚƚ ‘pim*per*nell*a*’	pimꝑnella	pīꝑnella ‘pi*mper*nella’	pīꝑnella ‘pi*mper*nella’
Epistolary Burgundy (10), Bethazaer	pimꝑnella	pimꝑneƚƚ ‘pim*per*nell*a*’	pimꝑnella	pimꝑneƚƚ ‘pim*per*nell*a*’	pimꝑnella

Table 13: Pimpernel in Middle English recipes

Manuscript	2320	3566	Boston	Trinity	Gonville
John of Burgundy (9), recipe 1	pimꝑnelł	pymꝑnelle	pimꝑnelł	pymꝑnel	pymꝑnelł
John of Burgundy (9), recipe 2	pimꝑnelle	pymꝑnelł	pimꝑnelł	pymꝑnelle	pimꝑnelł

Table 14: Bol Armonyac in Latin recipes

Manuscript	2320	3566	Boston	Trinity	Gonville
John of Burgundy (8), Emplaustrum quod dicitur Emanuel	boli armoniaci	boli armoniaci	boli armoniaci	boli armoniaci	boli armoniaci
John of Burgundy (8), Bethazaer	boli armoniacᵓ	boli armoniacᵓ	boł armoniacᵓ	boł armoniacᵓ	boli armoniacᵓ
Epistolary Burgundy (10), Bethazaer	cum bolo armoniacᵓ	cum bolo armoniacᵓ	cum bolo armoniacᵓ	cum bolo armoniacᵓ	cum bolo armoniacᵓ

manuscripts, with the slight exception that the Boston and Trinity scribes abbreviate *boł* ‘bol*i*’ with a strikethrough in the last recipe (*Bethazaer*).

The scribal treatment of Latin medical vocabulary is mirrored by scribal treatment of measurements, which are only found in the Latin texts in the Voigts–Sloane Sibling Group. These too display a similar situation in which the scribes reproduced the forms in their exemplar. Table 15 shows the frequencies of all apothecaries’ weights in the Group.

Figure 6 shows how the most frequent apothecaries’ symbol, ʒ ‘drachm’, is distributed across three texts: 1, 8 and 10. The “abbr” column displays the number of instances where they use the symbol ʒ and the “expan” column where they write the word out.[10] It reveals that the scribes followed their exemplar nearly all the time. In texts 1 and 8, containing laxative and purgative remedies and the long Latin version of John of Burgundy’s plague treatise, they always use the symbol. In the epistolary version of John of Burgundy they write the word expanded, with only a single exception made by the scribe of Sloane

---

10 In this case, I include abbreviations such as *dragmā* ‘drachmam’ (found once in Sloane 2320) and *dragmatᵓ* ‘drachmate’ (found once in Sloane 2320) in the word count for ‘expan’.

Table 15: Apothecaries’ weights

2320		3566		Boston		Trinity		Gonville	
ʒ ‘drachm’	33	ʒ ‘drachm’	30	ʒ ‘drachm’	28	ʒ ‘drachm’	27	ʒ ‘drachm’	32
℞ ‘recipe’	28	℞ ‘recipe’	30	℞ ‘recipe’	27	℞ ‘recipe’	26	℞ ‘recipe’	29
an^{a} ‘of each’	25	℥ ‘ounce’	25	℥ ‘ounce’	25	℥ ‘ounce’	24	an^{a} ‘of each’	25
℥ ‘ounce’	21	an^{a} ‘of each’	23	an^{a} ‘of each’	24	an^{a} ‘of each’	21	℥ ‘ounce’	23
ß ‘semis’	16	ß ‘semis’	17	ß ‘semis’	15	ß ‘semis’	16	ß ‘semis’	17
lb ‘libra’	5	lb ‘libra’	5	lb ‘libra’	3	lb ‘libra’	5	lb ‘libra’	5
m ‘manipulus	3	m ‘manipulus’	3	℈ ‘scruple’	2	m ‘manipulus’	3	m ‘manipulus’	4
℈ ‘scruple’	3	℈ ‘scruple’	2	m ‘manipulus’	2	℈ ‘scruple’	3	℈ ‘scruple’	3

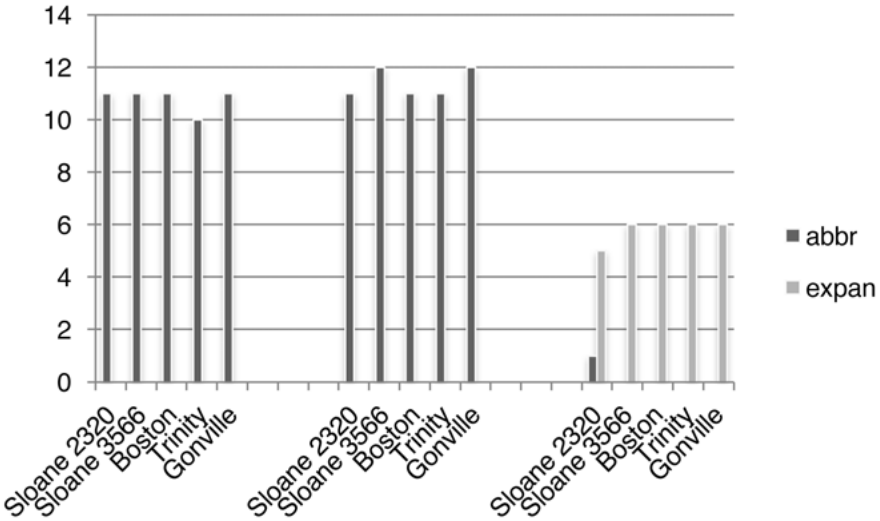


Figure 6: Abbreviation and expansion of 'drachm' in the Group (1. Manipulus medicine de digestivis et laxativis, 8. John of Burgundy: plague treatise [long version], 10. and John of Burgundy: plague treatise [the epistolary version]).

2320. Minor differences in word count are due to small textual differences and slight variability between the symbols for ʒ 'drachm' and ℥ 'ounce' which the scribes sometimes confuse.[11]

4 Conclusions

The analysis uncovered considerable differences in the frequencies of words that get abbreviated in Latin and Middle English. When it comes to function words in Latin and especially Middle English, the scribes have an individual inventory of forms that they use, and the types can vary considerably from manuscript to manuscript. When it comes to Latin names of ingredients, apothecaries' weights and some verbs, the scribes reproduce the forms found in the exemplar, and the frequencies are very close in all manuscripts, with only minor variation.

---

[11] In text 1, the Trinity scribe writes ℥ where others use ʒ. In text 8, Sloane 2320, 3566 and Trinity omit a drachm symbol in different recipes. In text 10, Sloane 2320 uses a ʒ, where other four always write the expanded word.

There are a number of possible explanations. First, these results fit in well with what is known of the provenance of one manuscript. According to a paleographical identification by A. I. Doyle (1956: 299–301), Boston MS 19 is the "litill booke of pheesyk" commissioned by John Paston (II) and copied by the free-lance scribe William Ebesham, from whom two letters survive as part of the Paston letters. Neither the patron nor the scribe, who was based in Westminster and whose hand can be found in several different kinds of texts, were medical practitioners. Thus, the results of the current study support the connection of the Voigts–Sloane Group to the commercial manuscript production in London or its metropolitan area. The scribes reproduced the forms in their exemplar very faithfully, in order to avoid mistakes with specialised subject matter like medicine, whereas with function words they had their own personal habits of abbreviating.

Another possibility is that the results can be explained by generic differences, and the regularity of abbreviations is a feature of recipes. Apothecaries' weights, in which the greatest amount of regularity were discovered, are a feature particular to this genre. Moreover, also names of ingredients in the Latin recipes display considerable regularity, and they do so in a context where their accurate transmission would have been of utmost importance. The effect of the genre is a point that needs to be investigated and clarified by another study, using a bigger corpus.

To conclude, despite being based on a small corpus, this paper managed to uncover interesting variation in the patterns of abbreviation in medical manuscripts. In addition to the work by Wright (2000, 2012) as well as Rogos (2012) or quantitative graphetic studies like Varila (2014), it demonstrates that there is definite potential in using a linguistic approach to abbreviations. For example, it would be possible to construct scribal profiles using a diagnostic set of graphemes, as proposed by Varila, for scribal identification (169). The pattern of abbreviating would presumably differ considerably with manuscripts copied in the medical faculties of universities. Furthermore, the XML encoding used in this study has the added advantage that it allows the encoding of both the symbol and the expansion. This combination creates interesting possibilities for further corpus-based studies, and it is quite likely that future studies will uncover additional regularities. All in all, manuscript abbreviations are data that we have barely begun to mine and which promise rich yields in the future.

References

Alonso-Almeida, Francisco. 2008. The Middle English medical charm: Register, genre and text type variables. *Neuphilologische Mitteilungen* 109(1). 9–38.

Alonso-Almeida, Francisco & Ruth Carroll. 2004. A new proposal for the classification of Middle English medical texts. In Alicia Rodríguez Álvarez & Francisco Alonso-Almeida (eds.), *Voices on the past: Studies in Old and Middle English language and literature*, 21–33. A Coruña: NetBiblo.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 2000. *Longman grammar of spoken and written English*, 3rd impression. Harlow: Longman/Pearson Education.

Carroll, Ruth. 1999. The Middle English recipe as a text-type. *Neuphilologische Mitteilungen* 100(1). 27–42.

Carroll, Ruth. 2004. Middle English recipes: Vernacularisation of a text-type. In Irma Taavitsainen & Päivi Pahta (eds.), *Medical and scientific writing in late medieval English*, 174–196. Cambridge: Cambridge University Press.

Doyle, A. I. 1956. The work of a late-fifteenth century English scribe, William Ebesham. *Bulletin of the John Rylands Library* 39(2). 298–325.

Driscoll, M. J. 2006. Levels of transcription. In Lou Burnard, Katherine O'Brien O'Keeffe & John Unsworth (eds.), *Electronic textual editing*, 254–261. New York: The Modern Language Association of America.

Driscoll, M. J. 2009. Marking up abbreviations in Old Norse-Icelandic manuscripts. In Maria Grazia Saibene & Marina Buzzoni (eds.), *Medieval texts, contemporary media: The art and science of editing in the digital age*, 13–34. Pavia: Ibis.

eTK = A digital resource based on Lynn Thorndike and Pearl Kibre, *A catalogue of incipits of mediaeval scientific writings in Latin* (Cambridge, MA: Mediaeval Academy, 1963) and supplements. http://cctr1.umkc.edu/search (accessed 3 June 2016).

eVK2 = An expanded and revised version of Linda Ehrsam Voigts and Patricia Deery Kurtz, *Scientific and medical writings in Old and Middle English: An electronic reference* CD (Ann Arbor, MI: University of Michigan Press, 2000). http://cctr1.umkc.edu/search (accessed 3 June 2016).

Görlach, Manfred. 1992. Text-types and language history: The cookery recipe. In Matti Rissanen, Ossi Ihalainen, Terttu Nevalainen & Irma Taavitsainen (eds.), *History of Englishes: New methods and Interpretations in historical linguistics*, 736–761. Berlin & New York: Mouton de Gruyter.

Green, Monica. 2000. The possibilities of literacy and the limits of reading: Women and the gendering of medical literacy. In Monica H. Green (ed.), *Women's healthcare in the medieval west: Texts and contexts*, 1–76. Aldershot: Ashgate.

Grund, Peter. 2003. The golden formulas: Genre conventions of alchemical recipes in the Middle English period. *Neuphilologische Mitteilungen* 104(4). 455–475.

Grund, Peter. 2006. Manuscripts as sources for linguistic research: A methodological case study based on the Mirror of Lights. *Journal of English Linguistics* 34(2). 105–125.

Honkapohja, Alpo. 2011. Multilingualism in Trinity College Cambridge manuscript O.1.77. In Jacek Fisiak & Magdalena Bator (eds.), *Foreign influences on medieval English*, 25–45. Frankfurt am Main: Peter Lang.

Honkapohja, Alpo. 2013. Manuscript abbreviations in Latin and English: History, typologies and how to tackle them in encoding. In Anneli Meurman-Solin & Jukka Tyrkkö (eds.), *Principles and practices for the digital editing and annotation of diachronic data* (Studies in Variation, Contacts and Change in English 14). Helsinki: Research Unit for the Study of Variation, Contacts and Change in English.

Honkapohja, Alpo. 2017. *Alchemy, medicine, and commercial book production: A codicological and linguistic study of the Voigts-Sloane manuscript group*. Turnhout: Brepols.

Jones, Claire. 1998. Formula and formulation: ‘Efficacy phrases’ in medieval English medical manuscripts. *Neuphilologische Mitteilungen* 99(2). 199–216.

Jones, Claire. 2004. Discourse communities and medical texts. In Irma Taavitsainen & Päivi Pahta (eds.), *Medical and scientific writing in late medieval English*, 23–36. Cambridge: Cambridge University Press.

Lass, Roger. 2004. Ut custodiant litteras: Editions, corpora and witnesshood. In Marina Dossena & Roger Lass (eds.), *Methods and data in English historical dialectology*, 21–48. Bern: Peter Lang.

Mäkinen, Martti. 2004. Herbal recipes and recipes in herbals – intertextuality in early English medical writing. In Irma Taavitsainen & Päivi Pahta (eds.), *Medical and scientific writing in late medieval English*, 144–173. Cambridge: Cambridge University Press.

Mäkinen, Martti. 2006. *Between herbals et alia: Intertextuality in medieval English herbals*. Helsinki: University of Helsinki.

Matheson, Lister M. 2005. *Médecin sans frontières*? The European dissemination of John of Burgundy’s plague treatise. *ANQ* 18(3). 19–30.

Matheson, Lister M. 2006. John of Burgundy: Treatises on plague. In M. Teresa Tavormina (ed.), *Sex, aging, & death in a medieval medical compendium: Trinity College Cambridge R.14.52, its texts, language, and scribe*, 569–602. Tempe, AZ: Arizona Center for Medieval and Renaissance Studies.

Middle English Dictionary. http://quod.lib.umich.edu/m/med/ (accessed 3 June 2016).

OED = Oxford English Dictionary. http://dictionary.oed.com/ (accessed 3 June 2016).

Pahta, Päivi. 2004. Code-switching in medieval medical writing. In Irma Taavitsainen & Päivi Pahta (eds.), *Medical and scientific writing in late medieval English*, 73–99. Cambridge: Cambridge University Press.

Pahta, Päivi & Irma Taavitsainen. 2004. Vernacularisation of scientific and medical writing in its sociohistorical context. In Irma Taavitsainen & Päivi Pahta (eds.), *Medical and scientific writing in late medieval English*, 1–18. Cambridge: Cambridge University Press.

Robinson, Peter M. W. & Elizabeth Solopova. 2006. *Guidelines for transcription of the manuscripts of the Wife of Bath’s Prologue*. http://petermwrobinson.me.uk/canterburytalesproject.com/pubs/op1-transguide.pdf (accessed 3 June 2016).

Rogos, Justyna. 2011. On the pitfalls of interpretation: Latin abbreviations in MSS of the Man of Law’s tale. In Jacek Fisiak & Magdalena Bator (eds.), *Foreign influences on medieval English*, 47–54. Frankfurt am Main: Peter Lang.

Rogos, Justyna. 2012. Isles of systemacity in the sea of prodigality? Non-alphabetic elements in manuscripts of Chaucer’s “Man of Law’s Tale”. IsLE Richard M. Hogg Prize, honourable mention. http://www.isle-linguistics.org/resources/rogos2012.pdf (accessed 28 March 2016).

Schendl, Herbert & Laura Wright. 2011. Code-switching in early English: Historical background and methodological and theoretical issues. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English*, 15–45. Berlin & Boston: Mouton de Gruyter.

Stannard, Jerry. 1982. Rezeptliteratur as Fachliteratur. In William Eamon (ed.), *Studies on medieval Fachliteratur: Proceedings of the special session on Medieval Fachliteratur of the sixteenth International Congress on Medieval Studies, Kalamazoo, Michigan (U.S.A.), May 10, 1981*, 59–63. Brussels: Omirel.

Stenroos, Merja. 2007. Sampling and annotation in the Middle English Grammar Project. In Anneli Meurman-Solin & Arja Nurmi (eds.), *Annotating variation and change* (Studies in variation, contacts and change in English 1). Helsinki: Research Unit for the Study of Variation, Contacts and Change in English.

Taavitsainen, Irma. 2001. Middle English recipes: Genre characteristics, text type features and underlying traditions of writing. *Journal of Historical Pragmatics* 2(1). 85–113.

Tavormina, M. Teresa. 2005. The Twenty-Jordan series: An illustrated Middle English uroscopy rext. *ANQ* 18(3). 43–67.

TEI consortium (eds.). 2007–. TEI P5: Guidelines for Electronic Text Encoding and Interchange. http://www.tei-c.org/Guidelines/ (accessed 28 March 2016).

Varila, Mari-Liisa. 2014. Graphetic variation within one scribal hand as evidence on manuscript production. In Matti Peikola & Merja Kytö (eds.), *Manuscript studies and codicology: Theory and practice*. [Special issue]. *Studia Neophilologica* 86(supplement 1). 157–170.

Voigts, Linda Ehrsam. 1989. The character of the carecter: Ambiguous sigils in scientific and medical texts. In A. J. Minnis (ed.), *Latin and vernacular: Studies in late-medieval texts and manuscripts*, 91–109. Woodbridge: D.S. Brewer.

Voigts, Linda Ehrsam. 1990. The 'Sloane Group': Related scientific and medical manuscripts from the fifteenth century in the Sloane collection. *The British Library Journal* 16. 26–57.

Voigts, Linda Ehrsam. 1996. What's the word? Bilingualism in late-medieval England. *Speculum* 71(4). 813–826.

Voigts, Linda Ehrsam. 2003. The master of the king's stillatories. In Jenny Stratford (ed.), *The Lancastrian court: Proceedings of the 2001 Harlaxton symposium*, 233–252. Donington: Shaun Tyas.

Wright, Laura. 2000. Bills, accounts, inventories: Everyday trilingual activities in the business world of later medieval England. In D. A. Trotter (ed.), *Multilingualism in later medieval Britain*, 149–156. Woodbridge: D.S. Brewer.

Wright, Laura. 2012. On variation and change in London medieval mixed-language business documents. In Merja Stenroos, Martti Mäkinen & Inge Særheim (eds.), *Language contact and development around the North Sea*, 99–115. Amsterdam & Philadelphia: John Benjamins.

IV Contexts

Joanna Kopaczyk

13 Administrative multilingualism on the page in early modern Poland: In search of a framework for written code-switching

1 Introduction

At the beginning of his book on bilingualism in the ancient Roman world, J. N. Adams (2003) notices three important issues concerning the study of past multilingual communication. Firstly, even though linguists are aware of the fact that historical societies came in contact with each other regardless of linguistic differences, linguistics as a field, and – by extension – historical linguistics, has largely treated monolingualism as the norm and the primary context for textual study.[1] Secondly, in studies of historical language contact, priority is often given to lexical borrowing at the cost of other potential outcomes of multilingual communicative situations, for example code-switching.[2] Finally, if multilingualism is studied at all, it is often approached from the perspective of a particular dominant language group, but other languages belonging to the same communicative space and time are underplayed. In the context of this volume, the first and second issues are readily addressed. The third observation opens up interesting research avenues into areas of historical language contact which have not received much scholarly attention so far. This chapter draws on seventeenth-century manuscript material stored in the National Archives in Cracow, comprising municipal administrative and legal records pertaining to Scottish immigrants to early modern Poland-Lithuania. Thus, the languages coming into contact are Latin, Polish and Scots, a combination rarely encountered in the study of historical multilingualism. Excerpts from these records will serve

---

Acknowledgement: Joanna Kopaczyk would like to acknowledge the support of Adam Mickiewicz University in Poznań and the University of Edinburgh where she was affiliated while preparing this chapter for publication.

---

[1] In addition to shorter publications, edited volumes on historical code-switching in Europe include Braunmüller and Ferraresi (2003) and Schendl and Wright (2011b). Together with the present collection, these publications testify to a very welcome change of paradigm in historical linguistics.

[2] For a wider discussion of the differences and overlaps between borrowing and code-switching, see Myers-Scotton (1992), and for a more recent overview, see Gardner-Chloros (2009).

DOI 10.1515/9781501504945-013

to illustrate a pertinent problem in analysing multilingualism of the past: how does code-switching work on the written page and how can we approach it systematically?

The study starts with a discussion of code-switching in written texts as first-hand sources of information on the code-switching of the past (section 2). Sebba (2012: 2) points out that the written text works as a multi-layered semiotic object, which is why it is crucial to recognize the semiotic plains and linguistic structural levels on which code-switching can take place in writing. My intention is to marry this structural approach with the recent developments in the pragmatic analysis of visual aspects of text on the page (Machan 2011; Carroll et al. 2013; also mentioned by Sebba (2012: 5–6)). Before I explore how this can be done for the documents pertaining to Scots in Poland, I present the socio-historical background of early modern multilingualism in Poland-Lithuania (section 3), with special attention paid to the role of Latin in the context of urban centres and administration, taking the then capital of the kingdom, Cracow, as an example. In fact, Cracow is one of the many places where Scottish immigrants settled in the early modern period, and one of the few where records still survive to testify to this fascinating historical episode. Against this background, I analyse excerpts from the records which illustrate the interplay of the three languages on the page and in the structural frame of the message (section 4). In the concluding part, I sketch out a framework for the analysis of multilingual written texts from the code-switching perspective and suggest how it could capture the communicative choices presented in the case studies.

2 Code-switching in historical texts: Multilingualism on the page

It seems that present-day code-switching in bilinguals "remains largely stigmatized" (Bullock and Toribio 2009: 10); however it is hard to argue for the same attitude towards code-switching in the past. In some genres, switching between the codes was an expected part of the generic repertoire or indicated functional characteristics of the languages involved. In her work on account-keeping in medieval London, Wright (2001) comes to the conclusion that code-switching was predictable and expected. Talking about her earlier analyses of medieval prognostications, Taavitsainen (2017) noticed that code-switching can be regarded as a dimension of genre vernacularization when the scribe puts down a new genre in both the traditional code, e.g. Latin, and in the vernacular, with no change of hand or other visible cues. McLelland's (2004) study of an important seventeenth-century treatise on the German language shows that the combination of Latin and German fulfils discoursal and pragmatic functions, such as citing authorities, quoting speech (in the vernacular), summarizing, structuring discourse, or glossing. Nurmi and Pahta (2013: 197) distinguish three main functions of switching codes in legal genres, following Mattila (2006: 140–141): the rhetorical function which "rais[es] the level of the text and add[s] splendour to it", the display of the writer's status through his familiarity with the prestigious code, and reliance on Latin legal heritage in concepts and terminology, just as in scientific texts. In this way, code-switching enables more comprehensive communication and reflects cultural norms, literacy levels, genre features and communicative practices of the day (for overviews of current scholarship and discussion on code-switching as a historical phenomenon see Schendl 2010; Schendl and Wright 2011a).

When writing is the medium of communication, the book or the page becomes a communicating artefact (Jucker and Pahta 2011: 3). Apart from the extralinguistic aspects that multilingual written texts communicate, such as the economic or cultural value of the texts, their age, provenance, etc., they may contain various degrees of language mixing on multiple linguistic levels, from macro to micro. Code-switching may occur at document boundaries (Voigts 1996: 821–823) within the same manuscript. Languages may mix on the level of discourse moves, e.g. reported speech, as well as on the level of individual lexical items. In early printed medical texts, e.g. herbals, italics or spaces were used by printers to mark out codes, especially on the lexical level (McConchie 2011). Going deeper into structural levels, code-switching may also happen at morpheme boundaries and even in the choice of the orthographic representation. It is necessary to consider these complex patterns of language choices and relate them to the immediate visual impression that is created during the reading experience. Some of these potential types of written code-switching emerge in the seventeenth-century documents written by Cracow clerks in relation to various matters concerning the Scottish diaspora in the city. The languages in question are Latin, Polish and, to some extent, Scots, and while the general context is administrative and formal, the texts reveal individual communicative situations behind the written text.

3 Multilingual and multi-ethnic Poland-Lithuania in the seventeenth century

3.1 A historical overview

As Norman Davies (1996: 554) puts it, "[t]he realm of the last Jagiellons was absolutely the largest state in Europe" and experienced "its 'Golden Age' during the sixteenth and early seventeenth centuries". In 1618, the Commonwealth of the Kingdom of Poland and the Grand Duchy of Lithuania had around 11 million inhabitants[3] from various ethnic backgrounds with a long historical presence: Polish, Lithuanian, Ruthenian, German, Jewish, to name the most important groups (Samsonowicz 1993: 469). The country offered favourable conditions for immigration because of its religious tolerance, trading and military opportunities and family networks linking it with the rest of Europe. These were also the most important reasons why "[f]rom the later sixteenth century Scots moved into Poland in unprecedented numbers" (Devine 2003: 10).[4] In his thorough assessment of that Scottish migration, Bajer (2012: 343) estimates that "at its peak, in the 1640s, the community consisted of [...] perhaps as many as 5000–7000 individuals", which was about half of the Scottish population in Ulster at that time. The Scots added to the multilingual character of the places where they settled, and must have found ways of communicating in a country with no common linguistic background. For an early modern Scottish merchant, scholar or mercenary, Latin, the *lingua franca* of contemporary Europe, must have been the first resort.[5]

3.2 Latin-Polish bilingualism

Latin was the third language in the Polish-Lithuanian Commonwealth, after Polish and Ruthenian (Mikołajczak 1998: 184). It was the second language of the most influential section of the Polish society – the nobility (*szlachta*), which constituted about 10% of the whole population, a greater proportion than elsewhere in Europe. Latin was a "sociolect" of that group and a token of identity, if not snobbery, as evidenced by the famous saying from the period: *Eques Polonus sum, Latine loquor* 'I'm a Polish nobleman, I speak Latin', whatever its quality.

The influence of Latin on Polish had been present since medieval times but it really took off in the sixteenth century. On the wave of the Counter-Reformation, the Jesuits had an enormous impact on Latin in education, setting up colleges where Latin was present as a subject and a vehicle of instruction and debate. In the first half of the seventeenth century around 300 books were published in Poland in Latin, against 700 in Polish, but in the second half the proportion was equal (Mikołajczak 1998: 206). Examples of bilingual Latin-Polish texts from the period include official speeches and orations, correspondence among the nobility (even up to 25% of the text) and people of letters (Axer and Axerowa 2004), diaries and travelogues of the Polish gentry and nobility travelling in Europe. People coming to Poland also remarked on the presence of Latin because even simple people they encountered could communicate in that language. As the "last bastion of Christianity", Poland was keen to preserve the Latin-language ideology, *latinitas*, the common ground of European culture and identity.

---

[3] By comparison, the population of England in 1651 was about 5.2 million (Wrigley and Schofield 1981: 208–209).

[4] Unlike their Polish counterparts, British historians have only recently started to address this early emigration from Scotland to central Europe and countries around the Baltic in earnest (for a historiographic review, see Worthington 2015). Bajer (2012) is the most meticulous study of the Scottish migration to Poland-Lithuania, especially in the context of religious communities, and Kowalski (2010b) offers an examination of the Scottish diaspora in Cracow.

[5] Adding to the 1496 legislation requiring landowners to send their eldest sons to grammar schools, the Reformation in Scotland brought about a proposal to establish schools in every parish. Todd (2002: 59–62) reports that there were 405 parish schools, mainly in Lowland Scotland, before 1633, catering even for children of farm labourers. It is relatively safe to assume that an early modern Scot travelling through Europe would have at least rudimentary Latin.

