

Investigating variation in late medieval English

Sune Gregersen
Amsterdam/Copenhagen

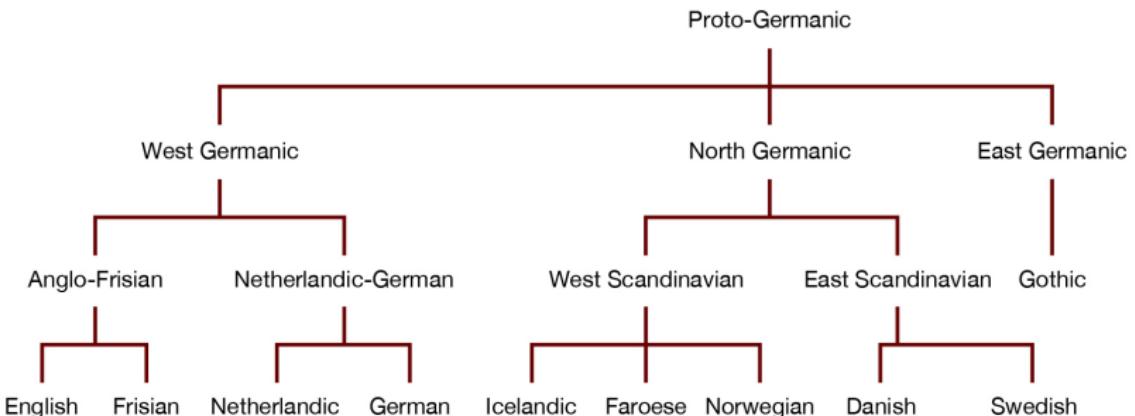
CIVIS course
Languages in Europe and their Diachronies
3 March 2021

Outline

- 1 A history of English (in 10 minutes)
- 2 Investigating (historical) language variation
- 3 Uncovering variation in Middle English
 - Phonology: initial fricative voicing
 - Grammar: Northern Subject Rule
 - Lexicon: the word *together*
- 4 Summary and discussion

Section 1

A history of English (in 10 minutes)



© Encyclopædia Britannica, Inc.

Figure 1: Germanic language family

Source: Encyclopædia Britannica

Periodization

- Old English (OE) *c. 800–1100*
- Middle English (ME) *c. 1100–1500*
 - Early (EME) *c. 1100–1350*
 - Late (LME) *c. 1350–1500*
- Modern English (ModE) *c. 1500–*
 - Early (EModE) *c. 1500–1700*
 - Late (LModE) *c. 1700–*

Old English (c. 800–1100)

- Comparatively good record (*c.* 3,000,000 words)
- Mainly religious texts, some epic and lyric poetry, laws, medical and other handbooks
- Three major corpora
 - Dictionary of Old English Web Corpus
 - York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE)
 - York-Helsinki Parsed Corpus of Old English Poetry (YCOEP)



Figure 2: Old English Hexateuch

British Library, Cotton MS Claudius B. iv, f. 7v



- [Simple Searches](#): Single word and phrase searches throughout the entire OEC
- [Boolean Searches](#): Find combinations of two or three words in a given paragraph, verse, or gloss
- [Proximity Searches](#): Find the co-occurrence of two or three words or phrases
- [Bibliography Searches](#): Identify works by author and title
- [Word Wheel](#): for Old English, Latin, and Runic
- [Word count broken down by text](#) || [Excel Spreadsheet](#)
- [About the DOE Web Corpus](#)

Figure 3: Dictionary of Old English Web Corpus

Middle English (c. 1100–1500)

- Norman conquest of England 1066–1075
- Almost no surviving English texts from the 12th century, from c. 1200 the record picks up again (mainly verse texts)
- From c. 1375 onwards substantial prose record – hundreds of manuscripts, many of them still unedited
- Several corpora available
 - Penn-Helsinki Parsed Corpus of Middle English (PPCME2)
 - Linguistic Atlas of Early Middle English (LAEME)
 - Corpus of Middle English Prose and Verse (CMEPV)
 - Corpus of Middle English Local Documents (MELD)
 - Middle English Grammar Corpus (MEG-C)

Penn-Helsinki Parsed Corpus of Middle English, 2nd edition, release 4

Citation

Kroch, Anthony, Ann Taylor, and Beatrice Santorini. 2000-. The Penn-Helsinki Parsed Corpus of Middle English (PPCME2). Department of Linguistics, University of Pennsylvania. CD-ROM, second edition, release 4 (<http://www.ling.upenn.edu/ppcme/ppcme-release-2016/PPCME2-RELEASE-4>).

