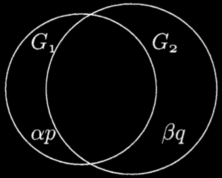


A variational model of the loss of English OV

Héctor Javier Vázquez Martínez & Beatrice Santorini
{hjvm, beatrice}@sas.upenn.edu

06.07.2023

Introduction

- Mechanism of change: Language acquisition
- Model of language acquisition: Variational Model of Learning (VM) (Yang, 2000; Yang et al., 2002)
- The aim of the work presented here is to add conceptual depth to analyses of the change by applying the Variational Model (VM) of Learning (Yang, 2000; Yang et al., 2002) to the historical data, which allow us to derive *s* from independently estimated advantages of the two competing grammars.

Overview

- **The Variational Model (VM)**
- Estimating the advantages α (VO) and β (OV)
- Predicted timecourse of the change
- Comparison of observed vs. predicted timecourse
- Discussion

A naïve model of language acquisition

- Naïve Reinforcement Learning paradigm
- A learner may entertain multiple grammars (*VO*, *OV*).
- Whenever the individual must parse an utterance *S*, one of the grammars (*VO* in below example) is sampled to do so.
- Linear Reward Penalty (LRP; Bush and Mosteller, 1951).

$$p' = \begin{cases} p + \gamma q & VO \to S \\ (1-\gamma)p & VO \not\to S \end{cases} \quad \Big| \quad q' = \begin{cases} p + \gamma q & OV \to S \\ (1-\gamma)p & OV \not\to S \end{cases} \tag{1}$$

Deriving the VM

α & β

Let α be the proportion of structures generated by the VO grammar that cannot be parsed by the OV grammar. Conversely, let β be the proportion of structures generated by the OV grammar that cannot be parsed by the VO grammar.

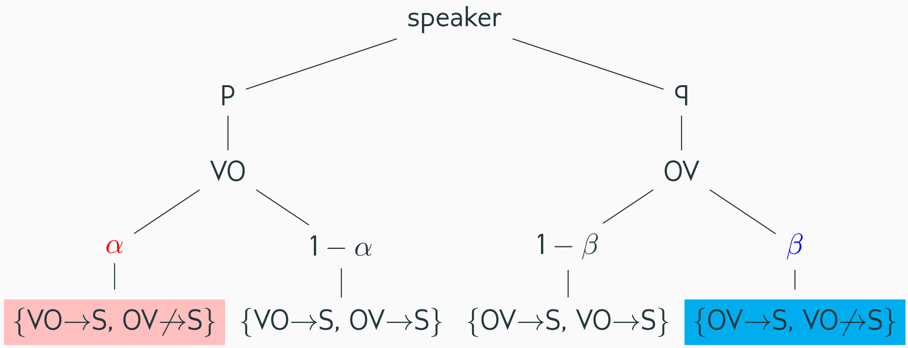
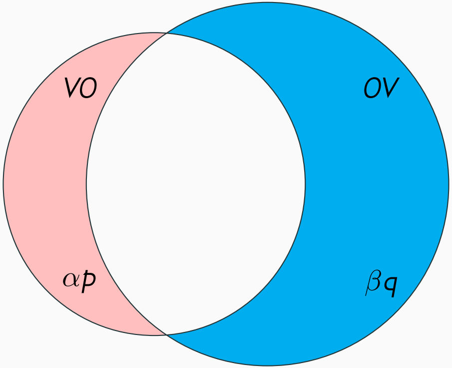


Figure 1: A tree diagram of the VM. An input utterance S is generated by the VO grammar with probability p or by the OV grammar with probability q. If generated by the VO grammar, with probability α, S cannot be parsed by the OV grammar. Conversely, if S is generated by the OV grammar, with probability β, it cannot be parsed by the VO grammar.

Deriving the VM



Penalty probabilities

Let $C_{VO} = \beta q$ define the probability that the VO grammar is penalized.

Let $C_{OV} = \alpha p$ define the probability that the OV grammar is penalized.

p & q

Let p_i be the proportion of linguistic input generated by the VO grammar at generation i. Conversely, let $q_i = 1 - p_i$ be the proportion of linguistic input generated by the OV grammar.