Among the genres produced in Latin to various degrees, we find documents and institutional records in urban contexts. Szczuczko (2004) remarks that even in Gdańsk, a city where the German language was a default means of communication since the thirteenth century, Latin enjoyed a stable position in the documents of the local institutions and city council acts and records. Even those written in German were organized into sections with Latin titles and inscriptions (for a wider historical background on Latin in secular administration in medieval Europe, see Bedos-Rezak 1996). Szczuczko concludes that the use of Latin on the one hand gave the records an elevated official character, and on the other carried more precision and was more convenient for legal discussions through links to Roman codes and terminologies.[6]

3.3 Multilingual historical Cracow

Cracow had been the seat of the royal court since 1320[7] and the greatest centre of Latinate culture in this part of Europe. In 1364 the first Polish university, the Academy of Cracow, was set up, while Latin began to oust German in civic institutions.[8] The seat of the Polish kings, Wawel, hosted the biscopric and the cathedral school responsible for the teaching of Latin. Because of its multinational character, the Wawel court used Latin in speech. In the fifteenth century, courtly poets and writers composed their works in Latin and the royal chancery employed c. 110 people to produce various Latin documents for envoys and diplomats going abroad and coming to Poland (Mikołajczak 1998: 86–127). City officials as well as local merchants knew Latin as the local schools catered for people who needed literacy in professional life. Around the university, a Latin Quarter was established, where Latin was used not only as a scholarly language but was also spoken in inns and taverns well into the sixteenth century. As Mikołajczak (1998: 126) notes, the use of Latin for expressive and emotive purposes brought about a specific kind of bilingualism.

---

[6] Latin in the legal context functioned in Poland much longer than in western Europe. Suffice it to say that the Napoleonic laws introduced in the Dukedom of Warsaw in the early nineteenth century were translated from French into Latin (Wołodkiewicz 1999: 251).

[7] The capital was moved to Warsaw in 1596.

[8] Magdeburg Law (*Magdeburger Recht*) was the original legal base for the foundation of many Polish towns, including Cracow.

The records of the Cracow city council (*acta consularia*) bear witness to the presence of immigrants, especially merchants and craftsmen from Germany, Italy, and Scotland (Kiełbicka and Wojas 1994: xxii). The Scottish diaspora had a significant and visible place. John Skene ([1597] 1641: 104), one of the most important scholars of law in Renaissance Scotland, defined a pedlar in his *De verborum significatione*, a late-sixteenth-century legal dictionary:

> Ane pedder, is called an marchand, or creamer, quha bearis ane pack or creame vpon his back, quha are called beirares of the puddill be the *Scottesmen* of the realme of *polonia*, quhairof I saw ane great multitude in the Town of *Cracowia*. Anno. Dom. 1569.
>
> [A pedlar is a name given to a merchant or trader who bears a pack or stall on his back, who are [also] called bearers of the pedlar's bag by the Scotsmen of the realm of Poland, of whom I saw a great multitude in the town of Cracow in the year of God 1569.]

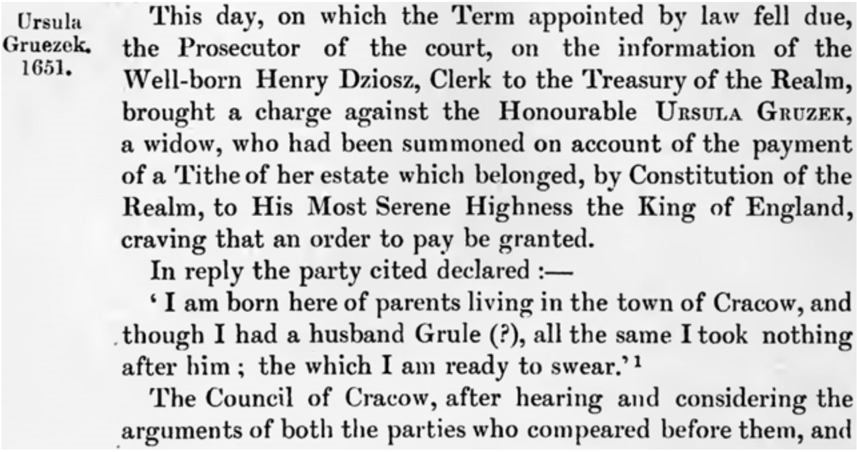
He refers here to his journey throughout central Europe where he must have encountered numerous compatriots, so that this memorable experience made its way to the dictionary definition, and indeed, with a reference to Scots in the Polish capital. Kowalski (2010b: 183–187) lists eighty Scottish merchants admitted to civic rights in Cracow between 1509 and 1655 (see also [2] below), and these are only males who may have had more extended families.

Below, I concentrate on the records related to Scottish immigrants, mostly merchants, who settled and worked in Cracow. I trace document entries pertaining to their quest for civic rights, their *post mortem* inventories and other contexts. The records create interfaces between the language of record-keeping (Latin), the local native language (Polish) and the language of the foreigners (Scots) and illustrate the intricate relationships between the different codes, their functions and their appearance on the manuscript page.

4 Case studies: Scots in Cracow and written code-switching in context

4.1 Editions versus manuscripts

To fully understand the complex interplay of various languages in the historical context one must go back to the original manuscript (or original print). From a historian's point of view, it is often enough to concentrate on the actual contents of a given passage, perhaps making a note of the languages employed. From a linguistic perspective this is highly unsatisfactory. A question can be posed whether the meaning of the passage can be gleaned only from the semantic value of words, perhaps altered in translation anyway, or whether its recovery depends on the full appreciation of the communicative context in its full multilingual complexity. The only available edition, or rather translation, of documents pertaining to Scots in Poland (Steuart 1915) indicates the languages involved in specific passages but does it inconsistently and much editorial intervention is involved in altering both the structure of the text and its appearance on the page (see also Kopaczyk 2013a, 2013b).



Ursula
Gruezek.
1651.

This day, on which the Term appointed by law fell due, the Prosecutor of the court, on the information of the Well-born Henry Dziosz, Clerk to the Treasury of the Realm, brought a charge against the Honourable URSULA GRUZEK, a widow, who had been summoned on account of the payment of a Tithe of her estate which belonged, by Constitution of the Realm, to His Most Serene Highness the King of England, craving that an order to pay be granted.

In reply the party cited declared :—

'I am born here of parents living in the town of Cracow, and though I had a husband Grule (?), all the same I took nothing after him ; the which I am ready to swear.'[1]

The Council of Cracow, after hearing and considering the arguments of both the parties who compeared before them, and

Figure 1: An excerpt from the testimony of Urszula Gruzkowa, translated and edited in Steuart (1915: 82)

The use of spaces, punctuation and capitalization in Fig. 1 relies largely on editorial decisions. The text in the margin seems to be editorial as well. There is a hint at the multilingual conditions in which this text was compiled but it does not inform our understanding of these conditions in a fruitful way. In spite of the patchy referencing and changes in catalogue structure in the Cracow archives, it was possible to locate almost all the documents where switches to Polish were indicated in Steuart (1915), and compare the originals with the translations. It becomes clear that a comprehensive study of historical multilingualism cannot be based on earlier editions.

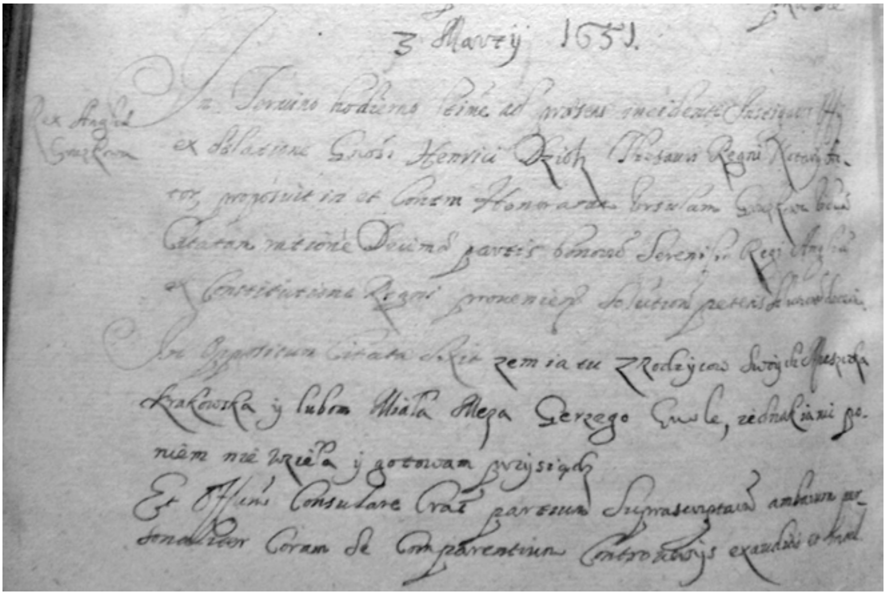


Figure 2: An excerpt from the testimony of Urszula Gruzkowa (1651), National Archives in Cracow, MS 521, p. 2566

Even a quick glance at Fig. 2, which is the original manuscript text of the translation in Fig. 1, reveals differences of linguistic interest between the two versions. First of all, the note in the margin contains a mixture of Latin and Polish to indicate the parties involved in court proceedings: *Rex Angliae* 'the king of England' and *Gruzkowa*, which already contains a switch in the codes. Another thing to notice on the manuscript page is the block arrangement of the Latin text and the continuing testimony in Polish, marked only with a slightly darker shade of ink, as if the scribe stopped and dipped the pen before launching the code-switch and reporting Gruzkowa's words. The switch back to Latin, opening a new paragraph, is marked by a line break. This passage and several other excerpts are discussed in more detail in 4.2, with the focus on the functions and forms of the switches between Latin, Polish and Scots and the appearance of these switches on the page. Since Latin is the matrix language, it is unmarked in the quotations and translations below, Polish is underlined and Scots given in bold.[9] The discussion will lead to more theoretical considerations in the final part of this paper.

---

[9] Admittedly, sometimes the exact language or the boundaries between codes were difficult to pinpoint. In particular, place names and personal names are problematic in this respect, see 4.5 below.

4.2 Code-switched reported speech

Passages of reported speech may frequently be quoted in the language of the actual speaker (for a theoretical background on conversational code-switching and quoting, see Gumperz [1982]; for historical examples, see for instance Pahta and Nurmi [2011: 236–237]),[10] regardless of the matrix code of the whole document. These instances of code-switching on the level of discourse may or may not be signalled by reporting verbs, and may or may not be flagged (Poplack 2004: Section 5.2) by means of specific expressions indicating the awareness that the switch is about to take place. Flagged code-switching to Polish was present in medieval Latin texts. Polish inserts were usually short and nominal, quoted in the nominative and flagged with *vulgariter*, *alias*, *in vulgari*, etc. (Kucała 1974: 91). There is no flagging involved in reporting the Polish words of Urszula Gruzkowa in (1).[11] In fact, the reported speech marker is already mixed: *dixit zem* 'said that+BE' (l. 6). When the text switches to Polish for the oath, the change of language also goes unmarked (ll. 13–14).

(1) In Termino hodierno ltiḿe ad praesens incidente Instigator Offij
ex delatione Gnośi Henrici Dziosz Thesauri Regni Notarij Ac_
tor, proposuit in et Contra Honoratam Ursulam Gruzkowa viduae
Citatam ratione Decimae partis bonoruʒ Serenisii Regi Angliae
5 ex Constitutione Regni prouenienꝝ Solutionʒ petens Soluendam
decimae.
In Oppositum Citata dixit zem ia tu Z Rodzycow Swoych Mieszczka
krakowska y lubom Miała Męza Gerzego **Gurle**, iednak ia nic po_
niem nie wzieła y gotowam przysiądz
Et Officium Consulare Cracꝰ partium Suprascriptaruʒ ambarum per_

---

[10] Culpeper and Kytö (1999) discuss cases of *code-meshing*, or instances of switches between dialects in courtroom discourse.

[11] All the transcriptions were done on the basis of the manuscripts in a diplomatic manner. Abbreviations were not expanded but rendered by means of a symbol closest to their appearance on the page. Line breaks and spaces are respected, spelling was not normalized for any language, punctuation and capitalization are as in the original. I would like to thank Šime Demo (University of Zagreb) for helping me transcribe the Latin handwriting and suggesting some translations. I intended for the translations to be as close to the original as possible, which may cause them to sound stylistically awkward at times. This is especially visible in reported speech switches to first person narration and in legal formulae which do not translate to English in a straighforward way. Any misinterpretations are, of course, my own responsibility.

sonaliter Coram se Comparentium Controuersijs exauditis et intel
[f.2567]
lectis eoq3 attento, q̨d Citata nihil Se bonor3 potest Maritum
Suum recepisse allegat et idipsum Iuramento Se Comprobaturam
offert, Ideo dicit Quatenus Comprobet in eam Iuramenti
Rotham. Iz po Mezu moim żadnych dobr zktorychbym po_
winna krolowi Angielskiemu dziesiecinę dac nie zostało Idq̨
eleuatis duobus Manus dextrae digitis in instanti DPB.
Ex quia Citata iuxta Rotham Suprascriptam Iuramentum
eiusmodi praestitit _ Ideo eandem Liberam ab Actione ꝑńti
ꝑnunciauit et pronunciat. DPB.

(National Archives in Cracow, *Acta Controversiae Cracoviensis*, MS 521, pp. 2566–2567)

[‘In today’s last term at present the official prosecutor, on the information of the well-born Henryk Dziosz, acting as the Clerk of the Royal Treasury, a charge was brought for and against the honorable Urszula Gruzkowa, a widow, summoned [to pay] a tenth part of her goods to His Most Serene Highness the King of England, as ordered by the constitution of the kingdom, seeking the payment of the tithe. In reply, the summoned [woman] said that I am of my parents a merchant of Cracow and indeed I had a husband George **Gurle**, however I took nothing after him and I am ready to swear it. And the Council of Cracow heard and understood the personal testimonies of the above mentioned parties carefully, that the summoned [woman] alleges she received no goods from her husband and she offers to confirm the same under oath. On that account she says insofar confirming in her oath that after my husband no goods are left from which I should pay tithe to the English king. Thus lifting two fingers of the right hand immediately. Then the summoned [woman] the above oath so performed. Therefore, she is pronounced free as regards this prosecution.’]

The reason why Urszula Gruzkowa had to stand before court was because in 1651 the exiled king Charles II persuaded several European kingdoms to back up his cause against Cromwell. The Polish parliament (*Sejm*) agreed to proclaim a property tithe on Charles’s subjects living in the realm of Poland to sponsor the king’s return campaign. “The subsidy of Charles II clearly shows that the British monarch regarded as his subjects not only first-generation migrants but also their descendants to the fourth generation, on both paternal and maternal sides” (Bajer 2012: 84). Some Scots claimed exemption precisely because they had been born in Poland, which raises interesting questions about identity and assimilation.[12] Bajer (2012: 101) estimates that about half of the Cracow Scots escaped the payment. The excerpt in (1) shows that Polish wives and widows of the Scots also faced the payment and had to deny any inheritance if they wanted to escape the tax.

---

[12] It has also been pointed out by Sebba (2012: 11) that switching codes may serve to construct the identity of the writer and the reader, and – by extension – of the person(s) whose words are being reported.

Another context for personal testimony, this time from the actual Scottish immigrants, is the quest for civic rights. Citizenship gave the right to conduct business so it was sought after by merchants and craftsmen. To gain this status one had to prove their legitimate birth, pay a fee and make an oath (Kowalski 2010b: 40). The social situation in Poland was favourable for foreigners who wanted to trade – Polish nobles could not engage in mercantile activities because it was below their status while the rest of society was dependent on the lords and tied to their land. In (2), Kaspar Hussen testifies to the good birth of William (or Wilhelm) Buchan who wants to acquire the status of a Cracow citizen.

(2) Primus compꝰt testis Fam: Casparus **Hussen** Scotus
Ciuis Clepardienꝰ sub iuramento errectis ad sacra duo:
bus manus dextrae digitis praestito, his formalibus
verbis recognouit.
ZEznawam o czijm mam wiadomoscz ij czom
[f.1242]
od wielu ludzi slijszal isz niebosczyk **Thomas Buchan** Z Sco
tiey ze dworu niedaleko miasta **Aberdijn**a Z Holzbietą mal
zonką swą wstanie sw malzenskim mieszkaiącz splodzil Sy:
now dwu braczi rodzonijch czo iest tego czo **Wilchelm**a na kto:
rego powod czo swiadecztwo zeznawam ij **Alexandr**a ktorij czo:
ras w tijm Roku ninieijszijm od kilkunastu niedziel w Sie:
dmiogrodzkieij Ziemi w miesczie Sebinowie czterij mile od
Bialograda vmarł, czo tak bijcz zeznawam.
Secundus testis Fam: Joannes **Armandt** Ciuis Cracꝰ Sub [. . .]

(National Archives in Cracow, *Advocatialia Cracoviensia*,
MS 226, p. 1241–1242)

['The first appeared witness the well-famed Caspar **Hussen**, a Scot, citizen of Kleparz [who] under oath raised two fingers of his right hand, and uttered these formal words. I testify to what I have knowledge of and what I have heard from many people that the deceased **Thomas Buchan** from Scotland from a manor not far from **Aberdeen** city[13] living with his wife Elizabeth in holy matrimony begat two sons, born brothers, namely this **Wilhelm** to whose call I testify and **Alexander** who now died this year several weeks ago in the land of Siedmiogrod in the town of Sebinow four miles from Białograd, that this be true I testify. The second witness the well-famed John **Armandt**, citizen of Cracow under [. . .]']

---

[13] For a discussion of Polish inflections appended to foreign roots, as in *Scotiey* and *Aberdijna*, see 4.4 below.

Here the reported speech passage is in Polish again, which raises the question as to what language the Scots were actually speaking before the court. This fascinating issue should be pursued further but for now let us concentrate on the visual cues for code-switching. This time the code-switched passage is placed in a separate paragraph and visually distinguished from the Latin, unlike the Polish testimony in (1). There is even quite pronounced capitalization at the beginning of the passage (l. 5). Apart from these visual cues there is no semantic flagging of the code-switch. The reporting verb *recognouit* 'recognized, confirmed' only signals the forthcoming speech passage but no reference is made to the change of language. Excerpt (3) contains a reported speech passage introduced by the reporting Latin verb *addere* 'to add' and an immediate switch to Polish with the relative pronoun *iż* 'that' (l. 11). A similar situation is visible in (4) but without any specific reporting verb, just the reference to the oath and the symbolic non-verbal behaviour (l. 11–13).

4.3 Other discourse moves

Reported speech passages usually comprise longer, coherent passages of discourse. However, it seems that these were not the only contexts in which code-switching on the level of discourse could take place in the Cracow records. Consider (3), an extract from the proceedings concerning a will which had been left behind by a Scottish merchant William Clelland (cf. Bajer 2012: 138).

(3) Primus testis Famatus Jacobus **Karmichel** Mercator
Ciuis Cracouien^{9} praestito de more per eleuationem duorum versus
caelum manus dextrea digittorum corporali Juramento ad Jnterroga_
torium primum punctum tale. Jezeli wie ze tęn **Clilent** był Fa_
5 ctorem y Sługą tego P Wilchelma Erdysza y Pana Jana **Kil**a
Miesczan Brodzkich, y iezeli tęn Erdysz co tu teraz stawa iest
prawdziwym Panem Nieboszczyka **Klilant**a. Respondit affirma
tiue. Ad idum Jeżeli wie ze te Towary y Rzeczy ktore są Ar
restowane v P **Blakal**a y P. **Vsiert**owey są własne P **Gwilchelm**a
10 Erdysza y **Kil**ego, y jezeli ten Pąn Erdysz y **Kilo** są spolni Towa_
rzysze handlu. Respondit similiter affirmatiue, eo addito iz przed
smiercią immediate powiadał przedemną tęn Nieb. **Klilent** zem ia Panu
memu Erdyszowi nic niewinien ani tesz on mnie, y to tesz wiem ze tego
Klilanta do wyciągania długow y Creditow, y uroznych handlach P
15 Erdysz poszyłał.

(National Archives in Cracow, *Acta Controversiae Cracoviensis*, MS 521, p. 2748)

> ['The first witness the well-famed Jacob **Karmichel** merchant, citizen of Cracow, customarily gives by the elevation of two fingers of his right hand to heaven, a bodily oath to the first point of the interrogation as follows. Whether he knows that this **Clilent**[14] was a factor and servant of that Mr Wilhelm Erdysz and Mr Jan **Kilo** Citizens of Brody, and whether this Erdysz that stands here now is a true master of the late **Klilant**. He replies affirmatively. Further, whether he knows that the merchandise and goods which are arrested at Mr **Blakal's** and Mrs **Usiert's** are owned by Mr **Gwilchelm** Erdysz and **Kilo**, and whether this Mr Erdysz and **Kilo** are together partners in trade. Similarly, he replies affirmatively adding that immediately before death that late **Klilent** told me that I to my Mr Erdysz owe nothing and neither he owes me, and I also know that Mr Erdysz sent this **Klilant** to extract debts and credits and in various trade deals.']

Here the matrix language of the proceedings is Latin but the actual question to the witness is reported in Polish (l. 4–7). Then, for indirect reporting of the affirmative answer the scribe switches to Latin (l. 7–8) which comes across as a metacode of the communicative situation and adds authority to the reported proceedings. The second question is rendered in Polish again (l. 8–11), with the same affirmative answer reported indirectly in Latin. Interestingly, the witness wanted to add something to his testimony, and that discourse move is also coded in Latin (l. 11).

The extent to which Latin served in early modern Polish records to render important discourse moves can be best appreciated in the last line of (4) below. The Scots testify that they do not know of any other compatriots living in Cracow (again, this is the context of the Polish subsidy to Charles II), and for the verb of confirming (*comprobare*) the scribe switches to Latin again. It is difficult to ascertain whether this verb simply belongs to the legal discourse of the recordings or whether it is quoted verbatim from the testimony of the Scots, who most probably showed some skill in Latin.[15]

From their research into multilingual practices in the *Lampeter Corpus*, Nurmi and Pahta (2013) draw the conclusion that "multilingual practices in the domain of law are highly conventionalised": switches to Latin happen at lexical and phrasal boundaries, within established formulaic conventions. The switches found in the corpus are usually short and aphoristic, like legal terms and maxims (2013: 193–197) and they may also come from outside the legal register, e.g. temporal adverbials like *Anno* (2013: 199–200; cf. Kopaczyk 2013a: 293–303).

---

[14] For a discussion of various degrees of polonization of personal names, see 4.5 below.

[15] In his study on Patrick Gordon of Auchleuchries, a Scottish soldier and traveller through central Europe, Kowalski says that "Gordon spoke Latin with Poles [. . .] In the first years of his time on Polish soil this was a necessity as Latin was the only foreign language he had brought with him from home" (2010a: 31). Other Scots may have exhibited a similar linguistic training.

In the present material, the situation is different because the language of the official administrative proceedings is not the vernacular but, indeed, Latin. Interestingly, the switches to Polish are conventionalized only to some extent (e.g. when the witness gives an oath) but sometimes the switches back to Latin may be used for important discourse moves or for shorter switches which are not that clearly distinguished from borrowings.

4.4 Code-switching across morpheme boundaries

In highly inflected languages such as Latin and Polish, code-switching may happen across morpheme boundaries, for example when adding a foreign inflectional ending to the native base (see Wright 1992, 1994, 1995). Kucała (1974: 91) reports that in medieval Latin-language documents there were cases where Polish words were given Latin inflections, depending on the grammatical context. In the Cracow records, the only context where that occurs in the Latin passages is onomastic in nature. The reverse happens to Latin and Scots onomastic material in the Polish passages. In (2), in the reported speech passage, we have already seen Polish inflections attached to place names: the Latinate term *Scotia* inflects for the Polish locative and *Aberdijn* 'Aberdeen' inflects for the genitive (ll. 6–7). Arguably, the personal names *Wilchelm* and *Alexander* had already been nativized in Polish, which is evidenced by Polish case endings (ll. 9–10). There are interesting differences between first names and surnames and their cross-linguistic behaviour which I address below.

The Polish reported speech passage in (3) contains Latin roots which serve as the base for Polish derivation and inflection, with a somewhat dubious status between a borrowing and a code-switch. For instance, the Latin verb *arrestare* 'to seize' gives the adjectival plural *arrestowane* (ll. 8–9), even though the root is still spelled according to the Latin orthography. The meaning of this verb seems to be restricted in seventeenth-century Polish to legal contexts (SXVII, *aresztować* 'to arrest'). Similarly, the inflected plural *Creditow* (l. 14) does not seem to belong to the general vocabulary of Polish at this time yet (it is missing from the list of over 23,000 entries in SXVII) but functions perfectly in legal and economic discourse as an adaptation of the Latin *creditum* 'debt'. Indeed, the word appears in a binomial *długow y Creditow*, where it repeats the meaning of the Polish długi 'debts'. In (4) below, the final line contains the Latin root *specificare* 'to specify', to which the Polish first person plural inflection is added: *specificuiemy* (l. 14). There is a problem in deciding whether the root of a given lexical item had already been nativized, which would mean, of course, that no code-switching is taking place at the morpheme boundary. However, in writing, the choice of the orthographic system may be indicative of the linguistic status of a given morpheme. It is uncontroversial to assume that the scribes were versed in the orthographic systems associated with particular languages. What is more, their proficiency in a given language could be limited solely to the written domain. So the choice of Latin spelling within a Polish passage is indicative of a transitional status of a particular lexical item, with code-switching on the level of morphology to adjust that item to the grammar of the recipient language.

4.5 Code-switching in personal names

First names, due to their etymology, seem to be more susceptible to latinization in terms of spelling and inflection than surnames. Consider (4), the list of Scots living in Cracow at the time of the taxation in favour of the exiled king, Charles II:

(4) Qui quidem Jacobus **Karmichel** et Jacobus **Blakal**
hos infrascriptos, vt ꝑote' Andream **Frazer** Vilhelmum **Tory**
Georgium **Kruiksang** Thomam **Dyxon** Petrum **Englis**
Joannem **Rets** Jacobum **Ciamer** Abrahamum **Frede**, Richardum
Gordon Abrahamum **Vsierd**, Jacobum **Grick**, Jacobum **Korbet**
et **Gaspar**um **Hunter**, absentiam nonnullorum ab vrbe vt ꝑote'
Jacobum **Ciamer** et Richardi **Gordon** tum inoꝑiaṁ Jacobi **Grik**,
deinde quod Abraham **Vsierd** Juramentum super Liquidoñe
Substantiae suae in Lubownia praestitit et Quietōnem de repo
sita Decima obtinuit denunciantes nominarunt et specificarunt
affirmantes neminem amplius praeter hos hic Cracouiae reperiri. Quod
et Juramento eleuatis ad Sydera duobus manus dextrae digitis
in eam Rotham. Iz tu wKrakowie niemasz więcey Szotow
[f.2632]
nad tych ktorych specificuiemy ani o tym wiemy ze
by ich więcey było comprobarunt.

(National Archives in Cracow, *Acta Controversiae Cracoviensis*,
MS 521, p. 2631–2632)

['Indeed, Jacob **Karmichel** and Jacob **Blakal**, these underwritten, and then Andrew **Frazer**, William **Tory**, George **Kruiksang**, Thom **Dyxon**, Peter **Englis**, John **Rets**, Jacob **Ciamer**, Abraham **Frede**, Richard **Gordon**, Abraham **Vsierd**, Jacob **Grick**, Jacob **Korbet** and Gaspar **Hunter**, some absent from the city, and then Jacob **Ciamer** and Richard **Gordon**, both absent, Jacob **Grik**, thereafter Abraham **Vsierd** under oath in Lubownia performed on

> the question of the tithe to be obtained declare and specify that no more than those are found in Cracow. And under oath they raised two fingers of the right hand to heaven to swear that here in Cracow you have no more Scots than these we specify and approve that we know not if there are more.']

Some first names are listed in their Latin versions, e.g. the inflected *Petrum* for *Peter* (l. 3) or *Joannem* for *John* (l. 4). Some 'international' first names, transmitted to European languages from and through Latin, for instance *Jacob* or *Abraham*, are inflected for case in the Latin passage, too. However, if we assume that some of these names still retain the native Scots base, and only attach the Latin inflection for case, we can treat them as instances of code-switching on the morpheme level, as in *Richardum* (l. 4). This practice goes hand in hand with other instances of macaronic writing where "Latin syntax dictates the macaronic structure [and] the relevant case and number [are] suffixed [...] regardless of etymology" (Wright 1992: 764–765). In the Polish passages the names of the Scots get Polish inflections, as in (2) (*Wilchelma*, *Alexandra*, ll. 9–10) or (3) (*Gwilchelma*, l. 9.), and they are spelled according to the orthographic conventions of Polish, and – most probably – according to how the scribe heard the name.

