

A variational theory of specialization in acquisition and diachrony

Joel C. Wallenberg

joel.wallenberg@ncl.ac.uk

Centre for Behaviour and Evolution, Institute of Neuroscience

School of English Literature, Language and Linguistics

Pre-publication Draft.

Abstract

stuff

Keywords: stuff; stuff;

Corresponding Author:

Joel C. Wallenberg

Percy Building

Newcastle University

Newcastle Upon Tyne, UK

NE1 7RU

Tel.: +44-(0)191-208-7366

joel.wallenberg@ncl.ac.uk

1 Introduction

This article presents an empirical case study in the diachronic specialization of morphological forms for different syntactic contexts, and uses it to test aspects of the theory of variational specialization in Wallenberg (2016) and Fruehwald and Wallenberg (in prep). Specifically, the study tests hypotheses about the speed of diachronic specialization, and answers a question I'll refer to as Yang's Paradox: how can we reconcile diachronic results showing that specialization is slow, with experimental results on acquisition showing that it's fast?

To do this, the main empirical problem I focus on is the diachronic specialization of the forms *melted* and *molten* in pre-modern English. Though the forms initially had the same meaning and syntactic distribution, they eventually specialized for perfect/passive participle and adjectival contexts, respectively, as shown in examples (1)-(3).

- (1) The gold was {melted / *molten} by the fire. (**passive participle context**)
- (2) The fire has {melted / *molten} the gold. (**perfect participle context**)
- (3) She shaped the {?melted / molten} gold into a ring. (**adjectival context**)

Being at the word-level (or morpheme-level) of linguistic structure, this case is a good match for the acquisition literature that focuses on children's learning of novel lexical items. This study tests (and ultimately supports) the hypothesis that specialization in a speech community is orders of magnitude slower than specialization for an individual child in an experimental setting, due to the problem of coordinating the dimension and direction of specialization among many speakers. I also show how Yang (2000)'s variational grammar learning model can be extended to the problem of specialization.

The first section below sets up the problem, providing some background on the idea of variational specialization and the set of empirical results that inspired it. It also introduces Yang's paradox as a potential problem for a unified account of specialization across domains of the grammar. Section 3 describes the *melted/molten* study, and section 4 presents results from it. Section 5 discusses the implications of this study for Yang's Paradox, and refines the theory of variational specialization in light of the additional empirical information the *melted/molten* study provides. Finally I discuss some limitations of the study and conclude.

2 Variational Specialization

2.1 Previous results

PrinCon

whether/if

extraposition - here the specialization has happened, but it may have occurred before the observed period, given statistical results (plausible constancy of the effect of weight over time)

2.2 Yang's Paradox

Wallenberg (2016) and Fruehwald and Wallenberg (in prep) suggest that all specialization, across domains of the grammar, derive from the pressure exerted in acquisition by the PrinCon strategy. It is certainly tempting to suggest a single mechanism for this range of acquisition and diachronic observations. However, in light of the results on specialization in diachronic syntax, Charles Yang (p.c.) questioned the plausibility of a unified explanation, citing the very gradual, slow pace of syntactic specialization in the cases mentioned above. He reasoned that, since children specialize lexical items for different available meanings in experimental settings, and can specialize words within the time-course of a single experimental trial (e.g. the classic study Markman and Wachtel, 1988), this very fast lexical specialization must proceed by a different mechanism from the very slow syntactic specialization we've observed in those cases. The theory, as stated, seems to create an empirical paradox, and so may not be right. (One caveat, however: the review in Bion et al. 2013 shows that lexical specialization even in experimental settings needs reinforcement over some time to be retained, and so may not truly be instantaneous.) As there are doubtless differences in vocabulary acquisition and syntactic acquisition (not least of which is the effect of age on the two processes), it's a highly reasonable suggestion that there might be different mechanisms in the different grammatical domains.

However, there is also a potential resolution to the paradox, in what I'll call the two **coordination problems** of specialization in a speech community. The diachronic studies I mentioned observed the behavior of speech communities, which have importantly different properties from individual speakers in experimental settings. One difference is: a dimension along which the specialization can occur is given in the experimental setup, and doesn't have to be sought out by trial and error in the course of day-to-day life. Another difference is that the dimension of specialization doesn't need to be coordinated across individuals in a speech community, which it does in diachrony; the diachronic cases are always observations about populations of speakers, who influence each other intra- and cross-generationally, and can nullify each other's acquisition hypotheses. They also need to coordinate the direction of specialization: even if the community agrees on what domain to specialize items along, they need to agree that item A is for context A and B for B, rather than A for B and B for A.

