

# Natural selection, language acquisition, and the specialization of variants over time

Joel C. Wallenberg

`joel.wallenberg@ncl.ac.uk`



SUM-UP

Potsdam Universität

`www.staff.ncl.ac.uk/joel.wallenberg/papers/sumup_`  
`Wallenberg29June2018.pdf`

# Outline

## Specialization and Replacement

The Principle of Contrast and dimensions of specialization

## Morpho-lexical Case Study

How fast does specialization take place?

## Variational Specialization

Extending Yang (2000, 2002)'s model to specialization

## Conclusion

## Diachronic Blocking Effect

### “Blocking Effect” (Aronoff, 1976)

- General pressure against two forms existing for one function (“doublet”), forcing them to resolve in **replacement** or **specialization** (Kroch, 1994).  
{*lough*, *laughed*} (laugh-PST; ME, Taylor 1994)  
{*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) (Clark, 1987, 1990, *inter alia*).
- Children hypothesize that novel words also refer to novel objects.

## The Principle of Contrast (PrinCon)

- Demonstrated in experiments such as Markman and Wachtel (1988); Bion et al. (2013); see also nuanced review in Bion et al. (2013).
  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.  
Clark (1995, 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. Yang (2000, 2002, and subs.), Heycock and Wallenberg (2013)
- Or neither variant has an advantage (or much of one), in which case neutral change, drift (which can also lead to **replacement**; Kauhanen 2016).
- In language learning, the PrinCon means learners can pull apart the contexts of the variants, removing the competition through **specialization**.

## Example: Embedded Polar Questions

In all stages of English (and in historical Icelandic), a disjunction favors *whether* (Bailey et al. 2012; Wallenberg 2016, Wallenberg *in press*)