The situation is even more interesting when it comes to the surnames. In the Polish passages, the spellings are polonized to a large extent. Consider (3) and the various attempts to spell the surname *Clelland* as *Clilent*, *Klilant* and *Klilent* (ll. 4, 7, 12), or *Blakal* for *Blackhall* (l. 9; also (4) l. 1). Some Scottish surnames in the Latin passage are spelled according to Polish conventions, e.g. *Ciamer* for *Chalmers* (ll. 4, 7), *Vsierd* for *Usher* or *Urquhart*[16] (ll. 5, 8), and *Korbet* for *Corbett* (l. 5). This could count as code-switching on the level of orthography, reflecting the local pronunciation of the foreign names. There are attempts to inflect the Scots' surnames but the scribes' choices are not always consistent. In (3) above, *John Kilgour* becomes *Jan Kilo* (cf. Bajer 2012: 579) and the genitive singular is *Jana Kila* (l. 5) or *Kilego* (l. 10). The ending *-a* is a noun inflection for the masculine genitive singular while *-ego* is taken from the masculine adjectival paradigm (cf. Długosz-Kurczabowa and Dubisz 2001: 202, 244), most probably because of the vocalic element at the end of the root.[17] Otherwise, the surnames inflect according to the most relevant morpho-phonological pattern and they even produce patronymics for the wives, e.g. *Vsiertowey* (genitive singular feminine, (3), l. 9) from *Vsierd* (cf. (4), ll. 5, 8). What is interesting, however, is that surnames never inflect for case in the Latin matrix text, which creates a structural contrast with the inflected first names of the same people. This concerns Scottish surnames throughout excerpt (4) but also the Polish surnames *Dziosz* and *Gruzkowa* in (1), ll. 2–3.

---

[16] Kowalski (2010b: 316) suggests that this is the Polish version of *Usher*, while Bajer (2012: 587) interprets it as *Urquhart*.

[17] A similar uncertainty as to the appropriate inflectional endings appears in texts written in Poland by the Scots themselves (Kopaczyk 2013a: 311).

5 Historical code-switching on the page: In search of an analytical framework

As pointed out in the introduction, research into historical code-switching has been developing since the early 1990s, but it is still very much centred on case studies (for an overview, see Schendl 2012: 29–30). Sebba (2012: 1–2) notes that the study of code-switching in written texts does not have its own framework but extends and adapts frameworks for spoken code-switching, which may not be the best approach. In search of an analytical framework, and with a view to complementing the existing approaches, it is helpful to think about code-switching in a written text in terms of structural linguistic levels and the visual aspects of the handwritten or printed page. By making reference to a general structural map and considering the role of visual cues (Fig. 3), we look at the text as a wholesale multimodal communicative event, and we are able to interrogate the levels in a systematic way: starting from the most general and easy to pinpoint (the level of genre), to those integrated with the matrix language (the level of orthography). This approach has been inspired by the idea of the manuscript as a communicative object (Pahta and Jucker 2011) and the visual pragmatic approach to written texts (Machan 2011; Carroll et al. 2013).

According to this map, switching codes may occur and be visually marked on all levels of linguistic analysis. Some manuscripts are, in fact, collections compiled out of various genres, as is the case with the so-called *Sammelhandschriften* (Kurz and Voigts 2011; Schipor 2014), commonplace books (e.g. Lockridge 2009) or *silva rerum* (Roszak 2004), so the switching of codes may already start at the macrogenre level (or the intertextual level, as Voigts [1996] has it). One would expect visual cues to mark out individual genres composed in different codes, for instance a line or page break or a change of scribal hand. Discourse is understood in this approach as a level of linguistic complexity which extends beyond a sentence. Code-switching on the level of discourse will happen when larger, functionally and textually coherent passages in a different code are inserted, especially for a particular communicative function, for instance for reported speech passages, lists, instructions, formulaic greetings or closing paragraphs. These may also be marked by means of visual cues on the page. On the clausal level, code-switching in written texts resembles the intersentential code-switching

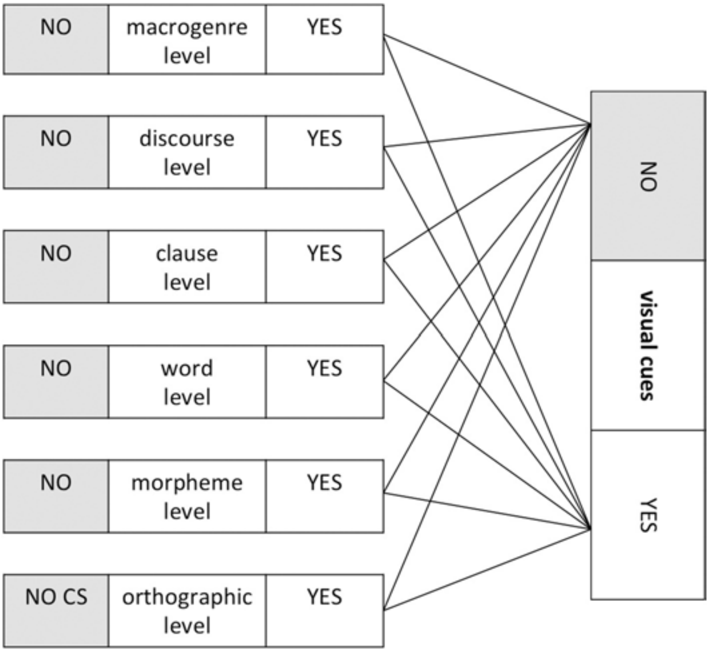

NO
macrogenre level
YES
NO
discourse level
YES
NO
clause level
YES
NO
word level
YES
NO
morpheme level
YES
NO CS
orthographic level
YES
NO
visual cues
YES

Figure 3: The interaction between visual cues and code-switching on different linguistic levels

in speech (Myers-Scotton 1993), whereby the switch occurs between complete clausal units.[18] Again, by means of, for instance, changing the hand or the ink colour, the scribe may choose to indicate such a switch on the page or, alternatively, may pass from one language to another across sentences without indicating it in any way. Intrasentential code-switching in speech happens on the level of a single phrase or word. This is also frequent in historical English written texts, as shown by Schendl for Old English charters (2012: 32–36) and sermons (2012: 37–39) or by Pahta (2004, 2011) for medieval and early modern medical texts. Within words, code-switching appears at morpheme boundaries. In her research on business records from medieval London, Wright (1992, 1994, 1995, 2001) draws attention to a mixture of linguistic codes: Latin, French and English, with scribal preferences for roots and inflections from a particular code. If a scribe chooses to underline, capitalize, abbreviate or otherwise indicate the difference in codes within a single lexical unit, we encounter a different communicative situation than when the same information is not conveyed by any visual cues. Finally, I would like to suggest that the switch of codes may also happen on the level of orthography. If we are dealing with a text written in a specific matrix language, and encounter vocabulary items which are spelled according to orthographic conventions taken from another code, we have a case of orthographic code-switching. It may also be indicated by visual cues on the page. McLelland (2004: 507) notes the relationship between font choice and language in the seventeenth-century linguistic treatise: German in Gothic type, Latin – in Roman font (see also Sebba 2012: 10; Angermeyer 2012; Kaislaniemi 2017). Since the script itself may be associated with a specific language, when orthographic conventions change, the script (the hand, the font) may also change, marking code-switching on the orthographic level with a visual cue.

---

[18] For clarity's sake, I do not introduce overlaps between structural levels. However, one should be aware of such possibilities, for instance when a single sentence functions at the same time as a coherent message on a discourse level.

If applied to the code-switching practice in the examined Cracow records, the model could be interpreted as in Fig. 4.

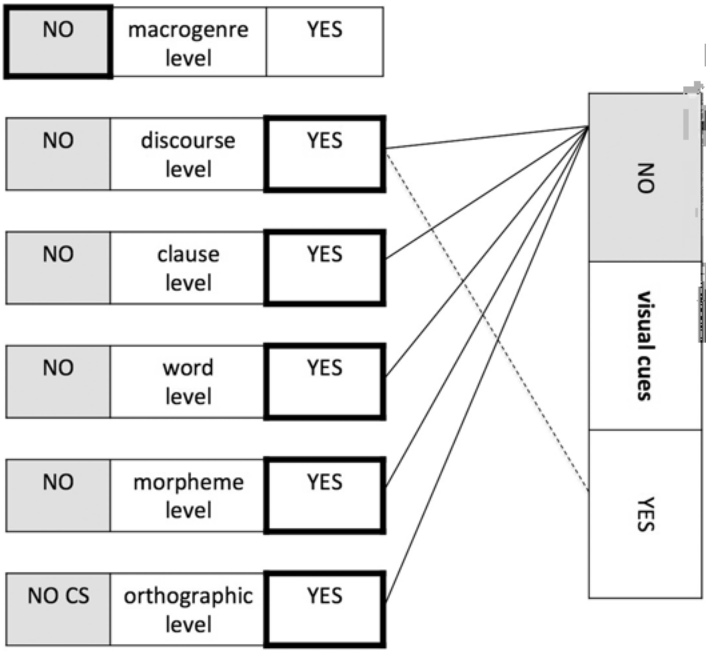


Figure 4: The interaction between visual cues and code-switching on different linguistic levels in documents pertaining to the Scots in early modern Cracow (National Archives in Cracow, MS 226, MS 521)

In the documents there is no switching on the macrogenre level because the handwritten codices in which the records were entered are not composed of different genres. One could, of course, look for stylistic differences between individual entries but the purpose and social function of these texts remains the same. This takes us to the discourse level, where code-switching between Latin and Polish does take place, for instance when moving from exposition or narration to reported speech. It may be unmarked visually (as in (1) l. 6), or marked with a visual cue, e.g. a line break (as in (2) l. 5).[19] Márton (2014) noticed such code-switching also in Hungarian witchcraft trials, where Latin activated a particular discourse type. On a clause level, code-switching also takes place in the Cracow records but in the fragments under analysis there were no visual cues to indicate it. Similarly on the level of individual lexical items, where the scribe simply continues writing in the same line, in the same hand, in the same ink, even though the code has been switched. On a morpheme level the situation seems the same, as code-switched inflectional endings are simply attached to the base without any visual cue. However, if one considers the ambiguous nature of abbreviated endings, which may suggest potential interpretations in more than one code, e.g. Latin and Scots, the possibility of having a visual cue for the switch of codes becomes apparent (cf. Wright 1992, 1995). In the present material no such cases were found but this is indeed a very likely phenomenon. Abbreviations can be positioned between morphology and the orthographic level. Here, I would like to argue, code-switching can also take place because a lexical item from one language becomes respelled according to the orthographic rules of another, without becoming a borrowing. In the Cracow records such respellings are mostly visible in proper names but also in lexical items which are on their way to becoming fully functional Latin borrowings in Polish, but have not yet been vernacularized in terms of spelling.

This model, or a map of structural complexity of a written text which contains various types of code-switching, is open to further refinement. The structural levels may be reconsidered, as more fine-grained or more general distinctions may be applicable to other texts, communicative contexts and other ways of text transmission in the written mode. Other factors to be implemented in this model should include the inventory and taxonomy of the visual cues, which may be diverse and may correlate with specific types of switches on specific structural levels. Ink and colour change, letter size, line breaks, the choice to use the margin, orientation, hands, pointers, etc. may all be used to mark code-switching, but it is equally significant if a scribe or printer does not resort to any of these cues. This is a promising research perspective for a better understanding of the interplay of multiple languages on a page, the multilingual proficiency of authors and audiences as well as the suitability of specific codes to fulfil certain functions with or without overt visual signals that they do so.

---

[19] Lower frequency of visual cues in this function is indicated by a dotted line in the figure.

Primary sources

Libri Iuris Civilis Cracoviensis, National Archives in Cracow

Acta Consularia Cracoviensia, National Archives in Cracow

SXVII = Elektroniczny Słownik Języka Polskiego XVII i XVIII Wieku [The electronic dictionary of 17th and 18th-century Polish], Polish Academy of Sciences, 1999–. http://sxvii.pl/ (accessed 10 January 2016)

References

Adams, J. N. 2003. *Bilingualism and the Latin language*. Cambridge: Cambridge University Press.

Angermeyer, Philipp Sebastian. 2012. Bilingualism meets digraphia: Script alternation and script hybridity in Russian-American writing and beyond. In Mark Sebba, Shahrzad Mahootian & Carla Jonsson (eds.), *Language mixing and code-switching in writing: Approaches to mixed-language written discourse*, 255–272. London: Routledge.

Axer, Jerzy & Anna Axerowa. 2004. Lektura tekstu dwujęzycznego: Przykłady [Reading a bilingual text: Examples]. In Jerzy Axer (ed.), *Łacina jako język elit* [Latin as the language of the elites], 161–176. Warszawa: OBTA / Wydawnictwo DiG.

Bajer, Peter P. 2012. *Scots in the Polish-Lithuanian Commonwealth, 16th–18th centuries*. Leiden: Brill.

Bedos-Rezak, Brigitte. 1996. Secular administration. In F. A. C. Mantello & A. G. Rigg (eds.), *Medieval Latin: An introduction and bibliographical guide*, 195–229. Washington, DC: The Catholic University of America Press.

Braunmüller, Kurt & Gisela Ferraresi (eds.). 2003. *Aspects of multilingualism in European language history*. Amsterdam & Philadelphia, PA: John Benjamins.

Bullock, Barbara E. & Almeida Jacqueline Toribio. 2009. Themes in the study of code-switching. In Barbara E. Bullock & Almeida Jacqueline Toribio (eds.), *The Cambridge handbook of linguistic code-switching*, 1–17. Cambridge: Cambridge University Press.

Carroll, Ruth, Matti Peikola, Hanna Salmi, Mari-Liisa Varila, Janne Skaffari & Risto Hiltunen. 2013. Pragmatics on the page. *European Journal of English Studies* 17(1). 54–71.

Culpeper, Jonathan & Merja Kytö. 1999. Investigating non-standard language in a corpus of Early Modern English dialogues: Methodological considerations and problems. In Irma Taavitsainen, Gunnel Melchers & Päivi Pahta (eds.), *Writing in non-standard English*, 171–187. Amsterdam & Philadelphia, PA: John Benjamins.

Davies, Norman. 1996. *Europe: A history*. Oxford: Oxford University Press.

Devine, T. M. 2003. *Scotland's empire and the shaping of the Americas 1600–1815*. Washington, DC: Smithsonian.

Długosz-Kurczabowa, Krystyna & Stanisław Dubisz. 2001. *Gramatyka historyczna języka polskiego* [A historical grammar of Polish]. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.

Gardner-Chloros, Penelope. 2009. *Code-switching*. Cambridge: Cambridge University Press.

Gumperz, John J. 1982. *Discourse strategies*. Cambridge: Cambridge University Press.

Jucker, Andreas H. & Päivi Pahta. 2011. Communicating manuscripts: Authors, scribes, readers, listeners and communicating characters. In Päivi Pahta & Andreas H. Jucker (eds.), *Communicating early English manuscripts*, 3–10. Cambridge: Cambridge University Press.

Kaislaniemi, Samuli. 2017. Code-switching, script-switching and typeface-switching in Early Modern English manuscript letters and printed tracts. In Matti Peikola, Aleksi Mäkilähde, Hanna Salmi, Mari-Liisa Varila & Janne Skaffari (eds.), *Verbal and visual communication in early English texts*, 165–200. Utrecht: Brepols.

Kiełbicka, Aniela & Zbigniew Wojas (eds.). 1994. *Libri Iuris Civilis Cracoviensis (1573–1611)*. Cracow: Secesja.

Kopaczyk, Joanna. 2013a. Code-switching in the records of a Scottish brotherhood in early modern Poland-Lithuania. *Poznań Studies in Contemporary Linguistics* 49(3). 281–319.

Kopaczyk, Joanna. 2013b. Scottish papers in early modern Poland: A new resource for historical linguists. In Aniela Korzeniowska & Izabela Szymańska (eds.), *Scotland in Europe / Europe in Scotland: Links, dialogues, analogies*, 85–98. Warsaw: Semper.

Kowalski, Waldemar. 2010a. Patrick Gordon in his own words: A soldier, a Scot, a Catholic. In Paul Dukes (ed.), *The Patrick Gordon diary and its context*. [Special Issue]. *Journal of Irish and Scottish Studies* 3(2). 19–38.

Kowalski, Waldemar. 2010b. *Wielka imigracja. Szkoci w Krakowie i Małopolsce w XVI – pierwszej połowie XVII wieku* [The great immigration: The Scots in Cracow and Małopolska in the sixteenth–early seventeenth century]. Kielce: Wydawnictwo UHP im. Jana Kochanowskiego.

Kucała, Marian. 1974. Łacińska fleksja rzeczowników polskich w tekstach średniowiecznych [Latin inflections of Polish nouns in medieval texts]. In Jerzy Kuryłowicz (ed.), *Studia indoeuropejskie* [Indo-European studies], 91–96. Wrocław: Ossolineum.

Kurtz, Patricia Deery & Linda Ehrsam Voigts. 2011. The significance of now-dispersed Bute 13: A mixed-language scientific manuscript. In Päivi Pahta & Andreas H. Jucker (eds.), *Communicating early English manuscripts*, 38–54. Cambridge: Cambridge University Press.

Lockridge, Kenneth. 2009. Individual literacy in commonplace books. In Harvey J. Graff, Alison MacKinnon, Bengt Sandin & Ian Winchester (eds.), *Understanding literacy in its historical contexts: Socio-cultural history and the legacy of Egil Johansson*, 236–242. Lund: Nordic Academic Press.

Machan, Tim William. 2011. The visual pragmatics of code-switching in late Middle English literature. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English*, 303–333. Berlin & Boston: Mouton de Gruyter.

Márton, Petykó. 2014. Discursive construction of witchcraft as a community and witch as an identity in 18th century Hungarian witchcraft trials. Paper presented at the 1st Poznań Historical Sociopragmatics Symposium, Adam Mickiewicz University in Poznań, 15–16 May.

Mattila, Heikki E. S. 2006. *Comparative legal linguistics*. Abingdon: Ashgate.

McConchie, Roderick W. 2011. Compounds and code-switching: Compositorial practice in William Turner's *Libellus de re Herbaria Novvs*, 1538. In Jacob Thaisen & Hanna Rutkowska (eds.), *Scribes, printers and the accidentals of their texts*, 177–190. Frankfurt am Main: Peter Lang.

McLelland, Nicola. 2004. A historical study of codeswitching in writing: German and Latin in Schottelius' *Ausführliche Arbeit von der Teutschen HaubtSprache* (1663). *International Journal of Bilingualism* 8(4). 499–523.

Mikołajczak, Aleksander W. 1998. *Łacina w kulturze polskiej* [Latin in Polish culture]. Wrocław: Wydawnictwo Dolnośląskie.

Myers-Scotton, Carol. 1992. Comparing codeswitching and borrowing. In Carol M. Eastman (ed.), *Codeswitching*, 19–39. Clevedon: Multilingual Matters.

Myers-Scotton, Carol. 1993. *Social motivations for code-switching: Evidence from Africa*. Oxford: Oxford University Press.

Nurmi, Arja & Päivi Pahta. 2013. Multilingual practices in the language of the law: Evidence from the *Lampeter Corpus*. In Jukka Tyrkkö, Olga Timofeeva & Maria Salenius (eds.), *Ex Philologia Lux: Essays in honour of Leena Kahlas-Tarkka*, 187–205. Helsinki: Société Néophilologique.

Pahta, Päivi. 2004. Code-switching in medieval medical writing. In Irma Taavitsainen & Päivi Pahta (eds.), *Medical and scientific writing in late medieval English*, 73–99. Cambridge: Cambridge University Press.

Pahta, Päivi. 2011. Code-switching in Early Modern English medical writing. In Irma Taavitsainen & Päivi Pahta (eds.), *Medical writing in Early Modern English*, 115–132. Cambridge: Cambridge University Press.

Pahta, Päivi & Andreas H. Jucker (eds.). 2011. *Communicating early English manuscripts*. Cambridge: Cambridge University Press.

Pahta, Päivi. & Arja Nurmi. 2011. Multilingual discourse in the domain of religion in medieval and early modern England: A corpus approach to research on historical code-switching. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English*, 219–251. Berlin & Boston: Mouton de Gruyter.

Poplack, Shana. 2004. Code-switching. In Ulrich Ammon, Norbert Dittmar & Klaus J. Mattheier (eds.), *Sociolinguistics: An international handbook of language and society*, 2nd edn, 589–596. Berlin & New York: Walter de Gruyter.

Roszak, Stanisław. 2004. *Archiwa sarmackiej pamięci: Funkcje i znaczenie rękopiśmiennych ksiąg* silva rerum *w kulturze Rzeczpospolitej XVIII wieku* [The archives of Sarmatian memory: Functions and significance of the handwritten *silva rerum* books in the culture of 18th-century Poland]. Toruń: Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika.

Samsonowicz, Henryk. 1993. Grupy etniczne w Polsce XV wieku [Ethnic groups in 15th-century Poland]. In Jacek Chrobaczyński, Andrzej Jureczka & Michał Śliwa (eds.), *Ojczyzna bliższa i dalsza* [The closer and further homeland], 461–469. Kraków: Secesja.

Schendl, Herbert. 2010. Multilingualism, code-switching, and language contact in historical sociolinguistics. In Juan Manuel Hernández-Campoy & Juan Camilo Conde-Silvestre (eds.), *The handbook of historical sociolinguistics*, 520–533. Oxford: Blackwell.

Schendl, Herbert. 2012. Literacy, multilingualism and code-switching in early English written texts. In Mark Sebba, Shahrzad Mahootian & Carla Jonsson (eds.), *Language mixing and code-switching in writing: Approaches to mixed-language written discourse*, 27–43. London: Routledge.

Schendl, Herbert & Laura Wright. 2011a. Code switching in early English: Historical background and methodological and theoretical issues. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English*, 15–45. Berlin & Boston: Mouton de Gruyter.

Schendl, Herbert & Laura Wright (eds.) 2011b. *Code-switching in early English*. Berlin & Boston: Mouton de Gruyter.

Schipor, Delia. 2014. Code selection and code-switching in the Beverley Town Cartulary. Paper presented at the Historical Code-Switching: The Next Step Symposium, University of Tampere, 11–13 June.

Sebba, Mark. 2012. Researching and theorising multilingual texts. In Mark Sebba, Shahrzad Mahootian & Carla Jonsson (eds.), *Language mixing and code-switching in writing: Approaches to mixed-language written discourse*, 1–26. London: Routledge.

Skene, John Sir, Lord Curriehill. [1597] 1641. *De verborum significatione: The exposition of the termes and difficill wordes conteined in the foure buiks of Regiam Maiestatem*… London: Printed by E.G. (Available from Early English Books Online.)

Steuart, A. F. (ed.). 1915. *Papers relating to the Scots in Poland (1576–1793)*. Edinburgh: Scottish History Society.

Szczuczko, Witold. 2004. Łacina w kancelarii Rady Miasta Gdańska [Latin in the chancery of the City of Gdańsk]. In Jerzy Axer (ed.), *Łacina jako język elit* [Latin as the language of the elites], 131–137. Warszawa: OBTA / Wydawnictwo DiG.

Taavitsainen, Irma. 2017. Meaning-making practices in the history of medical English: A socio-pragmatic approach. *Journal of Historical Pragmatics* 18(2).

Todd, Margo. 2002. *The culture of Protestantism in early modern Scotland*. New Haven, CT & London: Yale University Press.

Voigts, Linda. 1996. What's the word? Bilingualism in late-medieval England. *Speculum* 71(4). 813–826.

Wołodkiewicz, Witold. 1999. Łacina prawnicza w Polsce [Legal Latin in Poland]. In Jerzy Axer (ed.), *Łacina w Polsce. Pomniki prawa doby Renesansu w Europie Środkowo-Wschodniej* [Latin in Poland: The monuments of Renaissance law in central-eastern Europe], 251–253. Warszawa: OBTA.

Worthington, David. 2015. "Unfinished work and damaged materials": Historians and the Scots in the Commonwealth of Poland-Lithuania (1569–1795). *Immigrants & Minorities* 20(10). 1–19.

Wright, Laura. 1992. Macaronic writing in a London archive, 1380–1480. In Matti Rissanen, Ossi Ihalainen, Terttu Nevalainen & Irma Taavitsainen (eds.), *History of Englishes: New methods and interpretations in historical linguistics*, 762–770. Berlin & New York: Mouton de Gruyter.

Wright, Laura. 1994. Early modern London business English. In Dieter Kastovsky (ed.), *Studies in Early Modern English*, 449–465. Berlin & New York: Mouton de Gruyter.

Wright, Laura. 1995. A hypothesis on the structure of macaronic business writing. In Jacek Fisiak (ed.), *Medieval dialectology*, 309–321. Berlin & New York: Mouton de Gruyter.

Wright, Laura. 2001. Models of language mixing: Code-switching versus semicommunication in medieval Latin and Middle English accounts. In Dieter Kastovsky & Arthur Mettinger (eds.), *Language contact in the history of English*. Frankfurt: Peter Lang.

Wrigley, E. A. and R. S. Schofield. 1981. *The population history of England 1541–1871: A reconstruction*. Cambridge: Cambridge University Press.

Aleksi Mäkilähde

14 Approaching the functions of historical code-switching: The case of solidarity

1 Introduction

Previous research on historical code-switching has focused to a considerable extent on the functions of code-switching (henceforth CS), but there have been only few attempts to define the functions themselves in formal terms by applying a specific theoretical framework.[1] It has been shown, for example, that establishing what could be termed *solidarity* is an important function of CS, yet few researchers have endeavoured to define this concept explicitly. This gives rise to certain problems. For example, it is difficult to evaluate an analysis of the functions of CS if the classificatory principles are not stated explicitly. It is also impossible to borrow a framework if its application and form are unclear. Thirdly, it is hard to compare different studies if the various functions are not explicitly defined. Fourthly, and perhaps most importantly, the relationships between different functions, such as their relative hierarchies, do not become visible if the similarities and differences between the functions are not made clear.[2]

This chapter adopts a methodological/theoretical perspective, discussing the definition of a single function: *solidarity*. Although some studies of historical CS use the term *solidarity* to refer to a specific function, the term and concept, as noted above, are not themselves always clearly defined. Furthermore, many studies have approached the functions of CS only in terms of pre-theoretical

---

Acknowledgment: I wish to thank Esa Itkonen and Ville Leppänen for their helpful comments on various drafts of this chapter, the anonymous reviewers and the editors of the volume as well as Risto Hiltunen for their suggestions, and Ellen Valle for her assistance in polishing the language and style of the final version.

---

[1] For the purposes of the present chapter, I use the term *historical CS* to refer to the occurrence of multiple languages in texts written in languages which are either earlier stages of modern ones (e.g. Old English) or extinct (e.g. Gothic). The decision to focus on historical CS is in accordance with the scope of the present volume; it does not imply that historical CS as a phenomenon is essentially different from 'modern' CS. In fact, the discussion presented in this chapter should be applicable to both historical and modern contexts.

[2] *Function* itself is a notoriously unclear concept. It can be understood here as the purpose that an act serves or a goal that it attempts to achieve (see e.g. Itkonen 1983: 31, 156–157; 2013; Leech 1983: 13–14, 48; Givón 2013). An in-depth discussion is, however, beyond the scope of the present chapter.

DOI 10.1515/9781501504945-014

concepts, meaning that they have described those functions less formally.[3] Since it nevertheless seems clear that different researchers are talking about essentially the same function but in different terms, an informal analysis of this pre-theoretical concept is needed first. In other words, we need to understand what it is that we are talking about. A formal definition or explication of this function can then be attempted with the aid of a specific theoretical framework and previously defined theoretical concepts. Finally, this formalisation needs to be applied to individual instances of CS in order to test its adequacy. In this chapter, I deal with these issues by proposing answers to the following research questions:

1) How has the function here referred to as *solidarity* been understood in previous studies of historical CS?
2) How might *solidarity* be defined within a particular theoretical framework, and what problems does this entail?
3) How does the definition affect our analysis of the functions of CS, and what kind of problems are posed by concrete examples?

