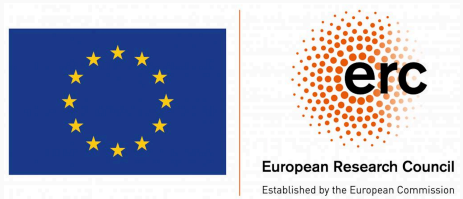
Quantify syncretism in subject-verb agreement marking

Yiming Liang & Alexandra Simonenko

November 7-9, 2024

NWAV52, Miami Beach, United States



Subject-verb agreement syncretism

Subject-verb agreement:

(1) gif hio of cealdum intingan cymð þonne sceal mon mid
if she of cold.DAT cause.DAT comes then shall one with
hatum læcedomum lacnian
hot.DAT leechdom.DAT heal.INF

'If it comes of cold causes, one should treat it with hot leechdoms' (YCOE, colaece,Lch_II_[1]:1.13.4.85) (Walkden, 2013: 173)

Syncretism: a morphological exponent corresponds to more than one combination of features (e.g., number, person)

person	singular	plural
1	hīere	hīerað
2	hīerst	hīerað
3	hīerð	hīerað

Table 1: Present-tense subject-verb agreement in West Saxon Old English for "to hear" (Walkden, 2021: 10).

Related to other linguistic phenomena

Subject-verb agreement syncretism has been argued to be related to various linguistic phenomena, like:

- **Subject expression:** Taraldsen's Generalization (TG): languages with rich agreement tend to allow pro-drop (Taraldsen, 1980; van Gelderen, 2000)
- **Verbal movement:** Rich Agreement Hypothesis (RAH): languages with rich subject verbal agreement morphology tend to have V-to-I movement (Kosmeijer, 1986; Kroch et al., 2000)

The problem

However, long-standing debates on the validity of those generalizations, largely due to the operationalization of ”syncretism”:

- Dichotomous classification without consensus → conflicting results

Definition	W. Sax. OE	Nth. OE	Sth. ME	Nth. ME	Ear. Mod. E
Platzack and Holmberg (1989)	Rich	Rich	Rich	Rich	Rich
Roberts (1993: 263–273)	Rich	Poor	Rich	Poor	Poor
Rohrbacher (1994, 1999)	Rich	Poor	Rich	Poor	Poor
Vikner (1997)	Rich	Rich	Rich	Rich	Rich
Koeneman (2000: 67–84)	Rich	Rich	Rich	Rich	Poor
Bobaljik (2002)	Rich	Rich	Rich	Rich	Rich
Koeneman and Zeijlstra (2014)	Rich	Rich	Rich	Rich	Rich?

Table 2: Classifications of richness of Historical English (Walkden, 2021: 14).

→ Koeneman and Zeijlstra (2014): counter-examples against RAH disappear under their criterion of “richness”

The problem

- Dichotomous classification without consensus → conflicting results
- Diachronic change of the agreement system results from morphophonological variation:
  → multiple endings for one person: late OE, plural persons were expressed by -að or -n, with the latter gradually replacing the former
  → a categorical distinction cannot capture the gradient nature of verbal agreement syncretism

The problem

- Dichotomous classification without consensus → conflicting results
- Diachronic change of the agreement system results from morphophonological variation:
  → multiple endings for one person: late OE, plural persons were expressed by -að or -n, with the latter gradually replacing the former
  → a categorical distinction cannot capture the gradient nature of verbal agreement syncretism

→ quantify the degree of syncretism of the agreement system as a continuum based on corpora

Quantify verbal agreement syncretism

Verbal ending: More certainty in predicting subject person&number
$\rightarrow$ less ambiguity in verbal ending (richer agreement)

For example, in modern English:

- P(3sg|-s) = 1
- P(1sg|$\varnothing$) = relative frequency of 1sg vs. other non-3sg subjects

Quantify verbal agreement syncretism

Verbal ending: More certainty in predicting subject person&number → less ambiguity in verbal ending (richer agreement)

For example, in modern English:

- P(3sg|-s) = 1
- P(1sg|∅) = relative frequency of 1sg vs. other non-3sg subjects

Overall certainty in predicting subject person&number = P(1sg|∅) · Freq(1sg, ∅) + ... + P(3sg|-s) · Freq(3sg, -s) + ... P(3pl|∅) · freq(3pl, ∅) ≈ Conditional entropy:

$$H(\text{person}|\text{ending}) = -\sum_{x \in \text{person}} \sum_{y \in \text{ending}} P(x,y) \log P(x|y) \qquad (1)$$