### 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.

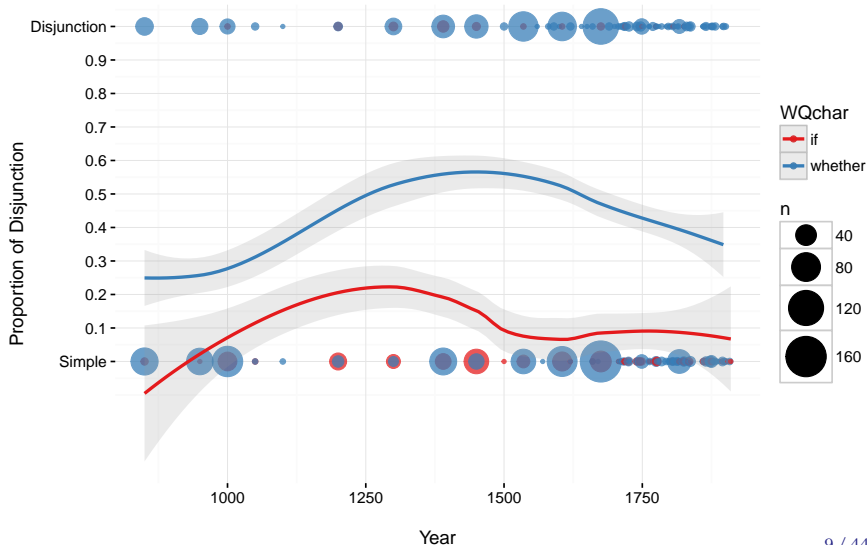
## Syntactic Variants

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



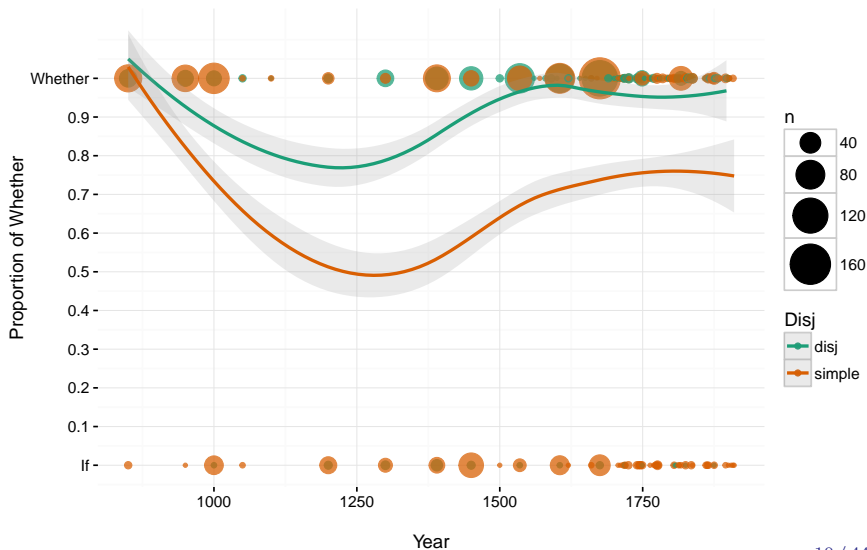
# Slow Specialization of *whether/if* (N = 1929 clauses)

Parsed Corpora: YCOE, PPCME2, PPCEME, PPCMBE



# *whether/if* replacement slowed/arrested

(N = 1929 clauses)



## Consequence: Blocking and Contrast

- 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).
- If categorical variants specialize along a categorical dimension, complete specialization should eventually result.
- Specialization and replacement can happen simultaneously.

## 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: Bion et al. (2013) show retention of the new mapping is not instantaneous, and not 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?

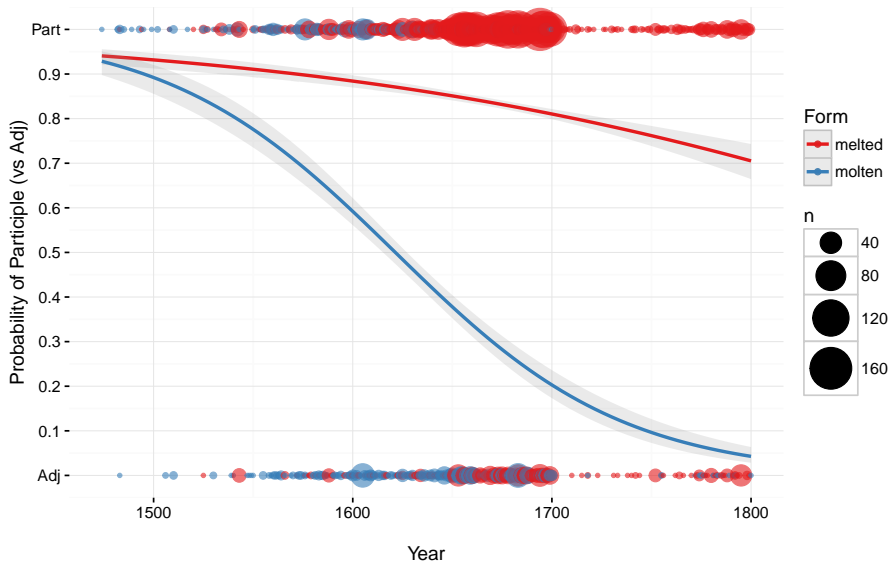
## *melted/molten* specialization

- Variation in participle forms going back to OE: *gemolten*, *gemielted* (West Saxon), *gemælted* (Anglian), with the first adnominal use of *(ge-)molten* in 1300 (MELT, OED).
- *molten* in PDE now seems to be fully specialized (and maybe *melted* as well):
  - (7) The gold was {melted / \*molten} by the fire.  
**((passive) participle context)**
  - (8) The fire has {melted / \*molten} the gold.  
**((past) participle context)**
  - (9) The {?melted / molten} gold flowed down the hill.  
**(adjectival or adjectival passive DP-internal context)**