I begin by briefly discussing the concept of solidarity in general, and by elaborating on the distinction between theoretical and pre-theoretical, followed by a comparison of (implicit) definitions of solidarity or corresponding concepts in previous research on historical CS. I then suggest a possible definition of this function based on pragmatic theory, making use in particular of concepts derived from Brown and Levinson's (1987) *politeness theory*, and illustrate the definition with examples drawn from early modern school drama.[4] I conclude with a discussion of certain problems involved in the proposed definition, with suggestions of ways to solve these problems, and with an evaluation of the approach.

2 Preliminaries concerning the concept of solidarity

Before moving on to the role of solidarity in previous studies, some of the central concepts introduced above need to be discussed briefly in more detail.

---

[3] For the distinction between theoretical and pre-theoretical (or atheoretical) and its relevance to the study of language, see Itkonen (1978: 103–104, 144–145, 198–208; 2008: 285, 292–294).

[4] For the present purpose, I interpret *pragmatic theory* broadly as a theory of acts or actions. This is consistent with a view of pragmatics as the study of the use of language in context (cf. e.g. Leech 1983: 1–5; Itkonen 2008; Huang 2014: 2). It should also be noted that I see no reason to limit the scope of pragmatics to purely linguistic acts (cf. Searle 1969: 17).

First of all, it is necessary to clarify what it is that we are in fact looking for in searching for instances of solidarity in previous studies. A good starting point is offered by Brown and Gilman's (1960) classic treatment of the semantics of T/V pronouns, in which they propose that the use of these pronouns is related to two social dimensions: power and solidarity. According to them, *solidarity* relates to the level of intimacy or (perceived) similarity between speaker (henceforth S) and addressee (henceforth H); they further note that "[t]he similarities that matter seem to be those that make for like-mindedness or similar behavior dispositions" such as "political membership, family, religion, profession, sex, and birthplace" (Brown and Gilman 1960: 258).[5]

As mentioned above, not all studies of historical CS use a specific term to refer to the concept which Brown and Gilman (and others) call solidarity. This is the case in particular with data-driven approaches, where the functions of CS may be explained in great detail but perhaps only in terms of pre-theoretical concepts. In essence, these are concepts that the analysts (as well as the interactants themselves) seem to be aware of, even if they cannot define them explicitly.[6] Since it is nevertheless quite clear that the various researchers are talking about more or less the same phenomena, it should be possible to analyse these shared concepts in order to arrive at an explicit definition and convert them into theoretical concepts. This is similar to explication as practised in philosophy. To sum up, I am here interested initially in those cases where the function of CS is analysable (or has been analysed by others) as establishing, indicating or maintaining 'like-mindedness' between S and H, in other words some form of in-group membership between them. This initial conception of solidarity will then be modified according to my analysis of previous studies.

3 Solidarity in earlier studies of historical CS

To my knowledge, the most in-depth discussion of solidarity and historical CS is provided by Adams (2003). Since he also uses the precise term, I will begin by quoting him at some length:

---

[5] Solidarity is of course an important concept in many different fields, including sociology (see e.g. Hechter 1987; Scott 2015, s.v. "social solidarity", "solidarism") and philosophy (see e.g. Feinberg 1968; Rorty 1989). A discussion of these other fields is beyond the scope of the present chapter. For a broader sociolinguistic perspective, see e.g. Labov's (1972: 36–40) analysis of diphthong centralisation in the English of Martha's Vineyard, although he does not use the term *solidarity*.

[6] A similar case can be made for such concepts as *politeness*, *insult*, or *threat*, to name just a few. Although people normally have an idea of what these concepts mean at the pre-theoretical level, providing a good definition for any of them is not simple.

> First, there is code-switching as a means of establishing a relationship with an addressee, as for example a sense of solidarity, or a position of dominance or aloofness. A speaker may diverge occasionally from the addressee's preferred language as a means of distancing or partial exclusion [. . .], or he may converge with the addressee by adopting as far as he is able the first language of the second person, though his ability to do so may be limited. Convergence can also be described as accommodation. Divergence and convergence by code-switching may represent language use as symbolising power on the one hand or solidarity on the other. Or again a mixed discourse, where speaker and interlocutor are equally bilingual, can reinforce the feeling of a shared, mixed identity or culture. This type of code-switching may often be interpreted as a mark of solidarity or as a demonstration of shared membership of an in-group. (Adams 2003: 301)

Adams does not provide any explicit definition of *solidarity*, but since he refers to Brown and Gilman's discussion of solidarity and power (Adams 2003: 301, fn. 12) and mentions in-group membership, it may be assumed that his definition would be close to theirs. He also emphasises the solidarity vs. power opposition by referring to the opposition of convergence vs. divergence. These terms have apparently been borrowed from communication accommodation theory (CAT) (see e.g. Giles, Coupland, and Coupland 1991 for an overview; see Adams 2003: 350–351, 576 for references to his sources), where *convergence* is defined as "a strategy whereby individuals adapt to each other's communicative behaviors in terms of a wide range of linguistic-prosodic-nonverbal features" (Giles, Coupland, and Coupland 1991: 7), while *divergence* is defined as "the way in which speakers accentuate speech and nonverbal differences between themselves and others" (Giles, Coupland, and Coupland 1991: 8). In the CAT model both strategies are types of accommodation, but Adams seems to consider only convergence as a subtype of accommodation, or even to equate these two.

Adams also mentions a situation where S and H share a second language (L2), and are able to indicate that they belong to a group where that shared feature is seen as desirable (in one way or another). This seems to be the case with Roman elite interactions in particular, where, for example, he argues that the use of Greek literary allusions "is a highly artificial form of intimacy or solidarity" (Adams 2003: 312). Similarly, it is said that CS in a particular context "established Cicero's (as well as Varro's) membership of an élite group, comprising those educated Romans with expertise in Greek philosophy; or, to put it another way, it gives Varro and Cicero a shared identity or cultural solidarity" (Adams 2003: 317). In the extended quotation given above, however, he also mentions those cases where S switches to the first language (L1) of H (which is not the L1 of S) in order to express solidarity through convergence. I consider the difference between the two types of situation to be that in the former the interactants are considered equal (to a considerable extent) by both S and H, while in the latter this may not be so. Furthermore, in the former case CS (or more specifically, the knowledge of a mutual L2) is the type of similarity which makes for like-mindedness, while in the latter case the language itself is not necessarily the similarity that matters.[7] In other words, the latter type of switch may imply for example that S is willing to use H's preferred language. S may have many different motives for such a choice, and it is clear that the power-relations between S and H are crucial to understanding these motives. A social superior adopting the language of his subordinates is not the same as a subordinate adopting the language of his superiors.

Although it was noted above that Adams seems to treat convergence as a subtype of accommodation, this is not quite clear either, since his use of the latter term is also somewhat inconsistent. For example, in one context he says that CS at a morpheme-boundary "would seem to imply an attitude of deference or accommodation" (2003: 125), although elsewhere he also discusses accommodation as a form of disparagement or for example admiration (2003: 350–356). He states that accommodation "may be defined as the act of modifying an utterance in some way in deference to the addressee" (2003: 295); he then adds that "accommodation is the act of modifying language use […] to suit the addressee, the referent or the circumstances" (ibid.). In another context he makes the following remarks:

> By 'accommodation' is meant […] the adjustment of speech […] in some way to suit that of the addressee. The employment by Greeks writing Greek at Delos of the Latin style of filiation […] in reference to Romans was described earlier as a form of accommodation on the part of the Greeks. It is also a form of imitation, in that a Latin pattern is deliberately imitated in Greek. But in this case the aim is different. The intention is not to make the Greek sound odd, but to accord Romans their full naming formula in a different language. Imitation thus becomes accommodation, with the intended effect on the reader the defining distinction between the two forms of imitation. The accusative of the honorand in Latin addressed by Romans to another Roman introduces a discordant note, whereas the use of υἱός addressed in Greek by Greeks to Romans is in harmony with the expectations of the Romans. Accommodation of this type expresses solidarity, whereas the other type of imitation has the appearance of a method of distancing writer from addressee slightly. (Adams 2003: 685)

Adams seems to view imitation as a type of activity which sometimes constitutes accommodation and can therefore have the function of expressing solidarity. However, if accommodation refers to modifying the expression to suit H or the circumstances, there is no clear reason why imitation used for purposes of distancing should not be categorised as accommodation as well. Obviously S considers choice of language to 'suit the circumstances' in such a case as well. If we are to take "to suit the addressee" as consisting merely of those instances where the aim is to benefit H, then I cannot see how accommodation differs from what Adams describes as convergence. Furthermore, as noted above, he also discusses cases where S accommodates for example to mock H. In the CAT model, imitation of H's language use would be a case of convergence, while imitation of a speech style alien to H would be divergence. Convergence could still be used to either express solidarity with or to mock H.

---

[7] The former situation represents Brown and Gilman's Equal and Solidary category (1960: 259, Fig. 1), while the latter may represent any of their six categories, including those where a semantic conflict in the use of T/V forms would be predictable (see Brown and Gilman 1960: 258–260). Cf. Labov (1972: 36–40).

To sum up: the definition of solidarity is not completely clarified, and although convergence seems to be a strategy used to express solidarity, the exact relations between solidarity and convergence or accommodation are not made explicit. Yet it seems obvious that the concept evoked here is similar to the one discussed by Brown and Gilman, in that it is connected to in-group membership.

The type of solidarity exemplified in Adams's examples of elite bilingual communication (e.g. Cicero using Greek in his letters to Varro), in other words a situation where the knowledge of a shared L2 creates like-mindedness, is also mentioned in some other historical CS studies.[8] Here I focus on a few relevant cases. Some studies make explicit reference to an in-group/out-group dynamism; Davidson, for example, notes that in *Piers Plowman* CS between Middle English and Latin is used to "restrict and solidify "we" membership" (2003: 476). She also points out that signalling membership in one group may signal non-membership in another, meaning that solidarity and distance are two sides of the same coin (2003: 483, fn. 7). Similarly, according to Schendl (2012: 39; cf. 2011: 87), CS (from Middle English to Latin) in bilingual sermons "was a way to establish or increase the feeling of group-membership in a multilingual educated elite". A case comparable to Roman elite correspondence is found in early modern letter-writing; Pahta and Nurmi (2009: 45) note that "[t]he use of French in the Burneys' letters also reflects the role of French in contemporary society, particularly among the upper classes, among whom the Burneys increasingly associated". In all of these examples the in-group coincides broadly with the educated, literate elite, but there are of course cases where the in-group is defined more exclusively (see e.g. Nurmi and Pahta 2012: 56).

In addition to S switching to the shared L2 (in the abovementioned cases, a high-status language), it is evident that other factors also affect whether solidarity is achieved or not. First, S must be able to follow the norms that dictate, for example, in what contexts and how one should switch in order to achieve solidarity; it is hard to imagine that Cicero would have been able to establish solidarity with Varro by using, for example, incorrect Greek. Second, S's actions must be consistent with S's previous actions and the *line* adopted by S more generally. A *line* can be defined as "a pattern of verbal and nonverbal acts by which [a person] expresses his view of the situation and through this his evaluation of the participants, especially himself" (Goffman [1967] 2005: 5). Davidson (2003: 475–476) discusses an example from the *Canterbury Tales* which illustrates this rather well, namely euphemistic CS into French by the Wife of Bath in speaking of the female genitalia. According to Davidson, "the switch briefly allows the Wife to construct herself as elevated above and delicately tip-toeing around potentially touchy topics; but since the Wife generally speaks unequivocally about sex, the switch discloses her temporary position as French-user; and, in the end, it emphasizes her outgroup status" (2003: 476).

---

[8] For solidarity in Cicero's correspondence in particular, see e.g. Dunkel (2000); Swain (2002); Mäkilähde and Rissanen (2016).

In this case, the Wife tries to emphasise her like-mindedness with the literate, French-speaking groups, including some of her interlocutors, but her CS in this context does not match the line of action she has chosen otherwise (i.e. it does not fit her character). Using Itkonen's (1983: 174–176) terminology, we could say that there is a contradiction between the context (C) and the function of the utterance – or, more precisely, the function of the switch. In the Wife of Bath's case, H resolves the contradiction by noting that H and S are in disagreement over C in an important way. More specifically, H notes that in the context C S has an irrational belief (B), which consists of being able to achieve a certain goal (G) by performing a specific act (A). In other words, H is here confronted with what is from his or her perspective (and in C) an *irrational act* (Itkonen 1983: 65). We could perhaps call this a case of *unsuccessful* solidarity. In Adams's framework, this would probably count as an instance of accommodation (cf. the quotation on Greek at Delos above), and at least within the CAT model it could be analysed as a convergence which receives a negative reaction from Hs, or perhaps in this case the audience (cf. Giles, Coupland, and Coupland 1991: 35–36).

As mentioned above, solidarity is sometimes treated as in opposition to power. This does not mean, however, that a claim for solidarity cannot at the same time be a claim for power. In the case of a medieval bilingual sermon, for example, the Latin parts clearly lend authority to S (and the sermon) since Latin was the language of religion (Diller 1997/1998: 519; Pahta & Nurmi 2011: 230; cf. Brown and Gilman 1960: 255–257). This function would have been central from the perspective of the (mostly) monolingual lay audience, but at the same time the preacher could demonstrate to the other clergy present that he was able to use

CS, more particularly quotations from the Bible, in an appropriate manner.[9] Furthermore, in order to achieve solidarity, S and H would have to be in general agreement in the current context, and the use of identical strategies in different contexts (for example a personal letter vs. a debate) may therefore have different functions. This is similar to the idea that the same sentence can be used (as an utterance) in different contexts to perform different acts.

To sum up: a number of studies of historical CS have identified a function which consists of establishing (creating, upholding, claiming) like-mindedness between S and H, although the function itself may not always be defined explicitly. Some general remarks can be made, however: first, while it is not quite clear whether solidarity occurs only between equals (but see Foley 1997: 314), most of the examples discussed above are of this kind. Second, since it is possible to find cases of unsuccessful solidarity, it is clear that there are other factors, apart from the switch itself, which contribute to the success of this function. Third, solidarity seems to be related in some way to strategies such as convergence and accommodation, and these connections need to be made more explicit.

4 Solidarity in pragmatic terms

4.1 Defining solidarity

Since solidarity has been defined preliminarily by reference to like-mindedness or in-group membership, a logical point of departure for a more explicit definition is *common ground*. This is an important element in, for example, Clark's (1996) overall model, but for him it is a broad concept which can be employed to explain how communication functions in general, and which is created in interaction. Solidarity would be closer to claiming or emphasising common ground than to creating it, since everything S or H does during an interaction contributes to their common ground (see especially Clark 1996: 92–121). To be even more precise, solidarity should be seen as referencing a certain aspect of common ground between S and H, not common ground *in toto*. Furthermore, something more is needed to explain why interactants claim common ground with their interlocutors.

I have chosen Brown and Levinson's (1987) politeness theory as a basis for my definition of solidarity. This is perhaps the best-known model of linguistic politeness (or face-work more generally), and it has also been applied to historical contexts (e.g. Nevala 2004; Hall 2009; Jucker 2011). Here I present only a very brief outline of the theory, focusing on the relevant concepts, and refrain from discussing all the possible problems of the theory in general (for criticism, see e.g. Watts 2003; Culpeper 2011). The most central concept of the theory is *face*, defined as "the public self-image that every [competent adult] member [of a society] wants to claim for himself" (Brown and Levinson 1987: 61). Face is further divided into two interrelated aspects: *positive face* and *negative face*. The former is defined as "the positive consistent self-image or 'personality' [...] claimed by interactants", while the latter is defined as "the basic claim to [...] freedom of action and freedom from imposition" (ibid.). Brown and Levinson further describe both aspects of face in terms of *wants*, positive face being "the want of every member that his wants be desirable to at least some others", and negative face being "the want of every 'competent adult member' that his actions be unimpeded by others" (1987: 62). The former could also be defined as the want to be liked and admired by others, the latter as the want to act freely and autonomously.

---

[9] For views on the audience constellation of bilingual sermons, see e.g. Schendl (2012: 39).

It is seen as beneficial for interactants to maintain each other's face, since this is the only way for an individual participant to maintain his or her own face. Yet it is inevitable that participants will threaten each other's face or their own, and politeness is seen primarily as an instrument for softening these *face-threatening acts* (FTAs). Brown and Levinson (1987: 74–78) introduce three variables which can be used to calculate the severity or weightiness (W) of an FTA: the social distance (D) between S and H, the relative power (P) of S and H, and the absolute ranking (R) of the imposition. Depending on the factor causing the weightiness of an FTA, different politeness strategies may be chosen. The three variables can also be used to explain, for example, why interactants use different strategies when requesting something from their friends (low D and P values) or for instance from their boss (low or moderate D-value, high P-value).[10]

Within politeness theory, solidarity would seem to fit one particular category of positive politeness strategies, in which "[r]edress consists in partially satisfying [H's positive face wants] by communicating that one's own wants (or some of them) are in some respect similar to the addressee's wants" (Brown and Levinson 1987: 101). In other words, like-mindedness would mean a low D-value between S and H. Prima facie the more specific category would be *claiming common ground*, which involves "S claiming 'common ground' with H, by indicating that S and H both belong to some set of persons who share specific wants, including goals and values" (Brown and Levinson 1987: 103). The further qualifications for these strategies, however, only partly match our informal analysis of solidarity

---

[10] Depending on the type of workplace, these values could of course differ (e.g. high D-value and high P-value).

(sections 2 and 3), since "S can claim common perspective with H without necessarily referring to in-group membership" (ibid.). In other words, solidarity (as understood in the previous sections) cannot be equated with this whole group of strategies (i.e. claiming common ground). There is one particular substrategy, *using in-group identity markers*; these include such devices as address forms and CS. Some of the strategies, however, actually imply a low P-value rather than a low D-value; according to Brown and Levinson (1987: 108), "[u]sing such in-group kinds of address forms with imperatives [. . .] indicates that S considers the relative P (power, status difference) between himself and the addressee to be small" and that "even when used to children, it turns a command into a request".

It could well be maintained that this is close enough to the required definition (cf. Foley 1997: 314); however, at least two problems can be detected.[11] First, there is nothing to stop this function (i.e. the politeness strategy) from being successful if S is insincere in his statements. This would make it possible to have a situation where S does not, in fact, want to belong to the same in-group as H, but is merely pretending that he does. In the CAT model (and presumably in Adams's view), this could still be analysed as convergence or accommodation. Second, and related to this, the abovementioned politeness strategies aim at redressing a threat to H's face; as understood here, however, solidarity does not require a previous FTA, and should affect the face of both S and H.

In light of the discussion thus far, the following formal definition for solidarity can be proposed, using the concepts introduced in this section (cf. Mäkilähde 2012: 59):

> An act has the function of establishing solidarity iff (i) it implies a low D-value between S and H through shared membership in a social group, and (ii) at the same time enhances the positive face of both S and H.

Solidarity can also function as a politeness strategy, but it should be noted that in such a situation there needs to be an FTA which is mitigated. It is clear that, for example, in bilingual sermons, CS does not mitigate threats towards H's face, nor does it do so in the other cases discussed above (in section 3). I will now illustrate this definition with some examples.

4.2 Analysis of select examples

The following examples are drawn from the *Orationes* manuscript (Canterbury Cathedral Archives Lit. Ms E41, Canterbury Cathedral Library). It contains drama and speeches performed, and partly composed, by the students of the King's School, Canterbury, between approximately 1665 and 1684.[12] Some of the performances were basically monolingual (in either Latin or English), but most of the texts contain varying amounts of CS (mainly) between English, Latin and Greek. The contemporary audience consisted at least of the teachers, the other students, the Cathedral clergy, and the Dean of the Cathedral. The audience and the actors (the boys) were highly competent in Latin, and at least the boys in the upper forms were also proficient in Greek. In fact, one of the aims of these plays was to demonstrate the boys' proficiency in Classical languages and their knowledge of Classical literature (Mäkilähde 2012 and forthcoming; Mäkilähde, Alho, and Johnson 2016).

---

[11] To clarify the point, these are problems only when trying to account for solidarity by means of existing categories in the model. They are not problems within the confines of the model.

Since we are dealing with drama, we have to take into consideration what Clark (1996: 353–384) calls *layering*, or the fact that actions take place at a number of *layers* (cf. Diller 1997/1998: 507–508). Serious (or 'real', 'actual') actions are situated at the bottom layer (layer 1), while non-serious actions are situated at a layer constructed by the co-participants on top of this (layer 2); sometimes more layers may be added (Clark 1996: 354–355). In our case, this means that we have to account for communication at least between the dramatis personae (layer 2, within the play), and between the play and the audience (layer 1).[13] Due to this complexity, drama texts provide an interesting context for illustrating the concept of solidarity and testing our definition.

Quite often in the plays, CS occurs in conjunction with allusions to or quotations from the Classical authors who formed the core of the curriculum, including Vergil, Terence, Ovid, Homer and Cicero.[14] The following is a typical example of a schoolboy character citing one such author:[15]

---

[12] For the history of the King's School, see Woodruff and Cape (1908). For information on the *Orationes* manuscript, the texts and the context of their performance, see Mäkilähde (2012 and forthcoming); Mäkilähde, Alho, and Johnson (2016).

[13] Cf. Short (1989: 148–149). It should be noted that in the Canterbury performances the actors were also (at least partly) the playwrights, and sometimes the line between actor and character becomes blurred as well.

[14] In for example Genette's (1982: 8) terminology, both would be instances of *intertextuality*. For a discussion of the King's School curriculum, see Woodruff and Cape (1908); Mäkilähde (2012 and forthcoming); Mäkilähde, Alho, and Johnson (2016).

[15] Latin abbreviations have been expanded within square brackets. Abbreviated nomina actorum and Greek ligatures have been silently expanded. Latin parts have been italicised, although in the manuscript they are not visually marked in any way. All nomina actorum have been bolded, and the punctuation marks following them have been normalised as colons. Spelling and punctuation (but not layout) have otherwise been reproduced for the most part as in the original. All translations are by the author.

(1) **Dan:** O Will, those usefull rules, which but lately wee heard, doe soe reflect upon my former negligence that I cannot but cry out with that time-abusing Truant, *O mihi præteritos referat si Iupiter annos* ['oh, if Jupiter gave me back the lost years']! Besides, that Distich of Horace sticks close upon my heart. Namely,

Qui cupit optatam vitæ contingere metam
Multa tulit, fecitq[ue] puer, sudavit, et alsit.

['He who is inclined to reach a pleasant goal in life,
Has, when still a boy, accomplished and endured a great deal, sweated and frozen.'] (*Orationes*, ff. 182r–182v)

The first Latin phrase is derived from the *Aeneid* (Verg. *Aen.* 8,560), while the other two lines come from Horace (Hor. *Ars* 412–413). This example (at layer 1) is analysed in Skaffari and Mäkilähde (2014: 272) as having the function of claiming solidarity, which "is achieved if the audience recognize that these are references to books which they, too, have read, which in turn makes both the actors and the audience members of the same social group, that of educated people". In this sense, example (1) is very similar to some of the examples discussed in section 3 above, and it seems to accord with the definition of solidarity proposed in the previous section. When the dramatis personae are schoolboys, as in example (1), switches such as these can also be analysed as having the function of establishing solidarity at layer 2.[16] However, the language indicating like-mindedness is not always Latin, as can be seen from the following example:

(2) **Eugenius:** [. . .] *Et*, ni fallor, maximè sollicita est, ne expectationibus vestris injuriam faceret. (Intrat Philaster) Quod si – ['And if I am not mistaken, she is anxious about doing injustice to your expectations. (Enter Philaster) But if']

Philaster: *Quod si* – Put on your hat my submissive peice of learned hypocrisie. I marvel what makes you soe humble this morning.

Eugenius: Phil, welcome. (*Orationes*, f. 283v)

Both English and Latin are 'we-codes' for the schoolboy characters (i.e. at layer 2), and differences in the choice of language reflect differences in the particular aspect of D, which is indicated as being low. In other words, the similarity which causes like-mindedness may differ depending on whether the student characters use Latin or English. Latin implies that they have a similar educational background, while English implies friendship outside of the schoolroom as well as partnership in crime, since the students were expected to use only Latin on school premises. However, at layer 1, a switch to English would clearly not have a similar solidarity function between actors and audience.

---

[16] For the use of such quotations in early modern drama in general, see e.g. Ryan (2013); Delabastita and Hoenselaars (2015: 2–3).

Although the use of Latin quotations may often constitute a solidarity strategy between the dramatis personae (layer 2), this is not always the case. In one play, a student named Dan goes around asking other students what they aspire to be in the future, and criticising their choices. In the following example, he has been arguing with Herb, who wants to become a priest:

(3) **Dan:** [. . .] Farwel; take heed of Latine. *Cave ne titubes, mandataq[ue] frangas*. ['Be careful that you do not stumble and break your commission.']

Herb: But, Friend, remember, if you miss of your mark – *Plus fati valet hora benigni Quàm si te Veneris commendet Epistola Marti.* ['A moment of benignant fate is of more avail to you than a letter of recommendation from Venus to Mars.'] Therefore mock on. (*Orationes*, f. 147v)

Dan's line has been taken from Horace (Hor. *Epist.* 1,13,19), while Herb's line is from Juvenal (Iuv. 16, 4–5). Both of them use these quotations to bolster their own argument; this can be interpreted as S enhancing his own positive face by claiming a low D-value with an eminent person. In a sense, this is very similar to solidarity, but note that the ancient authors are not part of the interaction and do not have the role of H in this context; according to the definition proposed above, solidarity takes place between S and H. Rather, these can be classified as instances of claiming authority, similar to the example of using Latin quotations in medieval sermons (discussed in section 3). In a non-confrontational context, they can be analysed as instances of *generalised advice* (see Brown and Levinson 1987: 226).

In conjunction with (3), it also becomes clear why we dismissed the possibility of equating solidarity with claiming common ground in a broader sense. The quotations would be situated in the common ground shared by Dan and Herb, both of whom rely on this shared knowledge in their use of the quotations. Yet this does not achieve solidarity between them; instead, both acts constitute FTAs towards the other character, since the implication in both cases is that H is wrong while S is right, as shown by the fact that an eminent author agrees with S. However, at layer 1 the Latin quotations function to establish solidarity between the actors and their audience, just as in (1).

Although these quotations can almost always be analysed as instances of solidarity at layer 1, in some cases there are other functions which can be seen as primary. In the following example, a student has been addressing the audience directly (in Latin), arguing that the boys are deserving of a holiday:

(4) *Nimium certè ocij non petimus. Vetat id Hesiodus noster dum sic præcipit*, Μέτρα φυλάσσεσθαι. Καιρὸς δ’ ἐπὶ πᾶσιν ἄριστος. [‘Certainly we do not ask for too much leisure. Our Hesiod forbids it, when he commands thus: Maintain due measure. Proportion is best in everything.’]
(*Orationes*, f. 146v)

As indicated, the Greek quotation is from Hesiod (Hes. *Op*. 694), who was also part of basic grammar school reading matter. The purpose of the whole utterance in (4) is to assure the audience that the boys are not asking for excessive freedom, and that the advice of an eminent author prevents them from doing so. Both can be further analysed as negative politeness strategies: the former as *minimising R* (Brown and Levinson 1987: 176–178) and the latter perhaps as *indicating reluctance* (Brown and Levinson 1987: 188). The reference to Hesiod can also be analysed as a positive politeness strategy in that it asserts shared values between S and H, making it one type of a common-ground strategy. Other interpretations are of course possible. The point here is that although the same kind of strategies are employed in examples (1), (3) and (4), there are notable difference between each case, which may be manifested on both layer 1 and layer 2.

