

# How do you spell that? Intra-document variation in Middle English document

Lynne Cahill  
University of Sussex

# Outline

- Background: Middle English, document collections
- Standardisation:
  - as a general linguistic process
  - as applied to spelling
  - two key indicators: intra-document variation (IDV) and reduction in variants
- Corpora used: MELD, MEG-C, (ACE, LALME)
- Data analysis:
  - one frequently occurring word with four widely used variants: *said*
  - one frequently occurring suffix with two widely used variants: *-ing*
- Findings

# Middle English

- English divided into approximate periods
  - Old English (850-1150)
  - Middle English (1150-1500)
  - Modern English (1500-1800)
- Further sub-divisions (early, mid, late) but no clear boundaries
- Focus on Middle English
- Period characterised by influence of French, loss of inflections and emergence of official use, replacing Latin and French
- Availability of increasing numbers of fully digitised documents

# Middle English documents

- Various collections available
- Many originally transcribed, printed, digitised in later times
- Useful for historians, not for linguists (especially interested in spelling)
- LALME: Linguistic Atlas of Late Middle English
- ACE: Anthology of Chancery English
- MELD: Middle English Local Documents
- MEG-C: Middle English Grammar Corpus

# Middle English documents

- Various collections available
- Many originally transcribed, printed, digitised in later times
- Useful for historians, not for linguists (especially interested in spelling)
- LALME: Linguistic Atlas of Late Middle English
- ACE: Anthology of Chancery English
- **MELD: Middle English Local Documents**
- **MEG-C: Middle English Grammar Corpus**

# Standardisation

- Process by which a variety, a set of features, or a single feature become accepted as having a higher status
- Not (only) linguistic factors
- May be socio-political or practical; may be bottom up or top down
- Theories of broad language standardisation (e.g. Haugen 1997) usually focus on selection of one variety
- Propose stages of elevation and acceptance as standard
- E.g. Standard Southern British English, emerging from the speech of higher status social groups in the South-East of England

# Standardisation of spelling

- Standardisation of spelling different
- Accepted writing system/orthography of a language usually abstracts over different social or regional varieties
- Spelling standards/conventions more usually top down
  - emerge from dictionaries, education
  - may even be decreed by authority/committee
- Emergence of standard English spelling interesting
  - availability of lots of documents from the time before standardisation
  - still unclear **when** a standard really emerged

# Standardisation of spelling

- Two key factors indicate a move towards a standard:
  1. A reduction in the number of variants (and/or a move towards a clearly preferred variant)
  2. A reduction in intra-document variation (IDV)
- In this talk I will focus on 2 (1 is addressed in Cahill (2023))
- Crystal (2004:227) says “A standard language will only arise if a community is cognitively ready for it, and this state of mind was very likely present by 1400”
- Presence of IDV shows:
  - scribes do not consider that there is a single correct way to spell a word
  - hence the “state of mind” above is not present
  - crucially, this variation cannot be explained by other variables (time, person, place etc.)

# Corpora: MELD

- University of Stavanger
- Stenroos, Thengs and Bergstrøm (2017)
- Transcriptions of large numbers of documents (over 2000) ranging in time from 1400 to 1525
- Mostly official documents, from accounts and leases to letters and wills
- Full orthographic transcriptions

# Corpora: MEG-C

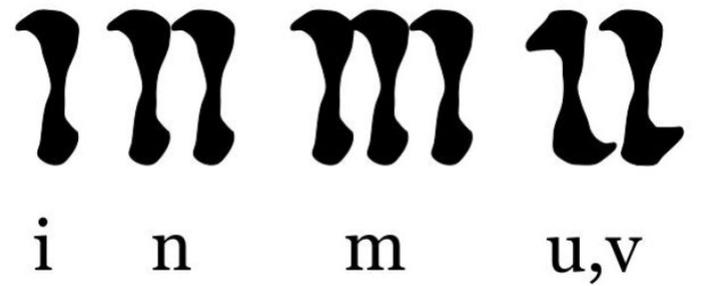
- Also University of Stavanger, in collaboration with University of Glasgow
- Stenroos et al (2011)
- Time period: 1325-1500
- Around 400 texts, around 650,000 words
- Wider range of text types: documentary as well as “literary” (including religious texts, poetry, fiction etc.)

