Towards a model of variational specialization in acquisition

Joel C. Wallenberg
joel.wallenberg@ncl.ac.uk

7 október 2016
Háskóli Íslands

Outline

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, Wallenberg, & van der Wurff 2012).

English

Disjunction:

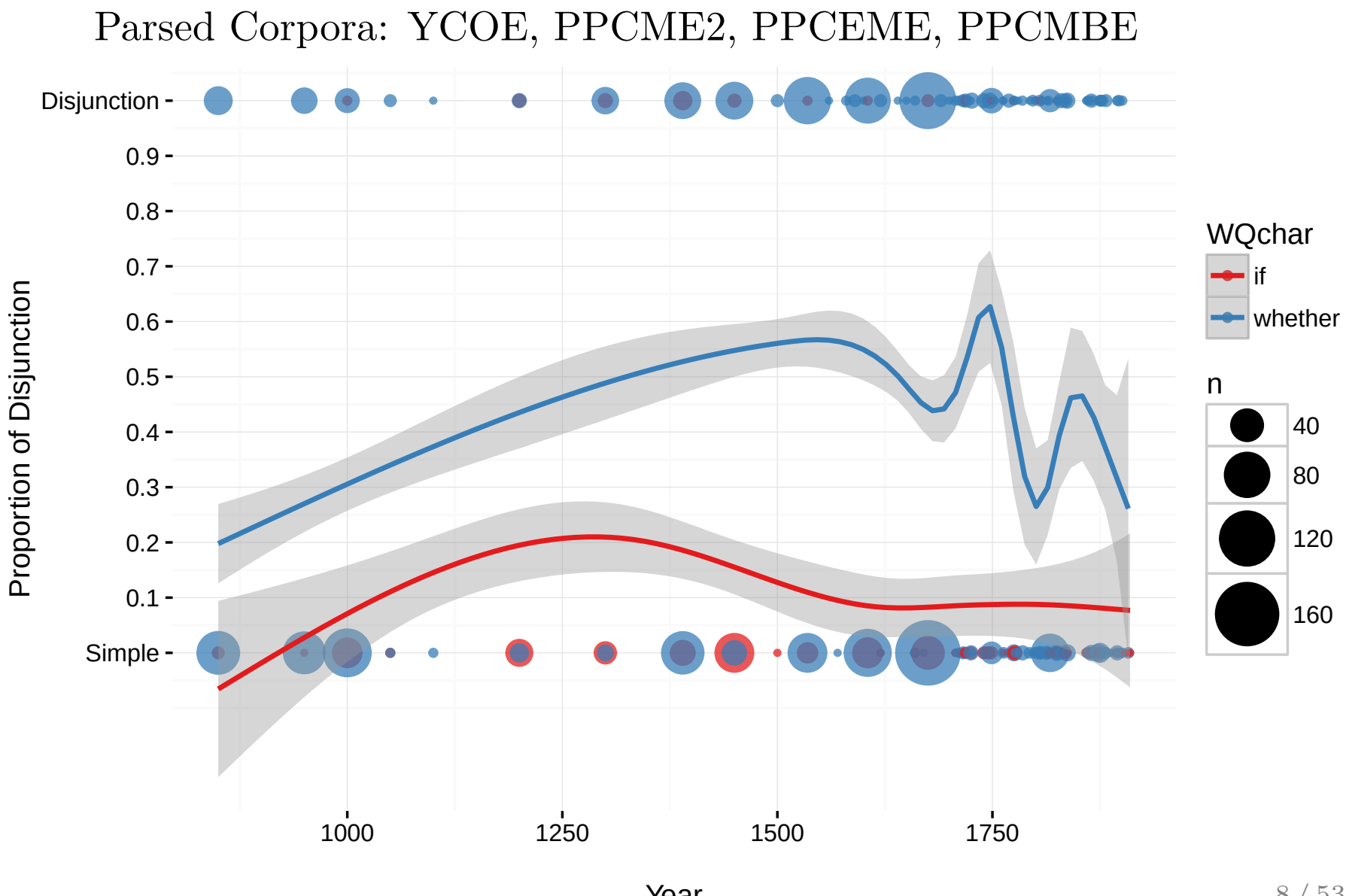
(2) I wonder {**whether**,if} John or Bill is bringing coffee.