So, a speech community needs to solve two coordination problems that do not need to be solved in the experimental context: speakers need to converge on a dimension of specialization, and to converge on which variant specializes in which direction along the dimension of specialization. These differences suggest that Yang's paradox is not truly a paradox, and suggest a simple hypothesis: if we can observe a case of word-specialization that *includes* these two coordination problems, i.e. occurs in a large speech community, that case of word specialization should also be slow. Furthermore, we should see some evidence of the coordination problems in the behavior of individuals in the populations.

3 Methods

3.1 Choice of Phenomenon

The study focuses on the morphological doublet *melted/molten*. These forms were purely formal variants of the MELT participle when the doublet arose during the Old English period (forms *gemolten*, *gemælted*, with first adnominal use of *(ge-)molten* dated to 1300 (OED XXX)). The forms then specialized over time such that *molten* became a pre-nominal adjective (or perhaps a very restricted adjectival passive), while *melted* remained a true participle (the same form occurs for passive and perfect participle contexts). These are shown above in (1)-(3).

I chose a morphological doublet rather than lexical forms with no etymological or paradigmatic relation, e.g. *shit* and *excrement*, because the latter type of doublet almost always come about under conditions that immediately suggest a dimension and direction of specialization; they most often originate in borrowings, which have a built-in social context that biases the specialization. Morphological doublets, on the other hand, often arise through overgeneralization in child language acquisition, and so can enter the speech community without any initial difference in context or meaning.

With a pair of this kind, we are more likely to observe something closer to an entire trajectory of specialization, from near total synonymy to complete specialization. This allows us to observe the speed of specialization that is the product of the PrinCon acquisition strategy, hopefully removed from strong initial biases, and we can also observe intergenerational speech community tackling both coordination problems: agreeing on a dimension of specialization, and a direction of specialization.

3.2 Study Design

This study uses the Penn-York Computer-annotated Corpus of a Large amount of English (PYCCLE-TCP; Ecay 2015), which consists of ~1 billion part-of-speech-tagged words, and is based on the Early English Books Online (EEBO) and Eighteenth Century Collections Online (ECCO) corpora. This large dataset allows sufficient time-depth to see a great deal of the specialization change as it progresses, and sufficient resolution to identify some individual speaker systems for the forms in question.

I searched PYCCLE with *Weihnachtsgurke*, a regular expression-based query language for PYCCLE (see PYCCLE citation and site). The forms *melted*, *molten*, and their spelling variants, were extracted and coded automatically for adjective or participial (passive or perfect) use, by using the part-of-speech tags in the surrounding context. I randomly sampled portions of the output to check by hand to ensure that any errors were few, randomly distributed, and due to occasional mistakes in part-of-speech tagging and not due to a systematic bias in the query. The resulting data was then analyzed statistically using R and the *lme4* package (Bates et al., 2014), and plots used *ggplot2* (Wickham, 2009). (See “Data, code and materials” below for queries and scripts.)

4 Results

The data clearly shows that during the period covered by the corpus, XXX to 1800, the probabilities of *melted* and *molten* occurring in the two syntactic contexts diverge over time; the forms specialize for the two contexts such that, by the end of the period, the chances of finding one of the forms in a given context is very different from the chances of finding the other in the same context, and the distance between those probabilities increases over the period. In Figure 1, *melted* and *molten* begin the period under investigation with very similar distributions, both being used primarily in the participle contexts. Over time, there's a decrease in the use of both forms in the participle contexts, i.e. an increase in the adjectival context, and this may simply be an overall property of the corpus. What is important for this study is that the decrease is not the same for both forms: *molten* drops in the participle uses in a way that *melted* does not, and *molten* is nearly absent from that context by the end of the time period.