The final extract to be discussed is similar to the Wife of Bath example presented above (in section 3). In this scene, two elderly country bumpkins (*senes rustici*), Credulio and Trunks, are talking to Grammatulus, a student. Grammatulus has just arrived onstage, and has been giving a soliloquy in Latin. Credulio is hoping to enter his son into the school, so he decides to address Grammatulus and ask for his opinion on the matter:

(5) **Credulio:** And verily he is a wonderous ready scholard at it. Young gentleman, may I presume to spur you a question, or two?

Grammatulus: *Mene, si placet, alloqueris? aut num quid me vis, obsecro?* [‘Please, are you talking to me? And pray, what do you want from me now?’]

Credulio: *Immò ego vult habere aliquid res, dic latinè* with you. [‘Nay, me wants to do some business, say in Latin “with you”’]

Trunks: Out, M^{r} Credulio, out, all to be out. You have forgotten those toys long since.

> **Credulio:** Why truely to my knowledge I have not look'd on my Grannum these twonty good years.
>
> **Grammatulus:** I am confident therefore in my hopes, Grave S^{r}, you will excuse the rudeness of my answer. Our schoole statutes confine me to noe other dialect. (*Orationes*, f. 285v)

Credulio does not manage to establish solidarity with Grammatulus because he uses grammatically incorrect Latin; he basically fails in his CS. In other words, this is another case of unsuccessful solidarity. This is also an FTA towards Credulio's own positive face, since he makes a fool of himself (Brown and Levinson 1987: 68; cf. Goffman [1967] 2005: 108). However, at layer 1 this could again be classified as solidarity between the actors and the audience, since the implication is that the audience *qua* members of the educated elite would find Credulio's failure to speak correct Latin amusing, which would in turn enhance the sense of like-mindedness between them and the actors.

5 Discussion

Above I pointed out that solidarity takes place only between S and H, who need to both be part of the interaction currently taking place. It would be possible to argue that someone's positive face is not enhanced if he is completely oblivious to an interaction taking place somewhere else, and that this provision is therefore redundant. However, if we focus, *à la* Goffman ([1967] 2005), on the public quality of face, it is undeniable that an individual's face may be affected even if they are not physically present in the interaction, in the sense that others may be prone to sustain a particular kind of face for the individual in future encounters. This is similar, for example, to the distinction between insulting someone directly and bad-mouthing them behind their back. It therefore seems reasonable to distinguish between "direct" and "indirect" solidarity as well. Whether both cases should be termed *solidarity* is of course another matter.

Another complication relates to example (1). If we consider the interaction situated at layer 2, there should be no objection to our analysis of the function of CS as establishing solidarity, since S and H are equals. A closer look at the interaction between actors and audience (layer 1), however, shows clearly that Ss are able to enhance their (collective) positive face, but what about Hs? Part of the audience, especially the teachers and the Dean, had power over the students, and in many respects the D-value between them and the students would also be high. It is fair to ask whether solidarity among the students is the same thing as solidarity between the students and this part of the audience;

in other words, whether these should be classified as two separate functions. In the examples discussed in section 3, solidarity always seems to affect both S and H positively. I think that in problematic cases we need to decide whether the high P-value (which is in this case quite constant) is in any way referred to or made relevant to the interaction. If it is not, then we can classify such instances of CS as solidarity. In practice, of course, and depending on the type of data available, it may be almost impossible to analyse these variables at all, and the drama data illustrate this rather well. An in-depth examination of these problems will form one starting point for my own future studies.[17]

In the previous section reference was made to other discourse strategies, such as the use of quotations, rather than CS alone. I think that this is both unavoidable and desirable. In the case of Latin/Greek quotations, it is both the switch and the instance of quoting that in combination have the function of establishing solidarity (see also Mäkilähde and Rissanen 2016). Similarly, in (5) Grammatulus combines several different strategies in his final turn to enhance Credulio's face: he begs for forgiveness, gives reasons for his FTA (i.e. having forced Credulio to use a language which he has not fully mastered), and uses a deferential address form (*Grave S^r^*). All of these can be classified as negative politeness strategies (Brown & Levinson 1987: 178–190). Similarly, in (2), Eugenius's use of Philaster's nickname is likewise part of the solidarity function. A related matter is that linguistic strategies such as CS can also be (and often are) multifunctional in any given context (cf. Itkonen 1983: 173). For instance, in (1) CS is used both for solidarity and to facilitate the discourse, namely by enabling the use of the quotation in its original form. Similarly, in (4) the Greek quotation can be seen both as creating solidarity and as a negative politeness strategy.

One clear advantage of a theoretically-oriented approach to the functions of CS is that it enables us to operationalise intuitively obvious differences between different functions. This is perhaps most obvious in the case of unsuccessful solidarity, discussed in detail above. In at least some such cases, the act may end up as an FTA towards S's positive face. Example (3) was a case where the act served to enhance S's own positive face, at the same time constituting an FTA against the positive face of H.[18] In a full-blown theoretical model of the functions of CS, these functions would need categories of their own. However, as discussed above, the question whether solidarity between equals and between non-equals should be analysed as different functions is perhaps less clear. Since we are dealing with human activities, it is only to be expected that there will be many unclear cases and much fuzziness regarding the formal categories. In other words, a theoretically-oriented approach will unavoidably rely on prototypical cases, abstractions and idealisations. However, "[w]ithout abstraction and idealization there is no systematization" (Searle 1969: 56).

---

[17] Another matter that deserves separate discussion concerns the difference between for example *establishing* and *expressing* solidarity.

[18] A discussion of such terms as *self-enhancement*, *self-face enhancement*, *self-promotion* or *boasting* is beyond the scope of the present chapter.

Finally, it needs to be emphasised that this is just one of the possible ways to define solidarity at a theoretical level. Another possibility would be for example to attempt to combine Brown and Levinson's and Goffman's conceptions of face, and to define solidarity as follows (cf. Goffman [1967] 2005: 5–12):

> *An act has the function of establishing solidarity iff (i) it implies a low D-value between S and H through shared membership in a social group, and (ii) both S and H feel that they are in face.*

It is, however, important to note that attempting to evaluate individual definitions in isolation is futile if they capture precisely the same idea from slightly different perspectives, as I think is the case with these two possible definitions of solidarity. Instead, we need to construct comprehensive frameworks or models which will be able to account for different kinds of pragmatic functions of CS, comparing these as we would with any other theoretical models. Depending on one's preferred viewpoint and the aims of the models, relevant features might include for example adequacy, elegance, simplicity, illuminating power, or psychological realism (cf. e.g. Chomsky 1965: 37–47; Itkonen 1978).

6 Conclusion

To sum up: in this chapter I set out to explicate the concept of *solidarity*, in three stages. I first pointed out that a number of previous studies had identified a function of CS which consisted in establishing, indicating or maintaining like-mindedness between S and H, but that the function itself was usually not explicitly defined. I then proposed a formal definition of solidarity, based on concepts derived in particular from the classic politeness theory of Brown and Levinson. Finally, I demonstrated an application of this definition to the case of early modern school drama, and discussed various issues and complications raised by this analysis.

Based on these analyses and discussion, I suggest that we can arrive at a better understanding of the functions of CS (as well as other discourse strategies) by combining a data-driven approach, such as philological close reading, with a theory-driven one, based for instance on pragmatics. Without the former we cannot arrive at any meaningful understanding of the phenomena under scrutiny, and it is therefore primary. However, devising formal definitions for the functions we have identified intuitively has several purposes. First of all, once a function is explicitly defined, it becomes possible to identify possible errors in categorising instances of CS, by determining whether or not they fit the proposed criteria. This may lead either to correcting the analysis, or, if the definition itself seems to be deficient, to its reformulation; the problematic cases discussed above were intended to demonstrate this. Secondly, by referring to pre-defined variables in our definitions, it becomes possible to show in formal terms how two functions that are intuitively different do in fact differ from each other. If the theoretical framework chosen is unable to distinguish between such functions, it may have to be modified accordingly. Finally, the whole point of engaging in the theoretical scrutiny of a concept is to systematise our understanding of the phenomena we are analysing (cf. e.g. Itkonen 1978: 211); this is precisely why it is useful to connect the discussion to a broader theoretical framework. As indicated at the end of the previous section, the next step is to account for other functions of CS, and to illuminate the formal relationships between the different functions.

References

Adams, J. N. 2003. *Bilingualism and the Latin language*. Cambridge: Cambridge University Press.

Brown, Penelope & Stephen C. Levinson. 1987. *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Brown, Roger & Albert Gilman. 1960. The pronouns of power and solidarity. In T. A. Sebeok (ed.), *Style in language*, 253–276. Cambridge: MIT Press.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.

Clark, Herbert H. 1996. *Using language*. Cambridge: Cambridge University Press.

Culpeper, Jonathan. 2011. *Impoliteness: Using language to cause offence*. Cambridge: Cambridge University Press.

Davidson, Mary Catherine. 2003. Code-switching and authority in late medieval England. *Neophilologus* 87(3). 473–486.

Delabastita, Dirk & Ton Hoenselaars. 2015. 'If but as well I other accents borrow, that can my speech diffuse': Multilingual perspectives on English Renaissance drama. In Dirk Delabastita & Ton Hoenselaars (eds.), *Multilingualism in the drama of Shakespeare and his contemporaries*, 1–16. Amsterdam & Philadelphia: John Benjamins.

Diller, Hans-Jürgen. 1997/1998. Code-switching in medieval English drama. *Comparative Drama* 31(4). 506–537.

Dunkel, George E. 2000. Remarks on code-switching in Cicero's letters to Atticus. *Museum Helveticum* 57(2). 122–129.

Feinberg, Joel. 1968. Collective responsibility. *The Journal of Philosophy* 65(21). 674–688.

Foley, William A. 1997. *Anthropological linguistics: An introduction*. Malden, MA: Blackwell.

Genette, Gérard. 1982. *Palimpsestes: La littérature au second degré*. Paris: Seuil.

Giles, Howard, Nikolas Coupland & Justine Coupland. 1991. Accommodation theory: Communication, context, and consequence. In Howard Giles, Justine Coupland & Nikolas Coupland (eds.), *Contexts of accommodation: Developments in applied sociolinguistics*, 1–68. Cambridge: Cambridge University Press.

Givón, T. 2013. On the intellectual roots of functionalism in linguistics. In Shannon T. Bischoff & Carmen Jany (eds.), *Functional approaches to language*, 9–29. Berlin & Boston: De Gruyter Mouton.

Goffman, Erving. [1967] 2005. *Interaction ritual: Essays in face-to-face behavior*. New Brunswick, NJ & London: AldineTransaction.

Hall, Jon. 2009. *Politeness and politics in Cicero's letters*. Oxford: Oxford University Press.

Hechter, Michael. 1987. *Principles of group solidarity*. Berkeley, Los Angeles & London: University of California Press.

Huang, Yan. 2014. *Pragmatics*, 2nd edn. Oxford: Oxford University Press.

Itkonen, Esa. 1978. *Grammatical theory and metascience: A critical investigation into the methodological and philosophical foundations of 'autonomous' linguistics*. Amsterdam: John Benjamins.

Itkonen, Esa. 1983. *Causality in linguistic theory: A critical investigation into the philosophical and methodological foundations of 'non-autonomous' linguistics*. London & Canberra: Croom Helm.

Itkonen, Esa. 2008. The central role of normativity in language and linguistics. In Jordan Zlatev, Timothy P. Racine, Chris Sinha & Esa Itkonen (eds.), *The shared mind: Perspectives on intersubjectivity*, 279–305. Amsterdam & Philadelphia: John Benjamins.

Itkonen, Esa. 2013. Functional explanation and its uses. In Shannon T. Bischoff & Carmen Jany (eds.), *Functional approaches to language*, 31–69. Berlin & Boston: De Gruyter Mouton.

Jucker, Andreas H. 2011. Positive and negative face as descriptive categories in the history of English. *Journal of Historical Pragmatics* 12(1). 178–197.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Leech, Geoffrey. 1983. *Principles of pragmatics*. London & New York: Longman.

Mäkilähde, Aleksi. 2012. *The forms and functions of multilingual practices in early modern school drama from the King's School, Canterbury*. Unpublished MA thesis. University of Turku. http://www.doria.fi/handle/10024/77009 (accessed 6 February 2016).

Mäkilähde, Aleksi. Forthcoming. Some reasons to study quotations, references and allusions in Neo-Latin school drama: Examples from the King's School, Canterbury. In Astrid Steiner-Weber et al. (eds.), *Acta conventus Neo-Latini Vindobonensis: Proceedings of the sixteenth international congress of Neo-Latin studies (Vienna 2015)*. Leiden: Brill.

Mäkilähde, Aleksi, Tommi Alho & Anthony W. Johnson. 2016. Performative grammars: Genre and allusion in a Restoration manuscript. *Neuphilologische Mitteilungen* 117(2). 315–344.

Mäkilähde, Aleksi & Veli-Matti Rissanen. 2016. Solidarity in Cicero's letters: Methodological consideration in analysing the functions of code-switching. *Pallas* 102. 237–245.

Nevala, Minna. 2004. *Address in early English correspondence: Its forms and socio-pragmatic functions*. Helsinki: Société Néophilologique.

Nurmi, Arja & Päivi Pahta. 2012. Multilingual practices in women's English correspondence 1400–1800. In Mark Sebba, Shahrzad Mahootian & Carla Jonsson (eds.), *Language mixing and code-switching in writing: Approaches to mixed-language written discourse*, 44–67. New York & London: Routledge.

Orationes = Canterbury Cathedral Archives Lit. Ms E41, Canterbury Cathedral Library.

Pahta, Päivi & Arja Nurmi. 2009. Negotiating interpersonal identities in writing: Code-switching practices in Charles Burney's correspondence. In Arja Nurmi, Minna Nevala & Minna Palander-Collin (eds.), *The language of daily life in England (1400–1800)*, 27–52. Amsterdam & Philadelphia: John Benjamins.

Pahta, Päivi & Arja Nurmi. 2011. Multilingual discourse in the domain of religion in medieval and early modern England: A corpus approach to research on historical code-switching. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English*, 219–251. Berlin & Boston: De Gruyter Mouton.

Rorty, Richard. 1989. *Contingency, irony, and solidarity*. Cambridge: Cambridge University Press.

Ryan, Cressida. 2013. An ignoramus about Latin? The importance of Latin literatures to George Ruggle's *Ignoramus*. In Philip Ford & Andrew Taylor (eds.), *The early modern cultures of Neo-Latin drama*, 159–174. Leuven: Leuven University Press.

Schendl, Herbert. 2011. Beyond boundaries: Code-switching in the leases of Oswald of Worcester. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English*, 47–94. Berlin & Boston: De Gruyter Mouton.

Schendl, Herbert. 2012. Literacy, multilingualism and code-switching in early English written texts. In Mark Sebba, Shahrzad Mahootian & Carla Jonsson (eds.), *Language mixing and code-switching in writing: Approaches to mixed-language written discourse*, 27–43. New York & London: Routledge.

Scott, John (ed.). 2015. *A dictionary of sociology*, 4th online edn. Oxford: Oxford University Press. http://www.oxfordreference.com/view/10.1093/acref/9780199683581.001.0001/acref-9780199683581 (accessed 6 February 2016).

Searle, John. 1969. *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.

Short, Mick. 1989. Discourse analysis and the analysis of drama. In Ronald Carter & Paul Simpson (eds.), *Language, discourse and literature: An introductory reader in discourse stylistics*, 139–168. London: Unwin Hyman.

Skaffari, Janne & Aleksi Mäkilähde. 2014. Code-switching in historical materials: Research at the limits of contact linguistics. In Robert Nicolaï (ed.), *Questioning language contact: Limits of contact, contact at its limits*, 252–279. Leiden & Boston: Brill.

Swain, Simon. 2002. Bilingualism in Cicero? The evidence of code-switching. In J. N. Adams, Mark Janse & Simon Swain (eds.), *Bilingualism in ancient society: Language contact and the written text*, 128–167. Oxford & New York: Oxford University Press.

Watts, Richard J. 2003. *Politeness*. Cambridge: Cambridge University Press.

Woodruff, C. E. & H. J. Cape. 1908. *Schola Regia Cantuariensis: A history of Canterbury School. Commonly called the King's School*. London: Mitchell Hughes & Clarke.

Richard Ingham

15 Medieval bilingualism in England: On the rarity of vernacular code-switching

1 Contexts for code-switching in medieval England

In medieval England three major languages, English, French and Latin, were in regular use, and their coexistence had important consequences for the English language. So much is uncontroversial: what is more a matter for dispute is the nature of the communicative practices in which their speakers interacted. Here we encounter the familiar problem in historical linguistics, that only written texts survive, whereas the overwhelming majority of language use was oral, even more so in premodern times than now. Only an imperfect and indirect reflection can ever be obtained from such sources of the capacities that individuals may have had, to employ more than one language in spoken use. As noted by Schendl and Wright (2011: 27), the numerous cases they and other researchers have found of apparent code-switching in pre-modern texts prompt the question of whether they constitute a written language phenomenon only, or whether they give us a window on to multilingual speech of the period. In this article, different genres of later medieval writing in which more than one language appears, asking what they tell us about mono-/multilingualism among language users in England at that time. In keeping with current practice, the fact of mixing languages in discourse will be referred to here as code-switching, defined by Poplack (1980: 583) as “the alternation of two languages within a single discourse, sentence or constituent”. Religious and lay texts are examined in order to elucidate the types of code-switching manifested. Two broad types are distinguished, between Latin and a vernacular, and between the two principal vernaculars. It will be seen that these differ in terms of their impact on the textual record, in ways that point to their differing prestige status.

To talk of multilingualism always requires us to consider the relative status of the languages in the society of the time, in effect asking who spoke which language, and for what purpose. The approach taken in older studies (Baugh 1935; Thomason and Kaufman 1988) was to envisage a “Norman French” speech community co-existing at first with an English-speaking one, then losing their competence in French and shifting to English after the 12th c. More recent work suggests that that approach failed to recognise the persistence of functional proficiency in French going well beyond that date. The insular dialect of French, Anglo-Norman, remained in common use until the late 14th century (Rothwell

DOI 10.1515/9781501504945-015

2001; Trotter 2003). It was maintained in existence not by a monolingual French-speaking community but via a substantial number of bilingual language users. Ingham (2012) argued that the linguistic situation among the educated classes was quite generally one of vernacular bilingualism, thanks to the use of French as a medium of instruction in education until the mid-14th century, and that it was this state of affairs, particularly among the educated elite, that brought about the extensive contact influence of French on English. Thousands of words entered English from French in the later medieval period, especially up to the mid-14th century (Dekeyser 1986).[1] This outcome can only be plausibly explained if the degree of interaction between the languages was fairly intense, a phenomenon which in modern societies at least is accompanied by a certain amount of language-mixing within a community of bilingual speakers. In historical periods, however, the spoken language in which that language-mixing would be expected to have taken place is irretrievably lost. Evidence for the interaction between languages that provided contact influence, if this primarily took place orally, is thus deficient.

I have argued elsewhere, however, that indirect evidence for spoken vernacular bilingualism in England is available in certain kinds of administrative texts produced by professional communities (Ingham 2009, 2011, 2013). In this article further data will be examined that support the position that the code-switching we find in these texts arises from the routine use of both vernaculars within communities of practice sharing socio-cultural and linguistic identities.

The concept of code-switching as an act of identity (Poplack 1980; Auer 2005; Schendl 2013) is a well-established interpretive framework in contemporary contexts. How far it can be applied to medieval ones is an important question to ask. In this study it will be shown that the distribution of vernacular–Latin code-switching and between vernacular code-switching in medieval texts written in England diverges sharply according to the acts of identity performed by the composer of the text.

For Auer (2005) acts of identity associated with code-switching are to do with what membership category the speakers seek to place themselves in. Accordingly, code-switching in the sources to be studied here will be related to two medieval social categories with sharply distinctive identities, the clergy and lay professional scribes working in legal and administrative roles. Religious prose works created by the clergy are known to display a certain amount of code-switching (Schendl 2013). Their discursive practices brought into play important contextual factors – the need to communicate doctrine recorded in a language,

---

[1] This was around the time when the standard of French clearly declined in England (Ingham 2010, 2012), linking the borrowing process with maintained competence in the source language.

Latin, not generally known to lay recipients, and also the presence among lay people of habitual users of one or other the two vernaculars. I then proceed to consider the evidence for code-switching in the textual record left by lay scribes working in the domain of land management, in the form of charters, accounts and other such documentation. It will be seen that code-switching in these two domains patterned quite differently. Religious prose code-switching attests to Latin-vernacular proficiency among the clergy, whereas documentary records will be argued to offer evidence of vernacular code-switching.

Before proceeding further, a few terminological points are in order. Standard accounts of code-switching distinguish intrasentential code-switching, where words or phrases switch from one language to another within a sentence, from intersentential code-switching, where the switch takes place between sentences. An intuitively plausible although sometimes elusive concept is that of the "matrix language" of the sentence, that is, the one used for its main grammatical structure (Myers-Scotton 1993). Intrasentential code-switching will then be a matter of inserting the non-matrix language items into that structure. Another relevant term will be "lexical gap" code-switching, where, in the absence of a corresponding item in the language of the discourse at the point of the switch, the discourse switches to an expression from another language. An occasion for lexical gap code-switching occurs when speakers of an imperfectly learned L2 may resort to their L1 for a concept for which they do not know the L2 expression.

2 Code-switching in religious texts

Medieval clergy in England constituted a community in which more than one language was routinely used. Latin was the language of religious worship, but substantial French-language material of a religious nature survives that was written by English clerics, some collected on the Anglo-Norman Hub (Trotter 2007), as well as plentiful English texts of a devotional or homiletic nature written throughout the medieval period. Works written by members of the clergy thus offer opportunities for the extent and nature of bilingualism to be gauged in this community of practice. In this section, four religious texts, two gospel commentaries in Anglo-Norman and in English, and two sermon collections in Anglo-Norman and English will be studied, in order to investigate aspects of code-switching in texts produced by religious writers.[2] The Anglo-Norman texts, an anonymous commentary on the Proverbs of Solomon, and an anonymous set of sermons on the Book of Joshua, were written in the early 13th century. The English texts, Richard Rolle's Gospel commentaries and the anonymous sermons published as Middle English Sermons by the Early English Text Society, were written in the following century. Although the authors of these four texts are mostly not known, it is clear from the content that they must have had at least a clerical training, and probably a good background in theology.

---

[2] I do not consider the specialised text-type of macaronic sermons, whose circumstances of delivery remain controversial (Fletcher 2009).

In the sermons, code-switching is most often intersentential and takes the form of Latin quotations from the Bible or earlier religious writers, accompanied by a translation into the relevant vernacular, e.g. the following from the Anglo-Norman sermons on Joshua:

(1) Sicume dit l'Evangile Habent Moysen et Prophetas; audiant illos, co est en rumanz "Il unt Moysen et les prohetes; oient ceals".
'As the Gospel says: They have Moses and the prophets; may they hear them, that is in Romance "They have Moses and the prophets; may they hear them".' (Joshua 1, p. 8)

(2) Issi le truuns escrit Frangit Deus omne superbum, co est "Deus fraint chascun orguilluse chose".
'Thus we find it written God breaks down every proud thing, that is "God breaks down every proud thing".' (Joshua 1, p. 26)

The same is the case in the Middle English Sermons (Ross 1940), a collection probably compiled around 1400, as in the following:

(3) Hic recepit peccatores. Frendes, þese wordes [. . .] ben þus much to seye on Englisch to youre understondynge: He resceveþ synneful men.
'He accepts sinners. Friends, these words are as if to say in English so you will understand: He accepts sinful men.' (MES 162, 21)

Now the use of another language from that of the running text, seen in many such examples in these texts, is quite unlike code-switching as practised in modern societies. It does not point to spoken language practices among bilinguals, but tells us that the sermons' lay audience, whether French or English-speaking, or both, was not expected to know Latin, whereas the sermon-writer knew both Latin and the relevant vernacular; membership categories were thus distinct. Latin expressions were introduced for their value as sources of scriptural authority, not as acts of shared identity construction. The sermon-writer's control of Latin set him apart from his audience, and can be seen in the Middle English Sermons, in which the author sometimes inserts untranslated Latin phrases that look like memoranda to himself:

(4) This name Clement aftur the exposicion of the Catholicon est idem quam nobilis, misericors pius et sanctus.
'This name Clement, according to the explanation in the Catholicon, means noble, merciful, pious and holy.' (MES 5, 15)

The English commentaries referred to here (Ogilvie-Thomson 1988) are attributed to Richard Rolle, a 14th-century spiritual counsellor. They are similar to the sermons in that they use Latin sentences which are immediately translated, as in:

(5) Crist [. . .] seith: Ite et predicate ewangelium Goth and precheth the Gospel.
'Christ says: Go and preach the Gospel (Go and preach the Gospel)'
(Rolle commentaries 78)

Also, the Latin sentences are often abbreviated, so the text works only if the reader knows, or the hearer is given, the full quotation,[3] e.g.:

(6) Heu michi Domine Wo to me lorde for I haue synned ourmykel.
'Woe is me Lord Woe is me Lord for I have have sinned greatly'
(Rolle commentaries 46, 366)

Another commentary attributed to Rolle, entitled Dirige Domine, on part of Matthew's Gospel, follows the same approach: Latin is used for a quotation, given in full, which is then always translated into English, and for that reason does not belong with bilingual code-switching as conventionally understood. In general, the English biblical commentaries tend to avoid intrasentential code-switching.

The Anglo-Norman Bible commentary, on the Book of Proverbs, is rather different. It is the only one among the text types sampled to use intrasentential code-switching, French to Latin:

(7) Riches quidet estre cil de l'Evangile qui chascun jur mangat **splendide**.
'Rich bethought himself the man in the Gospel who each day ate lavishly.'
(Par. Sal. 98)

---

[3] The rarity of silent reading (Saenger 1982), and the concomitant importance of reading aloud to an audience, in the medieval period should always be kept in mind.

(8) Tuit seient il turmenté de diverses turmenz et tribulaciuns **e diversis pressuris**.
'However tormented they may be by various torments and tribulations from diverse pressures.' (Par. Sal. 84)

(9) [. . .] a qui Deus ot duné si grant grace **quod ab omnibus amabatur**
'[. . .] to whom God had given such great grace that he was loved by all' (Par. Sal. 114)

Switching is always from French into Latin, never into English. The switches, as seen here, can be to a single word, a phrase, or a clause. Here we do have something approximating more to spoken code-switching as practised among contemporary bilinguals. In addition, such Latin phrases are clearly expected to be understood by the addressee without translation; they are not extracts from the biblical text in question. This can be seen in phrases such as:

(10) Cil **filii resurreccionis** qui apurtienent a la joie de la resurrecciun.
'These sons of resurrection who belong to the joy of resurrection.' (Par. Sal. 106)

(11) Mes alme fameiluse **sitiens justitiam** prendrat amer pur dulz
'But a starving soul thirsting for justice will take bitter for sweet' (Par. Sal. 1)

They are used to complete the sense of matrix language French sentences in a way possible only if writer and adressee were proficient in both languages. This text, then, suggests the existence of a French-Latin community of language practice at the time in question around 1200. Just such an erudite religious group, possibly monastic, or possibly featuring secular clergy, is known to have existed in Angevin and Plantagenet England, including such 13th century figures as Matthew Parris, Peter of Fetcham, and Bishop Grosseteste, all born in England but using Anglo-Norman French in religious works. The choice of this language rather than English is to be interpreted as an act of identity, badging both writer and audience as belonging to a transnational cultural elite, and the use of code-switching into Latin among this French-speaking group constitutes a further indication of category membership, that of a clerical subgroup within that elite. Hence a text such as the Anglo-Norman commentary on the Book of Proverbs has a different status from the other three. Code-switching in this type of religious work, as distinct from sermons, or Rolle's commentaries which were written for a non-clerical audience, served two kinds of discursive functions, one relating to the choice of French as an overall matrix language, the other to do with code-switching into Latin.

A monolingual French-speaking community of some sort presumably still existed in the 13th century too, as witness the translations into insular French of Ancrene Wisse (guidance for nuns), and in the previous century of English-tradition literary works such as the romances of Horn and Havelok.

