Observing and simulating changes in the Germanic past tense system

Dirk Pijpops & Isabeau De Smet

Quantitative Lexicology & Variational Linguistics, University of Leuven
Research Foundation Flanders (FWO)

How to combine empirical research with computer simulation and why?

Theorize

Assume

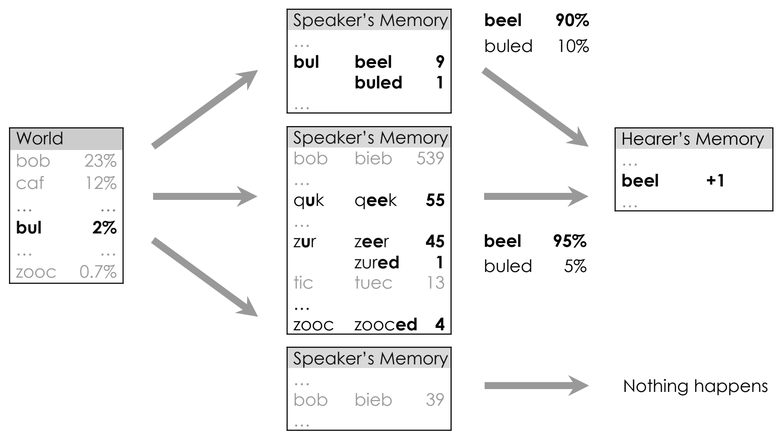
- Openness
- Single mechanism: exemplar-based
- Fledgling weak inflection

Do not assume

- Irregularity
- Memory constraints
- Segmentability

Why

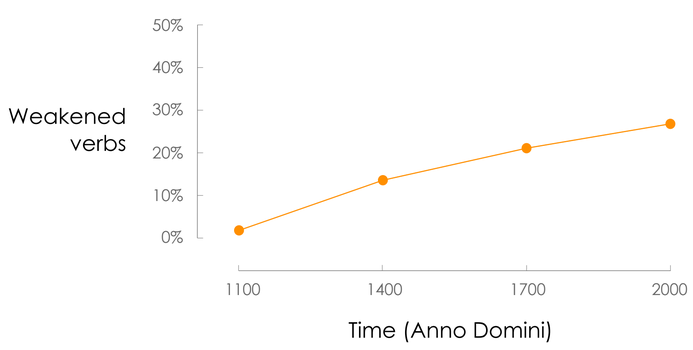
- Use Occam's Razor
- Shift the burden of proof
- It's ridiculously easy: Babel2



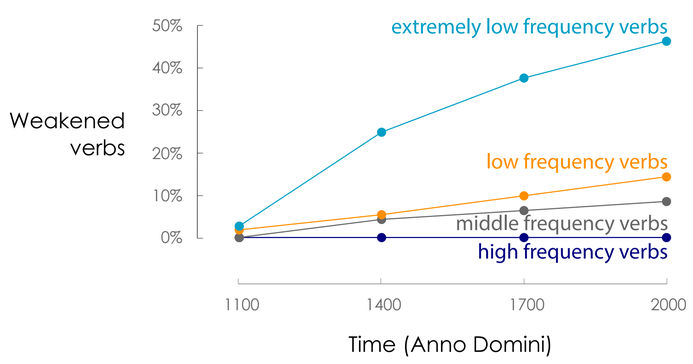
Observe

What does reality look like?