Acknowledgments

We are grateful to the following institutions and individuals for their support and assistance:

- The National Science Foundation for NSF Grants BNS 89-19701 and SBR 95-11368.
- The University of Pennsylvania Research Foundation for supplementary support.
- The users of the Penn parsed historical corpora for their financial support in purchasing the corpus.
- Our colleagues at the University of Helsinki, especially to Prof. Matti Rissanen, the project's emeritus director, for his permission to use the Middle English part of the Helsinki Corpus of English and for his enthusiasm toward our project, though it was far from his own interests; also to Dr. Merya Kytö.
- Mike Collins, now of the Computer Science and Artificial Intelligence Laboratory at MIT, for permission to use his Penn dissertation parser and for adapting it to our needs.
- Beth Randall for writing the CorpusSearch program, without which the corpus would not be useable for research.
- Rhona Alcorn and Robert Truswell for their help in parsing a few recalcitrant passages.
- The users of the corpus who reported errors to us, notably Hezekiah Bacovcin, Aaron Ecay, Robert Truswell, and Joel Wallenberg, thereby allowing us to improve the quality of the corpus over time.

Figure 4: PPCME2 front page

```

( (CODE <P_662.C1>)
( (CODE <heading>))
( (NP (D A)
      (IN TREATISE)
      (PP (P ON)
           (NP (D THE) (N ASTROLABE)))) (ID CMASTRO,662.C1.3))
( (CODE </heading>))
( (IP-MAT (NP-VOC (ADJ Lyte)
                    (NPR Lowys)
                    (NP|PRN (PRO$ my) (N sone)))
        (, ,)
        (NP-SBJ (PRO I))
        (VBP aperceyve)
        (ADVP (ADV wel))
        (PP (P by)
             (NP (ADJ certeyne) (NS evydences)))
        (NP-OB1 (PRO$ thyn)
                (N abilite)
                (IP-INF (TO to)
                        (VB lerne)
                        (NP-OB1 (NS sciences)
                                (RRC (VAG touching)
                                     (NP-OB1 (NS nombres) (CONJ and) (NS proporciouns)))))))
        (. ;)) (ID CMASTRO,662.C1.5))
( (IP-MAT (CONJ and)
          (ADVP (ADVR as) (ADV wel))
          (VBP considre)
          (NP-SBJ (PRO I))
          (NP-OB1 (PRO$ thy)
                  (ADJ besy)
                  (N praier)
                  (IP-INF *ICH*-1))
          (PP (P in)
              (ADJP (ADJ special)))
          (IP-INF-1 (TO to)
                    (VB lerne)))

```

Figure 5: PPCME2 file (Chaucer's *Treatise on the Astrolabe*)

Modern English (c. 1500–)

- Printing press in London (William Caxton, c. 1476)
- Standardization of the written language
- Expanding range of text genres, e.g. writings of a political and personal nature, novels, newspapers and magazines, etc.
- Colonial expansion and development of New Englishes

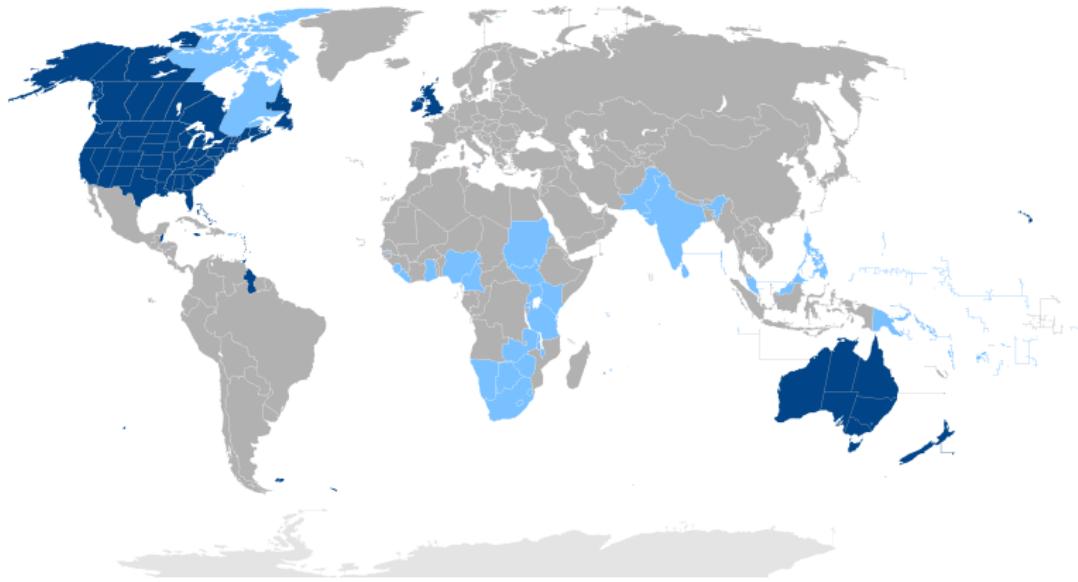


Figure 6: English-speaking world today

Source: Wikimedia Commons



Figure 7: English-speaking world, c. 1400

Original map from Wikimedia Commons

Middle English as a “diverse” period

“ME is, *par excellence*, the dialectal phase of English, in the sense that while dialects have been spoken at all periods, it was in ME that divergent local usage was normally indicated in writing.”

