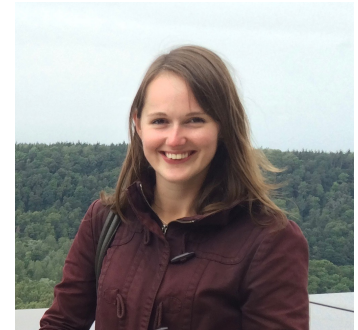Universität
Konstanz

Investigating interactional syntactic change in Middle English: Insights from visual analytics

Christin Schätzle & Hannah Booth

University of Konstanz

ICAME 41
HEIDELBERG
Digital Conference

Interactional language change

- Long-standing idea in historical linguistics: change can result from **multiple interacting factors** (e.g. Labov 1963, Malkiel 1967, Weinreich et al. 1968)

- Change can be the product of:
  - Interacting **language-internal** (i.e. system-driven) and **language-external** (i.e. socio-political) factors
  - Multiple interacting **exclusively language-internal** factors

Interactional language change

- Previous claim: syntactic change interacts with changes at other linguistic dimensions (e.g. phonology, morphology, semantics, information structure)
  $\longrightarrow$ Inherently an **interface phenomenon** (Keenan 1994, Longobardi 2001)

- And interacting changes **within** the syntax domain:
    - Principles & Parameters approach:
      underlying parametric change (e.g. Kroch 1989, Lightfoot 2013)
    - Usage-based paradigm:
      'multiple source constructions' (van de Velde et al. 2013)

Methodological challenges for historical linguistics

- Increasingly sophisticated corpus-based methodologies for syntactic change; many novel findings (e.g. Hilpert & Gries 2016, Pintzuk et al. 2017)

- **Standard procedure:** calculation of co-occurrence frequencies and statistical significances for different linguistic features across time stages

Texts	Indefinite NPs			Definite NPs			NPs as proper names		
	OV	VO	% OV	OV	VO	% OV	OV	VO	% OV
14th century	28	33	45.9%	11	57	16.2%	3	8	27.3%
15th century	23	30	43.4%	10	25	28.6%	1	3	25.0%
16th century	15	28	34.9%	17	26	39.5%	1	5	16.7%
17th century	28	59	32.2%	18	50	26.5%	0	20	0.0%
18th century	6	28	17.6%	7	31	18.4%	1	7	12.5%
19th century	34	425	7.4%	14	351	3.8%	4	68	5.6%
	134	603	18.2%	77	540	12.5%	10	111	8.3%

Definiteness distribution of NPs across different word orders in Icelandic (Hróarsdóttir 2000, 136)

- **Aim: identify** the factors involved in a change; understand **interactions across time**

HistoBankVis & ME

Schätzle & Booth

Methodological challenges for historical linguistics

- **But:** uncovering significant patterns and interactions is challenging:
    - Pair-wise comparison of the relevant bits of information across various tables
    - Data sparsity is an issue in historical linguistics
    - The factors causing a change are often unknown (or at least highly debated)
- Tools for investigating interactions in diachronic corpus-data are still lacking

- **Opportunity: Visual Analytics for Linguistics (LingVis)**

    - Turn complex data sets and their relationships into at-a-glance visualisations
    - Provide an interactive exploratory access to the data

"Analyze first, show the important, zoom, filter and analyze further, details on demand"
(Keim et al. 2008)

HistoBankVis & ME   Schätzle & Booth

This paper

- Investigating **interactional syntactic change in Middle English** (c.1100-1500)
  - ⟶ Substantial period of syntactic change, still not fully understood
  - ⟶ Loss of verb-second (V2) and rise of S(ubj)-V(erb)-O(bj) word order
- Various factors have been suggested (e.g. Los 2009, van Kemenade 2012)
  - ⟶ But **precise nature of interactions** remains elusive

- Penn-Helsinki Parsed Corpus of Middle English (PPCME2, Kroch & Taylor 2000)
  - Phrase-structure annotation, plus some functional information
  - Divided into 4 sub-periods: 1150-1250 (M1), 1250-1350 (M2), 1350-1420 (M3), 1420-1500 (M4)

- Method of investigation: **HistoBankVis** (Schätzle et al. 2017, 2019)
  - ⟶ LingVis system for historical studies

Clausal word order in Early English

- Clausal word order in Early English: highly complex, with a good deal of variation
- Overall: **subjects become increasingly prefinite**
    $\longrightarrow$ V2 gives away to SVO (decrease in 'subject-verb inversion')

- **Relevant factors** suggested for this change:
    - clause-initial constituent (e.g. van Kemenade 1987, Pintzuk 1999):
        - 'Group 1': *wh*-element/neg/discourse adverb
        - 'Group 2': adverbial/object noun phrase
    - subject type: pronominal/lexical (e.g. Haeberli 2002)
    - subject's information-structural (IS) status: given/new (e.g. van Kemenade & Westergaard 2012)
    - dominant dialect of text: north/west-midlands/east-midlands/south (Kroch & Taylor 1997, Kroch et al. 2000)

HistoBankVis & ME

Clause-initial category

- Old English: evidence that V2 was not fully consolidated
- Clause-initial category is one factor
- 'Group 1 contexts': initial *wh*-element, NEG or discourse adverb
  $\longrightarrow$ Subject typically postfinite ('subject-verb inversion')