# Data analysis

- Initial searches for variant forms of *said*
  - very frequent in MELD documents
  - known to have several variants (OED)
  - earlier most common variant different from modern standard spelling
- Four main variants for main vowel: <ai>, <ay>, <ei>, <ey>
- Mapped across England and across time (for subset of documents)
- Incidental finding that IDV continued into sixteenth century
- Crucially, dip in variation in mid 15<sup>th</sup> century followed by increase

# Data analysis

- Subsequent analysis of variable with only two variants:
  - <-ing> vs <-yng>
- Different from *said* vowel because:
  - only two possibilities
  - ultimate PDE standard is minority form in MELD and MEG-C
  - practical reason for choice: “minim problem” led to preference of <y> over <i>, especially before <n>
- If scribes used <y> in order to avoid ambiguity, we might expect more consistent use, at least intra-document



# Data analysis

- Subsequent analysis of variable with only two variants:
  - <-ing> vs <-yng>
- Different from *said* vowel because:
  - only two possibilities
  - ultimate PDE standard is minority form in MELD and MEG-C
  - practical reason for choice: “minim problem” led to preference of <y> over <i>, especially before <n>
- If scribes used <y> in order to avoid ambiguity, we might expect more consistent use, at least intra-document

# Data analysis

- AntConc used to extract all variants from documents
- Counties and regions:
  - In MEG-C, three regions coded:
    - North
    - East
    - West
  - In MELD, only counties coded, my own regions:
    - North
    - South
    - East Midlands
    - West Midlands

# Data analysis

So far:

- time\* and place analysis for *said* vowel in MELD and MEG-C
- (\*for documents where date information is available)
- place analysis for *-ing* in MELD
- Still to do:
  - time analysis for *-ing*
  - further time analysis of documents without date information
  - standardise grouping of regions\*
  - (\*counties don't currently map directly)