Figure 2 plots the same data, but with lines for syntactic context and proportion *molten* on the y-axis. From this view, one can see that *melted* is replacing *molten* in all contexts. However, the curve is steeper in the participle context, which becomes nearly entirely restricted to *melted* before the end of the time period. I fit a mixed-effects logistic regression with random intercepts for individual Text and Author, and main effects for Year¹ of text and syntactic Context, which confirmed these intuitions. A model comparison between models with and without an interaction between Year and Context showed that the model with an interaction fit significantly better ($p = 0.0003$). AIC decreased from 4777.4 to 4766.0 and BIC decreased from 4812.3 to 4807.8 for the model with the interaction. It seems safe to conclude that the effect of syntactic context on the frequency of *molten* vs. *melted* changed over the time period; the frequencies of the two forms behaved differently over time in the two contexts.

This dataset also provides the resolution to observe some individual speaker-systems with respect to this variable. There were 471 identifiable authors in this datanwhose birth and death dates were known, with an N of 3601 tokens for these speakers. Figure 3 shows the data aggregated by author for all of these authors, plotted over time by their mid-life years. For ease of interpretation, Figure 4 zooms in on just the subset of authors whose mid-life years occur during the century of most vigorous change, between 1570-1670. Green verticle lines appear on the graph for any speaker who used both *molten* and *melted* forms in this dataset, connecting the proportions of participle use for each form for that speaker.

To get a sense of how often individual speakers had fully specialized systems, I looked at the proportions of participle use for *molten* and *melted* for any speaker who used 5 or more tokens of both forms. This fairly conservative criteria ensured that every speaker under consideration had both forms in their inventory. Out of 25 identifiable writers who use more than 5 tokens of both forms, 12 (48%) were categorical in their use of **either** form, restricting either *molten* or *melted* to either the participle uses or the adjectival use. None of those speakers were fully specialized; none fully restricted one form to one context and the other to the other context. (See further discussion below.) It is also interesting to note whether these speakers, though stochastic in their behaviour, were following the community

1. The year of text was converted to a z-score, centered around the mean, to allow the model to be fit.

trend: 19 (76%) were, but 6 showed the opposite pattern, with more *molten* in the participle contexts than *melted*.

5 Discussion

That it is specialization

The intraspeaker variation on this variable bears very directly on Yang’s Paradox: the individual writers do not necessarily show the total specialization in their idioms that one might expect on the basis of the experimental literature. By looking at the green lines in Figures 3 and 4, it is possible to get a visual sense of how completely the two forms had specialized for any given speaker, and in which direction along the syntactic dimension of specialization. A casual inspection will show the reader that in a lot of cases, the green line extends across the entire plot, indicating that the particular speaker was categorical in their use of *molten* for one context and *melted* for another, typically in the direction of the community change: *molten* for the adjectival context and *melted* for the participle. However, many of these green lines are based on very few instances of either form, and the results in section 4 tell a different story.

The results for speakers who used more than 5 tokens of both *melted* and *molten* show that most of these speakers had not entirely specialized the forms for the adjectival and participial contexts, and so, as in other kinds of linguistic change, specialization changes show considerable intraspeaker variation while in progress (“competing grammars”; ? and subsequent). The pattern of change cannot be ascribed to variation between different speakers with very different inventories, but rather reflects stochastic behavior within speakers as well. These speakers are not specializing the forms for different contexts immediately upon hearing them in their early acquisition of the forms, as the experimental results might lead us to expect.

solution to yang’s paradox Different directions of inventories and the coordination problem

6 Limitations of the Study

For some speakers, it’s possible to see the difference between the true adjectival nature of *molten* and the passive participle *melted* in present-day English, even when both are used preminally, as in the contrast in (4).

- (4) That machine takes cocoa powder and milk and produces {molten / *melted} chocolate.
- (5) The {molten / *melted} rock flowed into the ocean.

melted, as a true passive participle, describes the outcome of an event that can be described by the verb MELT, and so is infelicitous when no melting event has taken place. *molten*, on the other hand, has specialized to be a pure adjective, and so simply describes a current state with no entailment of a melting event.

mistakes in pos-tagging

7 Conclusions

Conflicts of Interest

Conflicts of interest: none.

