

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

This article presents an empirical case study in the diachronic specialization of morphosyntactic forms for different syntactic contexts, and uses it to develop a theory of variational specialization. This theory links specialization in diachrony to specialization in language acquisition, sociolinguistic coordination in a speech community, and a general understanding of evolutionary dynamics. The case study illustrates these relationships with the specialization of *melted* and *molten* in Early Modern English, and tests the hypothesis that even diachronic specialization in a lexical domain will not take the same trajectory for different speakers, but that the community will nevertheless coordinate on a direction of specialization given multiple generations. In doing so, it answers a question referred 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? The study ultimately shows 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's (2000) variational grammar learning model can be extended to the problem of specialization, and that children plausibly do not play an active role in specializing linguistic forms: they only need to identify potential contexts that the forms could specialize for, and the learning analog of natural selection does the rest.

Keywords: language change; specialization; child language acquisition; sociolinguistics

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 A General Principle of Contrast

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 evolutionary dynamics on two variants competing for finite resources, e.g. mapping to a particular place in a child's memory. (These dynamics may be those of natural selection, where an advantage for a variant exists, or those of neutral processes, such as random death, as in Moran 1958; see Kauhanen 2016 for a discussion of neutral processes in language change.) In the long run, either a variant will eventually be removed (e.g. outcompeted by the other), or the competition will be removed. 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)¹ 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.

In short, Bailey et al. (2012) showed that the *whether* and *if* structures² have slowly been specializing 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 is very gradually being removed by specialization. (Note that in Icelandic, the same initial competition was removed by extinction of the *if* variant.)

One reviewer questions whether the difference between the *whether* and *if* forms is syntactic at all. Some syntactic difference is suggested by the possibility of the *whether or not* construction with the *whether* form but not the *if* form:

- (6) a. Mary wondered whether Sue was bringing coffee or not.
b. Mary wondered whether or not Sue was bringing coffee.
- (7) a. Mary wondered if Sue was bringing coffee or not.
b. * Mary wondered if or not Sue was bringing coffee.

1. Note that these are conference papers, so while this data has already been published in one sense, it appear in print here for the first time.

2. As to the question of how similar or different the *whether* and *if* structures are syntactically, see Larson (1985), though I do not believe there is a consensus on this, and the main thrust of the paper remains even if this case is not truly syntactic.

Perhaps this is simply the inability of a null operator to pied-pipe (if this construction is truly pied-piping; I do not know of a consensus on its analysis), in which case the difference between the constructions could be considered morphological, i.e. purely a Spell-Out issue. I will remain agnostic on this point for the time being. However, if there is no structural difference between the *whether* and *if* forms, then Yang’s Paradox (see the next subsection) actually applies even more strongly than if the variation above were syntactic: why should the specialization of morphological-forms proceed so slowly, given the experimental evidence? (See also Wallenberg 2016 for a case of even slower syntactic specialization, though in that case it is specialization along a continuous dimension.)

Building on Fruehwald and Wallenberg (2013), I 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 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 handled by (the language analog of) general evolutionary dynamics. (This may be an analog of natural selection, where an advantage for a variant exists in a particular context, or it may be neutral processes, such as drift in finite populations of utterances.)

2.2 Yang’s Paradox

Wallenberg (2016) and Fruehwald and Wallenberg (in prep) suggest that all specialization, across domains of the grammar, derives from the same Principle of Contrast strategy (and its consequences for variant competition). 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. For this reason, I've chosen to investigate a diachronic specialization trajectory in a plausibly word-level domain, the variation between *melted* and *molten* in Early Modern English. This specialization turns out to be much slower than the experimental evidence would suggest, which means that Yang's Paradox really does need to be resolved, even within the lexical domain alone. (Note that it is also possible that there is truly no lexical-syntax distinction, as per the Borer-Chomsky conjecture (Baker, 2008)...in which case Yang's Paradox also still needs resolved.)

3 Methods

3.1 Choice of Phenomenon

The study focuses on the morphological doublet *melted/molten*. These forms arose during the Old English period, initially with different etymologies (forms *gemolten*, *gemielted* (West Saxon) and *gemælted* (Anglian), with the first adnominal use of (*ge-*)*molten* dated to 1300 (citation MELT in *Oxford English Dictionary Online*)). The forms then specialized over time such that *molten* became a pre-nominal adjective, 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 comes 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 (this one probably arose through analogy and merger, though a full etymological study is beyond the scope of this paper), and so can enter the speech community without any initial difference in context or meaning. It is not certain that the members of this doublet entered into variation with no difference in meaning at all, but it is clear from the data below that they were not differentiated along the participle-adjective dimension at the beginning of the historical period I consider.