(1)   Hwi wolde **God** swa lytles þinges him forwyrnan
           why would God so small things him deny?
           'Why should God deny him such a small thing?' (ÆCHom I, 1.14.2)

(2)   Ne sceal **he** naht   unaliefedes don
           not shall he nothing unlawful   do
           'He shall not do anything unlawful' (CP 10.61.14)

(3)   þa   cwæþ **he** to him
           then said he to them
           'then he said unto them...' (BlHom_11:119.49.1511)

Clause-initial category

- 'Group 2 contexts': e.g. (non-discourse) adverbial or object NP
    - ⟶ Lexical subjects typically postfinite (inversion)
    - ⟶ Pronominal subjects typically prefinite (no inversion)

  (4) [On twam þingum] hæfde **God** þæs mannes sawle gegodod
       in two   things   had   God the man's   soul   endowed
       'With two things God had endowed man's soul' (ÆCHom I, 1.20.1)

  (5) [Be ðæm] [**we**] magon suiðe swutule oncnawan ðæt...
       by that, we   may   very clearly perceive that
       'By that, we may perceive very clearly that ...' (CP 26.181.16)

- OE: **clause-initial category** and **subject type** interact with subject-verb inversion
    - ⟶ In some contexts V2 is already giving way to SVO

HistoBankVis & ME

Schätzle & Booth

Middle English

- **Subjects overall becoming increasingly prefinite**

- But the Group 1/Group 2 distinction remains relevant
    - Group 1 contexts: subject-verb inversion persists
      $\longrightarrow$ 'residual V2' in Present-day English (Rizzi 1996)
    - Group 2 contexts: subject-verb inversion gradually decreases
      $\longrightarrow$ Lexical subjects increasingly prefinite (Haeberli 2002)

- Plus extra factors:
    - **Information structure**: discourse-new subjects increasingly prefinite (van Kemenade & Westergaard 2012)
    - **Dialect**: certain Northern texts are conservative; postfinite subjects generally persist (Kroch & Taylor 1997, Kroch et al. 2000)

Our methodology

- Data from PPCME2, extracted via CorpusSearch queries (Randall 2005)
  $\longrightarrow$ Restriction: matrix clauses which contain a finite verb and an overt subject

- **Investigated factors:**
  - subject position: prefinite/postfinite
  - subject type: pronominal/lexical
  - subject's information-structural status: given/new
  - clause-initial constituent: Group 1 (neg/discourse adverb); Group 2 (PP/non-discourse adverbial/object noun phrase)
    $\longrightarrow$ *wh*-elements excluded; inversion persists in questions
  - dominant dialect of text: north/west-midlands/east-midlands/south

- Previous studies: mostly binary comparisons
  $\Longrightarrow$ **LingVis** allows us to assess interactions between **several factors at once**

HistoBankVis & ME

Schätzle & Booth

HistoBankVis – Overview

- Generically applicable system for **historical linguistic research**
- Flexible investigation of a potentially high number of interacting factors
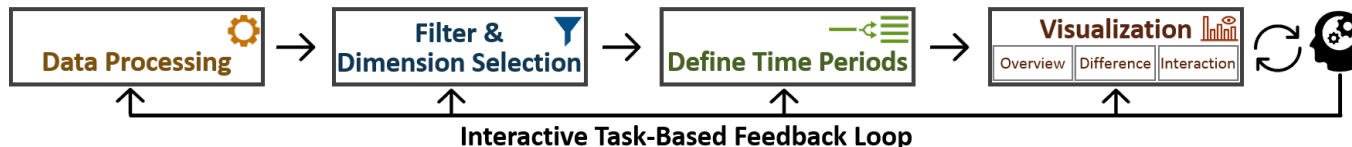
- Combination of several interlinked visualisation and filtering techniques
  $\longrightarrow$ exploratory access to complex data

- Three main components:

  - **Overview:** Compact Matrix
  - **Difference** Histograms
  - **Central to our investigations:** Dimension **Interactions**

Dimension Interactions

- Dimension interactions provide insights into the **interrelation between multiple features** of different dimensions

- Application of the **Parallel Sets** technique (Bendix et al. 2005, Kosara et al. 2006)
  - Feature frequencies are visualised as proportions of equally spaced vertical lines (data dimensions)
  - Dimensions are connected by coloured ribbons
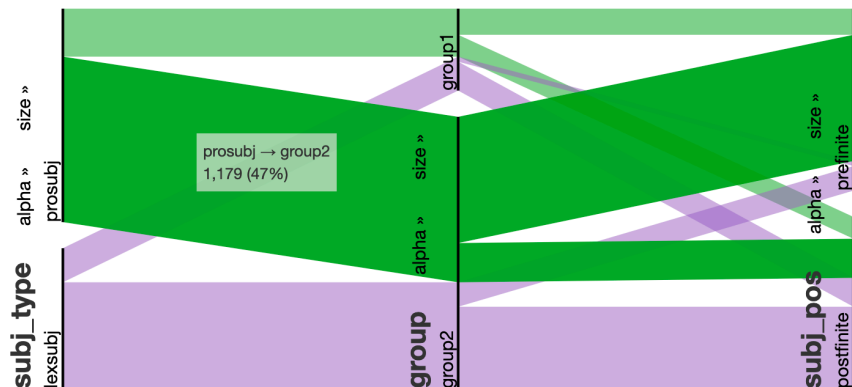 - Size of a ribbon: a feature's share of a feature from another dimension