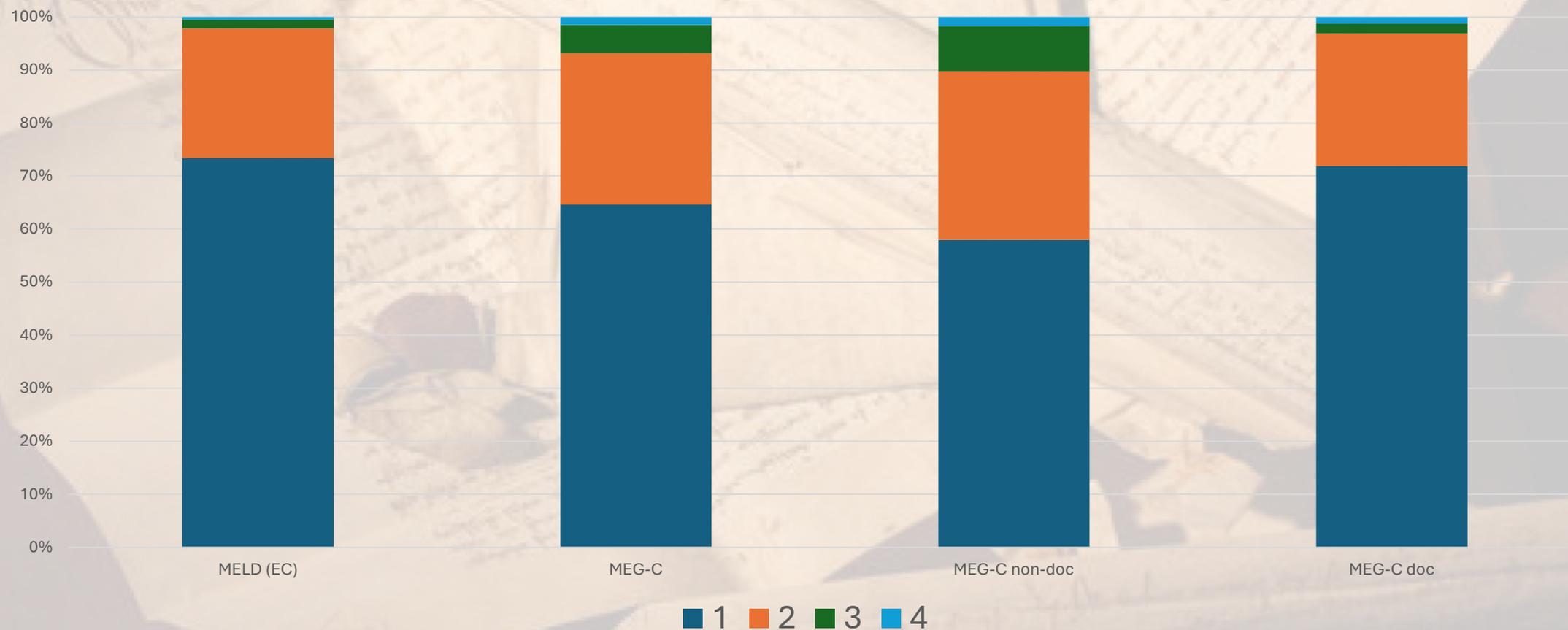
# Findings

- Overall, MEG-C has higher IDV than MELD
- Within MEG-C, non-documentary texts have higher IDV
- Documentary texts in MEG-C similar to MELD
- Different time periods, but MEG-C earlier start  
→ no significant difference in earlier texts