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, 1450 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. The distance between those probabilities increases over the period under consideration. Figure 2 shows the data organized so that the probability of the forms occurring in the participle context (out of participle+adjective contexts) is on the y-axis, Year on the x-axis, and red and blue colors identify the *melted* and *molten* forms, respectively. (The

dots represent occurrences of each form, with dot size indicating the N for any time point, and since each form can only be either a participle or an adjective, the dots appear at 0 and 1 in the plot.)

While this is a somewhat unusual way to display the data (cf. Figure 3 below), this display makes it easier to see how forms have diverged regarding their use for these two functions. The fall of both lines over the time period shows that there’s an overall decrease in the participle context, relative to the adjective one (for both forms taken together). I do not know the reason for this, but I suspect it’s an effect of increased genre diversity in the corpus over time, with more scientific, technological, and medical texts entering over time, which could increase the frequency of items like “melted/molten steel”.³ However, it is the differential behavior of the forms which primarily concerns us. *melted* and *molten* begin the period under investigation with very similar distributions, both being used primarily in the participle contexts, and with statistically indistinguishable distributions at the very beginning of the period (note the initial overlap in error regions in Figure 2. Over time, the frequency of participle use decreases for both forms, but the decrease is not by the same amount 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 in what readers are likely to find a more usual mode of display: 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 carried out a mixed-effects logistic regression with random intercepts for individual Text and Author, and main effects for Year⁴ 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 provided a significantly better fit ($p = 0.0003$). 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 data whose birth and death dates were known, with an N of 3601 tokens for these speakers. Figure 4 shows the data aggregated by author, plotted over time by their mid-life years, with the proportion of participle use for each form on the y-axis just as in Figure 2.⁵ Thus, Figure 4 is the subset of the data of Figure 2 for which we have author identities and dates of birth and death, displayed in the same way as in Figure 2. Green vertical lines appear on the graph for any speaker who used both *molten* and *melted* forms in this dataset, connecting the proportions of participle

3. An anonymous reviewer suggests that the industrial revolution could have had the effect of increasing discourse on such subjects, which seems plausible to me.

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

5. Mid-life year is not important for the analysis in any way, and is not necessarily analytically preferable to date of birth. Mid-life years are used simply to make it easier to display this data by individual speakers, given certain constraints of `ggplot2`, without accidentally aggregating any speakers with each other, and while connecting the dots with lines.

use for the two forms for a given speaker. If there is no green vertical line, then a given speaker did not produce both forms in his/her writing, and we cannot see an “inventory” in any meaningful sense. The green vertical lines simply make it easier to see an individual speaker’s usage of both forms, where a speaker did happen to use both forms (and every other aspect of the data is plotted as in Figure 2). For ease of interpretation, Figure 5 plots just the subset of data from Figure 4 that occurs during the century of most vigorous change, between 1570-1670.⁶ This invites the reader to look more closely at individual speaker inventories during the main period of community change, inventories which I discuss below in section 5.

To get a sense of how often individual speakers had fully specialized systems, I again subsetting the data, looking 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 criterion 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 **one or other 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. This means that speakers like Robert Almond, illustrated in the set of examples below, were not uncommon. (See further discussion below.)

- (8) a. Method of breeding Horses...Molten grease and fatning balls
b. ...which may bring away any melted grease
- (9) a. ...the grease is molten into them
b. ...considering that if grease should be melted
- (10) a. ...adding thereto some Honey; which being molten , give it the Horse
b. ...English Honey; and when these are melted, and well stirred together

(Robert Almond, *The English horsman and complete farrier...*, date: 1673)

It is also interesting to note whether these speakers, though stochastic in their behaviour, were following the community 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

6. This plot is intended as simply a zoomed version of Figure 4, though technically `ggplot2` refits the loess smoothing curves to the data when a subset is plotted. There is only minimal difference in the curves in Figures 3 and 4, however, and the curves are in any case merely a guide to the eye and not an analytical device in these plots.