Strang (1970, p. 224)

Middle English as a “diverse” period

“the ME period is, notoriously, the time when linguistic variation is fully reflected in the written mode. Thus the *Linguistic Atlas of Late Mediaeval English* [...] records no fewer than five hundred ways of spelling the item THROUGH”

Horobin and Smith (2002, p. 33)

Middle English as a “diverse” period

“Middle English: Dialects and Diversity”

Title of Corrie (2017)

Middle English as a “diverse” period

“Inde est quod Mercii sive Mediterranei Angli, tanquam participantes naturam extremorum, collaterales linguas arcticam et antarcticam melius intelligant quam adinvicem se intelligunt iam extremi. Tota lingua Northimbrorum, maxime in Eboraco, ita stridet incondita, quod nos australes eam vix intelligere possumus” (Higden, *Polychronicon*, c. 1350)

“þerfore it is þat Mercii, þat beeþ men of myddel Engelond, as it were parteners of þe endes, vnderstondeþ bettre þe side langages, norþerne and souþerne, þan norþerne and souþerne vnderstondeþ eiþer oþer. Al þe longage of þe Norþumbres, and specialliche at ȝork, is so scharp, slitting, and frotynge and vnschape, þat we souþerne men may þat longage vnneþe vnderstonde.” (Trevisa, translation of the *Polychronicon*, c. 1385)

(Babington, 1869, pp. 160–163)



Figure 8: Middle English dialect areas

Source: Lass, 1992

Section 2

Investigating (historical) language variation

Variationist linguistics

- Dialectology (from mid-19th century)
 - Focus on geographic (“diatopic”) language variation
 - Identification of boundaries between features (**ISOGLOSSES**)
 - Traditionally little interest in individual variation
- Sociolinguistics (from 1960s onwards)
 - Focus on all kinds of language variation – e.g. relating to social class/group, gender, sexuality, ethnicity ...
 - Much interest in individual and “intra-speaker” variation
- Both traditions investigate the different realizations (**VARIANTS**) of linguistically variable features (**VARIABLES**)

Dialectology – example

- Survey of English Dialects (SED)
 - University of Leeds, 1950s–1970s
 - Collection of words and dialect speech from elderly rural speakers
 - Questionnaire (some 1300 lexical items) and audio recordings
 - 313 localities in England
- Example of dialect feature: initial fricative voicing
 - Syllable-initial /f, s, θ, ð/ → /v, z, ð, ʒ/
 - E.g. recording from 1964 from Blackawton, Devon (British Library, C908/31)
 - *fifteen* [vɪfti:n]
 - *something* [zʌmðɪŋ]
 - *ourselves* [ə:zəlvz]



Figure 9: Initial fricative voicing in England

Source: Wikimedia Commons, based on Upton and Widdowson (2006)

Sociolinguistics – example

- Peter Trudgill, *The social differentiation of English in Norwich* (1974)
 - Investigation of four ‘speech styles’ in five social classes
 - Variation sensitive to both class and speech style
- Example of variable: realization of *-ing*: /ɪn/ or /ɪŋ/

Social class	Word list	Speech style		
		Reading	Formal	Casual
Middle middle	0	0	3	28
Lower middle	0	10	15	42
Upper working	5	15	74	87
Middle working	23	44	88	95
Lower working	29	66	98	100

Table 1: Incidence of /ɪn/ (percentages)

Challenges in variationist linguistics

- “Corpus size” and data collection
- Data quality (see e.g. Labov, 1972)
 - Observer’s paradox
 - Reliability of acceptability judgements
- Defining the variable
 - Especially a problem for syntax

Why study historical language variation?