Algorithm description

(1) Parents in gen. i use VO with probability p_i and OV with probability q_i

(2) Learners in gen. $i+1$ learn "forever" ($t \to \infty$);

- their grammar converges to a distribution of p_{i+1} and q_{i+1}, where

$$\lim_{t\to\infty} p_{i+1} = \frac{C_{VO_i}}{C_{VO_i} + C_{OV_i}}, \qquad \lim_{t\to\infty} q_{i+1} = \frac{C_{OV_i}}{C_{VO_i} + C_{OV_i}}$$

(3) Now parents in gen. $i+1$ use VO with prob. p_{i+1} and OV with prob. q_{i+1}

(4) New learners in gen. $i+2$ converge to a distribution of p_{i+2} and q_{i+2}

... ad infinitum

Deriving the VM

- Distribution of OV/VO in the next generation:

$$p_{i+1} = \frac{C_{VO_i}}{C_{VO_i} + C_{OV_i}} \tag{2}$$

- Recall from the Venn Diagram that $C_{VO} = \beta q_i$ and $C_{OV} = \alpha p_i$:

$$p_{i+1} = \frac{\alpha p_i}{\alpha p_i + \beta q_n} \tag{3}$$

relative advantage

Let s be the relative advantage of the VO grammar, defined in terms of the advantages of the VO/OV grammars as:

$$s = \frac{\alpha - \beta}{\alpha}$$

Deriving the VM

- Distribution of OV/VO in the next generation:

$$p_{i+1} = \frac{C_{VO_i}}{C_{VO_i} + C_{OV_i}} \tag{2}$$

- Recall from the Venn Diagram that $C_{VO} = \beta q_i$ and $C_{OV} = \alpha p_i$:

$$p_{i+1} = \frac{\alpha p_i}{\alpha p_i + \beta q_i} \tag{3}$$

- Rewrite p_{i+1} in terms of the relative advantage (s) of the VO grammar

$$p_{i+1} = \frac{p_i}{p_i + (1 - s) q_i} \tag{4}$$

Time course of change

The VM allows us to calculate the number of generations (n) to reach a target distribution (p_n & q_n) given the fitness (s) of the winning grammar and a starting distribution (p_0 & q_0).

$$p_{i+1} = \frac{p_i}{p_i + (1-s)q_i}, \qquad q_{i+1} = \frac{(1-s)q_i}{p_i + (1-s)q_i}$$

Divide q_{i+1} by p_{i+1}:

$$\frac{q_{i+1}}{p_{i+1}} = (1-s)\frac{q_i}{p_i} \qquad \rightarrow \qquad \frac{q_n}{p_n} = (1-s)^n \frac{q_0}{p_0}$$

Take the logarithm of both sides and solve for n:

$$n = \frac{(log(\frac{q_n}{p_n}) - log(\frac{q_0}{p_0}))}{log(1-s)} \tag{5}$$

Overview

- The Variational Model (VM)
- **Estimating the advantages α (VO) and β (OV)**
- Predicted timecourse of the change
- Comparison of observed vs. predicted timecourse
- Discussion

Estimating the advantages

Data sources

- Proxy for VO grammar: Appalachian English (AAPCAppE, Tortora et al., 2017) - oral history interviews from the 1930s, 1980s, and 1990s
- Proxy for OV grammar: Swiss German (WilKo, Schönenberger et al., in progress) - sociolinguistic interviews from Wil, Switzerland from the 2010s

Note: Middle Low German (Booth et al. (2020)) cannot be used to estimate β because it exhibits grammatical variation between OV and VO.

Estimating the advantages

Diagnostic contexts, preliminary considerations

- Exclude clauses with nonfinite verbs
- For present purposes, include clauses with negative and quantified objects
- Count clauses with short (1-2 words) postverbal and long (3+ words) preverbal objects toward α and β, respectively.

Estimating the advantages

Diagnostic context	AAPCAppE (VO)	Swiss German (OV)
verb > pronoun	6,281	
verb > particle	3,316	
verb > stranded prep.	776	
verb > short obj	8,647	
particle > verb		4,375
stranded prep. > verb		0
long obj > verb		2,058
Total unambiguous	19,020	6,433
Total clauses	144,923	109,525
Estimated advantage	**0.13** (α)	**0.06** (β)

Overview

- The Variational Model (VM)
- Estimating the advantages
- **Predicted timecourse of the change**
- Comparison of observed vs. predicted timecourse
- Discussion

Number of generations to completion

We use Equation 7 (repeated below) to calculate how long the OV→VO change would take from 99.8% OV to 0.2% OV.

$$n = \frac{(log(\frac{q_n}{p_n}) - log(\frac{q_0}{p_0}))}{log(1-s)} \tag{7}$$

The VM predicts $n \approx 17.4$ generations for English to converge at a categorically VO grammar.

