

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 conclude.

2 Variational Specialization

2.1 Research Context

This article takes up the proposal in Fruehwald and Wallenberg (in prep) that the “Principle of Contrast,” proposed in Clark (1987, 1990), and subsequent, is responsible for a general diachronic phenomenon of **specialization**. As originally proposed, the Principle of Contrast was stated for children’s lexical learning: that children during acquisition assign, wherever possible, contrasting word forms to contrasting meanings. It is possible for children to acquire synonyms, but Clark provides evidence that this is never the first hypothesis when a child is acquiring two phonologically distinct word forms. Experiments such as Markman and Wachtel

(1988) show that when children are presented with novel nonce words, they show a preference well above chance to associate those words with items for which they have no existing word (see also review of the experimental literature in Bion et al. 2013).

The proposal explored here is that The Principle of Contrast is much more general than originally proposed, and applies across the various modules of linguistic knowledge. I suggest that the Principle of Contrast is a general pressure on children to look for some dimension on which they can specialize forms in variation with each other, a **dimension of specialization**. Once this takes place, the dynamics of how the specialization proceeds can be called **variational specialization**, a special case of Yang (2000, 2002)’s variational learning, which is itself a special case of general evolutionary principles. One can see the Principle of Contrast as an acquisition strategy on the part of children, and surely it is, in a sense. But it is also one natural outcome of Darwinian selection on variation: whenever two variants compete for finite resources, e.g. mapping to a particular place in a child’s memory, either a variant will eventually be removed (i.e. outcompeted by the other), or the competition will. Specialization removes the competition for a particular form-function mapping.

This is the line taken by Bailey et al. (2012) and Fruehwald and Wallenberg (2013) with respect to the variation in embedded polar questions in English, as shown below:

- (4) Mary wondered whether Sue was bringing tea or coffee
- (5) Mary wondered if Sue was bringing coffee.

The reader should refer to the above references for more detail, but in short, Bailey et al. (2012) showed that the *whether* and *if* structures¹ slowly specialized for different syntactic contexts over the history of English (see Figure 1). The two contexts are clauses containing a disjunction (as in 4), and clauses without one (as in 5). The initial competition between the *whether* and *if* variants has very gradually been removed by specialization. (Note that in Icelandic, the same initial competition was removed by extinction of the *if* variant.)

Fruehwald and Wallenberg (2013) propose that all instances of specialization, including the lexical cases found in the experiments above, and the gradual syntactic case observed by Bailey et al. (2012), have the same mechanism, variational specialization, an extension of Yang (2000, 2002)’s variational learning. Yang suggested that a child learns forms A and B, and can track probabilities for them occurring, p and $(1-p)$. To this, I add that the variants are only licensed in a particular context, C. Specialization takes place when a child selects some dimension of specialization. This means that the child finds some way of dividing context C into sub-contexts, C_1, \dots, C_n , which will become important for the specialization. Next, they decouple the variants’ probability estimates in one context, e.g. C_1 , from those for the rest of the contexts and keeps track of the probabilities of each variant, A and B, for C_1 separately from, e.g. C_2 . This allows the probabilities of A and B in C_1 to diverge from those probabilities in C_2 , and allows, for instance, A to win the competition in C_1 , but lose in C_2 . This would be a case of complete specialization: both variants

1. As to the question of how similar or different the *whether* and *if* structures are syntactically, see Larson (1985) and Han and Romero (2004).

survive, but in entirely non-overlapping contexts, and so the competition between A and B is entirely removed. In this view, the Principle of Contrast is really the impetus for children to look for salient dimensions of specialization along which to divide C, and the rest of the process is straightforward Darwinian competition.