- Historical variation may help explain present-day distribution
- Information about textual origins useful e.g. for historians and literary scholars
- Language change always involves variation (as a minimum between an old and a new form)
 - Hence, the study of earlier variation may help explain why certain changes happen (or not)
- It's really interesting

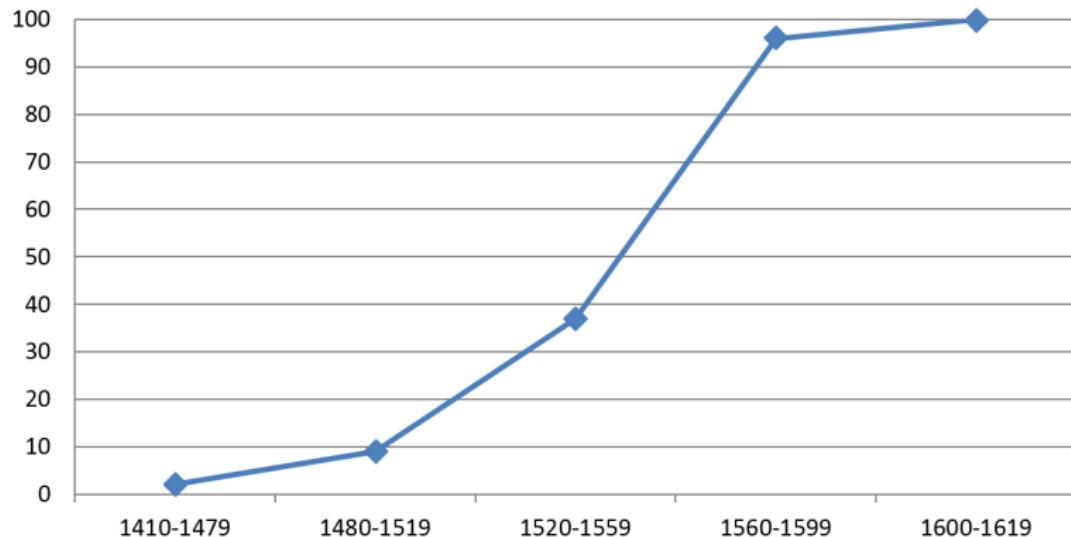


Figure 10: *NOM you* (vs. *ye*) in CEEC

Nevalainen, 2015, Fig. 3

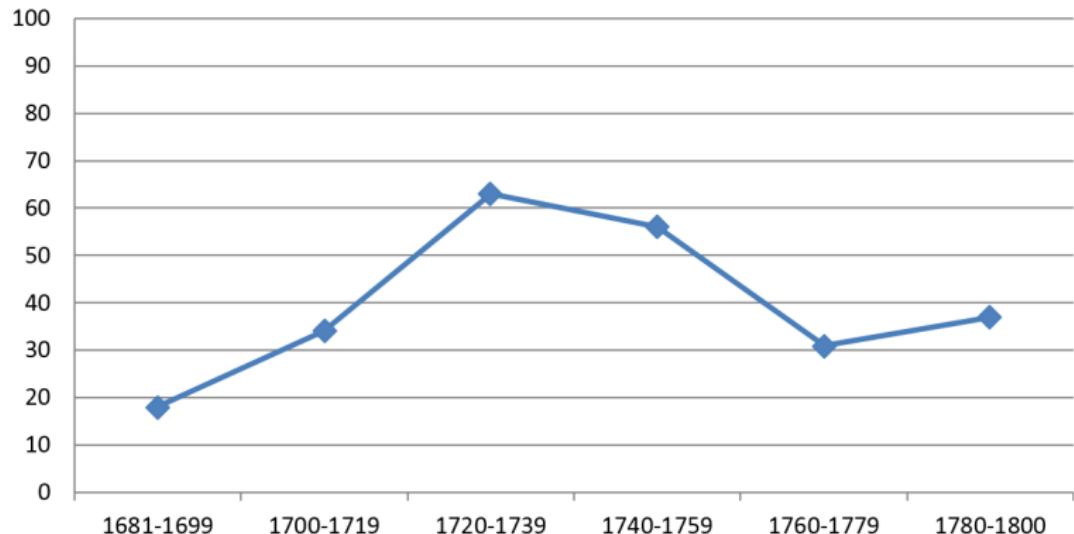


Figure 11: *you was* (vs. *you were*) in CEEC

Nevalainen, 2015, Fig. 4

Challenges in **historical** variationist linguistics

- Corpus size
- Data quality
- Defining the variable
- **No native speakers**
- **Lack of metadata**
 - Most medieval texts anonymous – serious problem for historical sociolinguistics
 - We often don't know the exact origins of texts – serious problem for historical dialectology

Section 3

Uncovering variation in Middle English

Sources

- ‘Public’
 - Official records
 - Religious texts (sermons, prayers, bibles)
 - Works of literature (usually verse)
- ‘Private’ (only late ME)
 - Letters
 - Diaries
 - Commonplace books
- All of these are handwritten

