

owards a model of variational specialization in acquisition

Joel C. Wallenberg

`joel.wallenberg@ncl.ac.uk`



**Newcastle
University**

**Centre for
Behaviour & Evolution**

7 október 2016
Háskóli Íslands

Introduction

Overall Hypothesis: all variation, including grammatical “optionality”, is the same as **competing grammars**, with the consequences (? , *In Prep*):

- We expect variation (apparent optionality) between multiple grammatical forms is diachronically unstable.
- The diachronic survival of forms depends on their ability to **specialize**, so that they no longer compete in use (**replacement**).
- The outcome of specialization depends on how/whether learners associate variants with some domain of specialization, and its mathematical character.
 - True(er) optionality = stable(er) variation: partial **specialization** of categorical variants along a continuous dimension.

Outline

Diachronic Blocking Effect

“Blocking Effect” (?)

- General pressure against two forms existing for one function (“doublet”), forcing them to resolve in **replacement** or **specialization** (?).

{*lough*, *laughed*} (laugh-PST; ME, ?)

{*melted*, *molten*} (PDE participle, adj pass)

{*jimmies*, *sprinkles*} (candy topping, Philadelphia)

“Principle of Contrast”

- A strategy that children use in acquiring language: assume that two forms have two meanings (or contexts)(??, *inter alia*).
- Children hypothesize that novel words also refer to novel objects.

The Principle of Contrast (PrinCon)

- Demonstrated in experiments such as ??; see also nuanced review in ?.
 1. 20 children
 2. 6 pairs of one familiar item (banana, cow, cup, plate, saw, spoon) and one unfamiliar item (cherry pitter, odd shaped wicker container, lemon wedgepress, radish rosette maker, studfinder, tongs).
 3. **Control:** “Show me one”
 4. **Test:** “Show me the X” (X = nonsense syllable)
- Control children pick the unfamiliar object at chance levels, but test children choose unfamiliar objects significantly higher than chance.

...and observational results

(1) Mo (at the fish-counter): That's a trout.

D (aged 2:5,1): That's a fish. That not a trout.

Mo: Well, a trout's a kind of fish.

D (pause, then pointing at a row of crabs): crabs are a kind of fish.

?, 97

Blocking = Contrast + Evolutionary Dynamics

- A doublet is two variants competing for finite resources (“competing grammars”), as in e.g. biological evolution.
 - Instead of competing for something like food, they are competing for use (time in the mouths/brains of speakers).
- Either one variant has a selectional advantage, and so **replaces** the other.
 - cf. ?, and subs., ?
- Or neither variant has an advantage (or much of one), in which case random walk and drift (which can also lead to **replacement**).
- But in language learning, the PrinCon means learners can pull apart the contexts of the variants so as to remove the competition.

Example: Embedded Polar Questions

In all stages of English and in historical Icelandic, a disjunction favors *whether* (Bailey, Wallenberg, & van der Wurff 2012).

English

Disjunction:

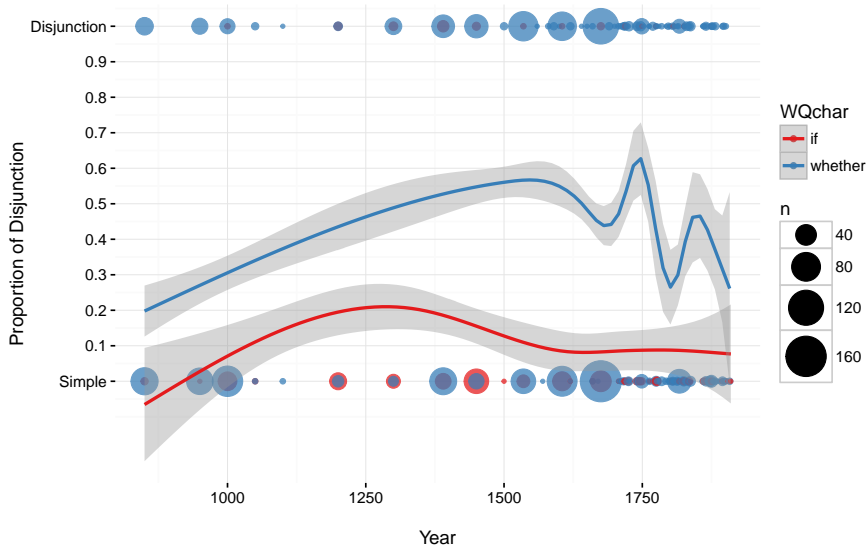
- (2) I wonder {**whether**,if} John or Bill is bringing coffee.
- (3) I wonder {**whether**,if} John is bringing tea or coffee.

Simple:

- (4) I wonder {whether, **if**} Bill is bringing coffee.