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 2, *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 3 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 Year² 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 4 shows the data aggregated by author for all of these authors, plotted over time by their mid-life years. For ease of interpretation, Figure 5 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

2. 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

The results overall show that there is specialization of the forms for different contexts or functions, specifically because the *molten* form becomes increasingly restricted to the adjectival context over time. The data also shows that the solution to Yang’s Paradox is not that there’s a different mechanism for specialization in the word domain than in the syntactic or structural domain; specialization in real historical data takes a long time in the morphological/lexical domain as well. The *melted/molten* dataset is nearly unique among historical datasets, in taking us from the very beginning of the specialization change, in the mid 15th century when there’s no detectable difference in usage between the forms, through to the end of the change in the late 18th century, when *molten* is almost entirely restricted to adjectival contexts. Thus, we can see that complete specialization of *molten* took over 300 years, which is not the near-instantaneous specialization observed in the acquisition literature. The paradox still holds, even in the word domain, if community coordination is not taken into account.

The persistence of the paradox, without an understanding of community coordination, is shown even more strongly in the intraspeaker variation on this variable: the individual writers do not necessarily show the total specialization in their idiolects that one might expect on the basis of the experimental literature. By looking at the green lines in Figures 4 and 5, 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* actually strengthen Yang’s Paradox, showing that most of these speakers had not entirely specialized the forms for the adjectival and participial contexts. So, as in other kinds of linguistic change, specialization changes show considerable intraspeaker variation while in progress (“competing grammars”; Kroch 1989 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 clearly not specializing the forms for the different contexts immediately upon hearing them in their early acquisition of the forms, as the experimental results might lead us to expect, but they are still partially specializing the forms in the direction of the ultimate community change.

The solution to the now strengthened Yang’s Paradox lies in the two coordination problems I outlined above in section 2.2. The intraspeaker variation suggests the presence of both coordination problems in this data. The fact that no speakers, of those who produced both tokens more than 5 times, had fully specialized them,

and a minority had specialized even one form for a particular context, is explained by the problem of coordinating the community on a dimension of specialization. Since the community cannot agree on participle vs. adjective as the dimension of specialization in a single generation, and various individuals of the first generation in which the *melted/molten* doublet came into existence will attempt specialization along a variety of idiosyncratic dimensions, the next generation cannot help but hear both forms in essentially all contexts. They will be true synonyms, which children can learn if they have to, competing forms for a single meaning. They can also learn probabilities for the use of competing forms, as described in the Variational Learning models of Yang (2000) and Yang (2002). As more and more speakers in the community come to converge on the adjective vs. participle dimension of specialization, the forms will gradually specialize, but there will still be plenty of evidence for both forms in both contexts for quite a few generations. At the moment, I do not have a theory of how the convergence on a dimension takes place, as there are many linguistically and socially salient dimensions that could become the coordinated dimension of specialization. I think it likely that random sampling of the forms by children in salient contexts can lead to an uneven distribution of forms per context by chance in some salient dimension, e.g. adjective vs. participle, and this begins the process of community coordination.

The best evidence that community coordination is gradually and imperfectly taking place during the period of change comes from the fact that speakers differ on the direction of specialization, even if they agree on the dimension. This is the second coordination problem. 24% of the speakers with more than 5 tokens of each form showed an idiolectal direction of specialization that went *against* the overall direction of the speech community. While the numbers are low and so do not reach statistical significance within speakers, it is at the very least clear that these speakers have not yet adopted the community's ultimate direction of specialization; community coordination on which form should specialize for which function is still in progress in this dataset.

One remaining aspect of the data deserves comment: at the same time as the specialization of *molten* is in progress, the *melted* form is replacing the *molten* form in all contexts. This is not a contradiction to the specialization result; the logistic model comparison above, identifying a significant interaction between Year and Context in the data, shows that there truly is specialization for context taking place, even as *melted* continues to replace *molten* in both contexts. The continuing replacement may be because *melted*, as a productive passive participle, can also be used adnominally as an adjectival passive. *molten*, on the other hand, in the adjectival context, is probably only an adjective. In this way, *melted* can be marshalled in productively to serve as an adjective at any time by any speaker who controls the participle *melted*, which in effect keeps its abundance high over time in the adjectival context, and creates a selectional pressure that allows it to outcompete *molten* in a way that a simple adjective competitor could not. But what is more important for the theory of specialization is simply the fact that *melted* can continue to outcompete *molten* even after they've begun to specialize along a dimension of specialization.