Tools

- Linguistic Atlas of Late Mediaeval English (LALME)
 - Covers *c.* 1350–1450
 - Questionnaire-based: “Linguistic Profile” with common items for each scribal text
 - 4 volumes originally published 1986
 - Online version (eLALME) allows users to generate maps themselves
- Linguistic Atlas of Early Middle English (LAEME)
 - Covers *c.* 1150–1325
 - Online corpus of diplomatic transcriptions
 - Does double duty – both a linguistic atlas and a searchable corpus
- Other tools under development (e.g. Corpus of Middle English Local Documents, MELD)

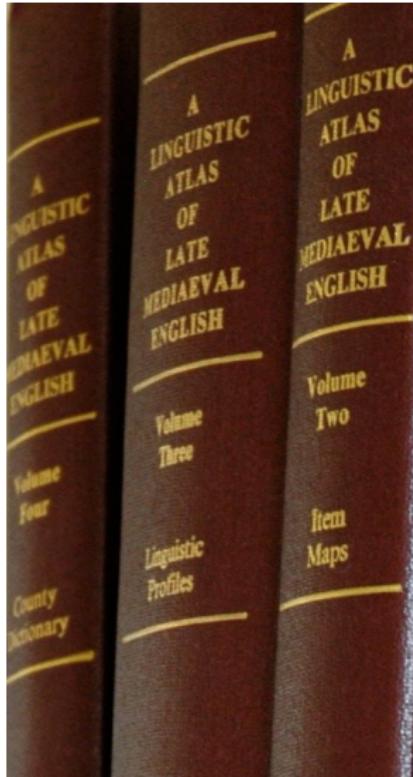


Figure 12: Original LALME

LP 512

Cambridge, Gonville and Caius College 175/96. First half 15th cent. MS in one hand. pp. 156: secular and religious poetry. Analysis from pp. 1-35, 107-118. LP 512. Grid 510 316. Lincolnshire.

1	THE:	þe ((the))
2	THESE:	þese ((þyse))
3	THOSE:	þoo
4	SHE:	sche
5	HER:	her, here
6	IT:	it
7	THEY:	þey ((þay))
8	THEM:	hem ((hem))
9	THEIR:	her ((here))
10	SUCH:	swylke, swylk
11	WHICH:	wyhche, þe-wyhche, whyche, þe-whylke
12	EACH:	ylke-a, ylke ((eche-a, ilke, euerylkone, euerylkon, ylkon))
13	MANY:	many, manye

Figure 13: eLALME, Linguistic Profile 512

Oxford, Bodleian Library, Laud Misc 471, fols. 128v-133v: Kentish Sermons

Date: C13b2

Location: 583 153

{=Central Kent=}

{=Sample represents all the text in English in this hand=}

{=Script - Textura semiquadrata=}

{=Status - MS punctuation done; tagging notes and textual notes up to date=}

{~f128va~}

{=English text starts half way down first column=} {'(s(cr)mo {-}}

In die epiphanie .(.) {(CVm natus esset ih(esus) i(n) bet(-)}

leem iude in dieb(us) hero{-}

dis regis ecce magi ab {-}

oriente ueneru(n)t ierosolima(m) {-}

dicentes . Vbi e(st) qui natus est rex iudeor(um) ?() {=Three-line coloured initial *C in *C*VM followed by coloured *V=} ¶ We redeth i þo ¶

holi godespelle of te-dai ase ¶

ure louerd god-almichti i{-}

-bore was of ure lauedi sei(n)te ¶

marie i þe cite of bethleem . ¶

Figure 14: LAEME, corpus file 142 (laud471kst)

Challenge 1: patchy record and lack of metadata

- Areas are not equally well-represented
 - ... and texts from different areas are not necessarily contemporaneous
- Medieval texts usually a) anonymous and b) not explicitly localized
 - “It is rather as if the compilers of a modern dialect atlas had access to any number of speakers, all willing to be interviewed but very few of whom divulged where they came from.” (LALME introduction, §2.3.1)
 - LALME solution: use localized texts as ‘anchor’ texts, then ‘fit’ other texts based on linguistic variables
 - LAEME solution: use ‘anchor’ texts and LALME distribution to fit other texts – unlocalizable texts included in a separate subcorpus

Challenge 2: medieval scribal practices

- Most surviving texts have gone through several rounds of copying
- Medieval scribes had different copying habits
 - ‘Literatim copyists’ copy texts exactly
 - ‘Translators’ rewrite texts in their own dialect
 - A third type does something inbetween, resulting in a MISCHSPRACHE consisting of elements from different dialects (Benskin and Laing, 1981)
- Sometimes it is possible to tease apart distinct dialect ‘layers’ in a text
 - ... at other times the texts are too mixed