Higher entropy → higher uncertainty of the system → more ambiguity in verbal endings → more syncretic agreement system

Quantify verbal agreement syncretism

New problem: conditional entropy is highly sensitive to the unbalanced distribution of subject's person&number in a dataset

- e.g., in correspondences, 1sg subject is predominant. So the conditional entropy will be low regardless of whether the agreement system is rich or not

Revise the metric: normalize the conditional entropy by the general uncertainty of subject's person&number distribution
$\rightarrow$ Verbal Agreement Syncretism Score (VASS):

$$\text{VASS} = \frac{H(\text{person}|\text{ending})}{H(\text{person})} \tag{2}$$

- the score is between 0 and 1
- higher VASS $\rightarrow$ higher syncretism in subject-verbal agreement
- comparable across datasets and languages

Case study: agreement in Historical English

Historical English: having lost its rich agreement system

pers.	sg	pl
1	hīere	hīerað
2	hīerst	hīerað
3	hīerð	hīerað

Table 3: West Saxon Old English.

$\Rightarrow$

pers.	sg	pl
1	hear	hear
2	hear	hear
3	hears	hear

Table 4: Modern English.

Case study: agreement in Historical English

Historical English: having lost its rich agreement system

pers.	sg	pl
1	hīere	hīerað
2	hīerst	hīerað
3	hīerð	hīerað

Table 3: West Saxon Old English.

⇒

pers.	sg	pl
1	hear	hear
2	hear	hear
3	hears	hear

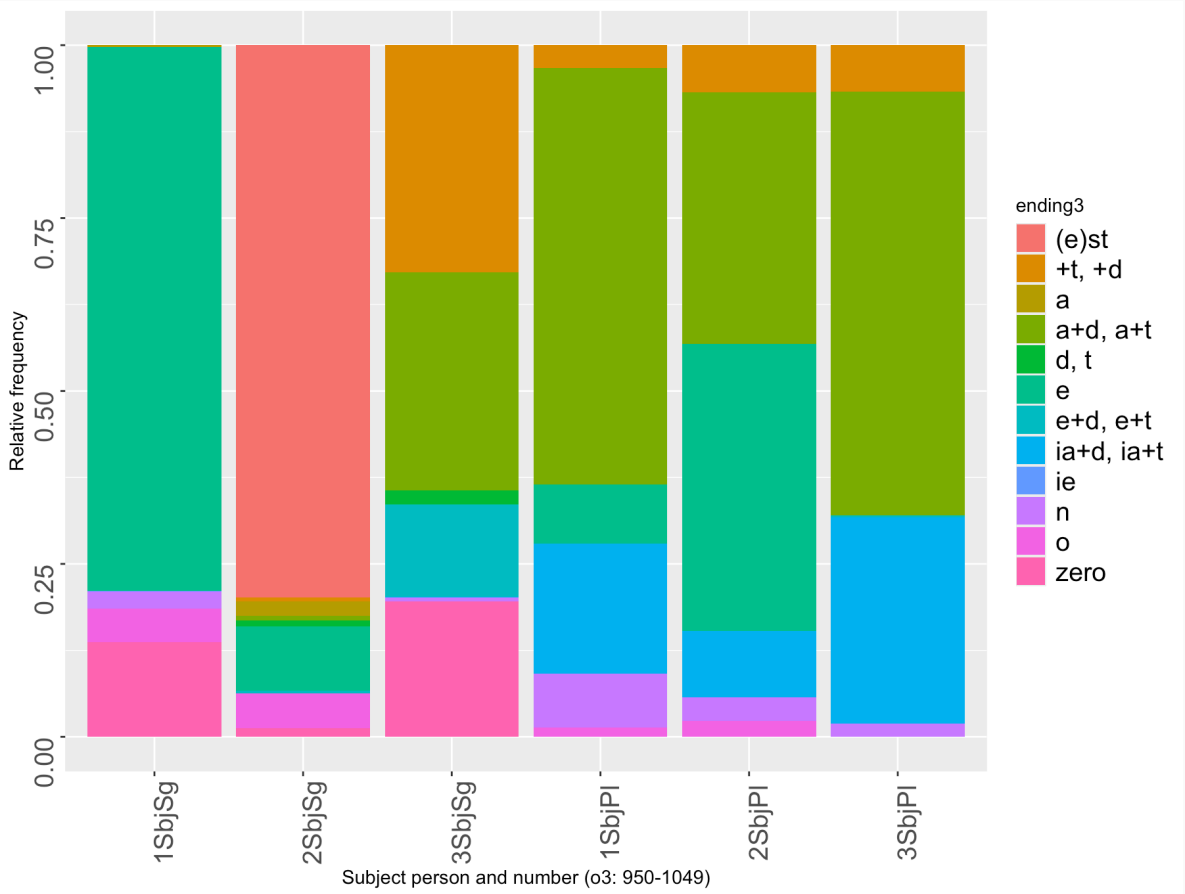
Table 4: Modern English.