prosubj → group2
1,179 (47%)
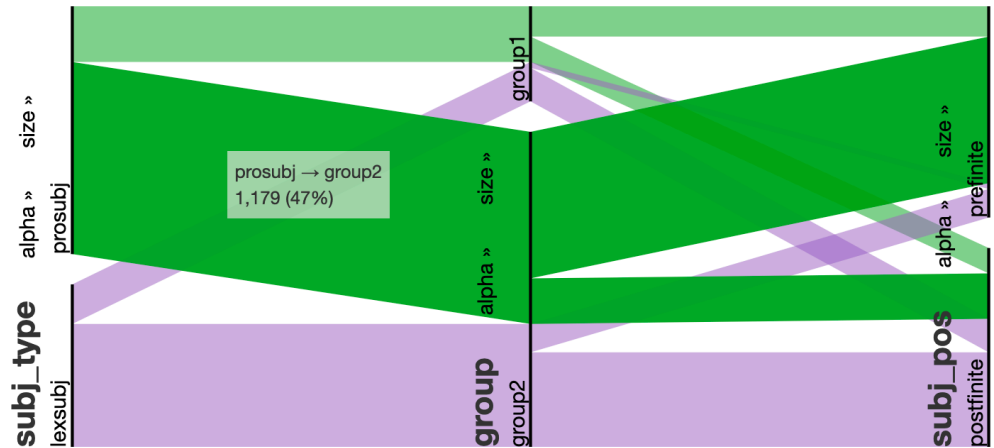
Interaction between subject type, group and subject position in M1

HistoBankVis & ME

Schätzle & Booth

Insights from the visualisation

Clause-initial category, subject type and subject position

- **Prediction**:
    - Pronominal subjects lead the change, becoming increasingly prefinite
    - Lexical subjects lag behind, but also become increasingly prefinite
    - Divergence between Group 1 and Group 2:
        - Group 1 contexts remain conservative (postfinite subjects)
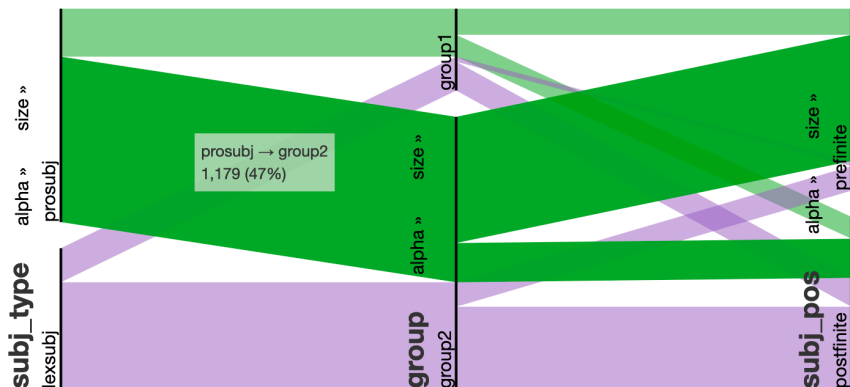   - Group 2 contexts is where the change mostly happens

HistoBankVis & ME

Schätzle & Booth

Dimension interactions: group, subject type, subject position

M1:
- Preference for pronominal subjects to be prefinite (weaker in Group1)
- Lexical subjects mostly postfinite

HistoBankVis & ME

Schätzle & Booth

Dimension interactions: group, subject type, subject position

M1:
- Preference for pronominal subjects to be prefinite (weaker in Group1)
- Lexical subjects mostly postfinite

M4:
- Pronominal and lexical subjects preferably prefinite
- Applies to both Group 1 and Group 2

HistoBankVis & ME

Schätzle & Booth

Insights from the visualisation

Clause-initial category, subject type and subject position

- **Findings**:
  - ✓ Pronominal subjects lead the change, becoming increasingly prefinite
  - ✓ Lexical subjects lag behind, but also become increasingly prefinite
  - ✗ Divergence between Group 1 and Group 2:
    - Group 1 contexts remain conservative (postfinite subjects)
    - Group 2 contexts is where the change mostly happens