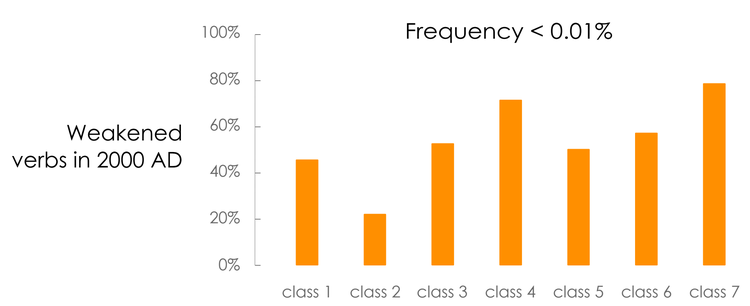
1. Gradual Rise



2. Conserving Effect



3. Class Resilience

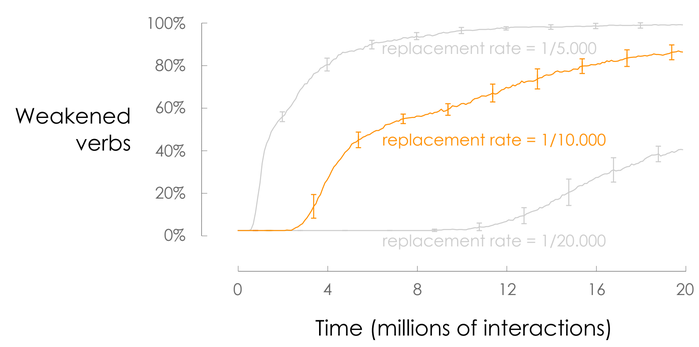


De Smet, Isabeau, 2016. De verzwakking van het preteritum in het Nederlands. Master's Thesis, University of Leuven.

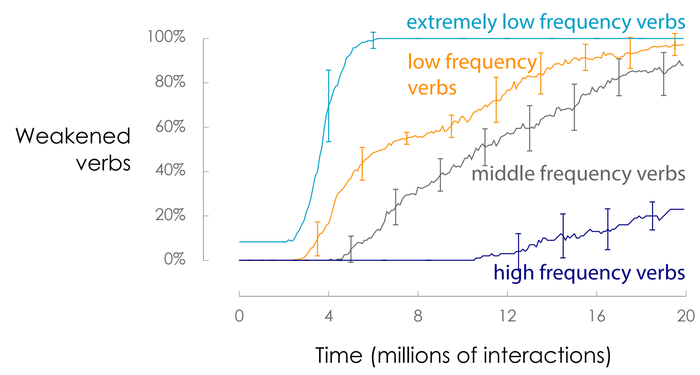
Simulate

What should reality look like?

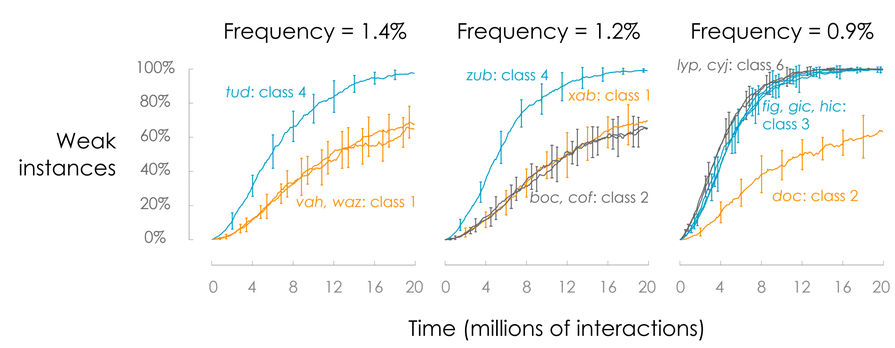
1. Gradual Rise



2. Conserving Effect



3. Class Resilience



Pijpops, Dirk, Katrien Beuls and Freek Van de Velde. 2015. The rise of the verbal weak inflection in Germanic. *CLIN Journal* 5. 81–102.

Dirk Pijpops holds Master's degrees in Artificial Intelligence and Linguistics. He is currently pursuing a PhD in Linguistics on alternations between Dutch direct and prepositional objects at the University of Leuven, under supervision of Dirk Speelman.

dirk.pijpops@kuleuven.be

Isabeau De Smet holds a Master's degree in Linguistics. She wrote her MA-thesis on the preterite morphology. This year she has continued that research in a one year project at the University of Leuven with her supervisor Freek Van de Velde.