Assuming a mean generation length of 30 years (Tremblay and Vézina, 2000; and others), the change is predicted to take place in $\approx$ 522 years.

Simulating the change backward

The precise date in which the VO grammar came into competition with the OV grammar is unknown.

We do know the change was completed around 1450-1500 C.E.

We fix the endpoint at 1500 C.E. and solve Equation 3 for p_i in terms of p_{i+1}:

$$p_i = \frac{(\beta p_{i+1})}{(\alpha - \alpha p_{i+1} + \beta p_{i+1})} \tag{6}$$

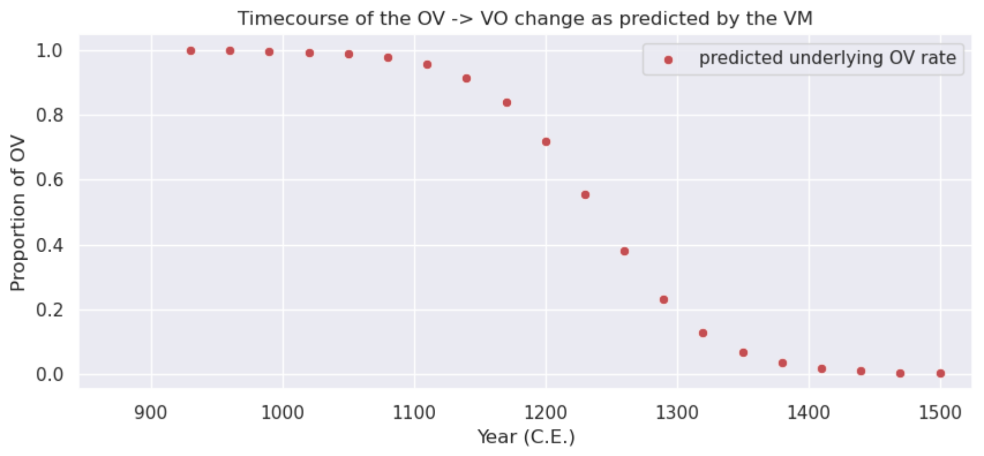


Figure 2: Predicted timecourse of the English OV→VO change by the VM. The endpoint was fixed at 1500 C.E. (generation *n*, 99.8% VO), and the model was allowed to run backward until it reached a distribution of 99.8% OV (generation 0). Each dot represents a generation.

Overview

- The Variational Model (VM)
- Estimating the advantages
- Predicted timecourse of the change
- **Comparison of observed vs. predicted timecourse**
- Discussion

Comparison of observed vs. predicted timecourse

YCOE (Taylor et al., 2003) - Old English prose, revised and corrected unpublished version of next release

Excluded:

- IP issues: apollo, lacnu, wulf
- Duplicate texts: chronC,D,E, exodusP, genesiC, inspolD, lsigewB, lwsigeT, nicodC,E
- Duplicate sentence tokens
- Certain or possible translations from Latin

Comparison of observed vs. predicted timecourse

PPCME2 (Kroch and Taylor, 2000) - Middle English prose, revised and corrected unpublished version of next release

Excluded:

ayenbi-m2, brut3-m3, orm-m1, reynes-m4, siege-m4

PLAEME (Truswell et al., 2019) - Middle English prose, revised and corrected unpublished version of next release

Included:

benetholm, buyfft, chertseyt, creditonat, creditonbt, gospatrict, hunctproct, ramseyat, ramseybt, ramseycott, thorneykt