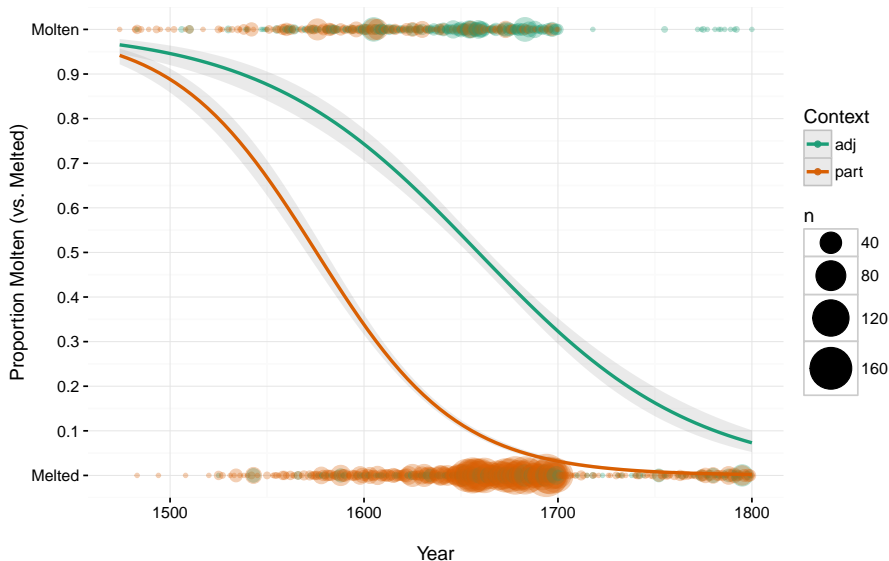
## *melted/molten* specialization

- (10) The gold was {melted / \*molten} by the fire.  
(**participle context**)
- (11) The {?melted / molten} gold flowed down the hill.  
(**adjectival context**)
- Question: how quickly did this morphological/lexical doublet specialize, in real time?
  - Question: how long did intraspeaker variation persist, in both contexts?
  - Using the Penn-York Computer-annotated Corpus of a Large amount of English based on the TCP (PYCCLE-TCP; Ecay 2015), roughly 1 billion words.

# *melted/molten* specialization N = 7946 tokens



# Simultaneous Replacement? $N = 7946$ tokens





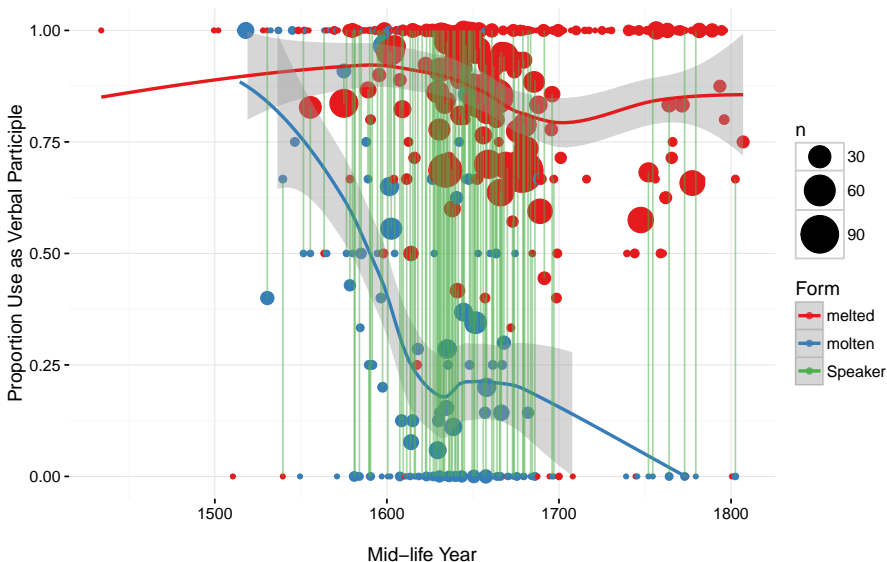
## Model Comparison: specialization by context