isabeau.desmet@kuleuven.be

Empirical Data

To allow for easy comparison with English (Lieberman et al. 2007) and German (Carroll et al. 2012), the data selection procedure replicated that of these earlier studies as closely as possible. 164 verbs were selected which were marked as strong in several dictionaries and reference grammars of Old Dutch (800-1200) and which could be tracked in dictionaries or reference grammars of Middle Dutch (1200-1500), Modern Dutch (1500-1900) and Contemporary Dutch (1900 onwards, see references for used dictionaries and grammars). These verbs were coded as strong (1), varying (0.5) or weak (0). Only base forms without suffixes were taken up, unless exclusively complex forms were attested. While coding, only preterite forms were considered, not participles. Not selected were the preterite-presents, irregular weak verbs, and verbs whose choice of preterite was dependent upon its meaning. The frequency of each verb was counted in the Corpus of Spoken Dutch, and divided by the total frequency of all verbs in the corpus. The 4 frequency bins shown in the graph above contain verbs with frequency > 1%, 1%-0.1%, 0.1%-0.01%, and < 0.01%.

Simulation Design

Before each interaction, a verb is selected from a set of 40 nonsense verbs. Each verb's chance of being selected corresponds to its frequency. These frequencies follow a Zipfian distribution, with the verb v of rank n having the frequency $freq(v_n) = \| \frac{100}{n} \| / \sum_{i=1}^{100} \| \frac{100}{i} \|$. Next, a speaker and a hearer agent are randomly selected from a population of 100 agents and interact according to the flow chart above. All starting agents are initiated with a memory of 39 strong forms for the 39 most frequent verbs and a single weak form for the least frequent verb. The initial memory count of verb v of rank n is $count(v_n) = \| \frac{100}{n} \|$. The 39 initially strong verbs are distributed across 7 ablaut classes as to create classes with equal token frequency, but different type frequency and vice versa. Every 10.000 interactions, 1 agent is replaced by a new agent with an empty memory. In the current settings, verbs are never replaced. The graphs show the running averages and standard deviations of 20 series of each 20 million interactions. The 4 frequency bins shown in the graph above contain verbs with frequency > 4%, 4%-1.5%, 1.5%-0.7%, and < 0.7%.

Acknowledgments

We cordially thank Katrien Beuls and Freek Van de Velde for their indispensable contributions to both the empirical study and the agent-based simulation. In addition, we would like to thank Remi van Trijp for interesting discussions and useful advice about the simulation, as well as the participants of the SLE-48 workshop *Shifting classes: Germanic strong and weak preterites and participles*.

References

Bailey, Christopher Gordon. 1997. The Etymology of the Old High German Weak Verb. University of Newcastle upon Tyne.

Ball, Christopher. 1968. The Germanic dental preterite. *Transactions of the Philological Society* 67. 162–188.

Boon, Ton den & Dirk Geeraerts (eds.). 2005. *Van Dale Groot woordenboek van de Nederlandse taal*. 14th ed. Antwerpen/Utrecht: Van Dale Lexicography.

Carroll, Ryan, Ragnar Svare and Joseph Salmons. 2012. Quantifying the evolutionary dynamics of German verbs. *Journal of Historical Linguistics* 2(2). 153–172.

Colaiori, Francesca, Claudio Castellano, Christine Cuskley, Vittorio Loreto, Martina Pugliese and Francesca Tria. 2015. General three-state model with biased population replacement: Analytical solution and application to language dynamics. *Physical review. E, Statistical, nonlinear, and soft matter physics* 91(1–1). 12808.

Fertig, David. 2000. *Morphological Change Up Close. Two and a Half Centuries of Verbal Inflection in Nuremberg*. Tübingen: Niemeyer.

Franck, Johannes. 1883. *Mittelniederländische Grammatik: Mit Lesestücken und Glossar [Middle Dutch Grammar: with reading excerpts and glossary]*. Leipzig: T. O. Weigel.

Franck, Johannes. 1909. *Altfränkische Grammatik [Old Franconian Grammar]*. 2nd editio. Göttingen: Vandenhoeck & Ruprecht.

Grauwe, Luc De. 1982. *De Wachtendonckse psalmen en glossen. Een lexikologisch-woordgeografische studie met proeve van kritische leestekst en glossaria, deel 2 [The Wachterdoncker psalms and glosses. A lexicological-wordgeographic study with examination of critical reading text]*. Gent: Koninklijke Academie voor Nederlandse Taal- en Letterkunde.