    ⟹ Divergence is less clear cut than expected:
    Group 1 is not static but follows suit eventually (at least in declaratives…)
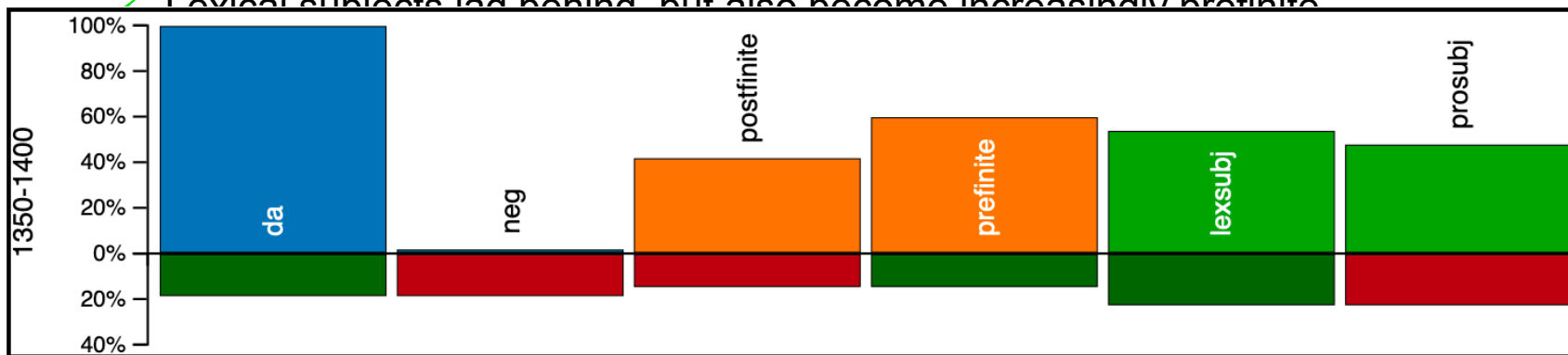   Perhaps due to exclusion of *wh*-questions?

    ⟹ Increase in prefinite subjects coincides with loss of clause-initial negation
    (data mainly clause-initial DAs)

HistoBankVis & ME

Insights from the visualisation

Clause-initial category, subject type and subject position

- **Findings**:
  - ✓ Pronominal subjects lead the change, becoming increasingly prefinite
  - ~~Lexical subjects lag behind, but also become increasingly prefinite~~

⟹ Increase in prefinite subjects coincides with loss of clause-initial negation (data mainly clause-initial DAs)

Insights from the visualisation

**Clause-initial category, IS-status of subject and subject position**
- **Prediction**:
    - Early ME: discourse-new subjects more frequently postfinite than discourse-given subjects
    - Late ME: IS effect weaker; little difference between discourse-new and discourse-given subjects
- **Unknown**:
    - Is the IS effect only relevant for Group 2 or is it for Group 1 too?
    - Is IS in fact the driving force behind what we have already seen?

HistoBankVis & ME

Schätzle & Booth

Insights from the visualisation

Clause-initial category, IS-status of subject and subject position

- **Findings**:
  - ✓ Early ME: discourse-new subjects more frequently postfinite than discourse-given subjects
  - ✓ Late ME: IS effect weaker; little difference between discourse-new and discourse-given subjects

- **Unknown**:
  - Is the IS effect only relevant for Group 2?
  - Is IS in fact the driving force behind what we have already seen?

⟹ Difficult to separate IS from subject type (given ~ pronominal; new ~ lexical)

⟹ But: indication that subject type is more important than IS (subject type seems to correlate stronger with position than IS)
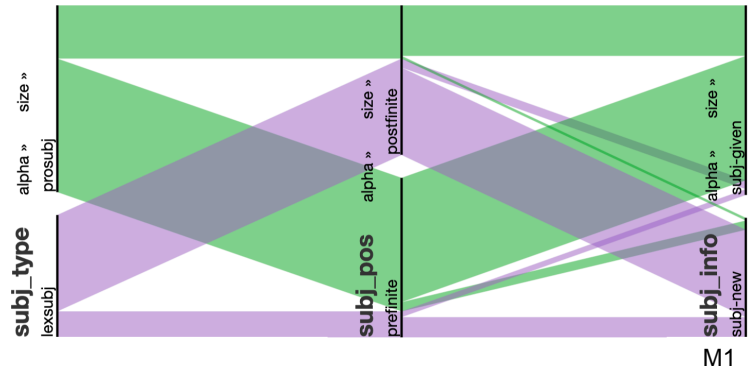
Insights from the visualisation

Are Northern texts special?

- **Prediction**:
    - Northern texts: higher frequencies of postfinite subjects

- **But**: only one Northern text is clearly dated
    $\longrightarrow$ *Northern Prose Rule of St. Benet* (1350-1420)
    $\longrightarrow$ exhibits a generalised V2 system (Kroch and Taylor 1997, Kroch et al. 2000)

- One step further: we include texts whose dating is less clear (~M34, M24, MX4)

Insights from the visualisation

Are Northern texts special?

- **Findings**:
  - ✓ *Rule of St. Benet*: higher frequencies of postfinite subjects
    - ⟹ Overall, subject-verb inversion is preferred with pronominal and lexical subjects; no group divergence!
    - ⟹ Indicates a 'conservative' V2-pattern

- When looking at Northern texts whose dating is less certain (M24, M34, MX4), the Group 1 issue is still cloudy
  - ⟹ Subject-inversion in Group 1 is not so marked

- Once again, factoring out questions gives us a more mixed picture
- HistoBankVis provides us with quick exploratory access to previously established hypotheses
  - ⟶ Data sparsity is an issue