The fact that this can be the case gives us insight into the passive, evolutionary nature of variational specialization. This result of simultaneous specialization and replacement of one form by another in all contexts is compatible with an analysis of specialization in which the only active role of the acquirer, the Principle of Contrast,

consists of choosing a dimension of specialization. Once that is chosen, and once the community converges on it, the result is that probabilities for the competing variants are tracked separately in the different contexts defined by the dimension of specialization. So, instead of tracking overall probabilities of *melted* and *molten*, the probabilities for the variants are stored separately for the adjectival and participle contexts. This allows the specialization to proceed, and would allow for *melted* to survive in one context and *molten* to survive in the other if the selectional pressures affecting the competition in each context allow for that.

However, in the data above, it seems that the selectional pressures are different in the two contexts such that *melted* outcompetes *molten* in both contexts, possibly because of the pressure I mentioned above. This means that specialization is taking place to an extent, as the rate of *molten*'s decline is different in the different contexts, but the selectional pressures of the environments are not sufficiently different to allow *molten* to actually win in the adjectival context. This result is decidedly not compatible with a view of specialization in which the acquirer actively forces both forms to survive in different contexts. Under my analysis, the acquirer merely creates the possibility for survival of the forms by choosing a dimension of specialization, and then evolutionary selection takes its course in the two contexts.

6 Conclusions

This study has introduced the idea of variational specialization, an extension of Yang (2000, 2002), and shown that it plausibly explains diachronic cases of specialization across domains of linguistic knowledge. These results are reconciled with experimental evidence on lexical specialization by accounting for the social coordination problems that need to be solved in the speech community case, but not in the experimental cases. The results taken together show that a unified theory of specialization is possible. Furthermore, the details of how *melted* and *molten* specialized over time support a hypothesis under which specialization is mostly passive on the part of the learning, and the product of evolutionary dynamics. The learner actively chooses a dimension of specialization, decoupling their tracking of probabilities of variants in a few contexts, but how the variation plays out in the different contexts is not actively decided by the learner.

Conflicts of Interest

Conflicts of interest: none.

Data, code and materials

Data sets available at:

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

Acknowledgments

I'd like to thank Charles Yang, Josef Fruehwald, Anthony Kroch, and audiences at the Workshop on Diachronic Stability at DiGS 18 at the University of Ghent, and at the 17th International Conference on Formal Linguistics at Nankai University.

References

- Bailey, Laura, Joel C. Wallenberg, and Wim van der Wurff. 2012. Embedded yes/no questions: reanalysis and replacement. Paper presented at Presented at the 2012 Annual Meeting of the Linguistics Association of Great Britain (LAGB), University of Salford.
- 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.
- Clark, Eve V. 1987. The Principle of Contrast: A constraint on language acquisition. In *Mechanisms of language acquisition*, ed. Brian MacWhinney, The 20th Annual Carnegie Symposium on Cognition. Lawrence Erlbaum Associates, Inc.
- Clark, Eve V. 1990. On the pragmatics of contrast. *Journal of Child Language* 17:417–431.
- 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. 2013. Optionality is Stable Variation is Competing Grammars. *Presented at 25th Scandinavian Conference of Linguistics, Formal Ways of Analyzing Variation (FWAV) Workshop*.
- Fruehwald, Josef, and Joel C. Wallenberg. in prep. Optionality is Stable Variation is Competing Grammars.
- Han, Chung-hye, and Maribel Romero. 2004. The syntax of whether/q... or questions: Ellipsis combined with movement. *Natural Language & Linguistic Theory* 22:527–564.
- Kroch, Anthony S. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1:199–244.
- Larson, Richard K. 1985. On the syntax of disjunction scope. *Natural Language & Linguistic Theory* 3:217–264.
- 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.
- Yang, Charles D. 2002. *Knowledge and Learning in Natural Language*. Oxford: Oxford University Press.