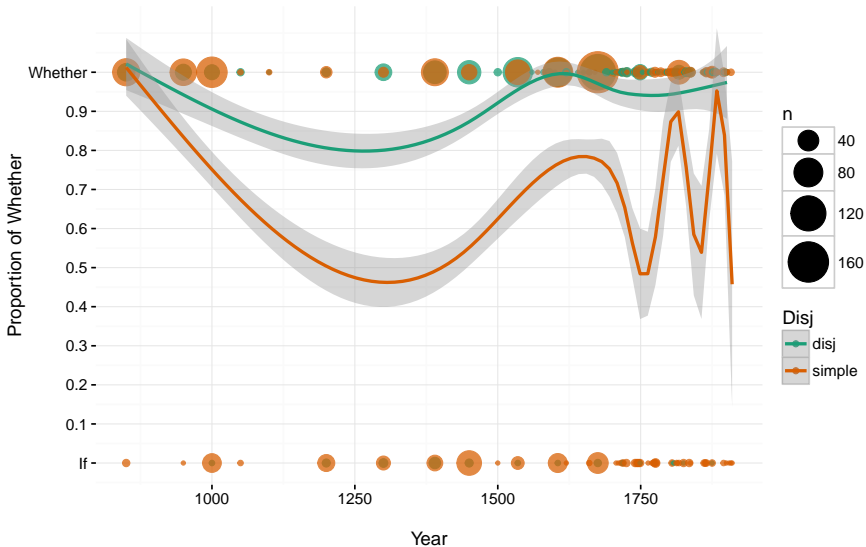
Slow Specialization of *whether/if* (N = 1929 clauses)

Parsed Corpora: YCOE, PPCME2, PPCEME, PPCMBE



whether/if replacement slowed/arrested

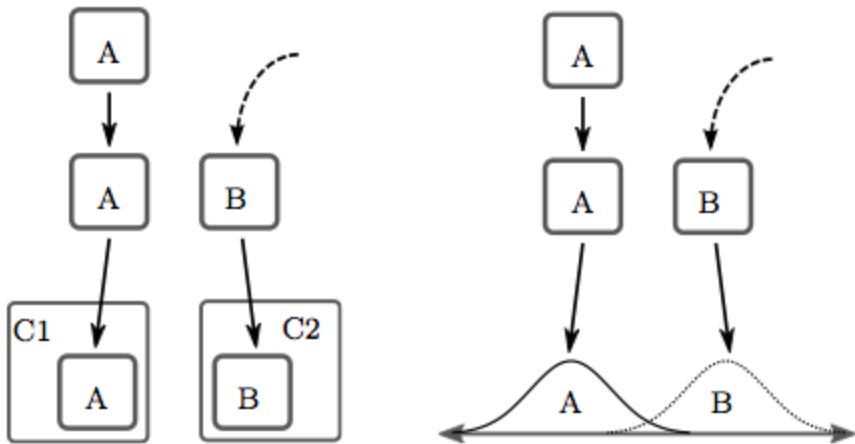
(N = 1929 clauses)



Consequence: Blocking and Contrast

- If specialization occurs, its effect depends on the **domain of specialization**. A change can be:
 1. A replacement change in progress (outright competition going to completion).
 2. A specialization change in progress (specialization for different functions going to completion).
 3. **“Stable” variation:** variants have **imperfectly specialized** along a continuous (or ordinal) dimension, e.g. style, prosodic weight.
- If categorical variants specialize along a categorical dimension, complete specialization should eventually result.
- If categorical variants specialize along a continuous or ordinal dimension, then complete specialization can **never** result, but replacement can be slowed by **imperfect specialization**.

Specialization along categorical and continuous dimensions



(figure from Fruehwald & Wallenberg *in prep*)

A Very Slow Change

- One consequence of our overall hypothesis is that some things that didn't look like change turn out to be.
- Relative clause extraposition is a change in progress, but a very slow one (???).
 - It has been mischaracterized as syntactic optionality.
- The study used the same coding query (with minor adaptation) on 7 parsed diachronic corpora (4 language histories).
- Both the time-depth and cross-linguistic dimensions were necessary in order to discover the change.
- Only because we had both dimensions were we able to observe (and confirm) the slowest syntactic change discovered to date.

Case Study: Relative Clause Extraposition

French

- (5) mais l'heure vient [que je ne parleray plus a
but the time comes that I NEG speak-FUT more to
vous en proverbes]
you in proverbs

“The time approaches when I will no longer speak to
you in parables”

(MCVF, 1523-NEW-TESTAMENT-P, A5V.2491)

English

- (6) All had now been tried [which either threats or promises,
forbearance or fatherly chastisement, could effect].
(PPCMBE, FROUDE-1830,2,2.20; date: 1830)

Hypotheses for the diachronic study

- **Hypothesis 1:** Relative clause position (a binary variable) is specialized along a continuous dimension, weight, and so it should be **nearly** stable, but not entirely stable.
- **Hypothesis 2:** All IE relative clauses derive historically from clause-adjoined relatives (?).
- **Hypothesis/Suggestion 2':** The old clause-adjoined kind are still around, in the form of extraposition, and the change Kiparsky proposed hasn't finished yet (?).