- **Data:** Parsed historical corpora of British English prose (YCOE, PPCME2, PPCEME, PPCMBE2) (Kroch, 2020; Taylor et al., 2003)
- **Extraction:** present-tense verb (VBP) co-occurring with an overt pronominal subject, using *CorpusSearch2* (Randall, 2010)
- **Exclusion:** BE, HAVE, auxiliaries, modals, subjects containing conjunction and subordinate clauses (to exclude subjunctives)
- **Dataset:** 20,692 datapoints, from 800 to 1913 (separated by 12 traditionally recognized periods of Historical English)

Measure VASS in Historical English

Old English Period 3 (950-1049):

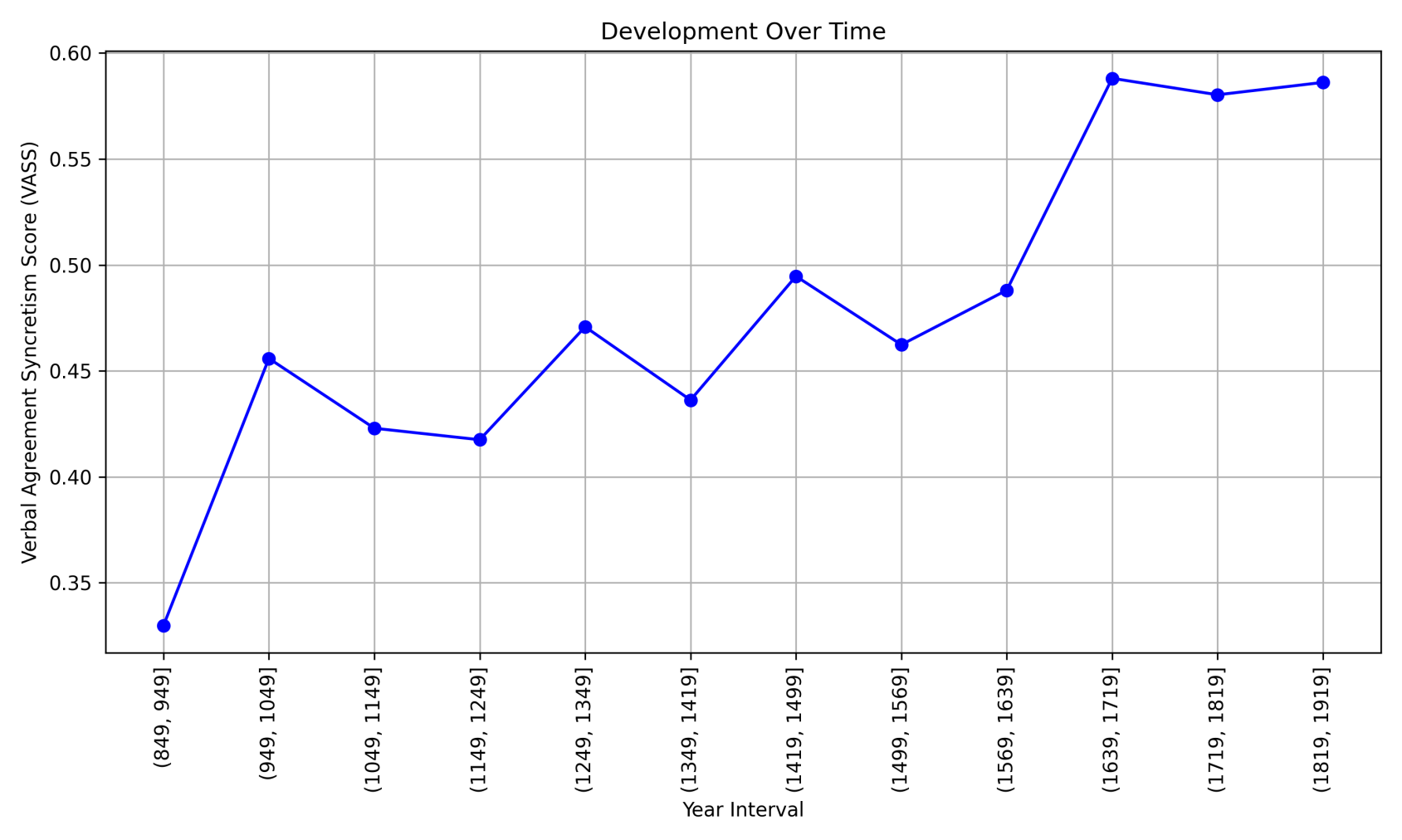
pers.	sg	pl
1	-e, -o, ∅	-θ (ð), -aθ (ð), -e, -iaθ(ð), -n
2	-(e)st, -e, -o	-θ (ð), -aθ (ð), -e, -iaθ(ð)
3	-θ (ð), -aθ (ð), -t (d), ∅	-θ (ð), -aθ (ð), -eθ(ð), -iaθ(ð)