Figure 1: Stuff XXX.

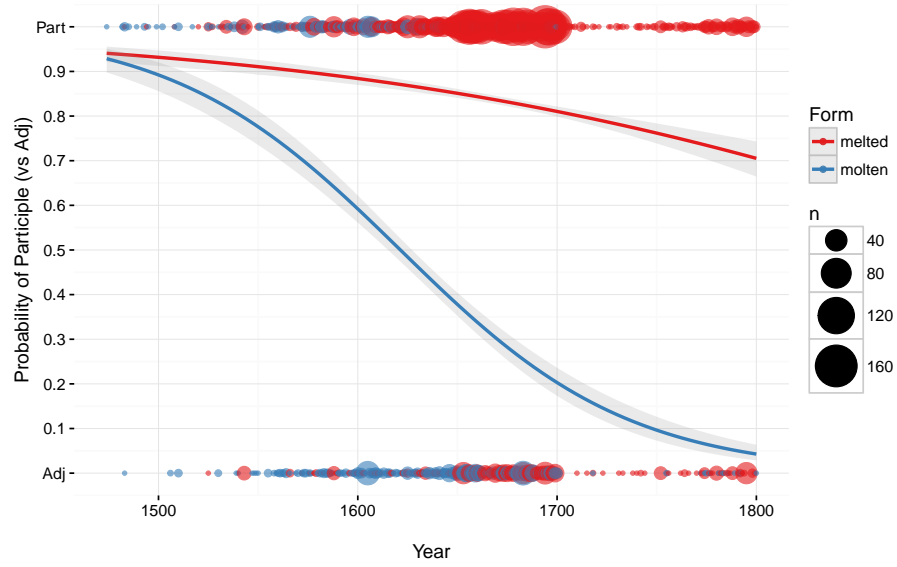


Figure 2: Syntactic context by year of text, for melted and *molten* forms over time. N = 7946 tokens.

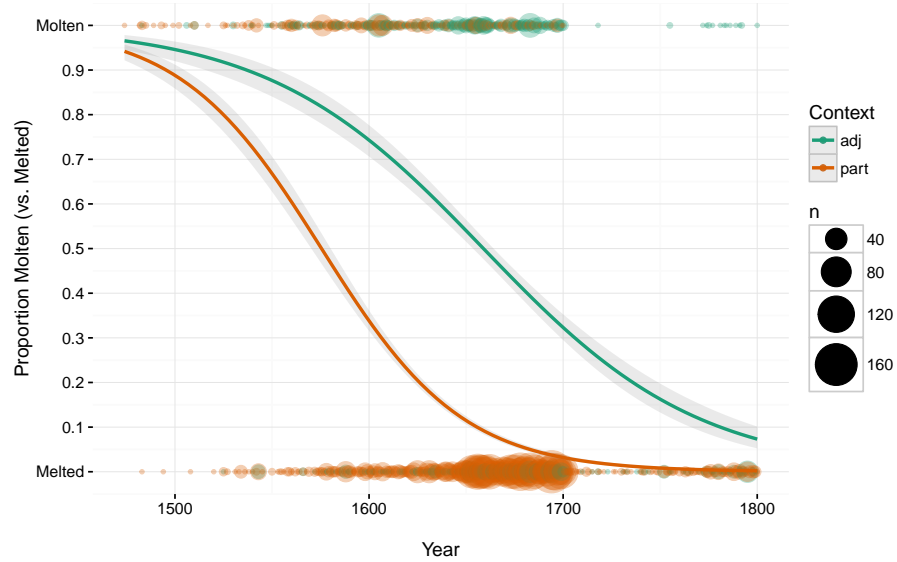


Figure 3: Proportion of *molten* uses by year of text, for both syntactic contexts over time. N = 7946 tokens.