Specialization within individuals

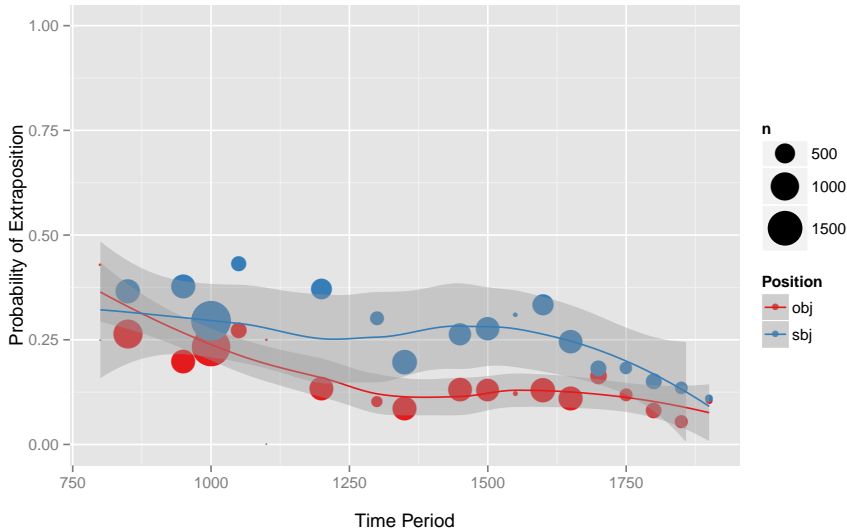
- Relative clause extraposition in the Parsed Corpus of Early English Correspondence (PCEEC; ?).
- Allows us to look at reasonable samples from individual speakers (letter-writers), as well as an historical sample from 1400–1700.
- Coded for prosodic weight of the relative clause, in number of words, from 0–50.

Hypothesis: individual speakers treat weight as a continuous variable, with extraposition specialized imperfectly along it (as suggested by ?).

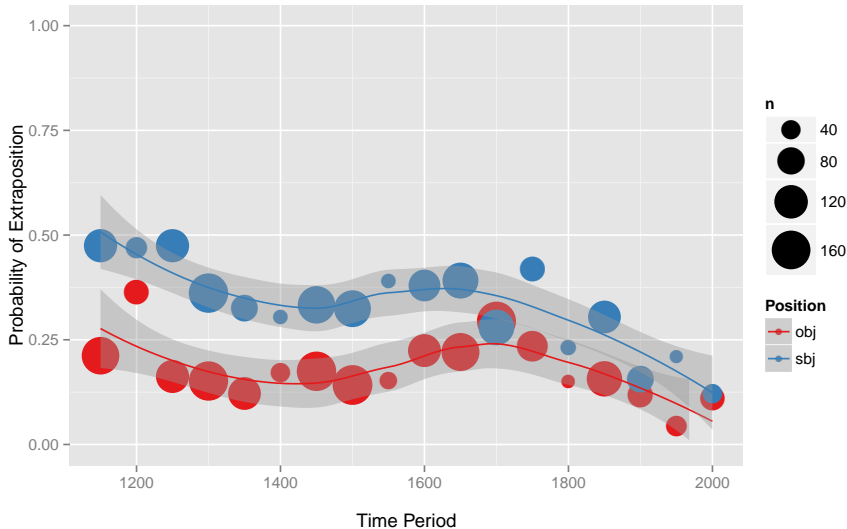
Diachronically, Crosslinguistically

- **English:** YCOE (?), PPCME2 (?), PPCEME (?), PPCMBE (?).
- **Icelandic:** IcePaHC (?).
- **Old/Middle French:** MCVF Corpus (Martineau, Hirschbühler, Kroch, & Charles Morin, 2010).
- **Historical Portuguese:** Tycho Brahe Corpus of Historical Portuguese (?).

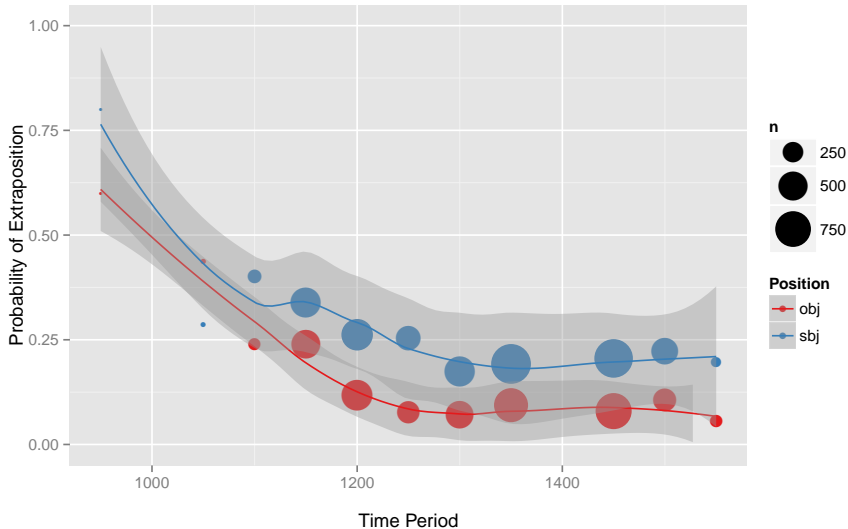
English, over time (N = 18530 clauses)