Are Northern texts special?
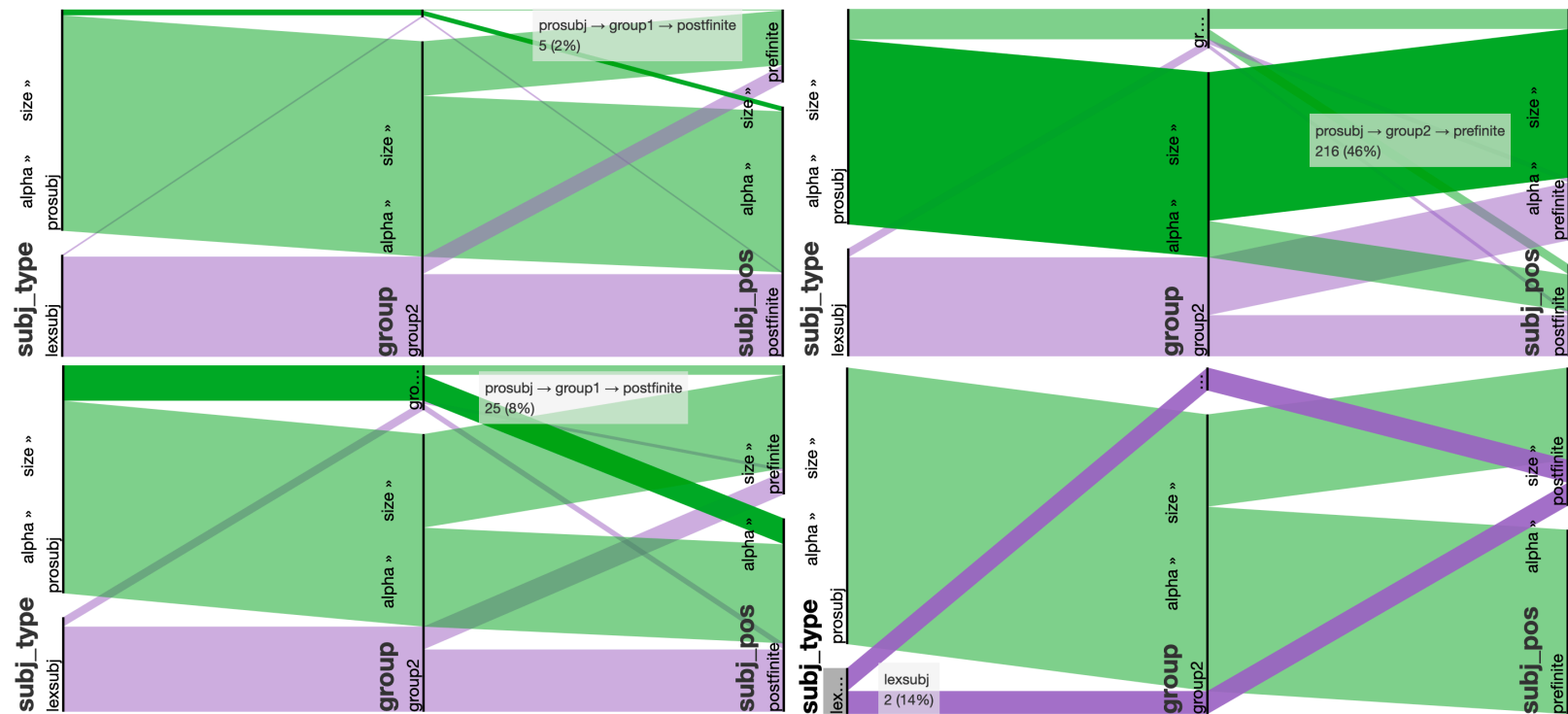
Time periods: M3 top left, M24 top right, M34 bottom left, MX4 bottom right

Conclusions

- HistoBankVis offers **new insights, even on a relatively well-studied change**:
    - Factoring out questions in Group 1 shows a more mixed picture than expected
    - Indication that subject type has a stronger effect on subject position than the information-structural status of the subject
    - But both subject type and subject-IS effect weaken over time
- HistoBankVis fosters an **iterative cycle of hypothesis testing and generation**
    - Confirmation/rejection of existing hypotheses
    - Generation of new hypotheses and ideas for future research
- **Future work**:
    - Clause-initial discourse adverbs contexts in Group 1
    - Effect of questions on Group 1
    - Information structure, via manual annotation
    - Continue to leverage data from texts whose dating is less certain

HistoBankVis & ME

Schätzle & Booth

Thank you!
Questions?

christin.schaetzle@uni-konstanz.de

hannah.booth@uni-konstanz.de

http://histobankvis.dbvis.de/

References

Bendix, F., Kosara, R., and Hauser, H. (2005). Parallel sets: Visual analysis of categorical data. In IEEE Symposium on Information Visualization INFOVIS, pages 133–140. IEEE.

Haeberli, E. (2002). Observations on the loss of verb-second in the history of English. In Zwart, C. and Abraham, W., editors, Studies in Comparative Germanic Syntax, volume 53, page 245. John Benjamins, Amsterdam.

Hilpert, M. and Gries, S. T. (2016). Quantitative approaches to diachronic corpus linguistics. In Kytö, M. and Pahta, P., editors, The Cambridge handbook of English historical linguistics, pages 36–53. Cambridge University Press, Cambridge.

Holmberg, A. (2015). Verb second. In Kiss, T. and Alexiadou, A., editors, Syntax. Theory and Analysis: An International Handbook, pages 342–383. de Gruyter, Berlin.