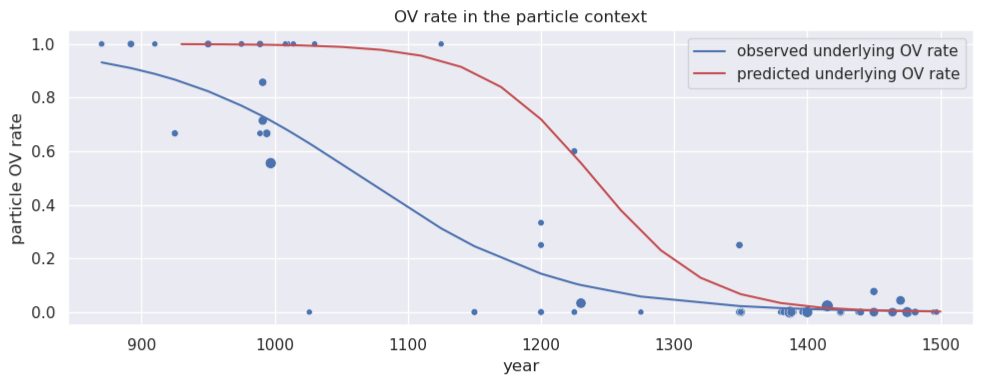


Figure 3: OV rate by date of composition in the particle verb context. The size of the points indicates the relative number of tokens considered in each text. The logistic curve is fit by weighting each text according to its number of tokens.

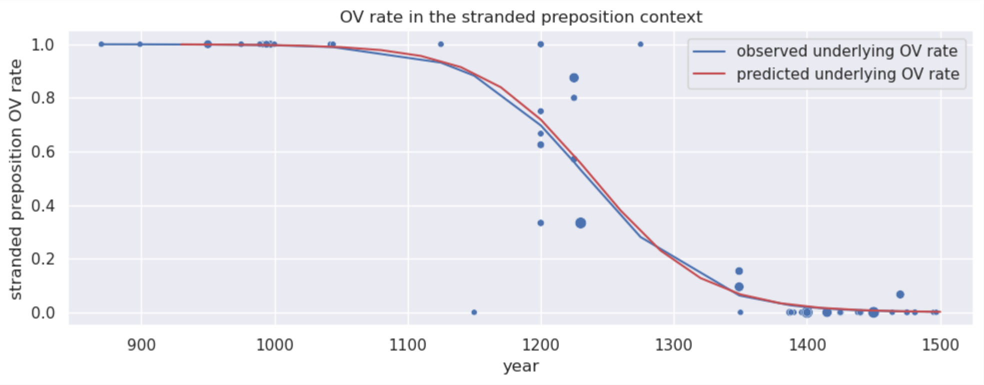


Figure 4: OV rate by date of composition in the stranded preposition context. The size of the points indicates the relative number of tokens considered in each text. The logistic curve is fit by weighting each text according to its number of tokens.

Overview

- The Variational Model (VM)
- Estimating the advantages *alpha* and *beta*
- Predicted timecourse of the change
- Comparison of observed vs. predicted timecourse
- **Discussion**

Discussion

- Duration of change (time depth)
- Different slopes for different contexts
- Difficulties with estimating β (preposition stranding, concurrent headedness changes)
- Sociolinguistic complications (Viking invasions, Norman Conquest)
- OV/VO change is basic, but not simple
- Are there simpler test cases?

Acknowledgments

Thanks to

- Ann Taylor, for making available a pre-release version of the next release of the YCOE.
- Robert Truswell, for making available PLAEME and for help with deciding which texts to include.
- Ann Taylor, Eric Haeberli, and especially Richard Zimmermann, for help with deciding which Old English data to include
- Donald Ringe, for reviewing our translations of the Old English examples
- The audience at ILST at the University of Pennsylvania, for their questions and comments

Thank you

Appendix A:
OV/VO variation in Old and Middle English

OV/VO variation in Old English, particles

Preverbal particle

(1) and mid geornfulre elnunge up arisende w+as
and with anxious zeal up arising was
'and arose with anxious zeal'

(coaelive,+ALS[Martin]_247.6125)

Postverbal particle

(2) and het hi gan ut
and commanded them go out
'and had them leave'

(coaelive,+ALS[Martin]_212.6101)

OV/VO variation in Old English, pronouns

Preverbal pronoun

(3)

for+dan	+te	Crist	wolde	hi	geedcucian	of	dea+de
because	that	Christ	wanted	them	revive		from	dead

‘because Christ wanted to raise them from the dead’

(coaelive,+ALS[Ash_Wed]_231.2835)

Postverbal pronoun

(4)

+Ta	wolde	he	+turh+tyn	hi	+twyres	mid	+tam	swurde
then	wanted	he	through-pierce	them	across	with	the	sword

‘Then he wanted to pierce them in the side with the sword’

(coaelive,+ALS[Ash_Wed]_225.2829)