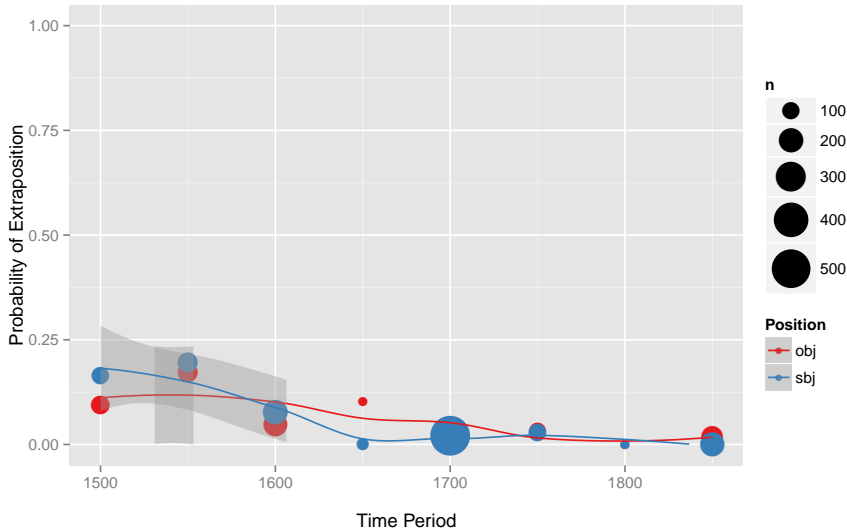
Icelandic, over time ($N = 3486$)



Old/Middle French, over time (N = 8207)



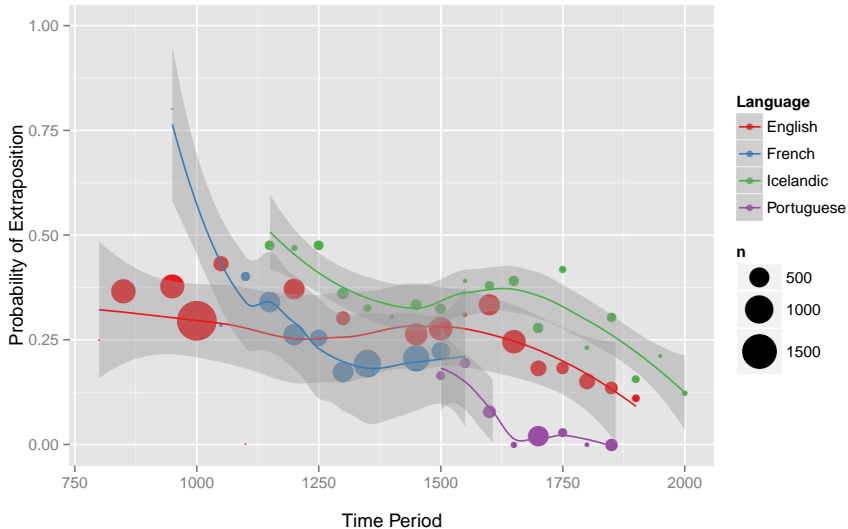
Portuguese, over time (N = 2398)



Statistical characteristics of the change

- The slope of the decline over time is shallow; slopes for Icelandic, English, French, and Portuguese = -0.37, -0.36, -0.32, -1.24 from Subject (based on mixed effects logistic regression controlling for weight and other factors).
- Weight has a significant effect in each language, but the effect doesn't change over time.
- Icelandic, English, and French show very similar same rate of change for EX (a kind of Constant Rate Effect?) (model comparison not possible for computational reasons with the mixed effects models; $p = 0.47$ for Icelandic and English with standard regression).

Four Languages (Subj Ex), over time



The origin of the slow change?

- (7) **yó** **mártyaḥ** **śísīte** **áty** **aktúbhir**,
which mortal sharpen-Mid-Sg overly nights-INSTR,
mā naḥ sá ripúr īsata
not us-GEN that trickster dominate-Subj3Sg
“As for the mortal who makes himself too sharp by
night, may that trickster not gain power over us”
(RV 1.36.16, cited in ?, 156)

The origin of the slow change?