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 (vertical) 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 vertical 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* remains a productive passive participle throughout the whole period, and can also be used adnominally as an adjectival passive, as it can in the modern language. *molten*, on the other hand, in the adjectival context, may have been reanalyzed as a simple adjective at some point (as it probably is in modern English); as a reviewer points out, this could be an instance of Kuryłowicz's Fourth Law (Kuryłowicz, 1945).⁷ It may be that as *molten* lost in the participle context, and became increasingly analyzed as an adjective in the adjectival context, it loses a competitive edge in the adjectival context. *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 as well, and creates a selectional advantage that allows it to outcompete *molten* in a way that a simple adjective competitor could not. *melted* is also the regular, weak verbal form, which means it may be innovated spontaneously in acquisition in all contexts at a certain rate, conferring an advantage on that form. But regardless of whether there is an identifiable advantage, 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

7. A further study of the adjectival context alone would be necessary to substantiate this suggestion, however. I leave that for future work, as this is a side point with regard to the theory of specialization presented here.

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 could allow for *melted* to survive in one context one and *molten* to survive only in the other, if the evolutionary dynamics affecting the competition in each context allow for that. For instance, if the two contexts are different such that one variant has a selectional advantage in context C_1 only, and the other has an advantage in context C_2 only, then e.g. *melted* would win in C_1 and *molten* would win in C_2 by the linguistic analogy of natural selection. Alternatively, if neither variant has an advantage at all, neutral processes, such as the linguistic analog of random death (i.e. not hearing a variant by chance at some moment), could allow one variant to win in C_1 and the other to win in C_2 .

However, in the data above, it seems that the selectional pressures in the two contexts are such that *melted* outcompetes *molten* in both contexts, possibly because of the potential advantages I mentioned above. The apparent advantage that *melted* has, however, seems to be different in the two contexts, even though the form has an advantage in both. 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 is an entirely expected possible outcome if the child's acquisition strategy really only involves choosing a dimension of specialization and separating the contexts, and any kind of differential pressures (or chance) are allowed to assert themselves in the contexts from there on out. This result is decidedly not compatible, however, 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 dynamics take their course in the two contexts without the acquirer doing anything actively aside from tracking variant frequencies (as she or he must do anyway).

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, Henri Kauhanen, Rachael Bailes, 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. I would also like to thank two anonymous reviewers. All remaining errors are of course my own.

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.
- Baker, Mark. 2008. The macroparameter in a microparametric world. In *The Limits of Syntactic Variation*, ed. Theresa Biberauer, 351–374. John Benjamins.
- 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 A.H., 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.

- Kauhanen, Henri. 2016. Neutral change. *Journal of Linguistics* 1–32. Draft: <http://henr.in/doc/neutral-change.pdf>.
- Kroch, Anthony, Beatrice Santorini, and Lauren Delfs. 2004. Penn-Helsinki Parsed Corpus of Early Modern English, release 3. Size 1.8 Million Words.
- Kroch, Anthony S. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1:199–244.
- Kroch, Anthony S., Beatrice Santorini, and Ariel Diertani. 2010. Penn Parsed Corpus of Modern British English. Size ~ 950000 words.
- Kroch, Anthony S., and Ann Taylor. 2000. Penn-Helsinki Parsed Corpus of Middle English. CD-ROM. Second Edition. Size: 1.3 million words.
- Kuryłowicz, Jerzy. 1945. La nature des procès dits «analogiques». *Acta linguistica* 5:15–37.
- 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.
- Moran, Patrick Alfred Pierce. 1958. Random processes in genetics. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 54, 60–71. Cambridge Univ Press.
- Taylor, Ann, Anthony Warner, Susan Pintzuk, and Frank Beths. 2003. The York-Toronto-Helsinki Parsed Corpus of Old English Prose. <http://www-users.york.ac.uk/~lang22/YcoeHome1.htm>.
- 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.