Keenan, E. (1994). Creating anaphors: An historical study of the English reflexive pronouns. Ms. UCLA.

Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., and Melançon, G. (2008). Visual analytics: Definition, process, and challenges. In Kerren, A., Stasko, J. T., Fekete, J.-D., and North, C., editors, Information Visualization, pages 154–175. Springer, Berlin.

Kosara, R., Bendix, F., and Hauser, H.(2006). Parallel Sets: interactive exploration and visual analysis of categorical data. IEEE Transactions on Visualization and Computer Graphics, 12(4):558–568.

Kroch, A. and Taylor, A. (2000). The Penn-Helsinki Parsed Corpus of Middle English (PPCME2). Department of Linguistics, University of Pennsylvania. Second edition.

Kroch, A., Taylor, A., and Ringe, D. (2000). The Middle English verb-second constraint: A case study in language contact and language change. In Herring, S. C. and van Reenen, P., editors, Textual Parameters in Older Language, pages 353–391.

Kroch, A. S. (1989). Reflexes of grammar in patterns of language change. Language Variation and Change, 1(3):199–244.

Kroch, A. S. and Taylor, A. (1997). Verb movement in Old and Middle English: Dialect variation and language contact. In van Kemenade, A. and Vincent, N., editors, Parameters of Morphosyntactic Change. Cambridge University Press, Cambridge.

References

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

Lightfoot, D. W. (2013). Types of explanation in history. Language, 89(4):e18–e38.

Longobardi, G. (2001). Formal syntax, diachronic minimalism, and etymology: the history of French chez. Linguistic Inquiry, 32(2):275–302.

Los, B. (2009). The consequences of the loss of verb-second in English: information structure and syntax in interaction. English Language and Linguistics, 13(1):97–125.5

Malkiel, Y. (1967). Multiple versus simple causation in linguistic change. In To honor Roman Jakobson: Essays on the occasion of his seventieth birthday, pages 1228–1246, The Hague. Mouton.

Pintzuk, S. (1999). Phrase structures in competition: Variation and change in Old English word order. Garland, New York.

Pintzuk, S., Taylor, A., and Warner, A. (2017). Corpora and quantitative methods. In Ledgeway, A. and Roberts, I., editors, The Cambridge handbook of historical syntax, pages 218–240. Cambridge University Press, Cambridge.

Randall, B. (2005). CorpusSearch2 User's Guide. Philadelphia: Dept. of Linguistics, University of Pennsylvania. http://corpussearch.sourceforge.net.

Rizzi, L. (1996). Residual verb second and the wh-criterion. In Belletti, A. and Rizzi,L., editors, Parameters and functional heads: essays in comparative syntax, pages 63–90. Oxford University Press Oxford, Oxford.

Schätzle, C., Dennig, F. L., Blumenschein, M., Keim, D. A., and Butt, M. (2019). Visualizing linguistic change as dimension interactions. In Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change, pages 272–278, Florence, Italy. Association for Computational Linguistics.

Schätzle, C., Hund, M., Dennig, F. L., Butt, M., and Keim, D. A. (2017). HistoBankVis: Detecting language change via data visualization. In Bouma, G. and Asedam, Y., editors, Proceedings of the NoDaLiDa 2017 Workshop on Processing Historical Lan-guage, pages 32–39, Linköping. Linköping University Electronic Press.
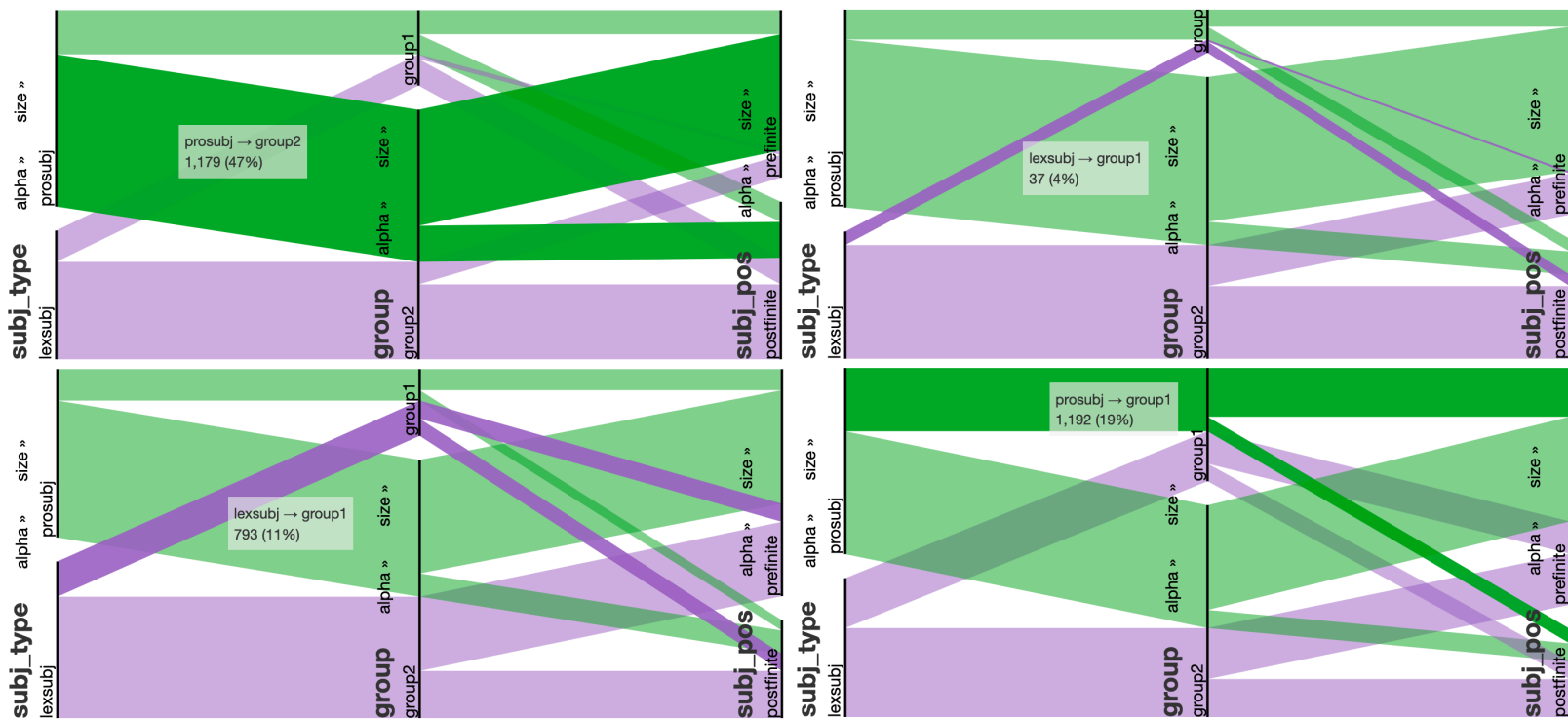
References