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

Simple:

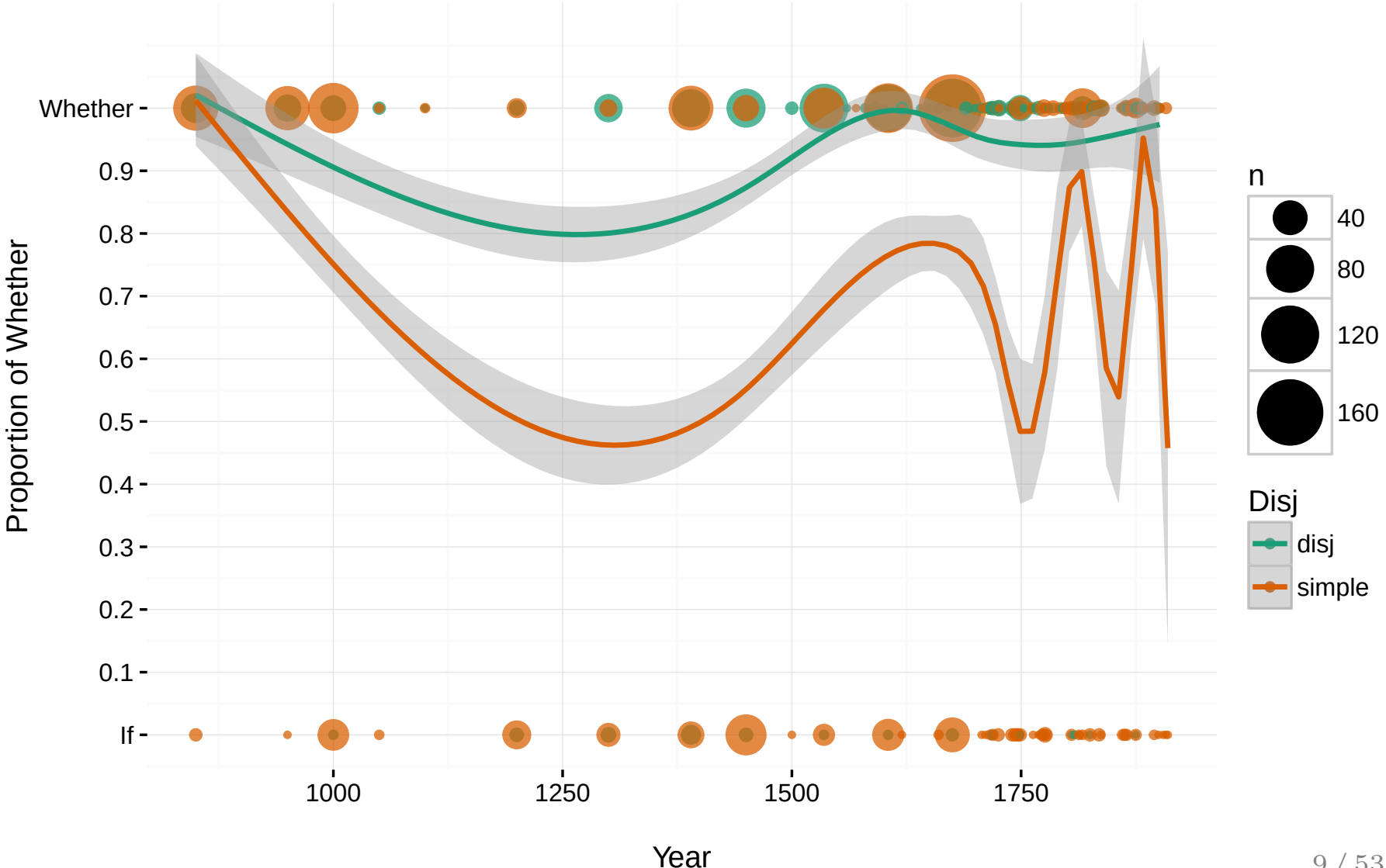
(4) I wonder {whether, **if**} Bill is bringing coffee.