Data, code and materials

Data sets available at:

<https://github.com/joelcw/molten>

Acknowledgments

Charles, Joe, Tony, Digs audience, Nankai audience

Funding

Appendix

query

References

- Bates, Douglas, Martin Maechler, Ben Bolker, Steven Walker, Rune Haubo Bojesen Christensen, Henrik Singmann, Bin Dai, Gabor Grothendieck, Peter Green, and Maintainer Ben Bolker. 2014. Package ‘lme4’. *R foundation for statistical computing, Vienna* 12.
- Bion, Ricardo AH, Arielle Borovsky, and Anne Fernald. 2013. Fast mapping, slow learning: Disambiguation of novel word–object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition* 126:39–53.
- Ecay, Aaron. 2015. The Penn-York Computer-annotated Corpus of a Large amount of English based on the TCP (PYCCLE-TCP). Public release 1. <https://github.com/uoy-linguistics/pyccle>.
- Fruehwald, Josef, and Joel C. Wallenberg. in prep. Optionality is Stable Variation is Competing Grammars.
- Markman, Ellen M., and Gwyn F. Wachtel. 1988. Children’s use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology* 20:121–157.
- Wallenberg, Joel C. 2016. Extraposition is disappearing. *Language* 92:e237–e256.
- Wickham, Hadley. 2009. *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. URL <http://ggplot2.org>.
- Yang, Charles. 2000. Internal and external forces in language change. *Language Variation and Change* 12:231–250.

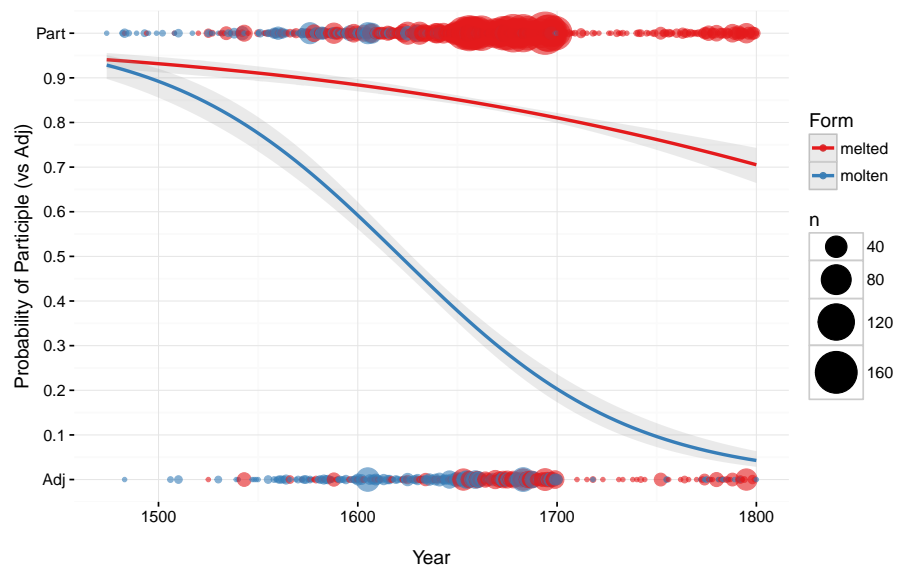


Figure 1: Syntactic context by year of text, for melted and *molten* forms over time. N = 7946 tokens.

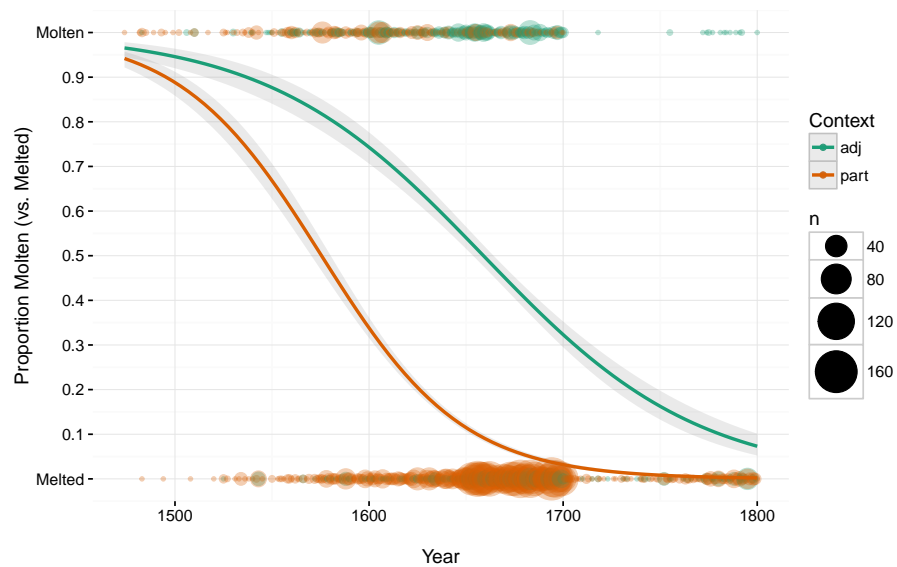


Figure 2: Proportion of *molten* uses by year of text, for both syntactic contexts over time. N = 7946 tokens.