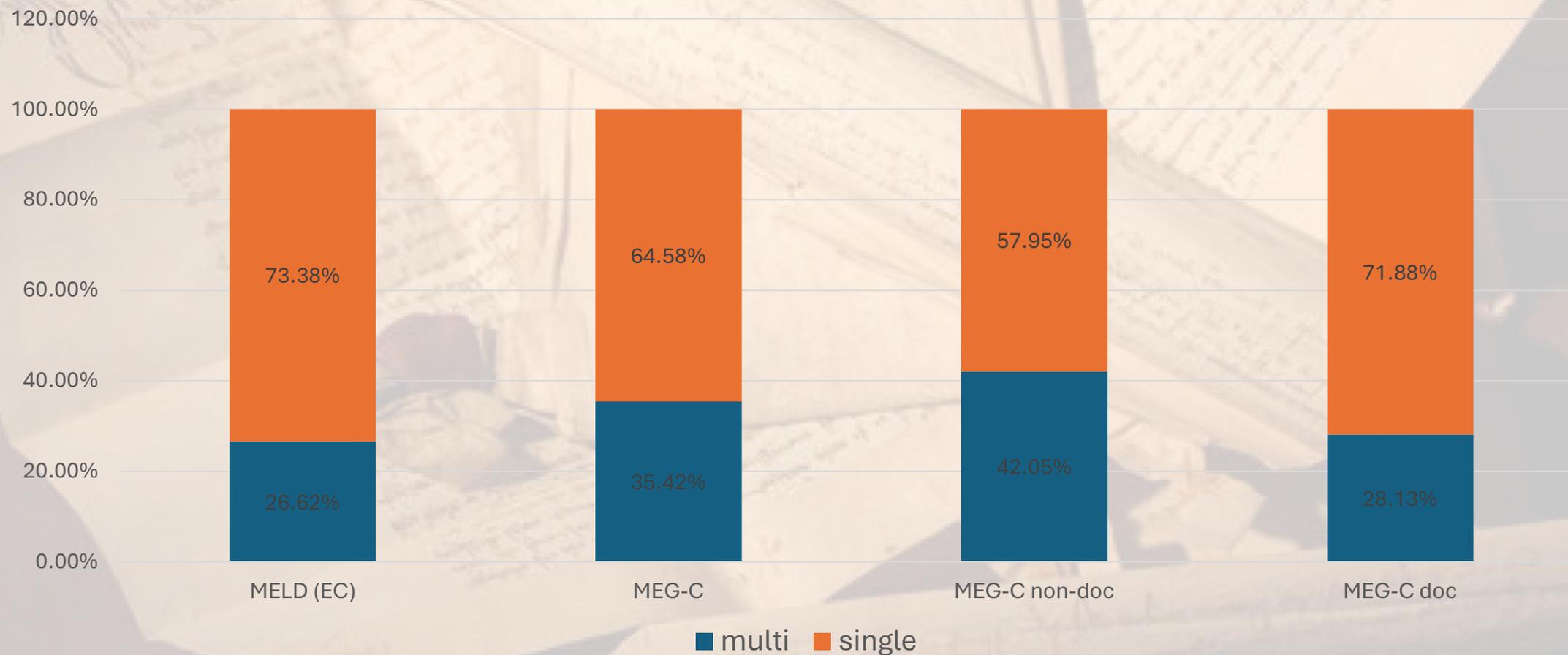
# Findings

- No linear increase in single variant docs in either MELD or MEG-C
- Peak is around 25 years earlier in MEG-C
- Same findings for *said* and *-ing*
- Across time period, less IDV in North (and South in MELD)

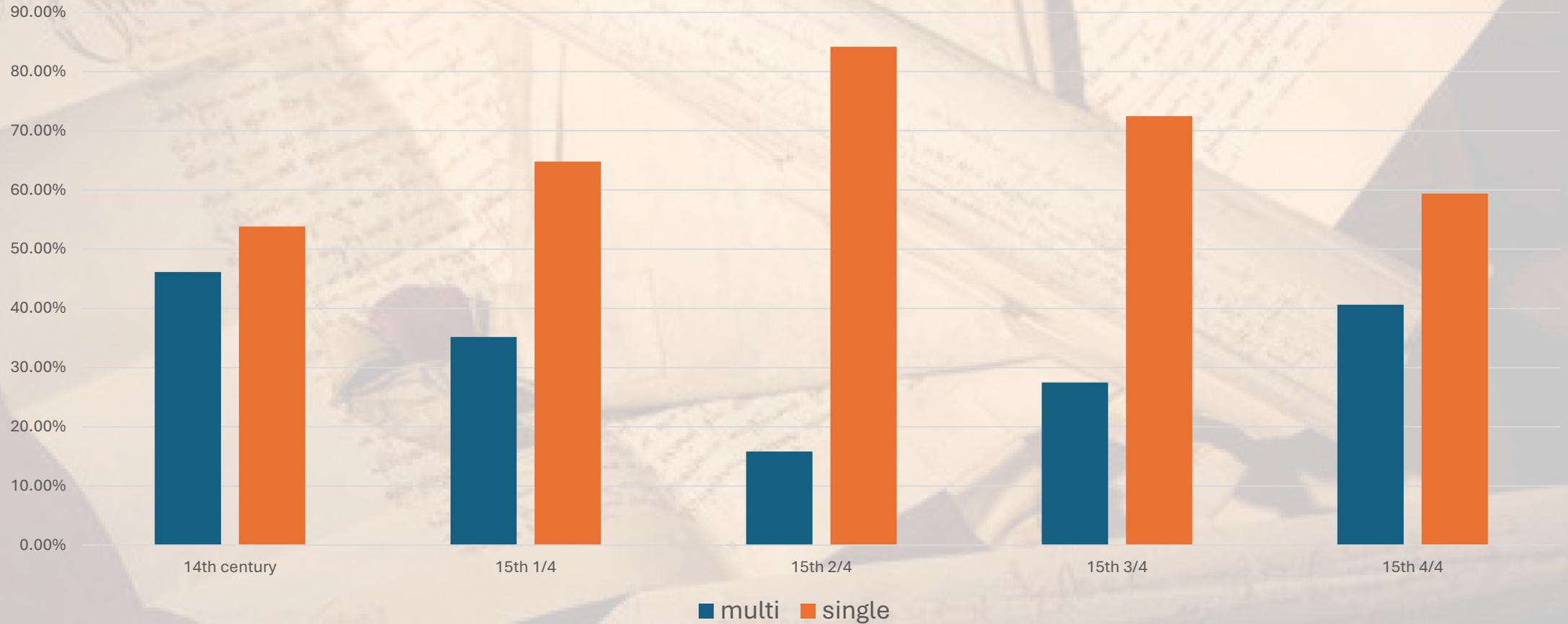
# Findings *said* (MELD vs MEG-C)