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

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

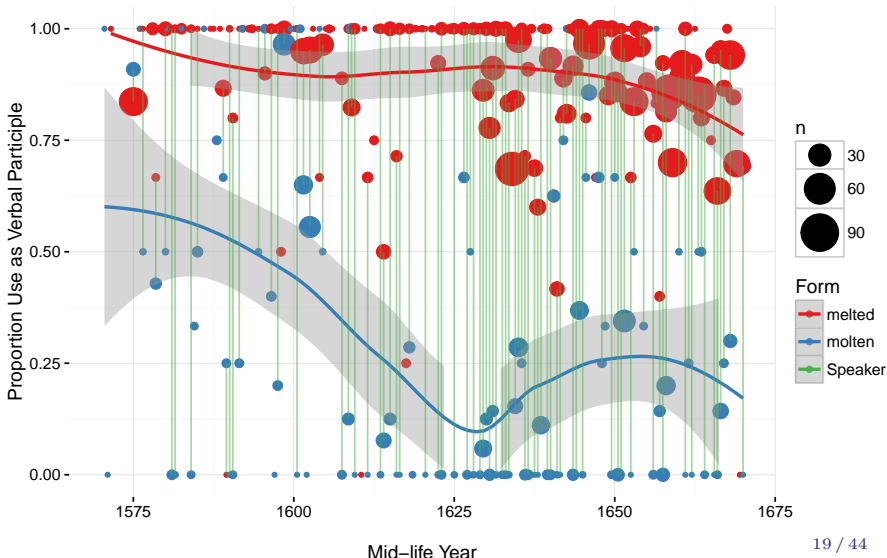
model	AIC	BIC	p-value (Chisq)
Constant Rate	4777.4	4812.3	—
with Date*Context	4766.0	4807.8	0.0003

471 identifiable speakers,  $N = 3601$  tokens



# Individual Speakers, 1570-1670 midlives

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



## Intraspeaker Variation

- (12) a. Method of breeding Horses...Molten grease and fatning balls
- b. ...which may bring away any melted grease
- (13) a. ...the grease is molten into them
- b. ...considering that if grease should be melted
- (14) 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)

## Intraspeaker Variation

Considering the 25 speakers who used > 5 tokens of both  
*molten* and *melted*:

- None were fully specialized; none restricted one form to one context and the other to the other context.  
13 (52%) were **not** categorical in **either** form.  
12 (48%) were categorical in their use of one or other form.  
19 (76%) followed the community trend (i.e. PDE).  
6 (24%) showed the **opposite** pattern, with more *molten* in the participle contexts than *melted*.
- Shows a process of community coordination on the **direction** of specialization is taking place (perhaps in addition to the **dimension**).

## Solving Yang's Paradox

- Perhaps the first generation to hear the innovation, Generation 1, does try to specialize completely.
- Generation 1 speakers will not necessarily converge on the same **dimension** of specialization (i.e. set of contexts), or the direction.
- 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, but there will be intra- and inter-speaker variation along the way.

## Yang (2000, 2002)'s Variational Learner

- Children regularly entertain multiple hypotheses about some linguistic feature, say **Variant A** and **Variant B**.
  - In times of change, both hypotheses may be valid; i.e. more than one variant for a parameter is present in the speech community.
  - E.g. *shirt* is present, *skirt* is introduced; OV is present, VO is introduced; the speech community (and individuals) will use both interchangeably for a time, **and this must be learned**.
- They track probabilities associated with each hypothesis, say  $p$  for A and  $(1 - p)$  for B.
- Learners update the probabilities online as they encounter linguistic data.

# Specialization and Yang's Variational Learning

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



### 3. Specialization completes

- 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  $C_1$  and augmenting Variant B in  $C_2$ ?
  - **Passively**, by allowing whatever evolutionary dynamics hold in the different contexts play out, whether the outcome is different or not?