Of monolingual English speakers, there were very large numbers among the lay population, for whom the French sermons of Maurice de Sully were translated around the same time as the Kentish Sermons. An Anglo-Norman manuscript of the French sermons dated to around 1200 also exists. Putting together facts such as these and the absence of code-switching in religious works from one vernacular to the other, we might suppose that the traditional picture of two monolingual communities, in each of which an educated elite also knew Latin, was essentially right, despite my earlier critique of this position (Ingham 2012). This would be a premature conclusion, however. It would seem, rather, that these religious texts are merely silent on the existence of vernacular bilingualism, so that for this purpose we would have to turn to other text types.

3 Code-switching in lay professional contexts

Previous research by Ingham (2009, 2013) and Wright (2002, 2005, 2011) has identified lay professional contexts in which French and English co-occur in texts produced by people working in these domains. Manorial accounts contain what can be referred to as embedded code-switching, that is, within a matrix Latin text occur instances of French such as definite articles preceding English nouns, e.g.:

(12) a. Pro le salthus (Framlingham 1324–1325: 68)
'For the salthouse'

b. In le nywemede (Cuxham 1358–1359: 594)
'In the new meadow' (Ingham 2009)

This written phenomenon, I have claimed, originated in the form of spoken code-switching. In the context of a manorial estate, English was routinely used to designate familiar landmarks and buildings by estate managers when conversing in French with those drawing up accounts. I hypothesised that their discourse included French matrix language utterances, within which the use of English terms would have produced instances of one-word switches visible in other documents that happen to have been drawn up in French rather than Latin, e.g. (Ingham 2013: 118)

(13) [. . .] de Tamyse tanque a les flodeyates del molyn de Egnesham.
'From the Thames to the floodgates of the mill of Eynsham'
(Eynsham Cartulary, 1328)

The otherwise unmotivated use of French definite articles in cases such as (12) is explained by positing that discourse forming the background to them was French matrix, of which all that remains are the definite articles, the rest having been put into Latin when the document was drawn up.

Wright's (2002, 2005, 2011) pathbreaking work on code-switching in mixed-language texts involved vermacular lexis in accounts documents. She found frequent switches between Latin function words such as the numeral *unum* and the preposition *pro*, and vernacular content words, e.g. (Wright 2011):

(14) a. Item in una alia parva camera unum yronbonde cheste ibidem.
'Item: in another small room an ironbound chest in the same place.'
(Guildhall will inventory, 1425)

b. Pro corde ad corpus ffrettad vj d.
'For the rope for tying the body 6d.' (Guildhall will inventory, 1425)

c. Pro xxix lb. cera ad dictum canvas cerandum.
'For 29lbs of wax for waxing the said canvas.'
(Guildhall will inventory, 1425)

As noted by Wright, because of heavy French-to-English borrowing, determining into which vernacular the switches took place is in many cases difficult, as here with canvas (orig. OFr. *kanevas*).

The switches in Wright's studies illustrated here are between Latin and a vernacular language, but in recent work she has observed mixed language use in London account rolls that brought three languages into play, as in (Wright 2010):

(15) a. Apud le fresshewharf
'At the freshwharf' (London Bridge accounts, 1429–1430)

b. Pro le shoying dictarum rotarum
'For shoeing the said wheels' (London Bridge accounts, 1461)

She considered, following Trotter (2010), that "le" was a way of signalling a switch into English (Wright 2010: 134). This is not the view of the present author; I believe these combinations of French article plus English nouns (henceforth FA+EN) are to be seen as "snippets of code switching" (Jefferson and Putter 2013) between the two vernaculars. Trotter (2010)[4] noted the same phenomenon in Exeter bridge accounts, and Ingham (2013) observed it in land charters from various locations in England as far back as the later 12th century, attesting to a time of spoken code-switching on the part of those drawing up the document in question when knowledge of French was unquestionably highly proficient. As suggested by Ingham (2009), by the mid-15th century that time was already past, however, and the FA+EN combination lived on only as a fossilised scribal convention, dying out in the 16th century. Now, an important point for the present study is that those keeping a record of expenditure on e.g. town bridges as in (15) operated in much the same way as their counterparts on a manorial estate as in (14). Generally, records such as accounts and inventories were not excerpted from texts already written in French, but rather, drawn up to impart new information that, in a world of expensive writing materials, would surely have originated as spoken communication. There is good reason to suppose, therefore, that behind the written record seen e.g. in (14) and (15) lay a piece of code-switched spoken communication.

Support for this hypothesis comes from administrative documents surviving from communities where spoken vernacular bilingualism is known to have been practised. Pryce (2005) identified comparable use of Welsh in intrasentential code-switching to designate topographical features in Latin charters from Wales, and code-switching in 14th century municipal documents from French-administered Flanders appears with French function word or words preceding Flemish content items referring to places in an urban environment:

(16) A le pardemarkt; Dou broethus
'To the horse market'; 'Of the bakery' (Ypres accounts [Ingham 2013: 116])

---

[4] Originally given as a Birmingham colloquium presentation in February 2007.

Here too, administrative texts reflected the spoken presence of two vernaculars in a bilingual community with switches between the latter embedded in a matrix Latin sentence, reinforcing the case for interpreting the phenomenon in English documents as a reflection of bilingual usage.

4 A further study of French function word and English noun combinations

The argument for bilingual code-switching as an interpretation of the FA+EN phenomenon by myself, Trotter and Wright so far rests entirely on the interpretation of the use of a single grammatical formative, the French definite article. If on the right track, the hypothesis outlined above predicts that English nouns preceded by other French function words, and not just definite articles, could be found embedded in documentary material such as Latin charters. The approach is therefore extended in the present study to investigate the preposition plus definite article combinations *del* meaning ‘of the’, and *al* meaning ‘to or at the’, which are undoubtedly French items, occurring in medieval Latin charters drawn up in England. The source for these is the electronic DEEDS corpus (Gervers et al. 1975) already used by Ingham (2013), which contains many hundreds of medieval charters, almost all in Latin, from the early 12th to the early 14th centuries.

Relevant data were indeed fairly numerous. In DEEDS a total of 85 occurrences of *del* and *al* were found to precede a vernacular noun, not counting cases when they were part of personal names (e.g. *Alani del Orchard*).[5] A few such nouns were French, e.g. *chemin* and *chasteignier* in:

(17) a. En le north del chemin dedi etiam [...]
‘To the north of the road I also gave [...]’
(DEEDS 00520511, 1171–1172)

b. [...] usque ad [...] terram del chasteignier.
‘As far as the [...] chestnut tree lands.’ (DEEDS 02860032, 1270)

The great majority, however, were English words designating a topographic feature on the land granted in the relevant charter, e.g.:

---

[5] Personal names arguably formed unanalysable units, and were discarded.

(18) Una al ho et tres alibi in campis de sudgiuele.
‘One at the mound (“ho”) and three elsewhere in the southgivel fields.’
(DEEDS 00200143, 1199)

The syntax of expressions with *del* can be conveniently studied thanks to the very frequent occurrence of this item. Most often it is found inside a Noun Phrase, with the function of introducing a postmodifying Prepositional Phrase.

(19) a. In prato del pol que contigue iacent inter pratum [. . .]
‘In the pool meadow which lie nearby between the meadow [. . .]’
(DEEDS 99909009, 1252)

b. Tam illa placea predicta del westmedue quam illa del berneker.
‘Both the aforesaid place of the westmeadow and that of the barnacre.’
(DEEDS 00970245, 1263)

Here the Prepositional Phrases *del pol* and *del westmedue* postmodify a common noun such as *pratum* (‘meadow’) or *placea* (‘place’), and supplied specific positional information, in the same way that in a non-code-switching passage a Latin PP such as those bolded below would, e.g.:

(20) a. Et Hulmede et pratum **ad caput crofte**.
‘And hulmeadow and the meadow at the top of the croft.’
(DEEDS 02976967, 1219)

b. Excepta [. . .] molendino et prato **iuxta idem molendinum**.
‘Except the mill and the meadow near the same mill.’
(DEEDS 02973928, 1227)

c. Totum ius quod habuimus [. . .] in quadam placea **iuxta terram Osmundi Molendinarii**.
‘Every right we had [. . .] in a certain place near Osmund the Miller’s land.’ (DEEDS 02710844, 1292)

d. Unam in cultura **super Netle Wido[nis]**.
‘One in the fields above Netle Wido.’ (DEEDS 00270085, 1187)

Since the syntax of the matrix language of the deed, Latin, accommodates a PP after the head noun, the construction we find in the code-switching cases such as (18)–(19) is in line with contemporary expectations of where code-switching can occur.

Another use of *del* is sometimes found in which the French preposition *de* contributes the sense of 'movement from', close to the original sense of Latin *de*, e.g.:

(21) Per quam itur del thweitis apud mariscum.
'By which there is a way from the clearing at the marsh.'
(DEEDS 00630085, 1293)

In all 85 cases observed, *del* and *al* were used with a syntactic function integrated into the matrix Latin sentence; the meanings of the prepositions *de* and *à* that contributed to the two combinations were always appropriate to the propositional content. They thus conformed to contemporary patterns of intrasentential code-switching as discussed e.g. by Myers-Scotton (1993) and others. However, the switch here was from Latin into French and displays code-switching at a scribal level, when the written Latin document was being drawn up. It is the immediately subsequent switch between French *del/al* and the English noun which appears informative as regards vernacular code-switching. Other than when part of personal names, which were excluded, as mentioned above, occurrences of *del* and *al* were always found in a sequence having the following format:

(22) . . . Matrix Latin + French function word + English content word . . .

Modulo the function word appearing as *del/al* rather than just the definite article, this is the same configuration as we had in e.g. example (12) above, and should be interpreted in the same way, as a small nugget of spoken vernacular code-switching preserved in the mainly Latin document. The interpretation offered is that this reflects bilingual discourse among those involved in land management who habitually referred to topographic features on the estate in English, but when discussing business with accountants and lawyers, did so in mainly French utterances in keeping with their higher-status occupation. Thus as regards the data in (18)–(19), my claim is that English words for mounds, pools, clearings and the like were employed by overseers working on the estate, and therefore familiar with its features, who were accustomed to using these English terms when conversing with monolingual English speakers. When preparing the content of written documents, however, with accountants or lawyers, they communicated in utterances that were mainly French, while leaving the topographic features expressed in English for easy reference.

5 Avoidance of code-switching in vernacular discursive texts

In medieval England two kinds of text, charters and accounts, present mixed-language phenomena that seem to point to the use of vernacular code-switching in ambient discourse in certain professional contexts. However, vernacular code-switching evidence is extremely scanty in discursive texts, i.e. outside documentary record-keeping. There is an early 15th century letter, given prominence in widely used textbooks such as Gardner-Chloros (2009: 40), that features code-switching between English and French. One could read this without appreciating its extreme rarity, among many hundreds of Anglo-Norman letters, none of which exhibits a clear example of French–English code-switching. No such letter is found in Legge (1941), Tanquerey (1916), or in the 14th c. section of the Stonor correspondence (Carpenter 1996). An electronic corpus of 13th and 14th century Anglo-Norman correspondence I created containing over 150 letters likewise has no such examples (http://wse1.webcorp.org.uk/anglo-norman/browse/). The salient fact in the genre of Anglo-Norman correspondence is thus the strict avoidance of code-switching. Indeed, one might say of the embedded code-switching with French function words and English nouns that this pattern is found in no other genre than accounts and charters. What is one to make, then, of the seemingly marginal status of code-switching between vernaculars at this time, when the etymological evidence massively suggests that the two languages were in regular contact with each other via bilingualism?

One reason why vernacular code-switching was so overwhelmingly avoided in written discourse might well be an inclination already at this time to see it as somehow too casual a practice to put into written form. However, the language choice factor must also have played a major part. We know that writing in English was a choice taken by those who were addressing a public lacking knowledge of French, e.g. with the 14th and 15th century English romances. Code-switching into French in such texts would threaten that very purpose. For its part, French was recognised as being useful because it did not require the addressee to know Latin, and also because it was not regionally variable within England. To code-switch into English in an insular French text would have raised the question: what dialect of English?[6] So for that reason alone, the user might well have been disposed to keep to French. The choice of language to be used for a discursive written text, it seems, outweighed the ability the writer may or may not have had to employ vernacular code-switching.

---

[6] A question whose pertinence is seen in the alternative dialectal versions of e.g. the Cursor Mundi, the Southern as against Northern versions of the Passion, and various other ME texts.

It is not until the very end of the lifespan of insular French that French–English code-switching becomes common in vernacular text genres such as architects' indentures (Salzman 1952) and legal debates such as those in Exchequer Chamber cases (Hemmant 1933). In the later 14th c. and mid-15th c. respectively, we start to see cases such as the following, in building indentures:

(23) a. Meisme le **gatehous** serra vowtz.
'The same gatehouse will be vaulted.' (1383; Salzman 1952: 463)

b. Le **wharf** serra del longure come il est meyntenant.
'The wharf will be of the length that it is now.'
(1387; Salzman 1952: 468)

c. En le secounde **flore** les ditz John et John ferrount une **drawying** chaumbre
[. . .] et le dit secunde Flore deuaunt serra fait ouesque une **seylingpece**.
'On the 2nd floor the said John and John will make a drawing room [. . .] the front of the said 2nd floor will be made with a projecting beam.' (1410; Salzman 1952: 484)

d. Les queuz measons pardesuis serrount faitz ouesque trois **roofes** gables.
'The which houses above will be made with 3 gabled roofs.'
(Salzman 1952: 484)

Likewise, lawyers who a hundred years earlier had conducted legal debates virtually without code-switching into English now used English quite indiscriminately for ordinary lexis, e.g.:

(24) a. Mez sil enble **hen** ou capon pur ce il serra pendu.
'But if he steals a hen or a capon, for that he will be hanged.'
(1456; Exch Ch. [Selden] p. 124)

b. [. . .] et certifie le Roy **pleynly** de chose dount il voille aver soun chartre.
'And inform the King plainly of a matter about which he might want to have his charter.' (1458; Exch Ch. [Selden] p. 164)

c. [. . .] sil ne soit pur money **lent** a Roy.
'Unless it be for money lent to the King.'
(1459; Exch Ch. [Selden] p. 159)

d. Le pleyntyffe averoit eue payment un masere, ouches et **rynges**.
'The plaintiff would have had (as) payment a goblet, clasps and rings.'
(Exch Ch. [Selden] p. 181)

The appearance here of what seems to be lexical gap code-switching in (24) is interesting, coming at a time when the use of French in England is known to have been in sharp decline. We do not really seem to be dealing here with normal bilingual code-switching, that is of switches between languages in which speakers/writers possessed a functional proficiency allowing them to access the relevant words for such ordinary vocabulary items as floor, wharf, hen, plainly, lend, and ring. Instead, English words frequently fill gaps in the users' knowledge of French. The qualitative change here is especially noticeable with legal texts, where a textual record in insular French is available going back to the mid-13th century, in which earlier cases such as (24a)–(24d) are essentially absent.

Yet the highly divergent nature of text type categories as to whether they show vernacular code-switching is probably not a matter for surprise, however, when it is remembered that code-switching in contemporary bilingual communities is characteristically a spoken phenomenon, which may, but need not, figure in its practitioners' written production. Documentation in the form of accounts, for all their formulaic style, must have originated in conversations when e.g. landlords' stewards met manorial bailiffs and reeves to assess how much had been spent, and on what. Likewise, land grant charters are thought to have been drawn up in the presence of the consenting parties, who would initially have made oral statements to the notary including reference to the characteristics and location of the land being granted. Bearing in mind an interactive spoken context in which such texts presumably originated helps us to understand how it is that they might, even when transmuted into Latin written records, still keep traces of their spoken origin in a milieu where code-switching was practised.

Fictional, religious, and regulatory texts, which form a large part of the mainstream material in later Middle English and Anglo-Norman, were generally composed as writing, however, and would not have originated in interactive spoken discourse between real-life participants. Under the deliberate control of the writer, they were far less likely to be influenced by factors such as the often unconscious sense of identity within a bilingual community that operates in contemporary spoken code-switching. Medieval vernacular code-switching was thus unlikely to be found outside text-types that are closely linked to oral language, in their provenance if not in the physical form in which they have come down to us.

6 Final remarks

Discursive texts requiring vernacular bilingual competence in their intended audience are remarkably rare in medieval England, almost to the point of non-existence. Given what we know of the scale of French contact influence on English from the sheer volume of French loans entering Middle English, this may seem surprising until we recall the highly skewed nature of our witnesses for medieval language use, entirely lacking in the kind of unplanned spoken discourse data that has provided the ordinary working material for code-switching researchers of present day bilingualism. Evidence for medieval vernacular bilingualism can only come in indirect form from what can be gleaned from clues in the textual record, and whatever language-mixing is visible in medieval writing must greatly understate the extent and importance of vernacular bilingualism in that period.

It has therefore been necessary to consider text types recording the praxis of particular professional groups for evidence that vernacular code-switching was a reality in the later medieval period. The rather austere written material of medieval accountancy and the like may best reveal the existence of spoken multilingualism in this period.[7] The clues exist in accounts and charter documents which record land management practices and changes in ownership of land that those in professional positions running manorial estates and the like would have needed to know. The verbal interaction that preceded the drawing up of these documents, it can be inferred, was conducted between professional group members in French, and featured frequent use of English content words relating to their topological context.

Whether this kind of discourse can be seen as having helped to create an identity is doubtful, however, since the postulated choice of French as the main language of discourse would already have served to identify speakers as members of an elite group. Switches into English for landscape terms could perhaps be viewed as ways of displaying membership of the local manorial community in which the interactions took place, but it seems likely that in many cases, such as *ho* and *thwaites* in examples (18) and (21) above, they arose as lexical gap switches by speakers not knowing the corresponding French term.

---

[7] Already pursued with respect to dialogic texts by Culpeper and Kytö (2010).

Code-switching in texts such as the Anglo-Norman commentary on Proverbs, on the other hand, did function as affirming membership of a social category, though here the language into which switches took place, Latin, created a shared clerical identity, not one involving vernacular bilingualism. In texts such as sermons or commentaries destined for a lay audience, no shared identity was created, for reasons having to do with the differing social identities of the originators of the texts and their audience.

One can thus partly agree with the perspective taken by Schendl (2013) that code-switching in the medieval period can be viewed as "acts of identity" involving interlocutors with a shared bilingual culture. But it has been argued here that this is usually the case only when the switch was into Latin, written by and for an audience knowing Latin. The indirect evidence of spoken vernacular bilingualism and code-switching that can be gleaned from administrative texts of the types discussed in this study does not seem to need interpreting in terms of the affirmation of category membership, but as rather as a reflection of usage within a professional community of practice making its choice of French the badge of identity, and resorting to English only when practical convenience warranted doing so. This in no way disqualifies data of this type from being considered as evidence of medieval bilingualism, but it seems to suggest that code-switching among such professional communities of practice was different in kind from contemporary conversational code-switching among bilinguals. Among lay people the status-affirming language practice was the choice of French: if code-switching into English served no identity-confirming function, and hence was generally avoided, the dearth of vernacular code-switching in surviving texts in medieval England, especially in French texts, is readily understandable.

Primary sources

MES = Ross, Woodburn O. (ed.). 1940. *Middle English sermons* (Early English Text Society Original Series 209). London: Oxford University Press.

Rolle commentaries = Ogilvie-Thomson, S. J. (ed.). 1988. *Richard Rolle: Prose and verse from Longleat 29 and related manuscripts* (Early English Text Society Original Series 293). Woodbridge: Boydell.

DEEDS = Gervers, Michael et al. (eds.). 1975–. The DEEDS Project. Toronto: University of Toronto. http://sites.utoronto.ca/deeds/.

Joshua = Hunt, Tony (ed.). 1998. *Sermons on Joshua*, vols I and. II. London, Birkbeck College: Anglo-Norman Text Society.

Par. Sal. = Hunt, Tony (ed.). 2012. *Les Paroles Salomun*. London, Birkbeck College: Anglo-Norman Text Society.

References

Auer, Peter. 2005. A postscript: Code-switching and social identity. *Journal of Pragmatics* 37(3). 403–410.

Baugh, A. C. 1935. The chronology of loan-words in English. *Modern Language Notes* 50(2). 90–93.

Carpenter, Christine (ed.). 1996. *Kingsford's Stonor letters and papers 1290–1483*. Cambridge: Cambridge University Press.

Culpeper, Jonathan & Merja Kytö. 2010. *Early Modern English dialogues: Spoken interaction as writing*. Cambridge: Cambridge University Press.

Dekeyser, Xavier. 1986. Romance loans in Middle English: A re-assessment. In Dieter Kastovsky & Aleksander Szwedek (eds.), *Linguistics across historical and geographical boundaries: In honour of Jacek Fisiak on the occasion of his fiftieth birthday I: Linguistic theory and historical linguistics* (Trends in Linguistics: Studies and Monographs 32), 253–265. Berlin & New York: Mouton de Gruyter.

Fletcher, Alan J. 2009. *Late medieval popular preaching in Britain and Ireland: Texts, studies, and interpretations* (Sermo 5). Turnhout: Brepols.

Gardner-Chloros, Penelope. 2009. *Code-switching*. Cambridge: Cambridge University Press.

Hemmant, M. (ed.). 1933. *Select cases in the Exchequer Chamber before all the justices of England*, vol. 51 (Publications of the Selden Society). London: Bernard Quaritch.

Ingham, Richard. 2009. Mixing languages on the manor. *Medium Ævum* 78(1). 80–97.

Ingham, Richard. 2010. Anglo-Norman as a contact language. In Richard Ingham (ed.), *The Anglo-Norman language and its contexts*. Woodbridge: York Medieval Press and The Boydell Press.

Ingham, Richard. 2011. Code-switching in the later medieval English lay subsidy rolls. In Herbert Schendl and Laura Wright (eds.), *Code-switching in early English* (Topics in English Linguistics 76), 95–114. Berlin & Boston: Mouton de Gruyter.

Ingham, Richard. 2012. *The transmission of Anglo-Norman: Language history and language acquisition* (The Language Faculty and Beyond, research monograph series no. 9). Amsterdam & New York: John Benjamins.

Ingham, Richard. 2013. Language-mixing in medieval Latin documents: Vernacular articles and nouns. In Judith A. Jefferson & Ad Putter (eds.), *Multilingualism in medieval Britain (c. 1066–1520): Sources and analysis* (Medieval Texts and Cultures of Northern Europe 15), 105–121. Turnhout: Brepols.

Jefferson, Judith A. & Ad Putter (eds.). 2013. *Multilingualism in medieval Britain (c. 1066–1520): Sources and analysis* (Medieval Texts and Cultures of Northern Europe 15). Turnhout: Brepols.

Legge, Mary Dominica (ed.). 1941. *Anglo-Norman letters and petitions from All Souls MS. 182* (Anglo-Norman Text Society 3). Oxford: Blackwell.

Myers-Scotton, Carol. 1993. *Duelling languages: Grammatical structure in codeswitching*. Oxford: Clarendon Press.

Poplack, Shana. 1980. Sometimes I'll start a sentence in Spanish *y termino en español*: Toward a typology of code-switching. *Linguistics* 18(7–8). 581–618.

Pryce, Huw. 2005. Uses of the vernacular in the acts of Welsh rulers, 1120–1283. In O. Guyot-Jeannin (ed.), *La langue des actes: actes du XIe congrès international de diplomatique* (Troyes, jeudi 11–samedi 13 septembre 2003). http://elec.enc.sorbonne.fr/CID2003/pryce (accessed 23 May 2016).

Rothwell, William. 2001. English and French in England after 1362. *English Studies* 82(6). 539–559.

Saenger, Paul. 1982. Silent reading: Its impact on late medieval script and society. *Viator: Medieval and Renaissance Studies* 13. 367–414.

Salzman, Louis Francis. 1952. *Building in England down to 1540: A documentary history*. Oxford: Clarendon Press.

Schendl, Herbert. 2013. Code-switching in late medieval macaronic sermons. In Judith A. Jefferson and Ad Putter (eds.), *Multilingualism in later medieval Britain (c. 1066–1520): Sources and analysis* (Medieval Texts and Cultures of Northern Europe 15), 153–169. Turnhout: Brepols.

Schendl, Herbert & Laura Wright. 2011. Code-switching in early English: Historical background and methodological and theoretical issues. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English* (Topics in English Linguistics 76), 15–45. Berlin & Boston: Mouton de Gruyter.

Tanquerey, Frédéric Joseph. 1916. *Recueil de lettres anglo-françaises (1265–1399)*. Paris: Champion.

Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language contact, creolisation, and genetic linguistics*. Berkeley: University of California Press.

Trotter, David. 2003. L'anglo-normand: variété insulaire, ou variété isolée? *Médiévales* 45: 43–54.

Trotter, David. 2007. Anglo-Norman hub textbase. www.anglo-norman.net.

Trotter, David. 2010. Bridging the gap: The (socio)linguistic evidence of some medieval English bridge accounts. In Richard Ingham (ed.), *The Anglo-Norman language and its contexts*, 52–62. Woodbridge: York Medieval Press and The Boydell Press.

Wright, Laura. 2002. Code-intermediate phenomena in medieval mixed-language business texts. *Language Sciences* 24(3–4). 471–489.

Wright, Laura. 2005. Medieval mixed-language business texts and the rise of Standard English. In Janne Skaffari, Matti Peikola, Ruth Carroll, Risto Hiltunen & Brita Wårvik (eds.), *Opening windows on texts and discourses of the past* (Pragmatics & Beyond New Series 134), 381–399. Amsterdam & Philadelphia: John Benjamins.

Wright, Laura. 2010. A pilot study on the singular definite articles *le* and *la* in fifteenth-century London mixed-language business writing. In Richard Ingham (ed.), *The Anglo-Norman language and its contexts*, 130–142. Woodbridge: York Medieval Press and The Boydell Press.

Wright, Laura. 2011. On variation in medieval mixed-language business writing. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English* (Topics in English Linguistics 76), 191–218. Berlin & Boston: Mouton de Gruyter.

Laura Wright

16 A multilingual approach to the history of Standard English

1 Introduction

This paper focuses on three developmental stages in the history of English which are apparent from a multilingual perspective but which are currently omitted from textbooks: the late medieval mixed-language business system, the fifteenth century tip-point when the switchover to English was imminent, and the subsequent shift to Proto-Standard English. I survey recent work which shows a disruption phase in the last few decades of the fourteenth century in both Anglo-Norman and mixed-language writing. Starting with this disruption phase around the 1370s and continuing to the tip-point to monolingual English around the 1480s (the dating is not concrete, it varies from archive to archive, but roughly fits these parameters), I argue that the intervening century constitutes a period of transition from Medieval Latin to Proto-Standard English.

2 Three developmental stages as viewed from a multilingual perspective

Prior to the fifteenth century, Londoners (and people elsewhere in Britain) kept accounts not in monolingual English, but in either a Medieval Latin or Anglo-Norman French matrix, with English (and sporadic words from other languages) embedded in a syntactically orderly manner. Certain linguistic elements were particularly resistant to representation in Medieval Latin or Anglo-Norman. Names of people, their social ranks and titles, place-names, currencies weights and measures, and the names of specific traded commodities were likely to be retained in their original form. Nouns and deverbal *-ing* forms could, non-categorically, appear in English. Illustration is given with some extracts from the London merchant and moneylender Gilbert Maghfeld's account-book for the years 1372–1395. The main matrix language (that is, the syntactic framework) is Anglo-Norman:

Acknowledgment: My grateful thanks to Jeremy Smith and Julia Fernández Cuesta for their stringent and constructive criticisms, and to José Miguel Alcolado Carnicero for permission to use his diagram.