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



whether/if replacement slowed/arrested

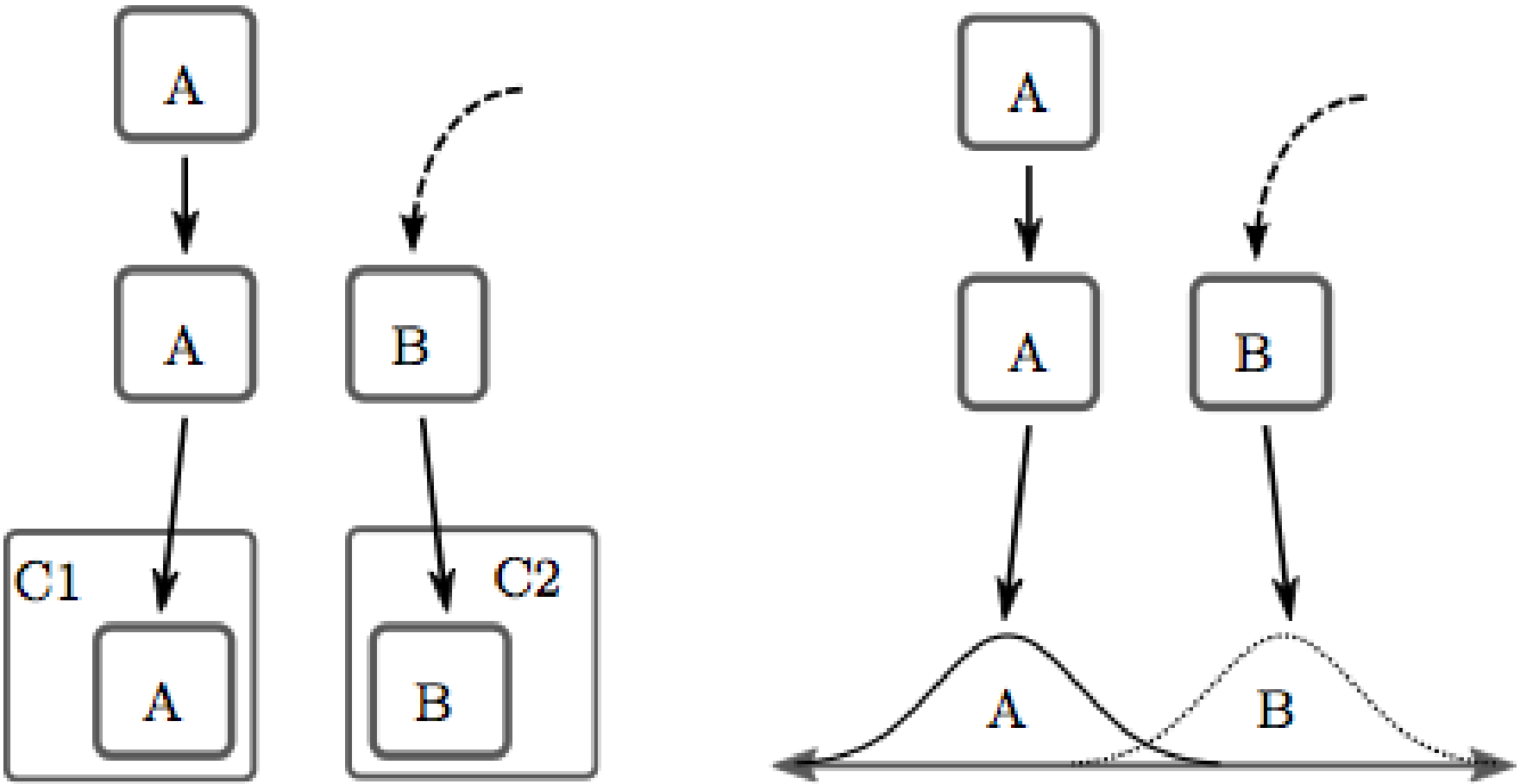
(N = 1929 clauses)



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.
- 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 (Wallenberg, to appear, 2013; Fruehwald and Wallenberg, in prep).
  - 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

Icelandic

(5) stjarna væri sén í landnorðri frá Jemen, [er
star was-SUBJ seen in northeast from Yemen that
Kómeta heitir]
Comet is-called
"A star would have been seen in the northeast from Yemen that's called a Comet
(1861.ORRUSTA.NAR-FIC,.784)

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)