OV/VO variation in Old English, full noun phrases

Preverbal ordinary noun phrase

(5)

[se	hiredes	ealdor]	wolde	hys	w+astmas	habban
the	family's	elder	wanted	his	harvest	have

'[the head of the household] wanted to have his harvest'

(coaelhom,+AHom_3_115.477)

Postverbal ordinary noun phrase

(6)

se	man	+te	het	+te	niman	+tin	bed
the	man	that	commanded	thee	take	thy	bed

'the man that told you to take your bed'

(coaelhom,+AHom_2_47.270)

OV/VO variation in Middle English, particles

Preverbal particle

(7) +tat +te brused blod may out ren
'that the bruised blood may run out'
(cmhorses-m3,119.325)

Postverbal particle

(8) +tat it may come out +the bettur
'that it [the blood] may come out the better'
(cmhorses-m3,119.327)

OV/VO variation in Middle English, pronouns

Preverbal pronoun

(9) what shold me lette
'What would prevent me?'

(cmreynar-m4,59.599)

Postverbal pronoun

(10) not to the thentente that men shold vse them
'not so that people should use them'

(cmreynar-m4,6.6)

OV/VO variation in Middle English, full noun phrases

Preverbal ordinary noun phrase

(11) Lauerd, we pray +te +tat we may suilke cunsal take, ...
'Lord, we ask you that we may take such counsel ...'

Postverbal ordinary noun phrase

(12) ... +tat we may do +ti wille.
'... that we may do your will.'
(both examples from the same sentence, cmbenrul-m3,8.239)

Appendix B:
OV/VO variation in Old Saxon and Middle Low German

Phrase structure variation in Old Saxon and Middle Low German

Diagnostic context	Old Saxon	Middle Low German
verb > pronoun	0	42
particle > verb	12	145
verb > particle	3	23
stranded prep. > verb	5	163
verb > stranded prep.	3	6

Source corpora: HeliPaD (Walkden, 2016), CLHG (Booth et al., 2020)

Appendix C: OV/VO advantages in other corpora

Estimates for the advantages, other corpora

Diagnostic context	CORAAL	MBE2	PCEEC2	Tueba-D/Z
verb > pronoun	5,276	14,205	19,538	
verb > particle	3,079	4,991	2,992	
verb > stranded prep.	828	938	1,051	
verb > short obj	8,407	18,675	17,453	
particle > verb	0	0	0	10,050
stranded prep. > verb	0	0	0	
long obj > verb				7,016
Total unambiguous	17,590	38,809	41,034	17,066
Total clauses	141,211	323,142	267,455	151,785
Estimated advantage	0.12	0.12	0.15	0.11

Appendix D: Observed OV rates by context

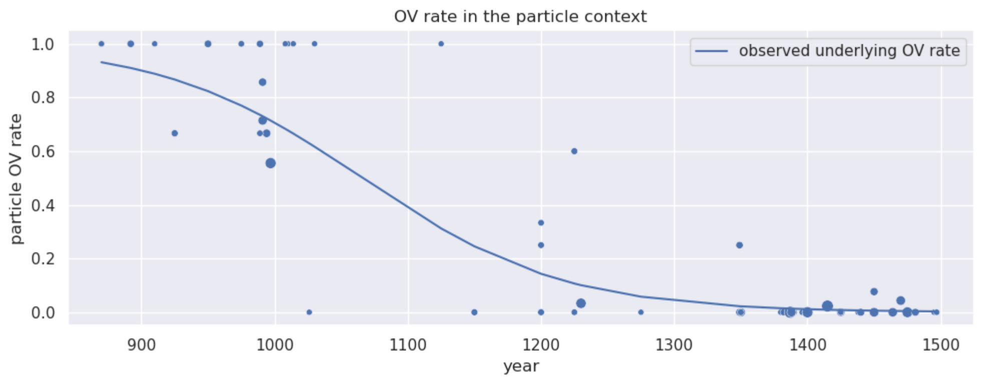


Figure 5: Observed OV rate by date of composition in the particle verb context. The size of the points indicates the relative number of tokens considered in each text. The logistic curve is fit by weighting each text according to its number of tokens.