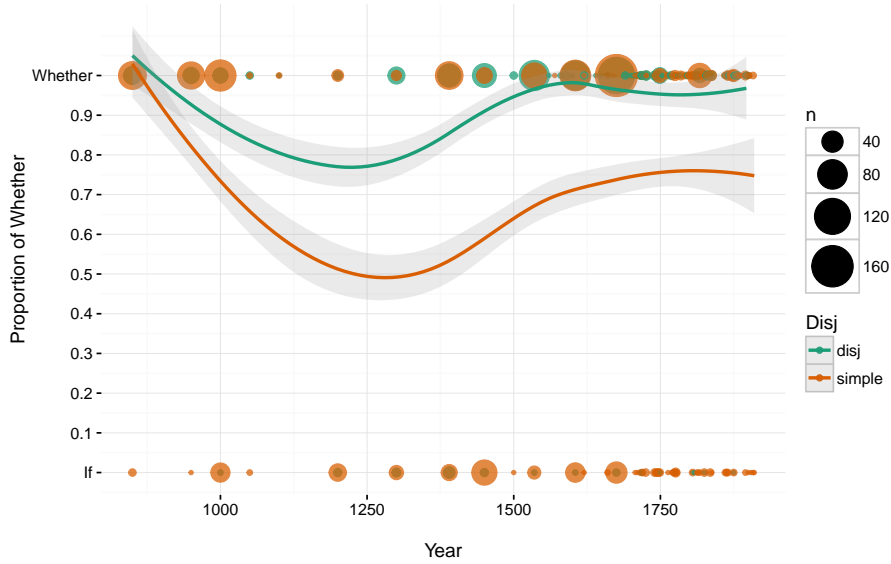


Figure 1: *whether/if* specialization, data from Bailey et al. (2012) using Taylor et al. (2003); Kroch and Taylor (2000); Kroch et al. (2004, 2010), N = 1929 clauses.

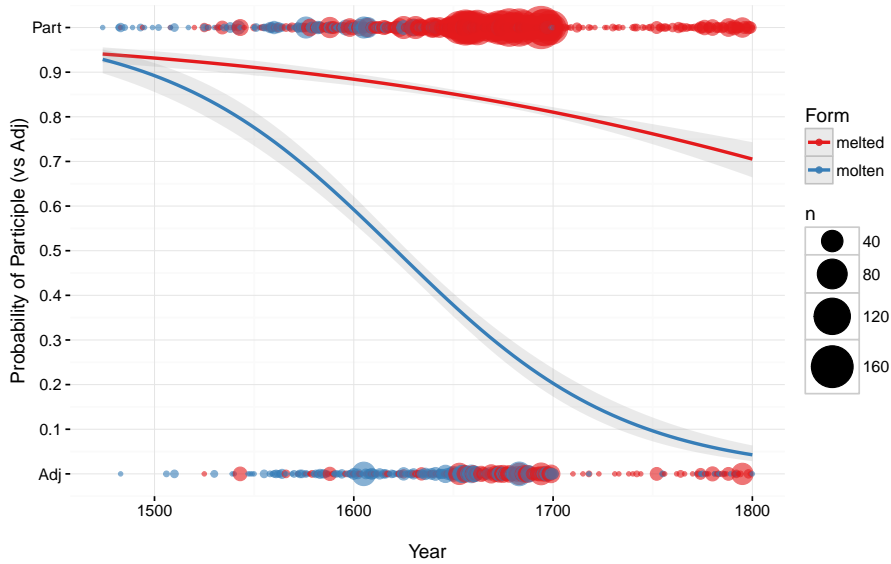


Figure 2: Syntactic context by year of text, for melted and *molten* forms over time. N = 7946 tokens.