- (8) By God's blessing I calculate that the Spirit of Dishonesty shall not get dominion over me; nor the Spirit of Despondency, nor any other evil spirit; **in which case all will and must be well.**
(Letter by Thomas Carlyle, date: 1835; ID CARLYLE-1835,2,266.176 in PPCMBE)
- (9) Nowadays, however, flowers can be arranged in various styles – some flat, some slightly raised, some bunched boldly in certain places and forming the piece de resistance of the whole work – **all of which variations depend upon the artistic perceptions of the operator.**
(*Commercial gardening...*, date: 1913; ID WEATHERS-1913,1,9.217 in PPCMBE)

Summary: Change in Extraposition

- **Why the change?** After actuation, extraposition and *in situ* are competing variants in use, so there can't not be a change, even with partial specialization.
 - Specialization can only be partial along the (continuous) weight dimension.
- The change is slow enough to be not observable without considerable time-depth.
- Perhaps ? identifies a change that goes back to Proto-Germanic, though hard to test.

Charles's Question (or Yang's Paradox?)

Experimental results on word-learning show the Principle of Contrast differentiates words nearly instantaneously. The PrinCon is too fast to produce the slow specialization we see in, e.g. syntax. Is there another pressure?

(Caveat: ? show retention of the new, differentiated mapping is not instantaneous, and reliable until after 24 months of age.)

So, is it really true that word/morpheme specialization happens very quickly? And if not, what about the experimental evidence?

melted/molten specialization

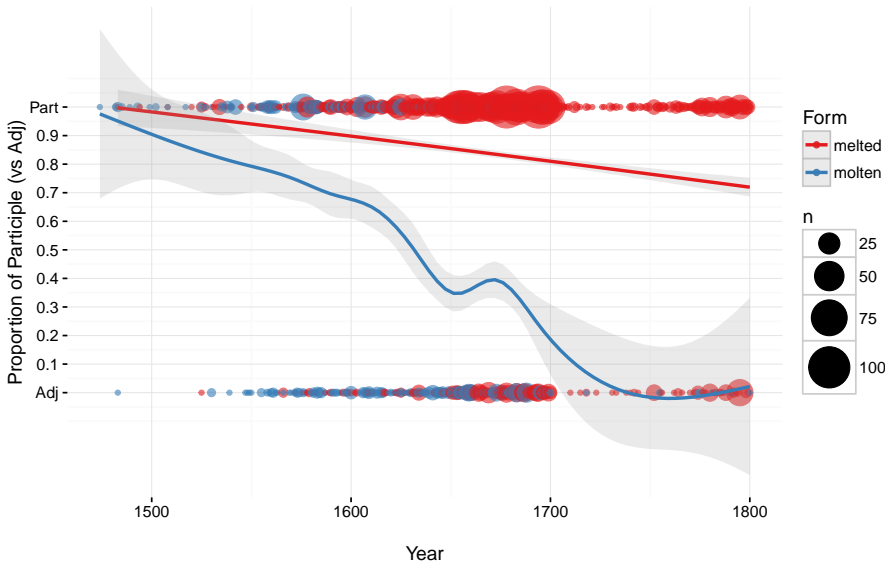
- Variation in participle forms *gemolten*, *gemælted* goes back to Old English, with first adnominal use of *molten* from 1300 (OED).
- *molten* in PDE now seems to be fully specialized (and maybe *melted* as well):

- (10) The gold was {melted / *molten} by the fire.
((passive) participle context)
- (11) The fire has {melted / *molten} the gold.
((past) participle context)
- (12) The {?melted / molten} gold flowed down the hill.
(adjectival or adjectival passive DP-internal context)

melted/molten specialization

- (13) The gold was {melted / *molten} by the fire.
(**participle context**)
- (14) The {?melted / molten} gold flowed down the hill.
(**adjectival context**)
- Question: how quickly did this morphological/lexical doublet specialize, in real time?
 - If you'd like to know why I didn't choose something more solidly lexical, just ask me...
 - Question: how long did intraspeaker variation persist, in both contexts?
 - Using the public section of the Penn-York Computer-annotated Corpus of a Large amount of English based on the TCP (PYCCLE-TCP; ?), roughly 600 million words.

melted/molten specialization N = 4881 tokens



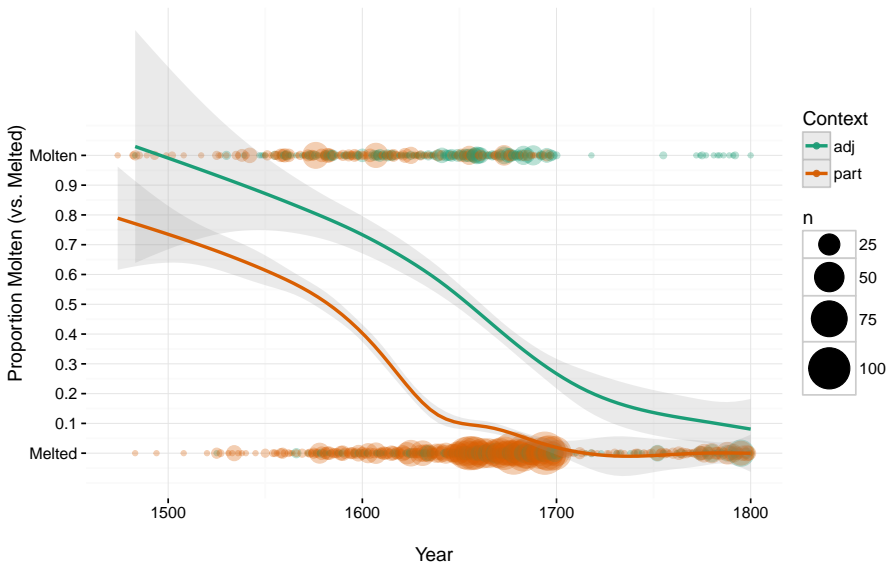
Model Comparison: specialization by context