## Prediction for an Active Hypothesis

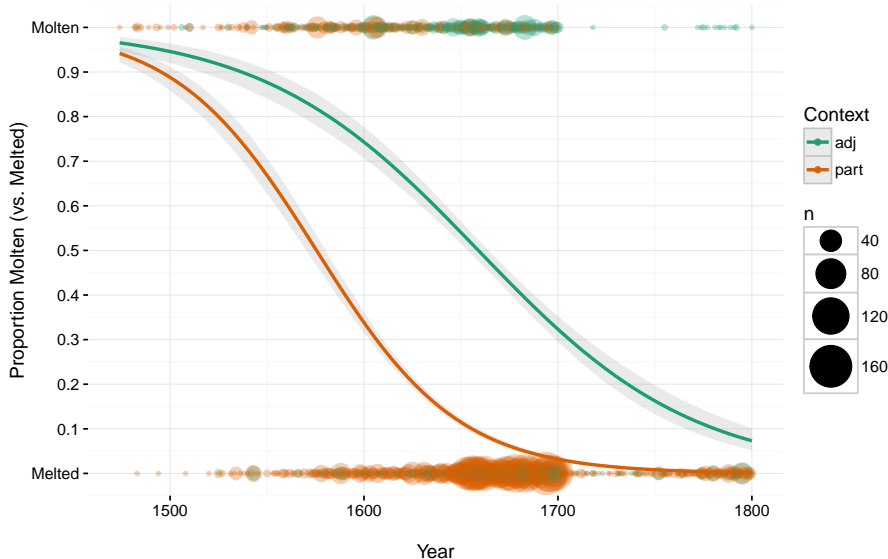
Once specialization begins to take place, it should be relentless, and **symmetrical**, and both variants should always survive.

- The frequency of Variant A in  $C_1$  is always being augmented in lockstep with the punishment of Variant A in  $C_2$ . Both will have to survive, and:

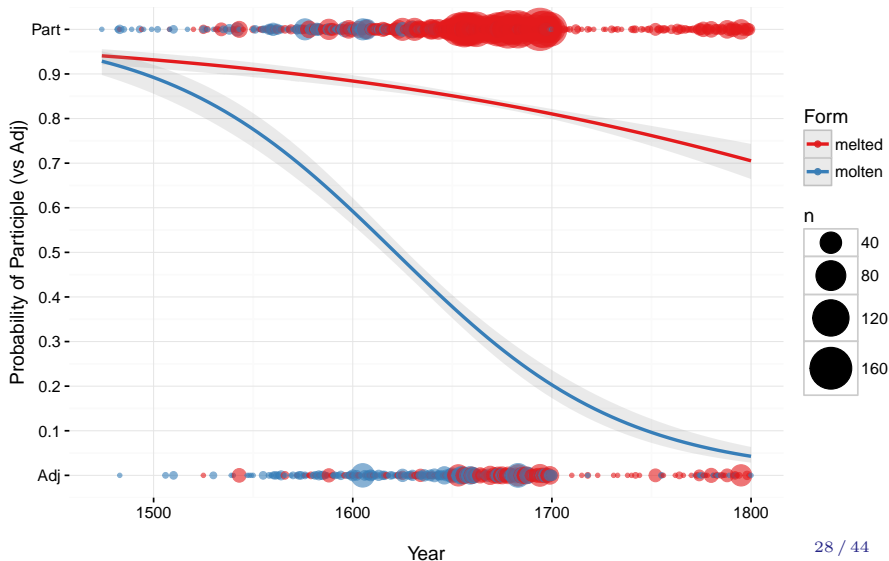
**Corollary 1:** the frequency of Variant A in  $C_1$  and Variant A in  $C_2$  will need to move away from each other.

**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?

**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 the child's probability estimates 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.
- Imagine a pond...

## Another Possible Pond Case

Indefinite article allomorphs from the numeral ONE.

- (15)    a. a table
- b. an ear
- (16)    a. a tish
- b. an eyer

Once {a, an/eyn} was innovated and the dimension of specialization chosen (following vowel)...

*a* would have a cluster-avoiding advantage

*an* would have a hiatus-avoiding advantage

## Conclusions

- **Specialization** can allow competing forms to survive, but only if their functions diverge.
- An extension of Yang (2000, 2002)'s **variational learning model** provides some specific mathematical hypotheses about specialization in acquisition, which we can test.
- **PrinCon** has a natural definition in this model, and can be reconciled with the speed of specialization.