Haeseryn, Walter, Kirsten Romijn, Guido Geerts, Jaap de Rooij and Maarten van den Toorn. 1997. *Algemene Nederlandse Spraakkunst [General Dutch Grammar]*. Groningen: Nijhoff.

Helten, Willem Lodewijk van. 1887. *Middelnederlandsche spraakkunst [Middle Dutch Grammar]*. Groningen: J.B. Wolters.

Kate, Lambert Ten. 1723. *Aenleiding tot de kennisse van het verhevene deel der Nederduitsche sprake. Eerste deel [Introduction to the understanding of the lofty part of the Dutch Language. First part]*. Amsterdam: R. & G. Wetstein.

Ketterij, Cornelis van de. 1980. *Grammaticale interpretatie van Middelnederlandse teksten: Instructiegrammatica [Grammatical interpretation of Middle Dutch texts: Educational Grammar]*. Groningen: Wolters-Noordhoff.

Koelmans, Leendert. 1978. *Inleiding tot het lezen van zeventiende-eeuwse teksten [Introduction to reading seventeenth century texts]*. Utrecht: Instituut De Vooys voor Nederlandse Taal- en letterkunde.

Lieberman, Erez, Jean-Baptiste Michel, Joe Jackson, Tina Tang and Martin Nowak. 2007. Quantifying the evolutionary dynamics of language. *Nature* 449(7163). 713–716.

Loetzsch, Martin, Pieter Wellens, Joachim De Beule, Joachim Bleys and Remi van Trijp. 2008. The Babel2 Manual. *AI-Memo 01-08*. Brussels: AI-Lab VUB.

Loey, Adolphe van. 1980. *Middelnederlandse spraakkunst. Deel I: Vormleer [Middle Dutch Grammar. Part I: Morphology]*. 9th edn. Groningen: Wolters-Noordhoff.

Oostdijk, Nelleke, Wim Goedertier, Frank Van Eynde, Louis Boves, Jean-Pierre Martens, Michael Moortgat and Harald Baayen. 2002. Experiences from the Spoken Dutch corpus project. *Proceedings of the third international conference on language resources and evaluation (LREC)*. 340–347.

Pijnenburg, Wilhelmus. 2001. *Vroegmiddelnederlands woordenboek: woordenboek van het Nederlands van de dertiende eeuw in hoofdzaak op basis van het Corpus-Gysseling [Early Middle Dutch dictionary: dictionary of 13th century Dutch mainly based on the Gysseling Corpus]*. Leiden: Instituut voor Nederlandse Lexicologie.

Pinker, Steven and Alan Prince. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28(1). 73–193.

Quak, Arend and Joop van der Horst. 2002. *Inleiding Oudnederlands [Introduction Old Dutch]*. Leuven: Leuven University Press.

Quak, Arend. 1981. *Die altmittel- und altniederfränkischen Psalmen und Glossen [The old middle and old lower Frankish psalms and glosses]*. Amsterdam: Editions Rodopi B.V.

Sanders, Wille. 1974. *Der Leidener Willeram [The Leiden Willeram]*. Munich: Wilhelm Fink Verlag.

Taatgen, Niels and John Anderson. 2002. Why do children learn to say "Broke"? A model of learning the past tense without feedback. *Cognition* 86. 123–155.

Tack, Paul. 1897. *Oudnederfrankische grammatica [Old Lower Franconian Grammar]*. Gent: A. Siffer.

Verwijs, Eelco and Jacob Verdam. 1991. *Middelnederlandsch woordenboek [Middle Dutch Dictionary]*. Zedelgem: Zedelgem Flandria Nostra.

Vriendt, Sera de. 1965. *Sterke werkwoorden en sterke werkwoordsvormen in de 16e eeuw [Strong verbs and strong verb forms in the 16th century]*. Brussel: Belgisch interuniversitair centrum voor neerlandistiek, 1965.

Vries, Matthias de and Lamert Allard te Winkel. 1998. *Woordenboek der Nederlandsche taal [Dictionary of the Dutch Language]*. 's-Gravenhage: Nijhoff.

Zipf, George Kingsley. 1932. *Selected Studies of the Principle of Relative Frequency in Language*. Harvard: Harvard University Press.