Model 1: $\text{Form} \sim (1 \mid \text{file}) + (1 \mid \text{author}) + \text{zDate} + \text{Context}$

Model 2: $\text{Form} \sim (1 \mid \text{file}) + (1 \mid \text{author}) + \text{zDate} * \text{Context}$

model	AIC	BIC	p-value (Chisq)
Constant Rate	3039.1	3071.6	—
with Date*Context	3032.3	3071.3	0.003

Simultaneous Replacement? N = 4881 tokens

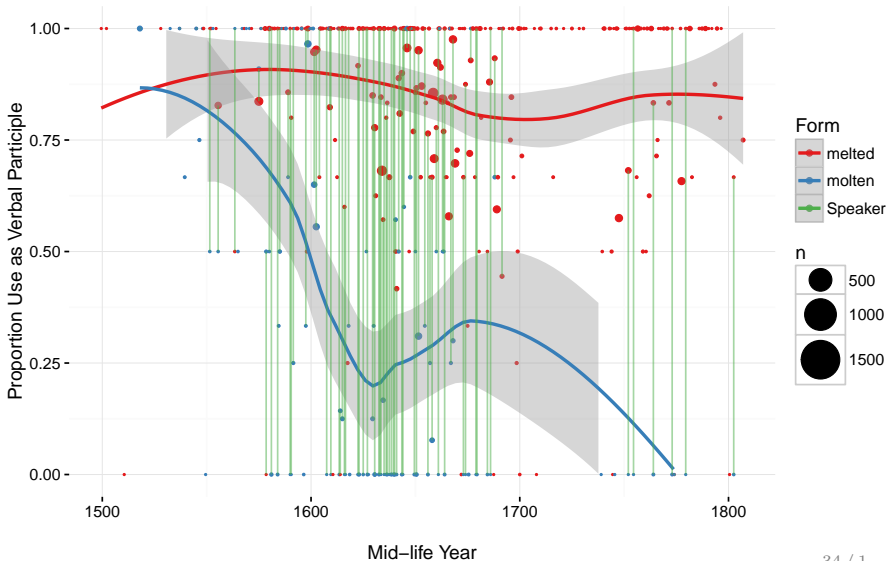


Simultaneous replacement or extreme specialization?

- (15) (***molten* implies heat in PDE:**)
Is silly putty molten rubber?
- (16) (***molten* implies liquidy/sludgy state in PDE:**)
melted spatula vs. molten spatula
- (17) (***both:***)
melted cheese vs. molten cheese
(J. Fruehwald, p.c., for examples above)
- (18) (***molten* implies recognizable substance in PDE:**)
...that the increase and augmentation of Nilus comes
of the snowe waters molten and thawed in those regions.
(attr Barnabe Riche, *The famous hystory of
Herodotus...*, date: 1584)

392 identifiable speakers, $N = 2514$ tokens

(Note: the differing lengths of green lines, and 1575, 1580, 1601.)



Intraspeaker Variation

- (19) a. Method of breeding Horses...Molten grease and fatning balls
- b. ...which may bring away any melted grease
- (20) a. ...the grease is molten into them
- b. ...considering that if grease should be melted
- (21) 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)

Reconciling Experimental Evidence

- Perhaps the first generation to hear the innovation, Generation 1, does try to specialize completely, if possible.
- Generation 1 speakers will not necessarily converge on the same dimension of specialization (and indeed, may mix categorical and continuous dimensions as well).
- Generation 2 cannot help but hear true synonyms, given the overlap of use in the community.
- Subsequent generations may converge on one dimension of specialization (or a few, again potentially mixing categorical and continuous), but there will be intra- and inter-speaker variation all the way.

Specialization and Yang's Variational Learning

1. Identify a domain of specialization:
 - **Actively**, by the child innovating *de novo*?
 - **Passively**, through random sampling of finite populations of utterances?
2. Allow the variants different (quantitative) representations for different contexts, along the domain of specialization:
 - a. For categorical variants along categorical dimensions, decouple tracked frequencies of variants for each context, C_1, \dots, C_n , in the dimension of specialization.
 - b. For categorical variants along continuous dimensions, decouple tracked mean values (or targets) of variants for the dimension of specialization.
3. Specialization goes to completion as the learner has variants behave differently in different contexts.