Three brief case studies

- Phonology, grammar, and lexicon
- All three examples based on eLALME
 - More texts and easier to use than LAEME
 - ... but LAEME is actually more powerful in many ways

Phonology: initial fricative voicing

- Known from dialects of Modern English (southwestern England)
 - In fact, used by Shakespeare to make characters seem ‘rustic’
- Relatively frequent in Late Middle English sources, and included as a separate item in LALME
 - Some 14th-century examples from the *MED*: *vrendes* ‘friends’, *visch* ‘fish’, *alzuo* ‘also’
- Comparison of ME and ModE shows that the [+voiced] variant has receded

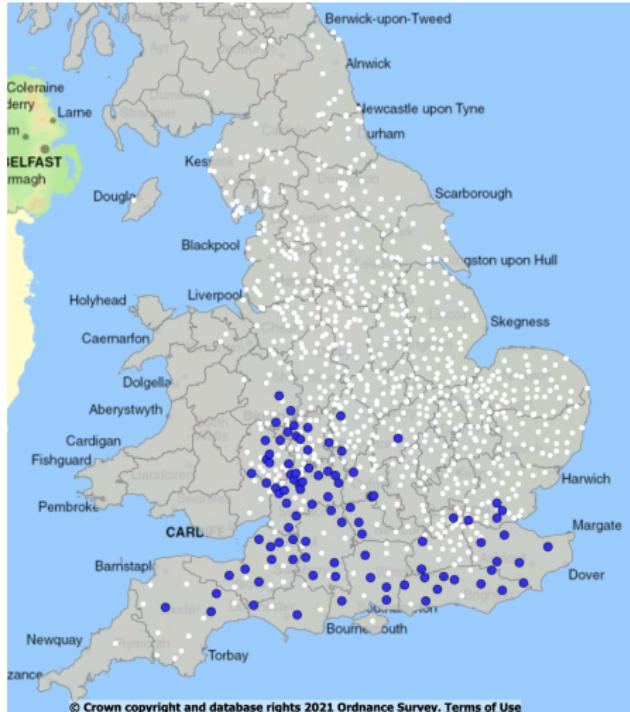


Figure 15: Initial fricative voicing in Late Middle English

eLALME, item 411



Figure 16: Initial fricative voicing in Modern English

Source: Wikimedia Commons, based on Upton and Widdowson (2006)

Grammar: Northern Subject Rule

- Different rules of subject-verb agreement in Northern dialects
 - Verbal present-tense ending always -s
 - ... except when the verb appears next to a 1SG or PL pronoun
- Feature not surveyed in LALME, but one can use a proxy
 - LALME item 160-50: PL forms of *have*

Grammar: Northern Subject Rule

Person	V next to PRON	Otherwise
1SG	-	-s
2SG	-s	-s
3SG	-s	-s
PL	-	-s

Table 2: The Northern Subject Rule

- (1) For we **knawe** wele þat oure goddis alwaye **helpes** vs
'For we know well that our gods always help us' (PLAlex. 43)
- (2) I couthe noȝte see that, þat now, thurgh scharpenesse of mekenes and
mescheffe, I **see** clerely & **knawes**
'I could not see that which now, because of the bitterness of
humiliation and misfortune, I see and know clearly' (PLAlex. 55)

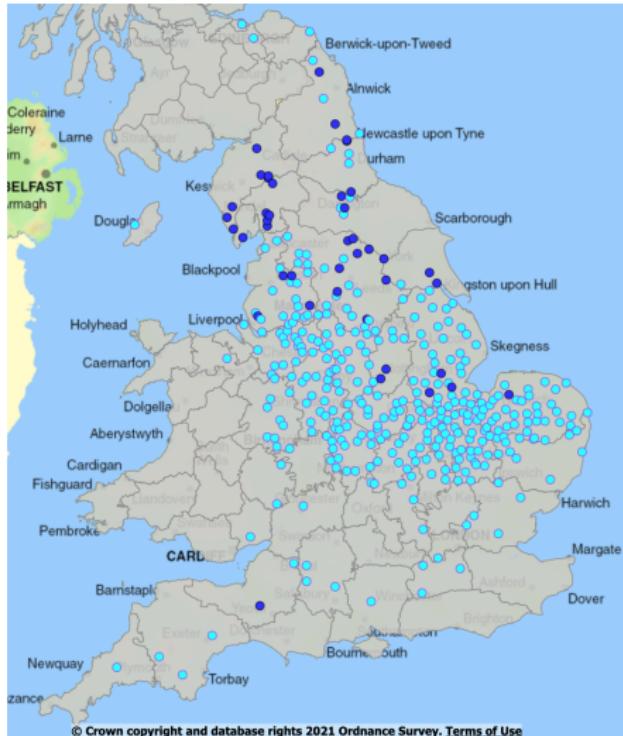


Figure 17: PRS.PL of *have*, forms with -s (dark blue)