DOI 10.1515/9781501504945-016

> fo 9v: *It͛ rec͛ ꝑ Marĝy M xiij tymƀ de cristy grey p̉c͛ le tȳƀ .vj. s͛*
> ‘Item, received from Margery M 13 timbers of cristy grey, price the timber: 6s’
>
> fo 10v: *John fflukke & John Joys de Wollewych doient en le .vj. io᷑ de Sept͛ pur j petit boot appelle Goodhale apayer al ffeste de Pascħ* […]
> ‘John Flukke and John Joys of Woolwich owe on the 7th day of Sept for 1 little boat called “Goodhale”, to pay at Easter […]’
>
> fo 14: *Meᵈ q̂ en le ij io᷑ de dit mois iay paye pur Rog͛ Ayscħbournhᵃm p᷑ le Scheryngg͛ de son lyu͛e achate de si taylo᷑ ijs͛ vijđ oƀ*
> ‘Memo, that on the 2nd day of the said month, I pay for Roger Ayschbournham for the shearing of his livery bought from his tailor, 2s 7d ½’
>
> fo 33v: *Meᵈ q̂ Joh ffluk de Wollewich doit en vaiłł de Seint Micħ p᷑ j verybot p᷑ chescun semain vijđ*
> ‘Memo, that John Fluk of Woolwich owes on St Michael’s eve for one ferryboat, for each week, 7d’
>
> fo 34: *Meᵈ q̂ Gybon Maufeld ad paye p᷑ John Gower Esquier a j Schippman p᷑ freit dune braspott mis ꝑ ŧre de lyne iesq̂s a loundrs*
> ‘Memo, that Gybon Maufeld pay by John Gower Esquire 1 shipman for freight of a brass pot, sent by letter from Lynn to London’
>
> (The National Archives, E101/509/19, Gilbert Maghfeld’s “A Merchant’s Account Book”)

The English components are:

names of people:	*Marĝy M, John fflukke/ffluk, John Joys, Rog͛ Ayscħbournhᵃm, Gybon Maufeld, John Gower*
names of places:	*Wollewich, lyne*
social ranks and titles:	*Esquier*
weights and measures:	*tymƀ* < OFr *timber* < Gmc ‘bundle of furs, usually 40 skins’
commodities (NPs):	*cristy grey,*[1] *boot, verybot, scheryngg͛, braspott*

---

[1] *Middle English Dictionary*, cristi-grei (n.) ‘A gray fur with tufts, or crests, of some kind’, first attested in 1378. See Wright (2002) for discussion of Gilbert Maghfeld’s placement of adjectives both fore and aft of the noun in the Noun Phrase.

Visual diamorphs (that is, words belonging to both English and French)[2] are: *It̕*, *rec̕*, *p̓c*, *&*, *Sept̕*, *ffeste*, *Me^{d}*, *lyu̓e*, *tayloꝰ*, *Seint Micħ*, *Esquier*, *freit*, *ƚre*. The distribution of the two languages is not random. Nouns, deverbal *-ing* forms and adjectives may optionally appear in English; but prepositions, conjunctions and pronouns must appear in the matrix language, which in this case is Anglo-Norman (Maghfeld could have chosen Medieval Latin as his matrix, as he occasionally did elsewhere in his account-book). So long as English words are assimilated into the Anglo-Norman text by taking inflections, by not taking up their native mutations, and by taking abbreviation and suspension signs which enable the reading of inflections as operative in both French and English, the morphological integrity of the base Anglo-Norman text is not compromised. This process was facilitated and enabled by visual diamorphs. The medieval abbreviation and suspension system optimised possibilities of visual diamorphs, and financial accounts show the highest frequency of visual diamorphs of any contemporaneous text-type. The principle of merging matter that can be merged was crucial to this variety, and this poses a challenge for theories of code-switching which depend on there being two distinctly-maintained opposing codes (e.g. Myers-Scotton 1993). Medieval business mixed-language writing is not alone in this; for example, a correlative phenomenon has been identified by Picone (1994) in spoken Louisianan French, where speakers produce forms that belong neither uniquely to English nor uniquely to French but to both at once, which Picone calls a "code-intermediate" state. Also relevant is Clyne's work (e.g. Clyne 1967, 1991, 2003) on words of similar phonetic shape in both languages triggering a switch from one to the other. Nonetheless, business mixed-language as written in medieval Britain is unlike the kind of data usually studied by synchronic linguists in that the switch-points are predictable, although not categorically implemented.

The mixed-language system changed over time, as do all languages in use. Ratios of English nouns premodified by English adjectives leading to multi-component English Noun Phrases increased gradually over the fourteenth and early fifteenth century. There was a tendency to move away from synthetic case inflexion and towards prepositional particles, following the drift towards analycity found in Medieval Latin, Anglo-Norman and English of the time. In the 1420s in the Latin-matrix accounts of London Bridge, the definite article *la* began to be used, although not signalling gender.[3] Here is an extract from a text at the tip-point (in Dorian's [1978, 1981] terminology, where a previously stable language undergoes a sudden change immediately before it dies out altogether), taken from a Port of London Medieval Latin-matrix customs account of 7 March 1481:

---

[2] The definition of visual diamorphs is the overlapping of two (or more) written codes into forms which are simultaneously both (or all) (see Wright 2011: 203).

> *Petro Segirʾ aƚ ꝑ j Cista j basket j ppᵅ & j wirkyn̂ Cû viij doß & ij ßtomħattꝭ vj brusshis .ijᶜ lƀ filyng vij doß fiƚ bloꝱ v doß & dī hatƭ Cardꝭ .vj parʾ volle Cardꝭ iiij Sôme nayle .xiiij doß pelƚ .ij panū depicƭ .xiiij doß ßtom̉hattꝭ ij gᵒs Cirotecꝭ & x hatter stockkys precʾ xxˡ̄ xiijˢ̂ iiijᵈ*
>
> 'Peter Segir, alien, 1 chest 1 basket 1 pipe 1 firkin with 8 doz. 2 Saint Omer hats, 6 brushes, 200 lbs filings, 7 doz. blue (or perhaps 'blood-coloured') thread, 5½ doz. hatters' cards, 6 pairs wool cards, 4 sums nails, 14 doz. skins, 2 painted cloths, 14 doz. Saint Omer hats, 2 gross gloves, 10 hatters' stocks, £20 13s 4d'
>
> (The National Archives, E122/194/25, translation in Cobb [1990: no. 93])

The English words are: *basket*, *wirkyn̂*, *ßtom̉hattꝭ*, *brusshis*, *filyng*, *hatƭ*, *volle*, *sôme*, *nayle*, *hatter*, *stockkys* and the visual diamorphs are: *aƚ*, &, *ƚƀ*, *bloꝱ*, *dī*, *cardꝭ*, *parʾ*, *doß*, *gᵒs*, *lī*, *ŝ*, *d*.[4] This leaves only monolingual Latin *Petro*, *ꝑ*, *cista*, *ppᵅ*, *cû*, *fiƚ*, *pelƚ*, *panū*, *depicƭ*, *cirotecꝭ*, *precʾ*, so that monolingual Medieval Latin is no longer the predominant language (at a ratio of 3:1, excluding numbers). The text consists of three-quarters of words that can be read as English and a quarter Latin, rather than the other way around. Here is another extract from the same day's entry:

---

[3] See Wright (2010a) for a discussion. In the [Medieval Latin + English] London Bridge House Estate records of the first half of the fifteenth century, the definite article *le* preceded a following English noun, and blocked a following Latin suffix on that noun. Definite article *la* occurred only in the prepostional phrase *de la*, without signalling a following feminine noun. In the fifteenth century [Anglo-Norman + English] London Merchant Taylors' Company documents, the definite article *le* signalled both a following English noun as well as a French one, but the definite article *la* was not usually followed by an English noun (French nouns could be qualified by both *le* and *la* in close proximity). Thus, gender was not marked by articles in this text type at this place and point in time.

[4] *volle card* antedates the *Oxford English Dictionary's* wool-card, *n.*, first attested 1564; *hatt(er(s card*, *hatter stock* have no *OED* entry, but cf. *OED* card, *n.1* "2. a. An instrument with iron teeth, used in pairs to part, comb out, and set in order the fibres of wool, hemp, etc., one of the cards being held in the hand, and the other fastened to a 'stock' or support."

> *Anthonio de la hay al ꝑ pp*^{a} *Cum .iiij. complet harneys . j barʾ Cû xxx parʾ wolle cardꝭ j bottom̂ of a basyn̂ j bar wyn lyes . j bras pottʾ . di barʾ Cû C ƚƀ fylingʾ p̉c x . ƚī . xj .*$^{\hat{s}}$ *iiij*^{d}
>
> 'Anthony de la Hay, alien, 1 pipe with 4 complete harnesses, 1 barrel with 30 pairs wool cards, 1 bottom of a basin, 1 barrel wine lees, 1 brass pott, half a barrell with 100 lbs filings, price £10 11s 4d'
>
> (The National Archives, E122/194/25, translation in Cobb [1990: no. 82])

The English words are: *complet harneys*, *wolle cardꝭ*, *bottom̂ of a basyn̂*, *wyn̂ lyes*, *bras pottʾ*; the visual diamorphs are: *ał*, *barʾ*, *parʾ*, *dî*, *ƚƀ*, *p̉cʾ*, *ƚī*, *s*, *d*; and monolingual Medieval Latin: *Anthonio*, *ꝑ*, *pp*^{a}, *cum*. The medieval system of Romance matrix-language plus English nouns, modifiers and stems of verbs is shifting. English premodified noun phrases are default, as opposed to earlier switching between English nouns and Latin or Anglo-Norman nouns, switching between English modifiers and Latin or Anglo-Norman modifiers, and switching between pre- and post-modification; and there is a prepositional phrase in English, *bottom of a basyn*, which is not part of the traditional system. It confirms that the tip-point has been reached. Previously, prepositions would always have been realised in the Romance matrix language, but here English is encroaching on function words too. The code-switched mixed-language system was about to be abandoned, and we see transgression of the switchpoint rules in the writing of the pre-shift generation, so our customs clerk of 1481 is just at that pre-shift point. However, because different institutions shifted from the mixed-language system to monolingual English in different decades between 1380 and 1480, the "pre-shift generation" cannot be dated to a specific set of twenty-five years; it took place gradually over more than a century, and several archives (including the Brewers', the Grocers', the Mercers' livery companies) show considerable toing and froing between mixed-language and monolingual English before eventually settling down to monolingual English.[5] Indeed several years' worth of monolingual English entries can be followed by a resumption of mixed-language writing, or sporadic return to Latin phrases and formulae, and all this characterises the tip-point, which looks a little different in each individual archive. It is not yet known how long this pre-shift disruption lasted, or whether it lasted for more than one generation in each case. The tip-point is not described in textbooks on the development of the history of the English language, and yet it is easily discernible, and there is plenty of evidence for it. The reason, presumably, is because only monolingual English texts have been deemed worthy of notice, even though mixed-language texts from the tip-point contain large amounts of sustained English.

---

[5] See Alcolado Carnicero (2014) and Metcalfe (2014) for descriptions of the Mercers' and Brewers' livery company records respectively, and Wright (2002) for analysis of the tip-point in the Mercers' archive. Miller (1997: 252–256) uses text from the tip-point of the Grocers' Company records as evidence of hybrid Anglo-Norman and English lexemes in order to build an argument about the preponderance of hybrid forms in Middle English.

By around 1500 the most laggardly, conservative institutions had finally tipped over into monolingual English (those in the van preceded them by a hundred years, so Gilbert Maghfeld's text of 1372–1395 was in keeping with all but the most radical of his contemporaries). When the tip-point had become a thing of the past, and the toing and froing between mixed-language accounts and monolingual English ones was over, monolingual English was the outcome – but it was not yet Standard English. Standard English is not regionally marked, and it does not admit of (much) spelling variation, whereas the kind of monolingual business English used in London immediately after the transition period was still southern in its morphology and graphies:

This is the acconte of Thomas kytson and Robart Browne Chyrche wardeins of the ꝑesch Chyrcħ of Seint Mary Mawdellens in Mylkestretʾ in london & Rule$^{\sigma}$ *is of the goods & ornaments that is to say from A*^{o} *dnī xv*^{C} *xix vnto anno dnī xv*^{C} *xx*^{ti} *which is a hole yere as apperyt by this Charge and discharge as heraftʾ ffollowith*

(Guildhall Library, London. MS 2596/1, fo 9. 1519–20. St Mary Magdalen Milk Street Cash Book, 1518–1606. Churchwardens' Accounts.)

Verb morphology is still southern *-th* (*apperyt*, *ffollowith*) in this cash-book, even though *-s* was a London variant by 1520, albeit at low frequencies, and spellings are mostly not those which would later become accepted as standard.[6]

In sum, the late medieval default was the business mixed-language system with its highly-regulated switchpoints, followed by a period of transition beginning in the late fourteenth century, culminating in transgressive switches at the tip-point, including shifts back and forth between monolingual English and mixed-language writing, followed by a tip to monolingual Proto-Standard English, all more or less completed by the late fifteenth century. It was a three-stage switchover: mixed-language, to transition period, to monolingual English. Then, there was a good two hundred years in which firstly elimination of variants took place, and subsequently selection of single, correct words and spellings (which process was still playing out over the eighteenth century).[7] Essentially, in mercantile writing in England before 1480, English was the foreign language, but because code-switching was default, English was visually present, and indeed obligatory in syntactically and semantically-regulated sites. But one needs to have a multilingual perspective in order to see it.

---

[6] Third-person indicative present-tense *-s* is found in London texts as early as the 1370s, but it remained a minority variant until the 1570s, not becoming dominant until the 1640s–1710s (Lass 1992: 138–139, 1999: 162–165; Nevalainen and Raumolin-Brunberg 1996: 104–107).

3 On textbook accounts of the origins of Standard English

None of this is mentioned in textbook accounts of the rise of Standard English. Wright (1996) surveyed textbook explanations of its origins, and pointed out some inconsistencies and unlikelihoods – which inconsistencies and unlikelihoods still tend to be repeated. The orthodox version goes like this: a pre-Standard written spelling-system called "Chancery Standard" evolved from either the East Midlands dialect (according to Ekwall [1956] and subsequent followers) or the Central Midlands dialect (according to Samuels [1963] and subsequent followers). It stemmed from scribes writing in the King's Office of Chancery, and it was the ancestor of Standard English. It was the result of wealthy and influential, although not numerous, East (or Central) Midlanders who had migrated to London, and caused the Londoners to shift their dialect. That this is an unlikely scenario can be argued thus:

Following earlier scholars Morsbach (1888) and Heuser (1914), Ekwall (1956: xi–lxviii) offered an explanation for fourteenth-century changes in London spellings for vowels in stressed syllables, which shifted from "Saxon" to "Anglian" (Wright 1996: 104). Ekwall wondered whether this change from Southern dialect to an apparently more Anglianised one might be due to Northerners' migration to London. To prove his theory one way or the other he culled locative surnames from early fourteenth-century London tax-lists and plotted them on a map to see where they came from. Obviously, this methodology is skewed: not all surnames are locative; of those that are, a parent, grandparent or godparent might have come from the place in question rather than the tax-list bearer; the bearer might have been an apprentice who received the locative surname from a master, having nothing to do with the master's family linguistically; and place-names often repeat, so that a Londoner (or ancestor, godparent or master) with *Sudbury* as part of their surname may have originated in Sudbury in Middlesex, or

---

[7] It is this latter stage where the new technology of printing, followed by the rise of prescriptivism, putatively had an effect.

Sudbury in Suffolk, with no way of knowing which. Ekwall was fully aware of all of these drawbacks:

> I came to the tentative conclusion that so far as can be judged from the Subsidy Rolls, the contribution to the London population from Southern counties about the beginning of the fourteenth century was larger than that from the Midland counties, and that the Midland character of the later London language could hardly be due to immigration on a larger scale from the Midlands than from the South (Ekwall 1956: xii–xiii).

This would seem to be categorical: "The aggregate figure for the Home Counties" (i.e. the counties around London – *LCW*) "is about 3,000 persons, thus about the same as that for the Midlands and the North taken together" (Ekwall 1956: lx). Ekwall was quite clear that his survey, imperfect as it was, did not support the theory of large numbers of immigrants from places North of London as an explanation as to why Northern spellings should have started to be used in London in the later fourteenth-century. "The question may then be raised whether it is probable that linguistic influence due to immigration from the Midlands and the North can have been sufficiently strong to affect the City dialect" (Ekwall 1956: lx). Ekwall suggested that the migration theory could still be salvaged, however, by positing post-1300 immigration of upper-class Midlanders into London. He assumed that mayors, aldermen and sheriffs would have belonged to the upper class (1956: lxii), and that wealthy, important and influential Midland and Northern mayors, aldermen, and senior livery-company men settled in London in the relevant period (albeit not in great numbers), and that they would not have understood Londoners' pronunciations:

> Pronunciations such as the Southern *eld*, *cheld*, for *old*, *cold* would not be readily understood in the Midlands and would be avoided. This accounts for the early disappearance of *cheld* from the London language. To sum up, the London language as we find it towards the end of the fourteenth century was a class dialect, the language spoken by the upper stratum of the London population (Ekwall 1956: lxiii).

A proto-sociolinguist, Ekwall identified some of these wealthy, influential and important Northern and Midland members of the Court of Aldermen and the Court of the Common Council, but he did not contextualise them against the more numerous and equally important members from London and elsewhere. Since 1956 there has been a wealth of sociolinguistic investigation revealing that speakers do not usually emulate the speech-ways of the upper classes (if being a City office-bearer or Common Councilman was indeed coterminous with being upper-class), and that the speechways of the population already *in situ* would have been more likely to have prevailed. Ekwall's theory is inherently unlikely, and he did, in fact, suggest an alternative: "the marked East Midland element in the London language may to some extent be bound up with the fact that this part of England was the old Danelaw" (Ekwall 1956: lxviii).

Seven years later, Michael Samuels published a highly influential paper which established the term "Chancery Standard" (Ekwall's phrase was "Westminster Standard") as the precursor of Standard English. Samuels was primarily interested in spelling variants of common words, because he was surveying monolingual English manuscripts in order to plot spelling variation across Britain, which was to culminate in the monumental *Linguistic Atlas of Late Medieval English* of 1986. He categorised fourteenth and fifteenth-century texts containing specific spelling practices into four "types". Type 1 consisted of religious Wycliffite writings of mainly Central Midland and Southern provenance. Type 2 spellings are found in a small group of fourteenth-century manuscripts produced somewhere near London which contain spellings commensurate with those found in the English Proclamation of Henry III, that is, of "early Essex-type" (Samuels [1963] 1989: 70). Type 3 spellings are found in the works of Chaucer, Gower and Hoccleve, and some of the civic documents in Chambers and Daunt (1931) and the wills in Furnivall (1882). Type 4 ("which I shall call 'Chancery Standard'") documents were those produced by the Office of Chancery, the royal secretariat; "that flood of government documents that starts in the years following 1430 [...] it is this type, not its predecessors in London English, that is the basis of modern written English" (Samuels 1989: 71). Samuels rejected Ekwall's suggestion of East Anglia (the wealthiest part of the Midlands, from which important and influential Midlanders might have been expected to come) as a likely provenance for these spellings, as Type 4 spellings do not match those in *A Linglistic Atlas of Late Mediaeval English* (*LALME*) from contemporaneous documents from Cambridgeshire, Norfolk or Suffolk. Instead, he proffered the Central Midlands as a more likely source, partly because many Type 1 documents came from there and there was a literary tradition, partly because the Central Midlands was in the centre of the country and hence not "peripheral" (an objection levelled against the dialect of Norfolk), and partly because the Central Midlands dialect was "progressive and easily understood all over the country" (1989: 74), as well as having good roads south.[8]

There are some assumptions here about what was and was not comprehensible outside a dialect area, about the progressive quality of the Central Midlands dialect, and about transparency in meaning being a property of a central (landlocked) location as opposed to the supposed opacity of the periphery (which logically ought to apply to all coastal dialects everywhere), but let us turn to why these two great and pioneering manuscript scholars (in whose debt I stand) should have made such claims in the first place. Their versions of events, which are essentially similar in seeking a locus of influence outside London as an explanation for substantial changes in late fourteenth-century London writing, were pre-variationist solutions. Subsequent work in sociolinguistics has shown that variation is a constant, and that dialects consist of ratios of variants, which are expected as default. If ratios of a given feature shift in favour of a feature found in a majority elsewhere, it is no longer automatically assumed that there must have been a movement of speakers from that area. Just as ideas can travel across relatively stable communities, so can linguistic features. The substantial fourteenth-century change perceived by Morsbach, Heuser, Ekwall and Samuels correlated not, I suggest, with a specific exodus of specific people from the Midlands (which has never been identified as such by historians), but with a massive change in Londoners' trading habits (which has),[9] causing an amount of dialect levelling,[10] and a corresponding major change in business writing habits.

---

[8] I have not summarised here my arguments against the claims of the next scholar in the field, John Fisher (1977, 1979); for these see Wright (1996: 108–109). For a fulsome refutation and rejection of Fisher's claims (and further elucidation of Samuels' types) see Benskin (2004).

4 Taking a multilingual perspective

Over the last quarter of a century an increasing number of scholars have been considering the dynamic relationship between Anglo-Norman and English, and Medieval Latin and English, and identifying various kinds of contact-induced change. William Rothwell has shown in a large number of publications that the later Anglo-Norman lexis in the British Isles was unlike that of France, that much of the Anglo-Norman word-stock was actually created in Britain and Ireland, and that many French words took on new meanings, unknown in France, in the British Isles.[11] Rothwell carefully documented this word by word, and his cumulative body of work shows that Anglo-Norman continued to be productive in Britain from the eleventh to the fifteenth centuries. Christiane Dalton-Puffer (1996) showed that various word-formation affixes hybridised in Britain, giving new hybrid forms not found on the Continent. Their findings beg the question: how did speakers in England come to create new French words so long after the Norman Conquest?

---

[9] Discussed in Keene (2000), Wright (2001, 2005, 2013) and not treated further here. Note the difference between exodus (speakers leaving region A and settling permanently in region B) and trading (repeated coming and going between regions A and B).

[10] Features that levelled include the reduction of adverbial *-liche* to *-ly* and the loss of regionally-marked verb plural indicative present-tense *-th*, *-n* and *-s* and subsequent adoption of zero. Regional *are*, the *th-* pronouns, verb third person singular *-s* and auxiliary *do* became unmarked universal forms by the end of the sixteenth century, with third person singular *-s* taking longer to go to completion. Another outcome of dialect contact is the creation of inter-dialect, where new features develop that did not previously exist: the new universal present participle *-ing* replaced regionally-marked *-and(e)*, *-end(e)*, *-ind(e)*.

Starting in 1992, I published a series of articles on mixed-language business writing, because that was what I found in archives when I went to look for what people in London, and indeed all England, actually wrote prior to the development of Standard English.[12] In a number of publications I have been figuring out the regularities of this code-switched mixed-language system, which acted as a written standard all over the country for keeping track of money transactions, at a time when the English dialects (which were very diverse) were not commonly written down. Accounts-keeping sounds like a very limited text-type, but keeping track of money was probably the main motivation for putting quill to parchment, of relevance to everyone who could write. Once alerted to the fact that code-switching was prevalent in business writing, Herbert Schendl searched for code-switched writing in other text-types, including those written before the Norman Conquest.[13] He established the fact that using French and Latin and English together was indeed a normal medieval practice across the board. The switchpoints, the size and type of the constituents, and the purpose of including other languages varied according to text-type and according to century, but the bald fact that medieval people did not usually write in monolingual English (or monolingual anything, for that matter) is starkly consistent. Written Medieval Latin is informed by the local spoken vernaculars, and it was not until the last quarter of the fifteenth century that any kind of consistent monolingualism emerged.

Richard Ingham documented the ways in which English Anglo-Norman syntax and morphology came to differ from Continental Anglo-Norman syntax and morphology, pinpointing the last generation before divergence as being that of the 1370s.[14] Children learning to write Latin at school from 1066 until towards the end of the fourteenth century learnt via the spoken medium of Anglo-Norman French. After that date, written Anglo-Norman in England displays the kind of grammatical levelling which occurs as the result of language acquired in adulthood rather than in childhood. Ingham deduced that the use of Anglo-Norman in England as a spoken vehicle for teaching in childhood must have ceased around the end of the fourteenth century, possibly as a result of lack of competent teachers, possibly mitigated by Black Death. Anyone who could write prior to the end of the fourteenth century was trilingual, and this continued to be the case into the fifteenth century, but without the same kind of native-like competence in Anglo-Norman. David Trotter, pondering the difficulties of deciding what was and was not French for the purposes of inclusion in the *Anglo-Norman Dictionary*, made the observation that *le/la/les* in Latin-matrix mixed-language writing signalled that the following Noun Phrase was not Latin; that is, that there was a different rule governing use of *le* in monolingual Anglo-Norman as used in France (where *le* changed to *la* and *les* according to the gender and number of the noun), to that governing use of *le* in mixed-language writing in England (where *le*, *la* and *les* did not mark gender and number, but gave meta-information about the make-up of the text).[15] I too found syntactic changes in mixed-language business writing around the end of the fourteenth century, with an increased usage of *le*, *la* and *les* in a Latin matrix and corresponding loss of Latin morphology. Thus, cumulatively and adducing different texts and different sorts of changes, a body of evidence noted by several scholars suggests that there was a sea-change in London writing in the final decades of the fourteenth century, which period of change lasted for a hundred years or so, until monolingual English became customary. What might have caused this period of change?

---

[11] For a partial bibliography see Trotter and Gregory (1997: xi–xiv); Schendl and Wright (2011: 19).

[12] See Wright (2011: 217–218) for a partial bibliography.

[13] See Schendl (2011: 93–94) for a partial bibliography.

[14] Ingham (2009a, 2009b, 2010a, 2010b, 2011a, 2011b, 2011c, 2013).

Professor Caroline Barron, a historian of fourteenth-century London, identified several monolingual English (or multilingual but with sustained passages in monolingual English) civic documents written in the 1370s and 1380s, a time when writing in monolingual English was still a radical choice.[16] One is a book of City of London ordinances in a late fifteenth-century copy of late fourteenth-century London English ("the Jubilee Book"); the others are ten English parish guild certificates of 1388. The book of ordinances caused great trouble at the time, and in March 1387, the Court of Common Council (a tier of the City of London's parliament) voted to burn it, causing an outcry from the City craft guilds. Barron hypothesises that the content survives in MS Trinity Coll. Cambridge 0.3.11, which is written in English and is of late fifteenth century date, but with sustained passages copied from a fourteenth-century original.[17] Writing in

---

[15] Wright (2010a, in press); Trotter (2010: 63). By "meta-information" I mean that *le* and *la* preceding the same noun in mixed-language texts signals "monolingual rules do not apply; this is not monolingual ML (or AN)" (Wright 2010a).

[16] Barron (2002), Barron and Wright (1995).

[17] Written by the Hammond scribe; see Wright (2012) for an analysis of negation strategies.

English presumably made the content more widely accessible, possibly aggravating the tensions and conflicts in London of the 1370s and 1380s. Barron finds that there was no one single cause for these conflicts ("The more I have studied London in the late fourteenth century, the more convinced I have become that there is no single 'big issue' that caused the turbulence of London in Richard II's reign" [Barron 2002: 1]). She cites shifts in patterns of overseas trade, rise in wages, growth of consumer spending, the drop in population, and rise in living standards all contributing to not just economic conflict, but perhaps also class conflict, and possibly long-term concern about the common good. These first sustained-English civic documents of the 1380s coincide with the first surviving use of passages of monolingual English in the City's records (three proclamations in Letter Book H from November 1383, following rioting at the October mayoral elections). These documents do not constitute a switch to English *per se*; rather they constitute the start of the period of transition, because civic documents continued to be written in Medieval Latin and Anglo-Norman for decades to come. It was not a top-down innovation, socially-speaking (as with the "influential and wealthy Midlanders coming to London" theory), nor was it exactly bottom-up (as events of 1381, the Peasant's Revolt, might lead one to expect). Rather, it was an amplification of access to information from being the territory of those who were literate in Medieval Latin and Anglo-Norman (to whatever degree), to those who were not. The citizens who belonged to the parish guilds and craft fraternities cut across the social scale, from City officials belonging to the grander livery companies, to local traders and craftsmen. Women belonged to parish fraternities, and women were traders. Presumably women would have been amongst the "symple & vnkonnyng men" who would have benefitted from the change to monolingual English.[18]

Finally, I mention the work of José Miguel Alcolado Carnicero, who analysed the records of the Mercers' Livery Company. The Mercers were one of the guilds who protested vigorously against the Jubilee Book reforms, presenting their petition known as the *Folk of Mercery* in monolingual English. Alcolado Carnicero shows that in the late 1300s and early 1400s the Mercers' Wardens wrote some of their records in monolingual Medieval Latin, and some in code-switched mixed-language. From 1440 onwards they wrote some of their records in mixed-language, and some in monolingual English. But between 1420 and 1440, everything they wrote (or at least, everything that has survived) was in mixed-language. Code-switching appears to have been their only linguistic system for these twenty years. They finally shifted to monolingual English in 1464.