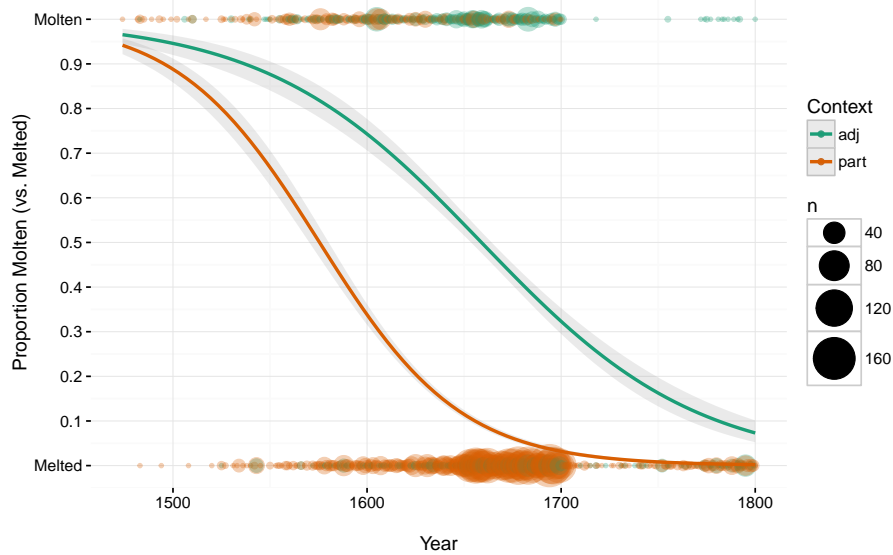


Figure 3: Proportion of *molten* uses by year of text, for both syntactic contexts over time. N = 7946 tokens.

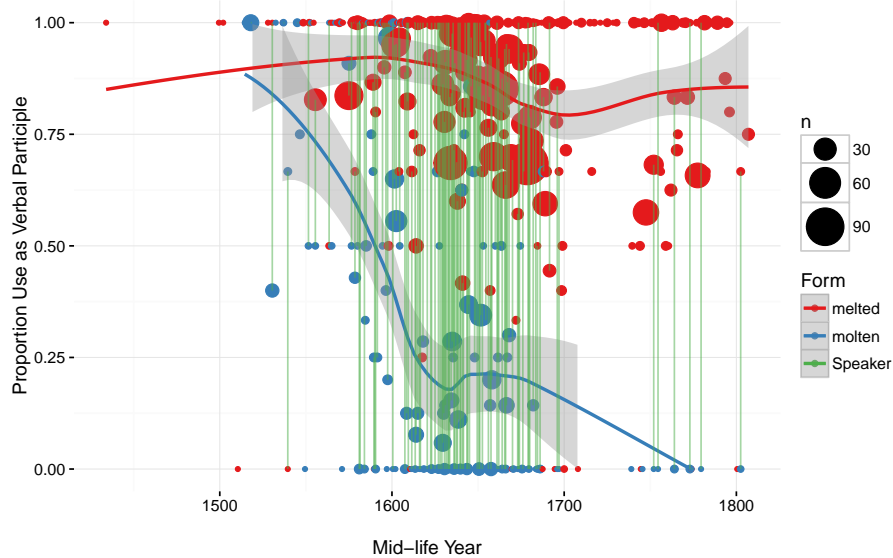


Figure 4: Syntactic context by mid-life year of author, for melted and *molten* forms over time. Green (vertical) lines connect proportions of participle use with melted and molten for speakers who used both forms. 471 identifiable speakers, N = 3601 tokens.

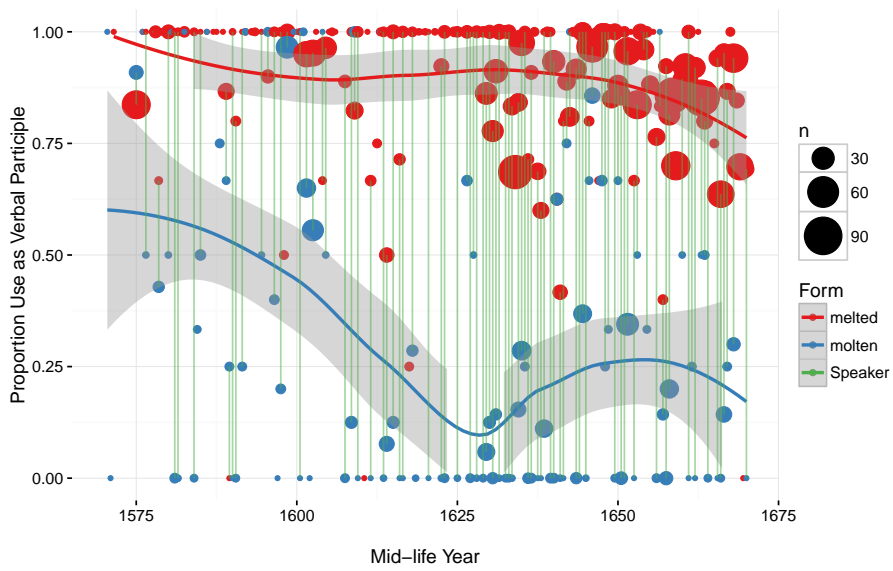


Figure 5: Individual author *melted*/*molten* systems between 1570-1670.