3a. Specialization completes in the categorical-categorical case

- Suppose Variant A is losing to Variant B due to global selective pressure, but they begin to specialize for C_1 and C_2 .
- Specialization completes in a categorical dimension:
 - **Actively**, by augmenting the represented frequency of Variant A in C_1 whenever Variant B is augmented in C_2 ?
 - **Passively**, by allowing whatever evolutionary dynamics hold in the different contexts play out, whether the outcome is different or not?

3b. Specialization completes in the categorical-continuous case

1. Suppose Variant A is losing to Variant B due to global selective pressure, but they begin to specialize along a continuous dimension C.
2. Learner allows their mean/target values for C to become distinct: μ_{C_A} , μ_{C_B}
3. Specialization completes in a continuous dimension:
 - **Actively**, by moving μ_{C_A} , μ_{C_B} away from each other?
 - **Passively**, by allowing μ_{C_A} , μ_{C_B} the possibility of moving away from each other?

Specialization in Acquisition: active or passive?

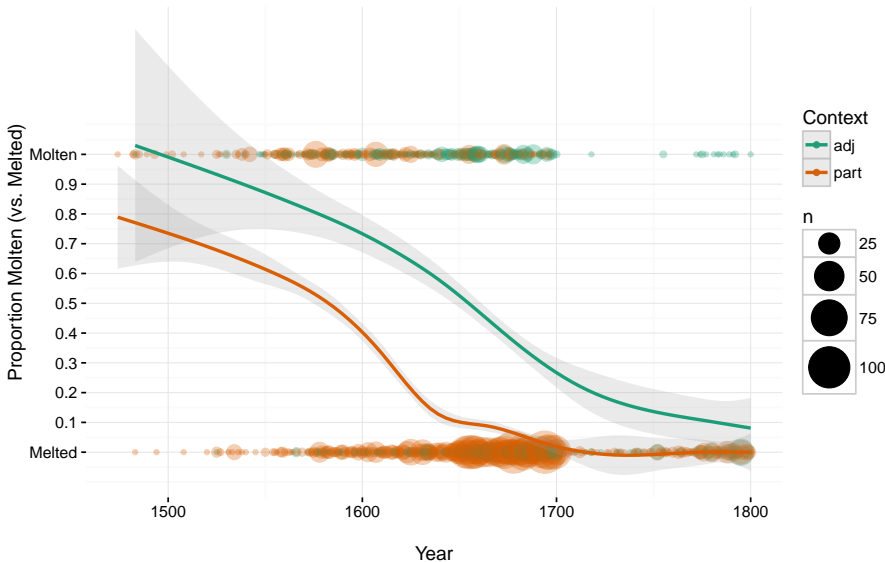
Prediction for a Strong Active Hypothesis: once specialization begins to take place, it should be relentless, and **symmetrical**, and both variants should always survive (in the cat-cat case).

- The frequency of Variant A in C_1 is always being augmented in reaction to, and in lockstep with, the loss of Variant A in C_2 due to independent selective pressures. Both will have to survive, and:

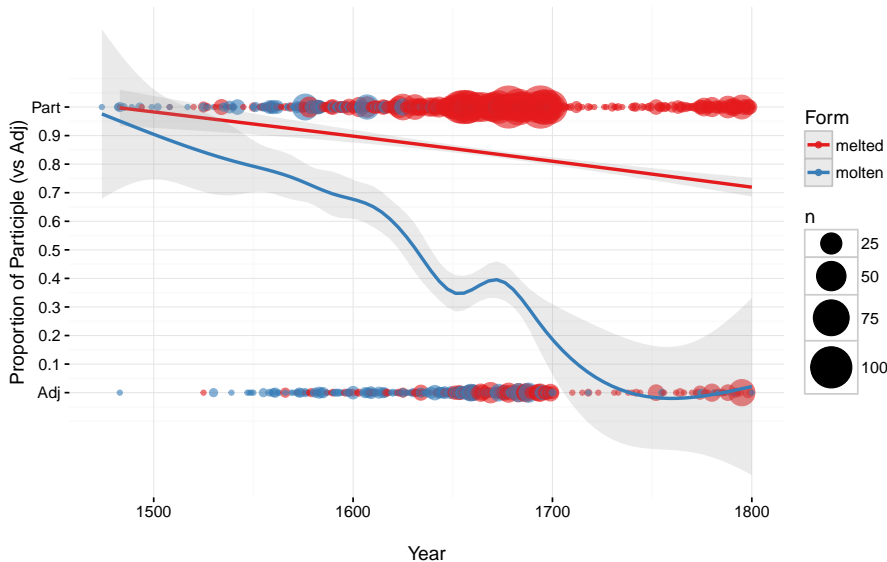
Corollary 1: the frequency of Variant A in C_1 and Variant A in C_2 will need to move away from each other.

Corollary 2: the probability of C_1 being expressed by Variant A will rise as the probability of C_1 being expressed by Variant B declines.

melted/molten: consider Corollary 1



melted/molten: consider Corollary 2



Specialization in Acquisition: active or passive?