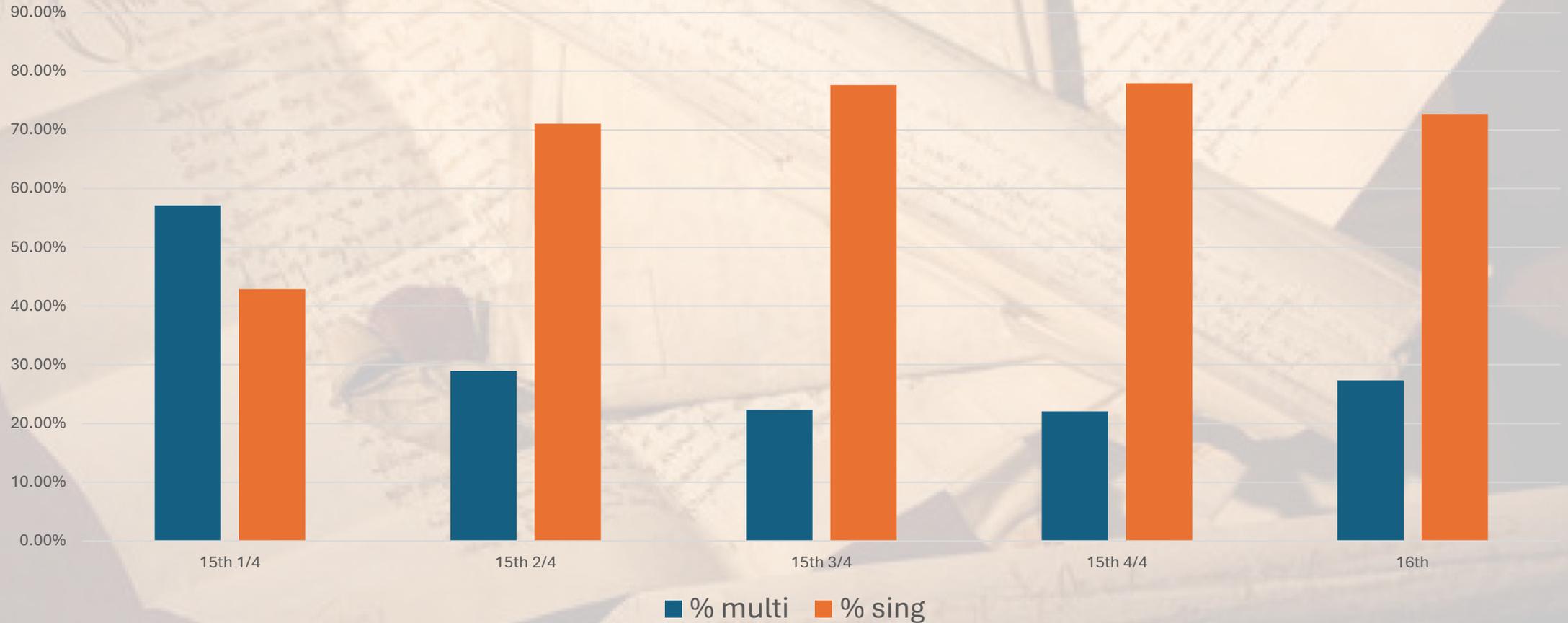
# Findings *said* (MELD vs MEG-C)