Lexicon: the word *together*

- Variation in ModE between words with/without ‘adverbial -s’
 - E.g. *beside/besides, forward/forwards, toward/towards*
- In earlier English also in the lexeme TOGETHER
 - MED records spellings such as *to-gideris, togideres, to gederys, to gybers*

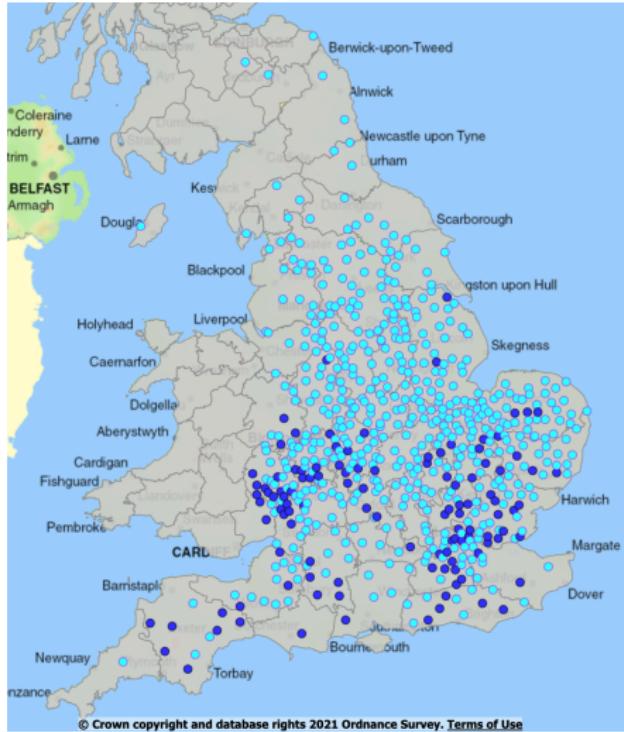


Figure 18: TOGETHER, spellings with -s (dark blue)

eLALME, item 268

Section 4

Summary and discussion

Summary

- Middle English universally recognized as a linguistically “diverse” period – good testing ground for historical variationist methods
- Insights and methods from modern dialectology and sociolinguistics can be applied to earlier periods
 - Systematic study of linguistic **VARIABLES** and their **VARIANTS**
 - Questionnaire-based and corpus-based approaches
 - Mapping of variants in space enables comparison with present-day distributions
- Some challenges are specific to historical variationist linguistics
 - Absence of native speakers
 - Lack of metadata (e.g. authors, geographic origin)
 - ‘Mixed’ scribal dialects (**MISCHSPRACHEN**)

Other avenues for research

- Variation relating to gender and social status
 - E.g. with letter collections from late 15th century
- Variation in language choice
 - Many bilingual (ME/Latin or ME/Anglo-Norman) or even trilingual texts
 - See e.g. recent volume ed. by Laura Wright, *The multilingual origins of Standard English* (2020).
- Stylistic, pragmatic, and genre-related variation

Discussion

- What role, if any, does linguistic variation play in your own studies and/or research?
- Have you noticed any variable features in the language(s) that you work on?
 - If yes: how might one investigate these variables?
- Any other thoughts/comments/ideas?

References I

- A *Linguistic Atlas of Late Mediaeval English* (2013). electronic. Edinburgh: University of Edinburgh. URL:
<http://www.lel.ed.ac.uk/ihd/elalme/elalme.html>.
- Babington, Churchill (1869). *Polychronicon Ranulphi Higden monachi Cestrensis*. Vol. 2. London: Longmans, Green, and Co.
- Benskin, Michael and Margaret Laing (1981). "Translations and Mischsprachen in Middle English manuscripts". In: *So meny people, longages and tonges. Philological essays in Scots and Mediaeval English presented to Angus McIntosh*. Ed. by Michael Benskin and M. L. Samuels. Edinburgh, pp. 55–106.
- Corpus of Early English Correspondence* (1998). Helsinki: Department of Modern Languages, University of Helsinki.
- Corrie, Marilyn (2017). "Middle English: Dialects and diversity". In: *The Oxford history of English*. Ed. by Lynda Mugglestone. 2nd ed. Oxford: Oxford University Press, pp. 106–146.