- **A Weak Active Hypothesis:** once specialization begins to take place, it is relentless, but not necessarily symmetrical: if Variant A is losing globally, and C_1 and C_2 are decoupled, the amount of augmentation of Variant A in C_1 can be $=$, $>$, or $<$ the global selective pressure against Variant A:
 - **amount of augmentation = selective pressure** \rightarrow variation is stable in C_1 and B wins in C_2 .
(THIS SCENARIO IS FULLY BIZARRE: we've now used the PrinCon to engineer stable variation.)
 - **augmentation < selective pressure** \rightarrow Variant A loses in both C_1 and C_2 but at different rates.
 - **augmentation > 2 x selective pressure** \rightarrow A wins in C_1 at the same rate B wins in C_2 .
 - **selective pressure < augmentation < 2 x selective pressure** \rightarrow A wins in C_1 , but more slowly than B wins in C_2 .

Specialization in Acquisition: active or passive?

- **Weak Active Hypothesis** is hard to disprove, but it may not really be testable at all:
 - Its implementation of PrinCon predicts stable variation, making it indistinguishable in detail from a theory without PrinCon.
 - There's no principled way to estimate the rate/amount of augmentation in acquisition, and it could vary from person to person and case to case, making it unobservable.
- Also incompatible with neutral change (?).

Specialization in Acquisition: active or passive?

Passive Hypothesis: PrinCon is the decoupling of contexts C_1, \dots, C_n by learners, and the rest is due to the selective pressures being different in the different contexts, **once the tracked frequencies are decoupled in the learner's representation of the variation.**

- Distinguished from the Constant Rate Effect by the decoupling of tracked frequencies for, e.g. Variant A in C_1, \dots, C_n .
- Contextual effects within the CRE can be thought of as transformations on a single tracked frequency for Variant A across all contexts.

Conclusions

- **Specialization** can allow competing forms to survive, but only if their functions diverge.
- **Imperfect specialization:** mismatch between categorical variation and continuous dimensions of specialization (or vice-versa) leads to long-term stochastic variation, though not quite stability.
- An extension of ??'s **variational learning model** provides some specific mathematical hypotheses about specialization in acquisition, which we should test.
- **PrinCon** has a natural definition in this model, and can be reconciled with the speed of specialization.
- Two things I didn't discuss because of time:
 1. Details of relative clause syntax.
 2. Phonetic changes that are a continuous linguistic variable specializing along a categorical dimension.

How does it relate to the project

- When varieties come into contact, new linguistic forms come into competition with older forms.
- Replacement and specialization both play out at an individual and speech community level, simultaneously.
 - We need a theory of this.
 - We need to test quantitative hypotheses.

Now for something completely different: gender

- Quick intro

How does it relate to the project?

- If there are subtle biases in social learning, based on continuous gender identity, this affects the spread of variants.
- If we catch new variants early, we may see these biases more clearly.
- Social networks.

Acknowledgements

Thank you first to Josef Fruehwald for working out many of these ideas with me, and to Paul Kiparsky, Anthony Kroch, Betsy Sneller, Charles Yang for much discussion of these issues.

Special thanks to Aaron Ecar for help with PYCCLE and Weihnachtsgurke, among other things. Also Anton Karl Ingason, Ricardo Bermúdez-Otero, Caitlin Light, Laurel Mackenzie, George Walkden, Becky Woods, and anonymous reviewers.

[https://github.com/joelcw/tyneside/tree/master/
extrapolation](https://github.com/joelcw/tyneside/tree/master/extrapolation)

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



Newcastle
University

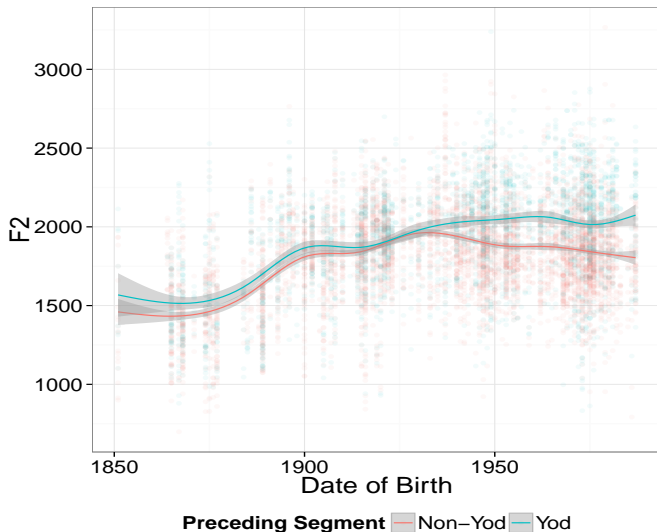


CENTRE FOR
BEHAVIOUR AND
EVOLUTION

References I

Phonological Specialization:

GOOSE-NEW split in New Zealand English (Seyfarth and Sneller 2014)



Spontaneous Phonologization: PRICE-raising in Philadelphia English (Fruehwald 2013)

(308 speakers)