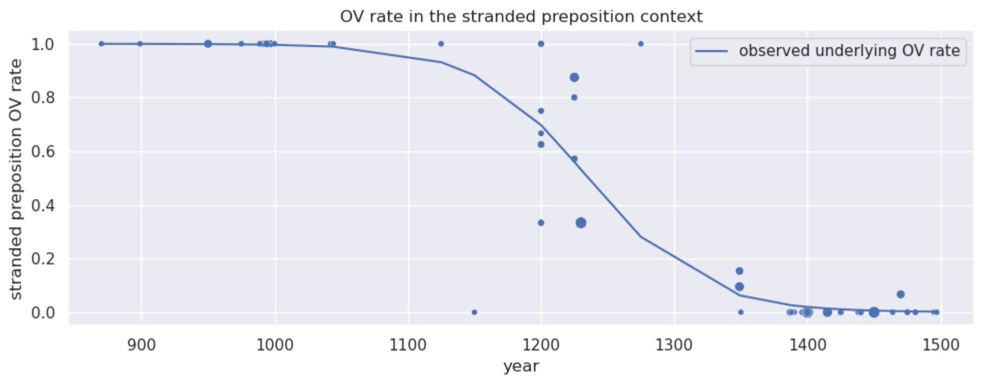


Figure 6: Observed OV rate by date of composition in the stranded preposition context. The size of the points indicates the relative number of tokens considered in each text. The logistic curve is fit by weighting each text according to its number of tokens.

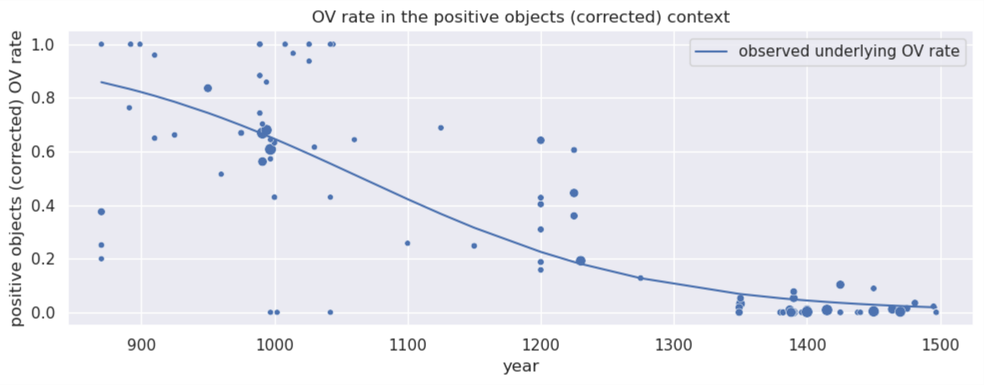


Figure 7: Observed OV rate by date of composition in the positive (nonquantified, nonnegative) context. The size of the points indicates the relative number of tokens considered in each text. The logistic curve is fit by weighting each text according to its number of tokens.

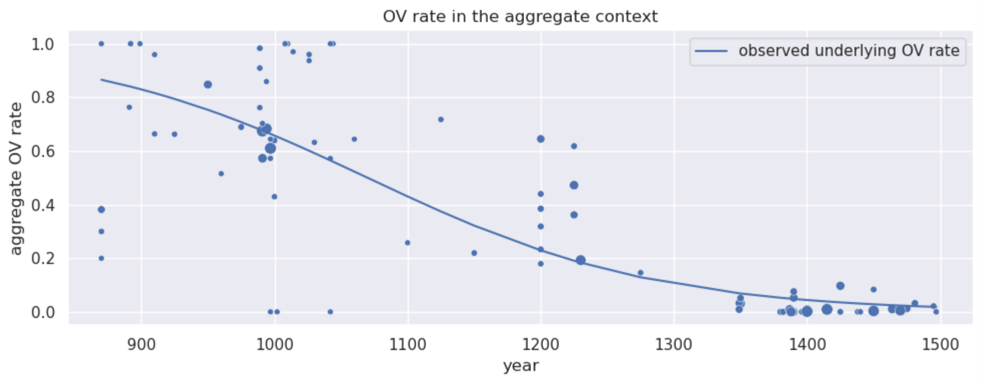


Figure 8: Observed aggregate OV rate by date of composition in all contexts: particles, stranded prepositions and positive objects. The size of the points indicates the relative number of tokens considered in each text. The logistic curve is fit by weighting each text according to its number of tokens.