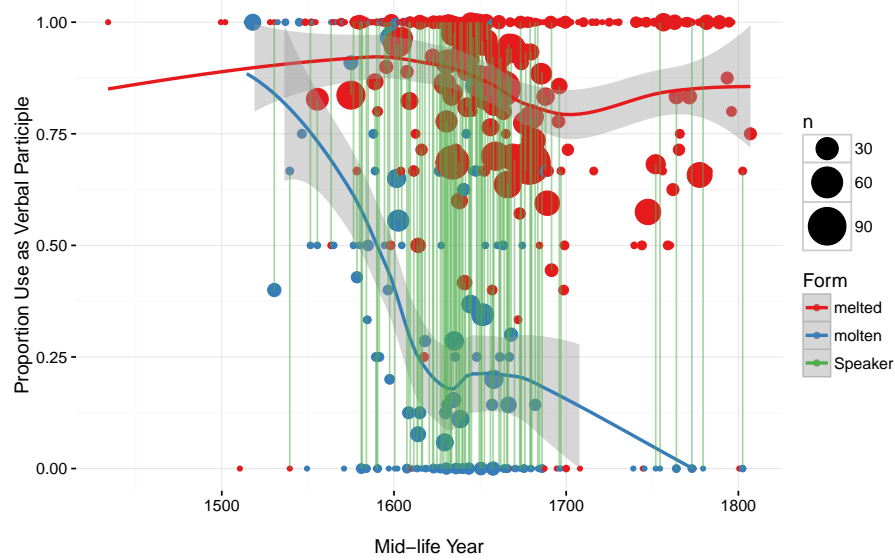


Figure 4: Syntactic context by mid-life year of author, for melted and *molten* forms over time. Green lines connect proportions of participle use with melted and molten for speakers who used both forms. 471 identifiable speakers, N = 3601 tokens.

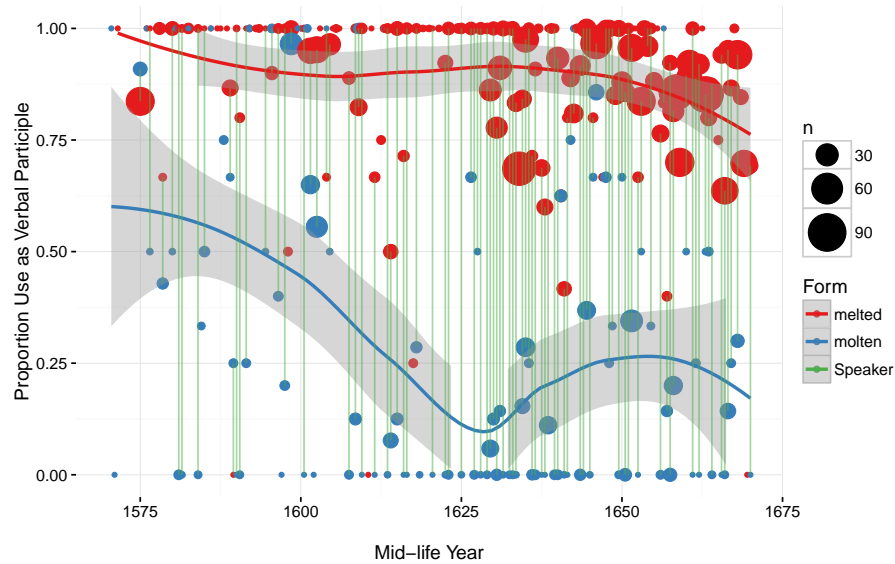


Figure 5: Individual author *melted*/*molten* systems between 1570-1670.