---

[18] The phrase comes from the Mercers' complaint to the king known as the *Folk of Mercery*, 1386 (Chambers and Daunt [1931: 36]).

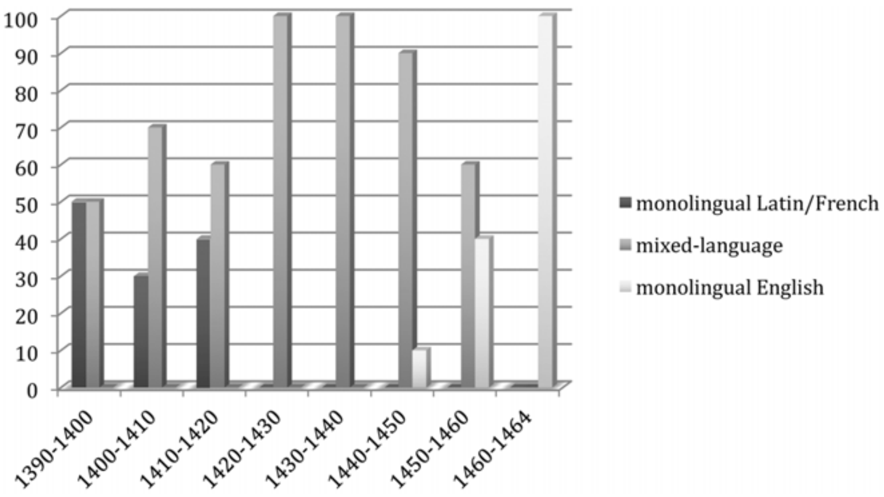


Figure 1: Percentages of Mercers' Wardens' Accounts written in mixed-language and monolingual language. Diagram adapted from Alcolado Carnicero (2013: 217)

A Mercers' clerk starting work in 1400, when the ratio was 30% monolingual Medieval Latin to 70% mixed-language documents, and ending his career fifty years later in 1450, when the ratio had shifted to 60% mixed-language and 40% monolingual English, might never have composed a monolingual document over a fifty-year career. The transition period was not brief. Resistance to change caused the shift from Latin to English to take a century to go to completion. It was not obvious in 1380 that a hundred years later, everyone would be keeping accounts and business documents in monolingual English. Alcolado Carnicero worked out who those early Mercers' Wardens adopters of language change were, and who their apprentices were (incorporators of more English), and who their apprentices were (incorporators of yet more English). He showed that small groups of specific Mercers led the shift from Latin and French to code-switching, and increased ratios of English. In subsequent work he has shown how loose-knit professional networks in general fostered innovation (Alcolado Carnicero 2015). This is a finer-grained approach to the "wealthy and influential citizens" theory of language change, and in this case, certain Mercers' Wardens may indeed have been wealthy and linguistically influential – but they were part of a trend, not its leaders, and other wealthy Mercers' Wardens were less influential. The ten London parish guilds who presented certificates in monolingual (or sustained) English in 1388/9 were self-confessedly not wealthy, nor were their members notably influential, but they were in the van of the shift to English.

Alcolado Carnicero has highlighted another difficulty with not taking a multilingual perspective: it causes at best unclarity, and at worst misinformation with regard to when the shift to English actually occurred. Just as Wright (1996) pointed out inconsistencies in the various textbook descriptions of the development of Standard English, Alcolado Carnicero (2015) points out inconsistencies in various datings of the shift to English. This is partly due to the Chinese-whispers effect of scholar repeating scholar and introducing error at some point along the chain, but mainly because there has been no distinction between the partial shift to English at the tip-point with subsequent documents in the archive then reverting to monolingual Latin, French or mixed-language; and total shift, at which the point of no return had been reached, and the tip-point passed.[19]

5 Conclusion

Pulling the above observations together, the transition period is characterised by:

a) morphological and phonological changes in monolingual London English between the fourteenth and fifteenth centuries (identified by Morsbach, Heuser, Ekwall, Samuels)
b) syntactic and morphological changes in British Anglo-Norman starting in the 1370s (identified by Ingham)
c) changes in mixed-language writing: *le/la* increase and commensurate loss of Latin morphology in the NP, increase in ratios of English NPs, increasing analytic word-order, and when the tip-point was finally reached (which varies from archive to archive), transgression of the syntactic switchpoint norms (identified by Wright)
d) the first monolingual English texts, or sustained passages of monolingual English, from the 1380s (e.g. those in Chambers and Daunt [1931]; the guild certificates and Jubilee Book text identified by Barron)
e) a period of code-switching as a norm, buffering between monolingual Medieval Latin and monolingual Proto-Standard English (identified by Alcolado Carnicero)
f) a period of political and social disruption in London in the 1370s and 1380s (of multiple causation, discussed by Barron)
g) a period of massive change in Londoners' trading patterns between the fourteenth and fifteenth centuries, from dealing with people in the surrounding hinterland, to dealing with countrywide networks and foreign merchants who made repeat visits to London (Keene [2000] and colleagues, summarised in Wright [2013]).

---

[19] See Wright (2015) for more on the recent work of the scholars discussed in this section.

Collectively, the transition period provides an answer to the question begged by Rothwell's finding that French words continued to be coined in England into the fifteenth century. A search of the *Oxford English Dictionary* yields 4,328 English words of French origin first attested between 1400–1500, a time when there were few mother-tongue speakers of French in England. Some of these may have been coined earlier, but this number allows the inference that French was still actively and productively used in England during this century. It is visible via the medium of mixed-language writing, and it is one of the reasons why mixed-language writing in general and the transition phase in particular is vital for understanding the development of Standard English. Standard English was an outcome of socioeconomic change, not in an "exodus from the Midlands" scenario, nor in an "imposition of Chancery house-style" scenario, but as a result of vastly increased weak-tie exchanges with people from elsewhere in Britain and the Continent. Levelling was one result (see footnote 10 above), but it is not the whole story. Trudgill (2009) sees non-levelled features such as third-person singular *-s*, or the inflected genitive phrase (*butlars frocke*, as opposed to northern *butlar frocke*) as an arrested – or perhaps just greatly delayed – drift towards analogical regularisation.[20] Hope (2000) sees it as a feature of written high-register language; that that which is high-register must be differentiated on the page from that which is low-register, and use of linguistically-unlikely features achieves this. But such matters were resolved considerably after the transition phase and the switch to Proto-Standard English. Proto-Standard English, as written around the country, still awaits detailed description.[21]

References

Manuscripts

Guildhall Library, London. MS 2596/1. St Mary Magdalen Milk Street Cash Book, 1518–1606, Churchwardens' Accounts.

The National Archive, Kew, MS E101/509/19, Gilbert Maghfeld: A Merchant's book of his debts, 25 January 1372–21 June 1395.

The National Archive, Kew, MS E122 194/25, Port of London Petty Customs Account. (Published in Cobb 1990)

Wren Library, Cambridge, MS Trinity Coll. Cambridge 0.3.11.

---

[20] For the uninflected genitive in London see Wright (2010b: 182–187); for a discussion of Trudgill's (2009) position, see Wright (2013). I am grateful to Julia Fernández Cuesta for raising the example of the inflected genitive phrase; see her discussions in Fernández Cuesta and Rodríguez Ledesma (2008, 2009).

[21] See Anita Auer's "Emerging Standards" project: http://www.emergingstandards.eu/.

Printed texts

Alcolado Carnicero, José Miguel. 2013. *Social networks and mixed-language business writing: Latin/French/English in the wardens' accounts of the Mercers' Company of London, 1390–1464*. Albacete: University of Castilla-La Mancha dissertation.

Alcolado Carnicero, José Miguel. 2015. Dating the shift to English in the financial accounts of some London livery companies: A reappraisal. *Multilingua*. 34(3). 373–404.

Barron, Caroline. 2002. The burning of the Jubilee Book. http://www.guildhallhistoricalassociation.org.uk/docs/The%20burning%20of%20the%20Jubilee%20Book%201376-1387.pdf (accessed 21 November 2014).

Barron, Caroline Mary & Laura Wright. 1995. The London Middle English guild certificates of 1388–9. *Nottingham Medieval Studies* 39. 108–145.

Benskin, Michael. 2004. Chancery Standard. In Christian Kay, Carole Hough & Irené Wotherspoon (eds.), *New perspectives on English historical linguistics: Selected papers from 12 ICEHL, Glasgow, 21–26 August 2002*, vol. II: *Lexis and transmission* (Current Issues in Linguistic Theory 252), 1–40. Amsterdam & Philadelphia: John Benjamins.

Chambers, R. W. & Marjorie Daunt. 1931. *A book of London English, 1384–1425*. Oxford: Clarendon Press.

Clyne, Michael. 1967. *Transference and triggering: Observations on the language assimilation of postwar German-speaking migrants in Australia*. The Hague: Martinus Nijhoff.

Clyne, Michael. 1991. *Community languages: The Australian experience*. Cambridge: Cambridge University Press.

Clyne, Michael. 2003. *Dynamics of language contact: English and immigrant languages*. Cambridge: Cambridge University Press.

Cobb, Henry S. (ed.). 1990. *The overseas trade of London Exchequer customs accounts 1480–1481* (London Record Society Publications 27). London: London Record Society.

Dalton-Puffer, Christiane. 1996. *The French influence on Middle English morphology: A corpus-based study of derivation* (Topics in English Linguistics 20). Berlin & New York: Mouton de Gruyter.

Dorian, Nancy C. 1978. *East Sutherland Gaelic*. Dublin: Dublin Institute for Advanced Studies.

Dorian, Nancy C. 1981. *Language death: The life cycle of a Scottish Gaelic dialect*. Philadelphia, PA: University of Pennsylvania Press.

Ekwall, Bror Eilert. 1956. *Studies on the population of medieval London*. Stockholm: Almqvist & Wiksell.

Fernández Cuesta, Julia & Maria Nieves Rodríguez Ledesma. 2008. Northern Middle English: Towards telling the full story. In Marina Dossena, Richard Drury & Maurizio Gotti (eds.), *English historical linguistics 2006*, vol. III*: Geo-historical variation in English* (Current Issues in Linguistic Theory 297), 91–109. Amsterdam & Philadelphia: John Benjamins.

Fernández Cuesta, Julia & Maria Nieves Rodríguez Ledesma. 2009. The northern echo: Continuities in contemporary Northern English. In Marina Dossena & Roger Lass (eds.), *Studies in English and European historical dialectology* (Linguistic Insights: Studies in Language and Communication 98), 157–189. Bern: Peter Lang.

Fisher, John Hurt. 1977. Chancery and the emergence of standard written English in the fifteenth century. *Speculum* 52(4). 870–899.

Fisher, John Hurt. 1979. Chancery Standard and modern written English. *Journal of the Society of Archivists* 6. 136–144.

Furnivall, Frederick J. (ed.). 1882. *Fifty earliest English wills in the Court of Probate, London: A. D. 1387–1439* (Early English Text Society Original Series 78). London: Oxford University Press.

Heuser, Wilhelm. 1914. *AltLondon: Mit besonderer Berücksichtigung des Dialekts*. Osnabrück: Liesecke.

Hope, Jonathan. 2000. Rats, bats, sparrows and dogs: Biology, linguistics and the nature of Standard English. In Laura Wright (ed.), *The development of Standard English 1300–1800: Theories, descriptions, conflicts* (Studies in English Language), 49–56. Cambridge: Cambridge University Press.

Ingham, Richard. 2009a. Mixing languages on the manor. *Medium Ævum* 78(1). 80–97.

Ingham, Richard. 2009b. The persistence of Anglo-Norman 1230–1362: A linguistic perspective. In Jocelyn Wogan-Browne (ed.), *Language and culture in medieval Britain: The French of England c.1100–c.1500*, 44–54. Woodbridge: York Medieval Press & The Boydell Press.

Ingham, Richard. 2010a. Later Anglo-Norman as a contact variety of French? In Richard Ingham (ed.), *The Anglo-Norman language and its contexts*, 8–25. Woodbridge: Boydell.

Ingham, Richard. 2010b. The transmission of Anglo-Norman. In Richard Ingham (ed.), *The Anglo-Norman language and its contexts*, 164–182. Woodbridge: York Medieval Press & The Boydell Press.

Ingham, Richard. 2011a. Code-switching in the later medieval English lay subsidy rolls. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English* (Topics in English Linguistics 76), 95–114. Berlin & Boston: Mouton de Gruyter.

Ingham, Richard. 2011b. Grammar change in Anglo-Norman and continental French: The replacement of non-assertive indefinite *nul* by *aucun*. *Diachronica* 28(4). 441–467.

Ingham, Richard. 2011c. The decline of bilingual competence in French in medieval England: Evidence from the PROME database. In Nils Langer, Steffan Davies & Wim Vandenbussche (eds.), *Language and history, linguistics and historiography: Interdisciplinary approaches* (Studies in Historical Linguistics 9), 71–292. Bern: Peter Lang.

Ingham, Richard. 2013. Language-mixing in medieval Latin documents: Vernacular articles and nouns. In Judith A. Jefferson & Ad Putter (eds.), *Multilingualism in medieval Britain (c.1066–1520): Sources and analysis* (Medieval Texts and Cultures of Northern Europe 15), 105–121. Turnhout: Brepols.

Keene, Derek. 2000. Metropolitan values: Migration, mobility and cultural norms, London 1100–1700. In Laura Wright (ed.), *The development of Standard English, 1300–1800: Theories, descriptions, conflicts* (Studies in English Language), 93–114. Cambridge: Cambridge University Press.

LALME = McIntosh, Angus, M. L. Samuels & Michael Benskin with the assistance of Margaret Laing & Keith Williamson. 1986. *A Linguistic Atlas of Late Mediaeval English*, 4 vols. Aberdeen: Aberdeen University Press.

Lass, Roger. 1992. Phonology and morphology. In Norman Blake (ed.), *The Cambridge history of the English language*, vol. 2: *1066–1476*, 23–155. Cambridge: Cambridge University Press.

Lass, Roger. 1999. Phonology and morphology. In Roger Lass (ed.), *The Cambridge history of the English language*, vol. 3: *1476–1776*, 56–186. Cambridge: Cambridge University Press.

Metcalfe, Caroline A. 2014. William Porlond, Clerk to the Craft and Fraternity of Brewers of London, 1418–1440. *Transactions of the London and Middlesex Archaeological Society* 65. 267–284.

Middle English Dictionary. http://quod.lib.umich.edu/m/med/.

Miller, D. Gary. 1997. The morphological legacy of French borrowed suffixes on native bases in Middle English. *Diachronica* 14(2). 233–264.

Morsbach, Lorenz. 1888. *Über den Ursprung der neuenglischen Schriftsprache*. Heilbroun: Henninger.

Myers-Scotton, Carol. 1993. *Duelling languages: Grammatical structure in codeswitching*. Oxford: Clarendon Press.

Nevalainen, Terttu & Helena Raumolin-Brunberg (eds.). 1996. *Sociolinguistics and language history: Studies based on the* Corpus of Early English Correspondence. Amsterdam & Atlanta: Rodopi.

OED = Oxford English Dictionary. http://www.oed.com.

Picone, Michael D. 1994. Code-intermediate phenomena in Louisiana French. In Katharine Beals, Jeannette Denton, Robert Knippen, Lynette Melnar, Hisami Suzuki & Erica Zeinfeld (eds.), *CLS 30: Papers from the 30th Regional Meeting of the Chicago Linguistic Society*, vol. 1, 320–334. Chicago: Chicago Linguistic Society.

Samuels, Michael Louis. [1963] 1989. Some applications of Middle English dialectology. *English Studies* 44. 81–94. Reprinted in Angus McIntosh, Michael Louis Samuels & Margaret Laing (eds.). 1989. *Middle English dialectology: Essays on some principles and problems*, 64–80. Aberdeen: Aberdeen University Press.

Schendl, Herbert. 2011. Beyond boundaries: Code-switching in the leases of Oswald of Worcester. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English* (Topics in English Linguistics 76), 47–94. Berlin & Boston: Mouton de Gruyter.

Schendl, Herbert & Laura Wright. 2011. Code-switching in early English: Historical background and methodological and theoretical issues. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English* (Topics in English Linguistics 76), 15–45. Berlin & Boston: Mouton de Gruyter.

Trotter, David A. 2010. Bridging the gap: The (socio)linguistic evidence of some medieval English bridge accounts. In Richard Ingham (ed.), *The Anglo-Norman language and its contexts*, 52–63. Woodbridge: York Medieval Press & The Boydell Press.

Trotter, David A. & Stewart Gregory (eds.). 1997. *De mot en mot: Aspects of medieval linguistics. Essays in honour of William Rothwell*. Cardiff: The University of Wales Press in conjunction with the Modern Humanities Research Association.

Trudgill, Peter. 2009. Vernacular universals and the sociolinguistic typology of English dialects. In Markku Filppula, Juhani Klemola & Heli Paulasto (eds.), *Vernacular universals and language contacts: Evidence from varieties of English and beyond*, 302–320. London: Routledge.

Wright, Laura. 1996. About the evolution of Standard English. In Elizabeth M. Tyler & M. Jane Toswell (eds.), *Studies in English language and literature: 'Doubt Wisely'. Papers in honour of E. G. Stanley*, 99–115. London: Routledge.

Wright, Laura. 2001. The role of international and national trade in the standardisation of English. In Isabel Moskowich-Spiegel Fandiño, Begoña Crespo Garcia, Emma Lezcano Gonzalez & Begoña Simal Gonzalez (eds.), *Re-interpretations of English: Essays on language, linguistics and philology*, vol. 1, 189–207. A Coruña: Universidade da Coruña.

Wright, Laura. 2002. Code-intermediate phenomena in medieval mixed-language business texts. *Language Sciences* 24. 471–489.

Wright, Laura. 2005. Medieval mixed-language business texts and the rise of Standard English. In Janne Skaffari, Matti Peikola, Ruth Carroll, Risto Hiltunen and Brita Wårvik (eds.), *Opening windows on texts and discourses of the past* (Pragmatics and Beyond New Series 134), 381–399. Amsterdam & Philadelphia: John Benjamins.

Wright, Laura. 2010a. A pilot study on the singular definite articles *le* and *la* in fifteenth-century London mixed-language business writing. In Richard Ingham (ed.), *The Anglo-Norman language and its contexts*, 130–142. Woodbridge: York Medieval Press & The Boydell Press.

Wright. 2010b. Eighteenth-century London non-standard spellings as evidenced by servants', tradesmen's and shopkeepers' bills. In Nicholas Brownlees, Gabriella Del Lungo & John Denton (eds.), *The language of public and private communication in a historical perspective*, 161–190. Cambridge: Cambridge Scholars Publishing.

Wright, Laura. 2011. On variation in medieval mixed-language business writing. In Herbert Schendl & Laura Wright (eds.), *Code-switching in early English* (Topics in English Linguistics 76), 191–218. Berlin & Boston: Mouton de Gruyter.

Wright, Laura. 2012. The Hammond scribe: His dialect, his paper, and folios 133–155 of Trinity College Cambridge MS 0.3.11. In Claudia Lange, Beatrix Weber & Göran Wolf (eds.), *Communicative spaces: Variation, contact, and change*, 227–258. Frankfurt am Main: Peter Lang.

Wright, Laura. 2013. The contact origins of Standard English. In Daniel Schreier & Marianne Hundt (eds.), *English as a contact language* (Studies in English Language), 58–74. Cambridge: Cambridge University Press.

Wright, Laura. 2015. On medieval wills and the rise of written monolingual English. In Javier Calle Martin and & Camilo Conde Silvestre (eds.), *Approaches to Middle English: Contact, variation and change*, 35–54. Frankfurt am Main: Peter Lang.

Wright, Laura. In press. On non-integrated vocabulary in the mixed-language Accounts of St Paul's Cathedral, 1315–1405. In Richard Ashdowne & Carolinne White (eds.), *Latin in Medieval Britain*. London: British Academy.

Index

abbreviations 227–228, 234–237, 243, 246–268, 294, 341
accommodation, *see* Communication Accommodation Theory
accounts, *see* text types
Anglo-Norman, *see* French
audience 44, 107, 127, 147–149, 156–157, 159, 174–175, 180, 195, 212–213, 305–306, 309–313, 322–325, 334–335

bare form 25, 231, 239
bilingualism 39–40, 49, 99, 148, 225–226, 275–280, 304, 319–320, 324, 327–328, 333–335
borrowing
– lexical 8, 47–49, 61–73, 78, 104–105, 133–134, 230–232
– nonce- 8, 78, 230
– vs. code-switching, *see* code-switching

charters, *see* text-types
clergy 13, 146–152, 157, 164, 174, 320–321
code-choice 89, 123, 136, 153, 173, 239, 277, 304, 310, 324, 331
code-intermediate 25, 341
code-mixing 7, 20, 230
code-switching
– definition of 3–4, 127, 319
– extrasentential 128, 135
– free 148, 159, 181, 189–190, 192
– functions of 30, 40, 42, 45, 54, 130, 148–149, 158–164, 186, 276–277, 299–316, 335
– intersentential 44, 130–131, 226, 228, 291–292, 321–322
– intrasentential 7, 125, 131, 133, 292, 321, 323, 327, 330
– prefabricated 44, 148, 164, 180, 193
– single-word 8, 48–52, 54, 61–65, 73, 77–78, 326
– tag switching 128, 136, 235
– vs. borrowing 8, 20, 47–50, 61–65, 78, 102, 207, 230–232, 288
Communication Accommodation Theory 302–308
congruent lexicalization 20, 78, 200, 226
content words 179, 232, 254–258, 262, 326, 334
contextualization cues 29
corpora 41–42, 50, 123, 134, 146, 150–153, 158, 165, 171, 173, 175–177, 184, 201–206, 246–247, 328, 331

diachronic change 20, 100, 134, 154, 164, 203
diamorphs 132, 226–240
– break 228–230
– homophonous 226, 238
– visual 179, 194, 210, 223–224, 226–228, 341–343
dictionaries 41, 48, 51, 56, 61, 65–69, 78, 84, 105, 113, 151, 280
drama, *see* text-types

education 39, 149–151, 172–175, 177, 190–195, 279–280, 309–313, 320, 349
embedded language 7, 25, 126, 128–129, 201, 208–214
emblems 227–228, 235–238, 240
English
– Chancery 345–347
– Early Modern 146, 173
– Late Modern 171–173, 175
– Middle 77, 80–94, 99–101, 108–115, 122–127, 134, 173, 226–227, 243–254, 260–265, 304, 322, 333
– Old 39–57, 99, 102, 105–106, 108, 122–125, 128, 134–135, 173
– Standard 344–347, 354

face-threatening act 307–308, 311–314
flagging 129, 131, 137, 283
foreignisms 48, 50, 55, 63–73
formulaic language 44, 80, 86, 130, 224, 235, 237–238, 256, 287, 291, 333, 343
French 73, 122–124, 133–134, 137, 174, 189, 304–305, 325, 327, 330, 334–335, 354

– Anglo- 77–78, 82–87, 90–92
– Anglo-Norman 26, 319–320, 324, 331–333, 339, 341, 348–351
– Insular 99–100, 319–320, 331–333
function words 20, 104, 179, 211, 227–228, 254, 256–257, 259–262, 267, 232, 326–330

genre, *see* text-types
German 20, 65, 149, 172, 194, 276, 279, 293
grammatical words, *see* function words
Greek 30, 149–150, 157, 172, 302–305, 309

Hebrew 149–150
homilies, *see* text-types

identity 4–5, 27–28, 30–31, 231, 278, 284, 302, 308, 320, 324, 333–335
in-group 226, 301–304, 308
indentures, *see* text types
ink, colour of 128, 132, 292, 294
intertextuality 127–128, 158–163, 245, 291, 309
Irish 105, 223–224, 228
Italian 25, 149, 172, 174, 187, 194

language choice, *see* code-choice
Latin 26, 39, 44–45, 77, 98, 122, 134, 149–150, 172, 199, 223, 243–244, 278–280, 288, 304–305, 311, 320–325, 335, 352
– Classical 207–208
– medieval 339, 341–342, 349, 351
– Neo- 200–203, 207
legal texts, *see* text types
lemmatiser 204, 206–207, 209
lexical gap 321, 333
lexical words, *see* content words
lingua franca 5, 149, 174, 278
loanwords, *see* borrowing, lexical

macaronic poems, *see* text types
manuscripts 106–110, 114–116, 127–132, 135–136, 223, 225, 244–246, 249–250, 254, 256–257, 260, 262–263, 265, 267, 280–286, 289, 308–309, 350
markedness 22–24, 29–30, 223, 344
martyrologies, *see* text-types
matrix language 7, 24–26, 31, 293, 321, 339
medical texts, *see* text-types
monolingual expectation 5, 127, 275

nonce borrowing, *see* borrowing
Norman Conquest 101, 121–122, 349

oral vs. written 21–24, 31, 147–148, 291, 319–321, 333–334

pamphlets, *see* text types
paraphrase 42–44, 158, 160
paratext 147
plays, *see* text types, drama
poems, *see* text types, verse
Polish 275, 277–280
politeness 306–308, 312, 314
Portuguese 66–67, 179
pronunciation 21, 65, 70, 72, 290, 346

quotation 42, 64, 86, 90, 93, 131, 160–163, 180, 314

readership, *see* audience
recipes, *see* text-types
repertoire 4–5, 15, 20–21, 24, 93, 195, 238, 260, 276
romances, *see* text-types

scientific treatises, *see* text-types
Scots 80, 86, 275, 277, 280
sermons, *see* text-types
social media 7, 21–22
Spanish 26, 30, 71, 173, 211
standardization 115, 172, 244, 248
status 87, 150, 173–174, 231, 277, 285, 305, 330, 335
suspensions 233, 243, 248, 341

technical terms 44, 46–47, 50, 54–55, 77, 93–94, 149, 245
TEI (Text Encoding Initiative) 204, 243
terminology 77, 79, 94, 132, 158, 163, 277
text-types
– accounts 243, 325–327, 331, 339–344
– administrative 79, 86, 275, 288, 320, 327–328

– business 25, 174, 226–227, 243, 292, 341, 349–350
– charters 40, 56, 106, 327–328, 331, 333
– drama 22, 25, 30, 176–177, 186–189, 308–311
– homilies 41–44, 125, 135, 146, 223, 232, 238
– indentures 332
– legal 78, 89–92, 174, 275, 277, 279, 287, 332–333
– letters 22–23, 27–29, 173–175, 268, 279, 304, 331
– macaronic poems 199–213, 216–219
– martyrologies 223, 225, 229
– medical 244–245, 248, 256–257, 262–265
– novels 22, 172–173
– pamphlets 148, 150
– petitions 78, 89–92
– recipes 243–254, 263–268
– religious 44, 109, 123, 125, 131–132, 148–151, 165, 225, 319–325, 333, 347
– romances 78, 89–92, 110, 325, 331
– scientific treatises 41, 44
– sermons 11, 127, 145–165, 225, 304, 321–325
– verse 44, 110–115, 202, 216–217
– wills 77, 78, 80, 84, 89–92, 106, 286, 347
trade 174, 223, 285, 339, 351
translanguaging 15, 30
translations 5, 39, 42, 83, 115, 135, 149, 245, 281, 325

vernacularization 149, 276
verse, *see* text-types

we-code 304, 310
weights 248, 265–268, 339–340
Welsh 97–98, 101–103, 105–112, 327

XML (eXtensible Markup Language) 203, 206, 216, 224, 243, 247, 268