Appendix E: Predicted OV→VO timecourse by context

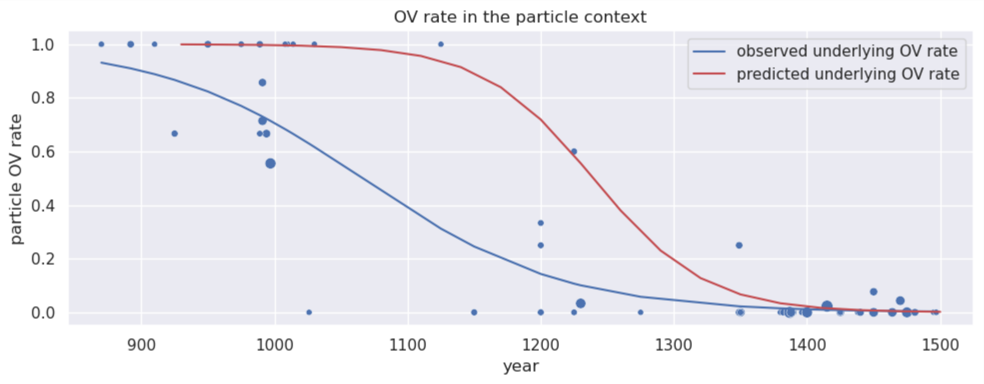


Figure 9: OV rate by date of composition in the particle verb context. The size of the points indicates the relative number of tokens considered in each text. The logistic curve is fit by weighting each text according to its number of tokens.

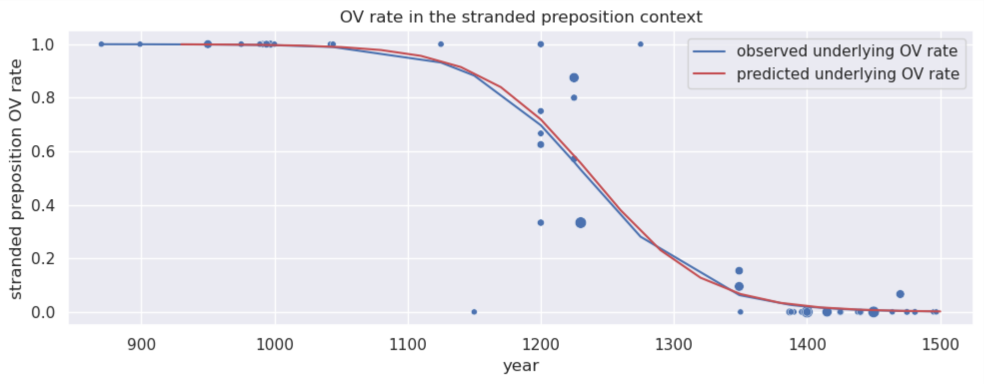


Figure 10: OV rate by date of composition in the stranded preposition context. The size of the points indicates the relative number of tokens considered in each text. The logistic curve is fit by weighting each text according to its number of tokens.

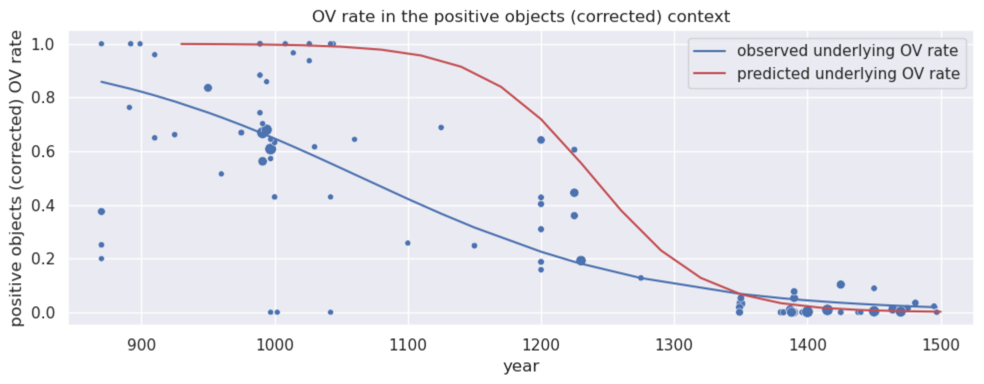


Figure 11: OV rate by date of composition in the positive (nonquantified, nonnegative) context. The size of the points indicates the relative number of tokens considered in each text. The logistic curve is fit by weighting each text according to its number of tokens.