VASS takes into account:
- relative frequency of the same ending across different person and number
- we ignored endings that occur less than 10 times in a cell

→ 0.46

Development of VASS of Historical English



A significant increase of VASS (i.e., syncretism) across time (pearson correlation: β = 0.89, p <0.001)

Conclusion

- VASS proves successful in quantitatively capturing the historical increase in English agreement syncretism
- Our metric opens the door to quantitative investigations of the relationships between agreement syncretism and other linguistic phenomena, both synchronically and diachronically
    - Relation between pro-drop and rich agreement: to investigate whether there is some correlation between pro-drop rate and VASS in corpora

Thank you!

The work is co-funded by the European Union (ERC, CAUSALITY, grant number 101042427). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

Yiming Liang, PhD
Ghent University
yiming.liang@ugent.be

References

- Bobaljik, J. D. (2002). **Realizing Germanic Inflection: Why Morphology Does Not Drive Syntax.** The Journal of Comparative Germanic Linguistics, 6(2), 129–167.
- Koeneman, O. (2000). **The flexible nature of verb movement.** Netherlands Graduate School of Linguistics.
- Koeneman, O., & Zeijlstra, H. (2014). **The Rich Agreement Hypothesis Rehabilitated.** Linguistic Inquiry, 45, 571–615.
- Kosmeijer, W. (1986). **The status of the finite inflection in Icelandic and Swedish.** Working Papers in Scandinavian Syntax, (26), 1–41.
- Kroch, A. (2020). **Penn Parsed Corpora of Historical English LDC2020T16. (Web download.).**

References II

Kroch, A., Taylor, A., & Ringe, D. (2000). **The Middle English Verb-Second Constraint: A case study in language contact and language change.** In S. C. Herring, P. van Reenen, & L. Schøsler (Eds.), Textual Paramete (p. 353). John Benjamins Publishing Company.

Platzack, C., & Holmberg, A. (1989).**The role of AGR and finiteness.** Working Pape (43), 51–76.

Randall, B. (2010). **CorpusSearch 2..**

Roberts, I. (1993). **Verbs and diachronic syntax: A comparative history of Eng** Kluwer.

Rohrbacher, B. W. (1994). **The Germanic VO languages and the full paradigm:** University of Massachusetts Amherst.

Rohrbacher, B. W. (1999). **Morphology-Driven Syntax.** John Benjamins Publishing Company.

Taraldsen, K. T. (1980).**On the NIC, vacuous application and the that-trace filter.** Indiana University Linguistics Club, Bloomington.

Taylor, A., Warner, A., Pintzuk, S., & Beths, F. (2003). **The York-TorontoHelsinki Pa**

References III

- van Gelderen, E. (2000). **A History of English Reflexive Pronouns.** John Benjamins Publishing Company.
- Vikner, S. (1997). **V-to-I movement and inflection for person in all tenses.** In Elements of grammar-handbook in generative syntax (pp. 189–213). Kluwer Academic Publishers.
- Walkden, G. (2013).**Null subjects in Old English.** Language Variation and Change, 25(2), 155–178.
- Walkden, G. (2021).**Do the wealthy stay healthy? Rich agreement and verb movement in early English.** Journal of Historical Syntax, 5(30), 1–28.

Appendix