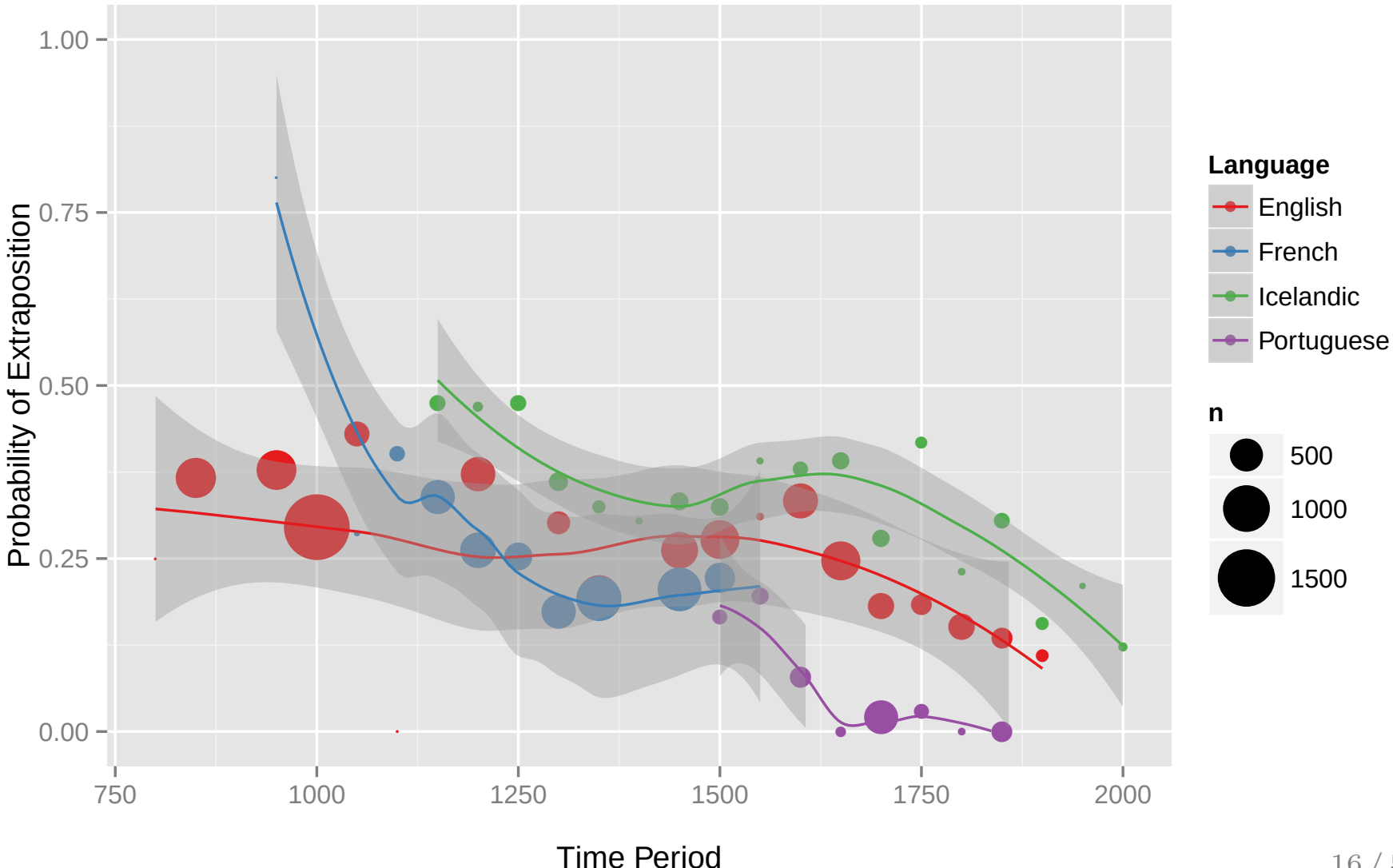
Hypotheses for the diachronic study

Hypothesis: Relative clause position (a binary variable) is specialized along a continuous dimension, weight, and so it should be **nearly** stable, but not entirely stable.

Diachronically, Crosslinguistically

- **English:** YCOE (Taylor et al., 2003), PPCME2 (Kroch and Taylor, 2000), PPCEME (Kroch et al., 2005), PPCMBE (Kroch et al., 2010).
- **Icelandic:** IcePaHC (Wallenberg, A.K. Ingason, E.F. Sigurðsson, and E. Rögnvaldsson, 2011).
- **Old/Middle French:** MCVF (Martineau et al., 2010).
- **Historical Portuguese:** Tycho Brahe Corpus of Historical Portuguese (Galves and Faria, 2010).