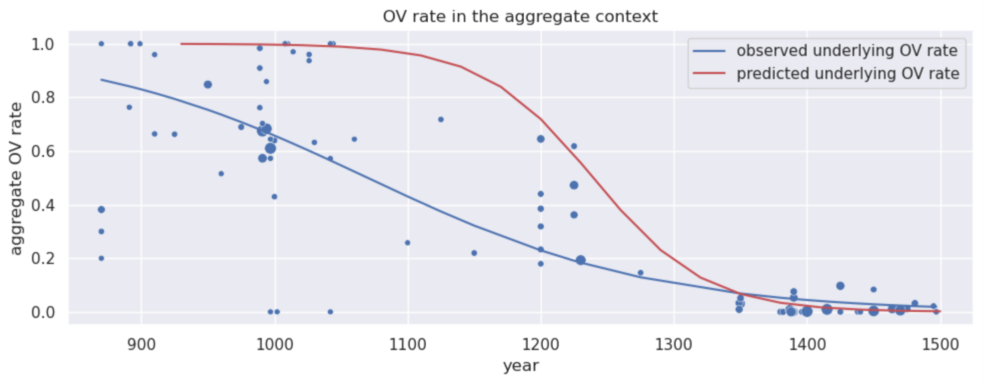


Figure 12: Aggregate OV rate by date of composition in all contexts: particles, stranded prepositions and positive objects. The size of the points indicates the relative number of tokens considered in each text. The logistic curve is fit by weighting each text according to its number of tokens.

Appendix F: Relative advantage vs. slope and intercept

From relative advantage to slope

The VM provides us a relative advantage s between the competing grammars, but no more. This is different from the slope (also s) and intercept (k) as in Kroch (1989). Howerer, we may derive the slope and intercept from the relative advantage:

$$p_{i+1} = \frac{p_i}{p_i + (1-s)q_i}, \qquad q_{i+1} = \frac{(1-s)q_i}{p_i + (1-s)q_i}$$

Divide p_{i+1} by q_{i+1}:

$$\frac{p_{i+1}}{q_{i+1}} = \left(\frac{1}{1-s}\right)\frac{p_i}{q_i} \quad \rightarrow \quad \frac{p_n}{q_n} = \left(\frac{1}{1-s}\right)^n \frac{q_0}{p_0}$$

Take the logarithm of both sides and express as $y = b + ax$:

$$log\frac{p_n}{1-p_n} = log\frac{q_0}{p_0} + log\left(\frac{1}{1-s}\right)n \tag{7}$$

Booth, H. M., Breitbarth, A., Ecay, A., and Farasyn, M. (2020). A Penn-style treebank of Middle Low German. In *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, pages 766—-775.

Bush, R. R. and Mosteller, F. (1951). A mathematical model for simple learning. *Psychological review*, 58(5):313.

Kroch, A. and Taylor, A. (2000). Verb-object order in Early Middle English. In Pintzuk, S., Tsoulas, G., and Warner, A., editors, *Diachronic syntax: Models and mechanisms*, pages 132–163. Oxford University Press.

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

Taylor, A., Warner, A., Pintzuk, S., Beths, F., et al. (2003). The York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE).

References ii

Tortora, C., Santorini, B., Blanchette, F., and Diertani, C. (2017). The Audio-Aligned and Parsed Corpus of Appalachian English (AAPCAppE), version 0.1.

Tremblay, M. and Vézina, H. (2000). New estimates of intergenerational time intervals for the calculation of age and origins of mutations. *American Journal of Human Genetics*, 66(2):651–658.

Truswell, R., Alcorn, R., Donaldson, J., and Wallenberg, J. (2019). A Parsed Linguistic Atlas of Early Middle English. In Alcorn, R., Kopaczyk, J., Los, B., and Molineaux, B., editors, *Historical Dialectology in the Digital Age*, pages 19–38. Edinburgh University Press, Edinburgh.

Walkden, G. (2016). The helipad: A parsed corpus of old saxon. *International Journal of Corpus Linguistics*, 21:559–571. Download site: https://zenodo.org/record/4395040.

References iii

Yang, C. D. (2000). Internal and external forces in language change. *Language Variation and Change*, 12(3):231–250.

Yang, C. D. et al. (2002). *Knowledge and learning in natural language*. Oxford University Press on Demand.