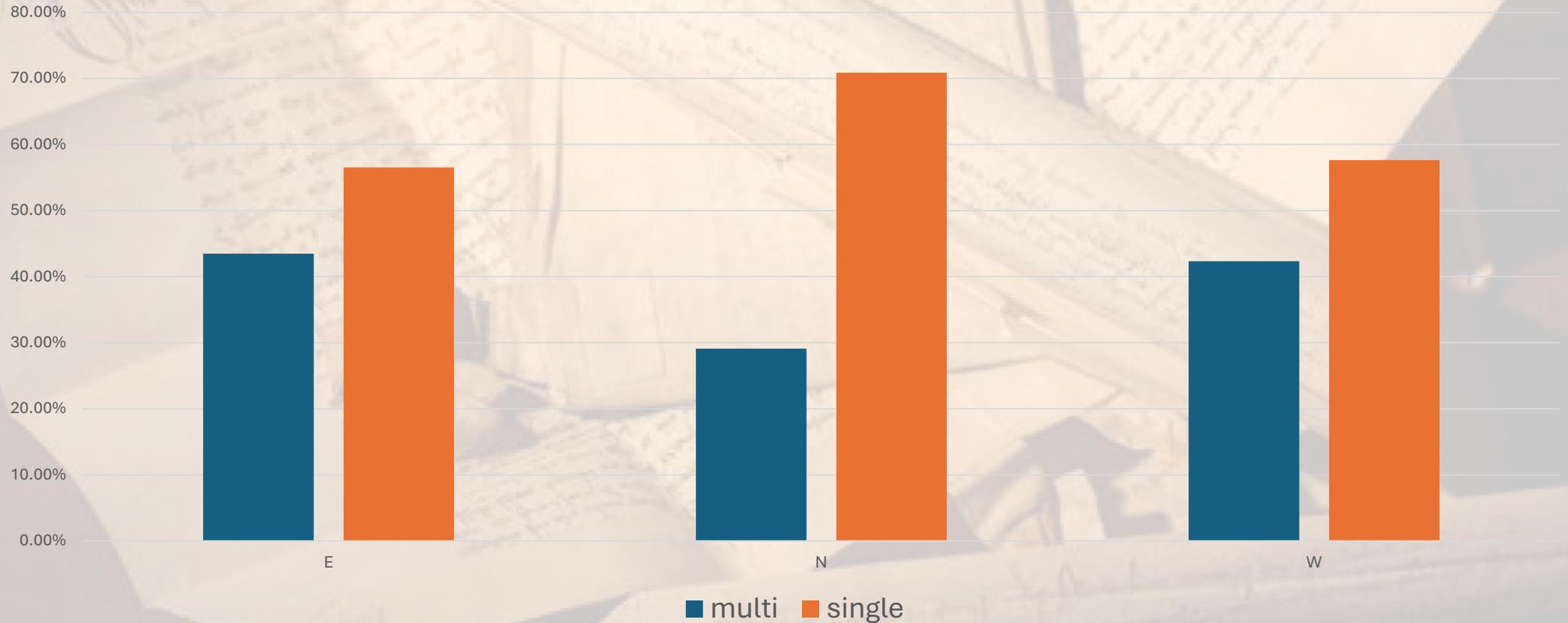
# Findings *said* (MEG-C) by date



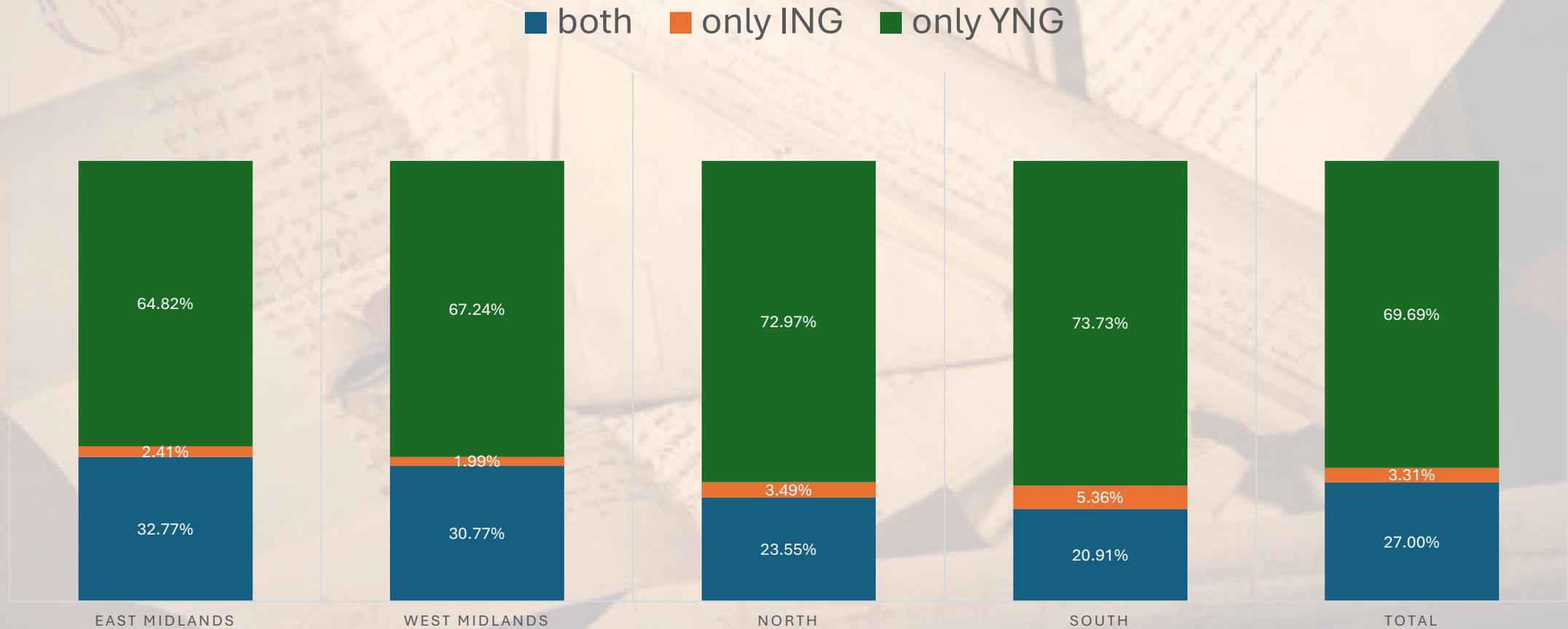
# Findings *said* (MELD) by date



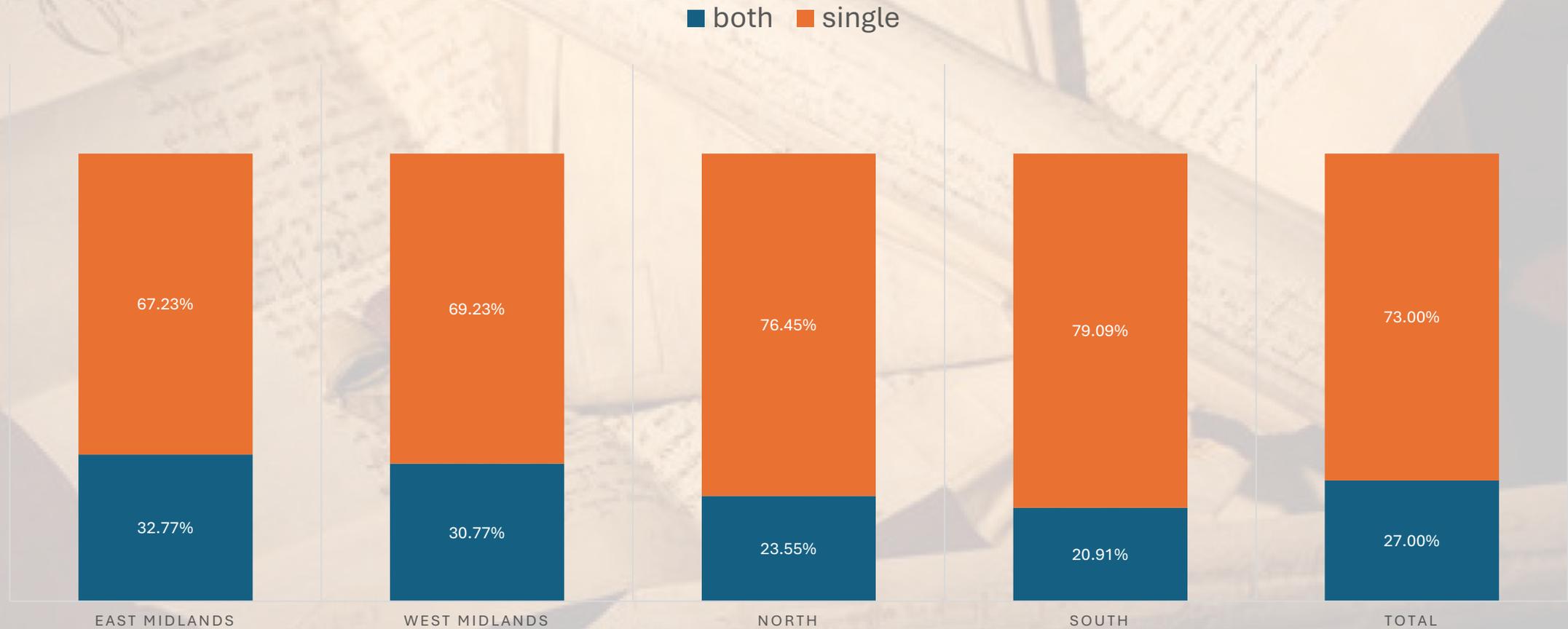
# Findings *said* (MEG-C) by region



# Findings *-ing* (MELD) by region



# Findings *-ing* (MELD) by region



# Conclusions

- No consistent move towards single form per document during fifteenth century
- More variation in non-documentary texts
- No difference between *said* (where consistency is unmotivated) and *-ing* (where consistency is motivated)
- Less IDV in documents from the North (and South)
  - more IDV in areas allegedly at the forefront of standardisation
- No evidence to support Crystal's idea about cognitive readiness by 1400

# References

- Cahill, Lynne (2023) The standardisation of spelling in Middle English: the case of *said*. In *Journal of Written Language and Literacy, Special Issue*
- Crystal, David (2004) *The Stories of English* London: Penguin.
- Haugen, E. (1997). Language Standardization. In: Coupland, N., Jaworski, A. (eds) *Sociolinguistics*. Modern Linguistics Series. Palgrave, London.
- Stenroos, Merja, Martti Mäkinen, Simon Horobin and Jeremy J. Smith (2011), *The Middle English Grammar Corpus*, version 2011.1. University of Stavanger. <http://www.uis.no/meg-c> [Accessed: 1 April 2020]
- Stenroos, Merja, Kjetil V. Thengs and Geir Bergstrøm (2017), *A Corpus of Middle English Local Documents (MELD)*, version 2017.1. University of Stavanger. <http://www.uis.no/meld> (Accessed: 31 December 2020)