van de Velde, F., De Smet, H., and Ghesquière, L. (2013). On multiple source constructions in language change. Studies in Language, 37(3):473–489.

van Kemenade, A. (1987). Syntactic case and morphological case in the history of English. Foris, Dordrecht.

van Kemenade, A. (2012). Rethinking the loss of verb second. In Nevalainen, T. and Traugott, E. C., editors, The Oxford Handbook of the History of English, pages 823–834. Oxford University Press, Oxford.

van Kemenade, A. and Westergaard, M. (2012). Verb-Second Variation in Middle English. In Meurman-Solin, A., Lopez-Couso, M. J., and Los, B., editors, Information structure and syntactic change in the history of English, pages 87–118. Oxford University Press, Oxford.

Weinreich, U., Labov, W., and Herzog, M.I.(1968). Empirical foundations for a theory of language change. In Lehmann, W. P. and Malkiel, Y., editors, Directions for Historical Linguistics, pages 95–195. University of Texas Press, Austin, TX.

Spaghetti junction, Birmingham (image on slide 3):
`https://historicengland.org.uk/services-skills/education/educational-images/spaghetti-junction-birmingham-10382`

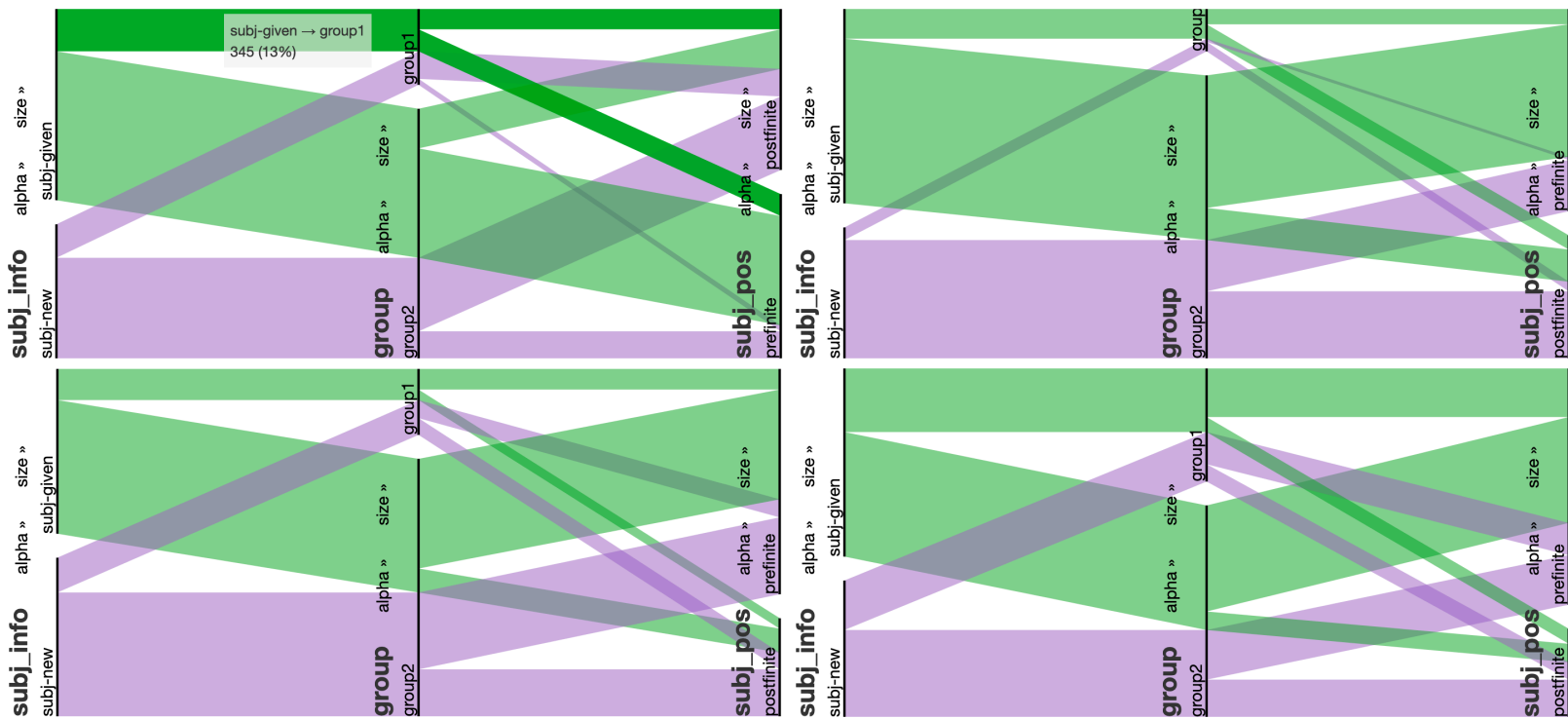# Dimension interactions: subject type, group, subject position