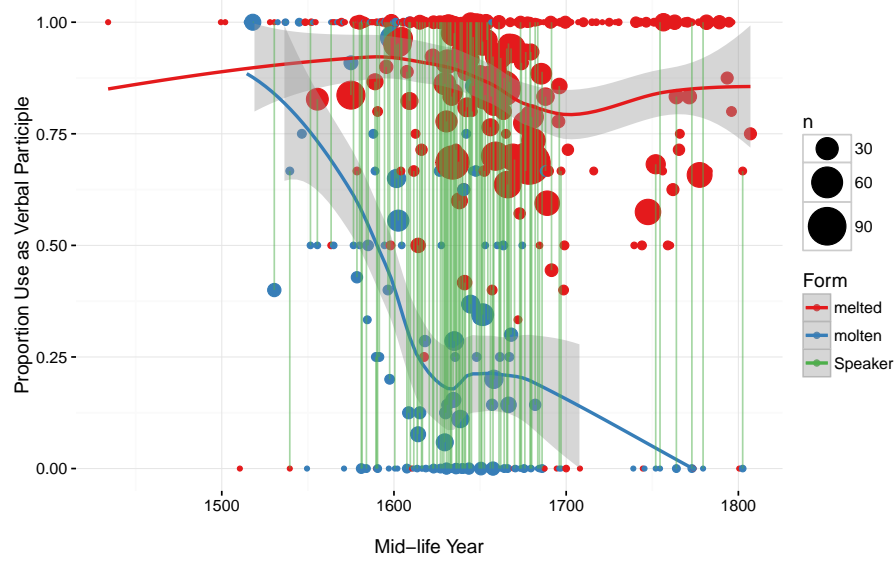


Figure 3: Syntactic context by mid-life year of author, for melted and *molten* forms over time. Green lines connect proportions of participle use with melted and molten for speakers who used both forms. 471 identifiable speakers, N = 3601 tokens.

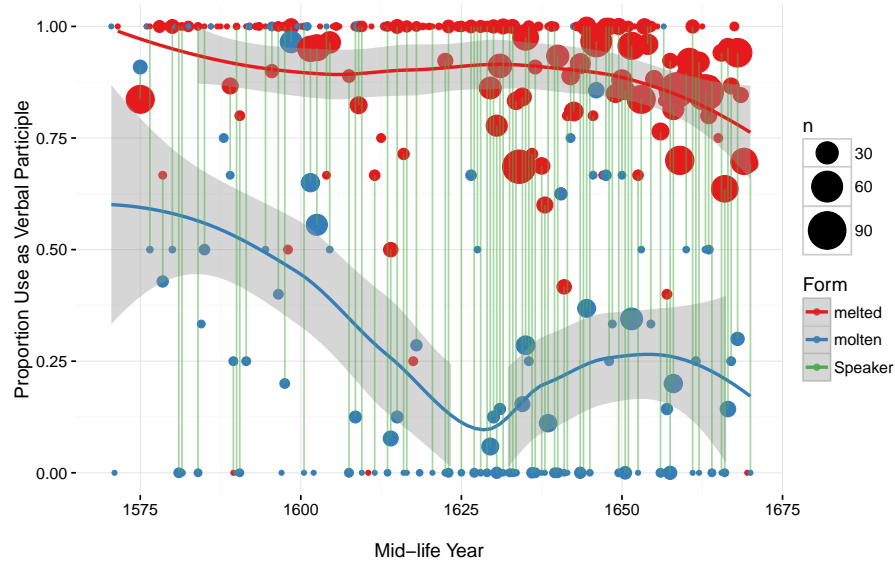


Figure 4: Individual author *melted*/*molten* systems between 1570-1670.