## Final Conclusions and Challenges

- Replacement and specialization both play out at an individual and speech community level, simultaneously.
- We can observe specialization diachronically, but can we observe the choosing of domains of specialization?
- Can we observe it in acquisition?
  - Production data may not be good enough here.



## Acknowledgements

Thank you first to Josef Fruehwald for working out many of these ideas with me, and to Anthony Kroch, Betsy Sneller, Charles Yang. Special thanks to Aaron Ecay for help with PYCCLE and Weihnachtsgurke, among other things. Thanks also to Anton Karl Ingason, Laurel Mackenzie, anonymous reviewers.

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



**Newcastle**  
University



CENTRE FOR  
BEHAVIOUR AND  
EVOLUTION

## References I

- Aronoff, Mark. 1976. *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.
- 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.
- 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.

## References II

- 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.
- Clark, Eve V. 1995. *The lexicon in acquisition*, volume 65 of *Cambridge Studies in Linguistics*. Cambridge University Press.
- 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>.

## References III

- Heycock, Caroline, and Joel C. Wallenberg. 2013. How variational acquisition drives syntactic change: the loss of verb movement in Scandinavia. *The Journal of Comparative Germanic Linguistics* 16:127–157.
- 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. 1994. Morphosyntactic variation. In *Papers from the 30th Regional Meeting of the Chicago Linguistics Society: Parasession on Variation and Linguistic Theory*, ed. K. Beals et al.

## References IV

- Kroch, Anthony S., Beatrice Santorini, and Ariel Diertani.  
2010. Penn Parsed Corpus of Modern British English. Size  $\sim$   
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.
- 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.
- Taylor, Ann. 1994. Variation in past tense formation in the  
history of English. *University of Pennsylvania Working  
Papers in Linguistics* 1:143–159.

## References V

- 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.
- 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.

## Simultaneous replacement or extreme specialization?

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

### 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:  $\mu_{C_A}$ ,  $\mu_{C_B}$
3. Specialization completes in a continuous dimension:
  - **Actively**, by moving  $\mu_{C_A}$ ,  $\mu_{C_B}$  away from each other?
  - **Passively**, by allowing  $\mu_{C_A}$ ,  $\mu_{C_B}$  the possibility of moving away from each other?



## Specialization in Acquisition: active or passive?

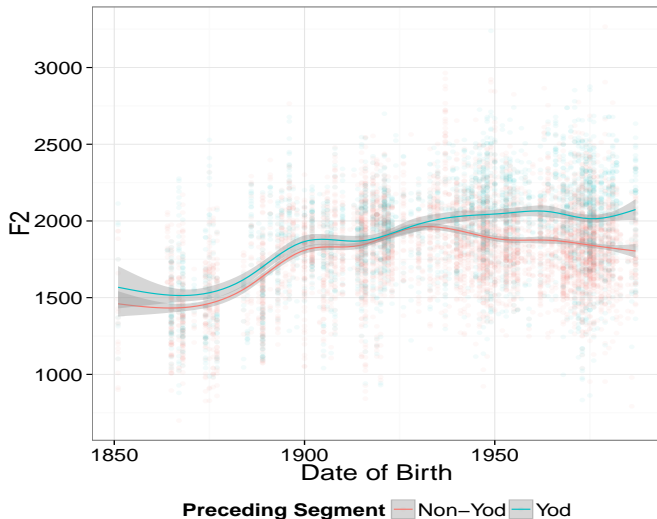
- Is there any way to distinguish the two, given that different linguistic cases may have different selectional pressures?
  - You can model a lot of scenarios assuming various selectional pressures interacting with various child-driven “accelerations” of specialization.
- **Possible hypothesis:** maybe the child-driven amount of manipulating A and B’s frequencies is the same per token in every linguistic case, and can be estimated.
- Maybe we can identify some true neutral changes, to abstract away from selective pressures (Kauhanen, 2016).

## But what if there's a global selective pressure for B?

- 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  $C_2$ .

# Phonological Specialization:

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



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

(308 speakers)