Time periods: M1 top left, M2 top right, M3 bottom left, M4 bottom right

HistoBankVis & ME

Schätzle & Booth

Difference histograms: Group 1, subject position, subject type

HistoBankVis & ME

Schätzle & Booth
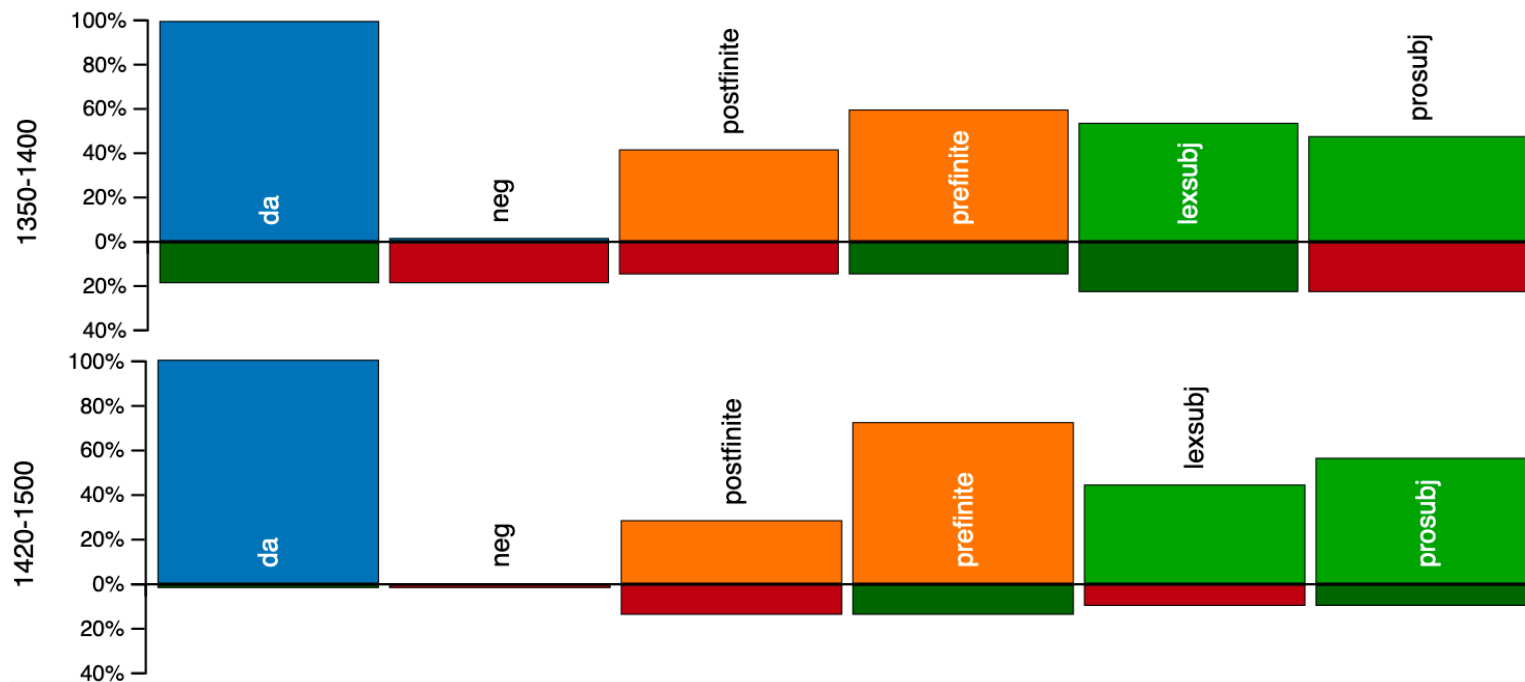
Dimension interactions: subject IS, group, subject position

Time periods: M1 top left, M2 top right, M3 bottom left, M4 bottom right

HistoBankVis & ME

Schätzle & Booth