References II

- Furnivall, Frederick James (1868). *The Ellesmere MS. of Chaucer's Canterbury Tales*. London: Trübner.
- Horobin, Simon and Jeremy Smith (2002). *An introduction to Middle English*. Edinburgh: Edinburgh University Press.
- Labov, William (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Laing, Margaret, ed. (2013). *A Linguistic Atlas of Early Middle English, 1150–1325*. Version 3.2. URL:
<http://www.lel.ed.ac.uk/ihd/laeme2/laeme2.html>.
- Laing, Margaret and Roger Lass (2006). “Early Middle English dialectology. Problems and prospects”. In: *The handbook of the history of English*. Ed. by Ans van Kemenade and Bettelou Los. Malden, MA: Blackwell, pp. 417–451.
- Lass, Roger (1992). “Phonology and morphology”. In: *The Cambridge history of the English language*. Ed. by Norman Blake. Vol. 2. Cambridge: Cambridge University Press, pp. 23–155.

References III

- Lass, Roger (2004). “*Ut custodiant litteras*: Editions, corpora and witnesshood”. In: *Methods and data in historical dialectology*. Ed. by Marina Dossena and Roger Lass. Bern: Peter Lang, pp. 21–48.
- Middle English Dictionary* (1952–). Ann Arbor: University of Michigan. URL: <https://quod.lib.umich.edu/m/med/>.
- Milroy, James (1992). “Middle English dialectology”. In: *The Cambridge history of the English language*. Ed. by Norman Blake. Vol. 2. Cambridge: Cambridge University Press, pp. 156–206.
- Mustanoja, Tauno F. (1960). *A Middle English syntax. Part 1: Parts of speech*. Helsinki: Société Néophilologique.
- Nevalainen, Terttu (2015). “Descriptive adequacy of the S-curve model in diachronic studies of language change”. In: *Can we predict linguistic change?* Ed. by Christina Sanchez-Stockhammer. Studies in Variation, Contacts and Change in English 16. Helsinki: Varieng. URL: <https://varieng.helsinki.fi/series/volumes/16/nevalainen/index.html>.
- Strang, Barbara M. H. (1970). *A history of English*. London: Methuen.

References IV

- | Upton, Clive and J. D. A. Widdowson (2006). *An atlas of English dialects.* Abingdon: Routledge.
- | Wright, Laura, ed. (2020). *The multilingual origins of Standard English.* Berlin: De Gruyter Mouton.



Thank you!
... and feel free to get in touch at s.gregersen@hum.ku.dk

Image source: British Library, Royal MS 17 E III, f. 145

Variation in Middle English syntax

- Usually acknowledged, but not dealt with systematically
 - “The assessment of dialectal features in ME syntax is made difficult by the lack of special studies. It seems that this important aspect of ME dialectology has received less attention than it deserves.” (Mustanoja, 1960, p. 41)
 - “It is very doubtful how far the corpus will yield a regional syntax [...] it may well be that syntax will performe remain the Cinderella of Middle English dialectology.” (LALME introduction, §3.10)

Chaucer's 'yonge poure clerkes'

"Iohn highte that oon / and Aleyn heet that oother
Of o toun were they born / that highte Strother
ffer in the North / I kan nat telle where [...]
'Symond,' quod Iohn / 'by god, nede has na peer
Hym boes serue hym **selne** / that has na swayn
Or elles / he is a fool / as clerkes sayn
Oure Manciple / I **hope** / he wil be deed
Swa **werkes** ay / the wanges in his heed
And forthy / **is** I come / and eek Alayn
To grynde oure corn / and carie it **ham** agayn
I pray yow / spedē vs heythen that ye may."

The Reeve's Tale, Ellesmere MS. (c. 1405), quoted from Furnivall
(1868, pp. 115–116)

An example of Northern Middle English

“Valerius tellis þat when Socrates þe wyse clerke wexid **alde** and taght his scolers, he wold not alway bynd **þaim** in at lernyng, bod som tyme to avanyssh away þer labur he wolde suffer þaim to have recreacion, to cauce þaim to be more dyligent & craftie in þer lernyng afterward, **at þai sulde** not be yrke of þer labur. And herefor þis wise philosophur Socrates [usyd] for to lope on a rede betwix his leggis, as **barnys rydys**, and ryde with þaim as þai do to make þaim sporte.”

Alph.Tales 671

An example of Northern Middle English

Alph.Tales 671	'Southern' ME	ModE gloss
<i>alde</i> (cf. German <i>alt</i>)	<i>olde</i>	'old'
<i>þaim</i> (3PL.OBL)	<i>hem</i>	'them'
<i>at</i> (COMP)	<i>þat</i>	'that'
<i>sulde</i>	<i>sholde</i>	'should'
<i>barnys</i>	<i>childre, children</i>	'children'
PL + <i>rydys</i>	<i>ryde, ryden</i>	'ride'

Table 3: Some Northern features in Alph.Tales 671