Four Languages (Subj Ex), over time



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.

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.

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: 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 *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):

(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)

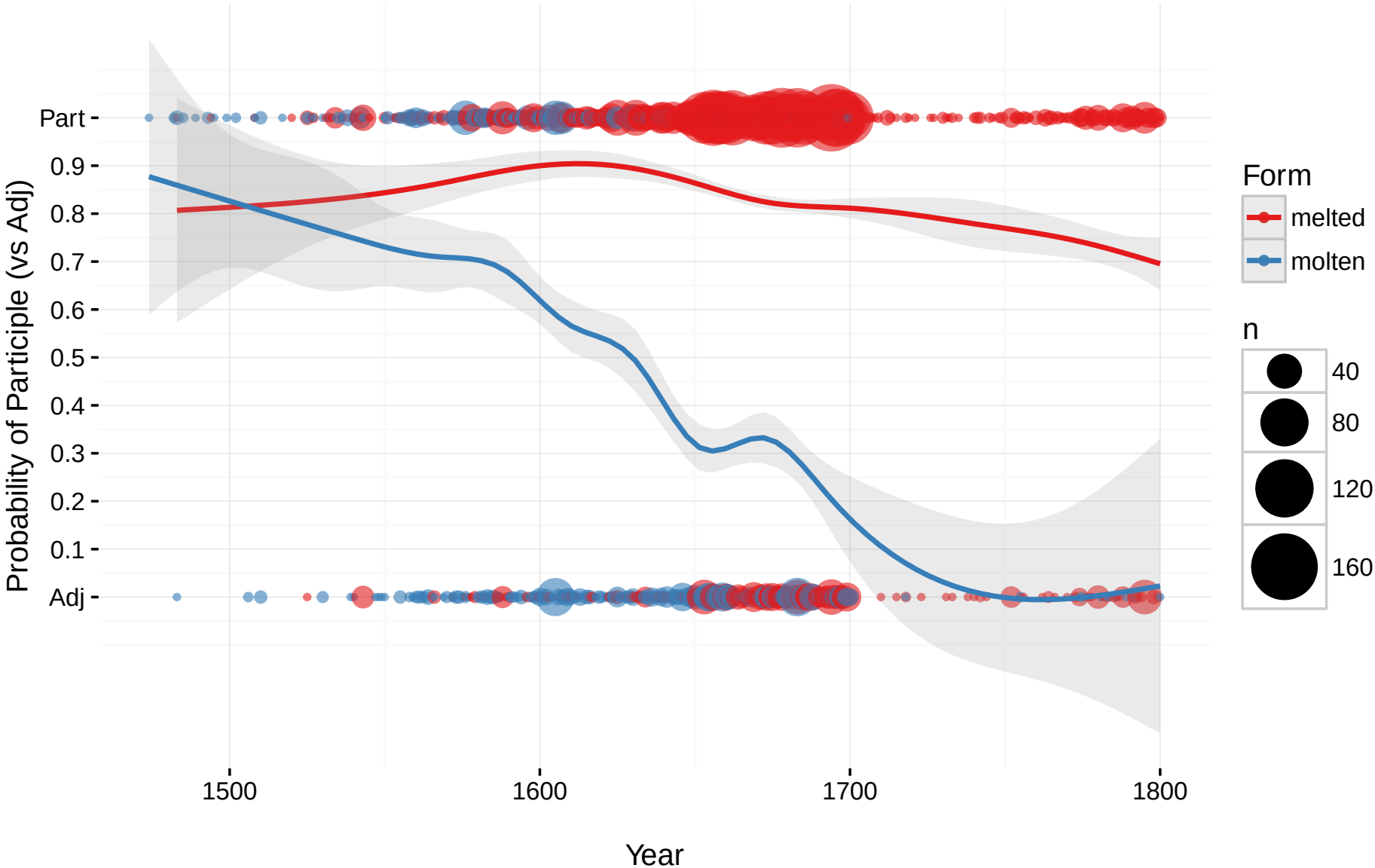
melted/molten specialization

(10) The gold was {melted / *molten} by the fire.
(participle context)

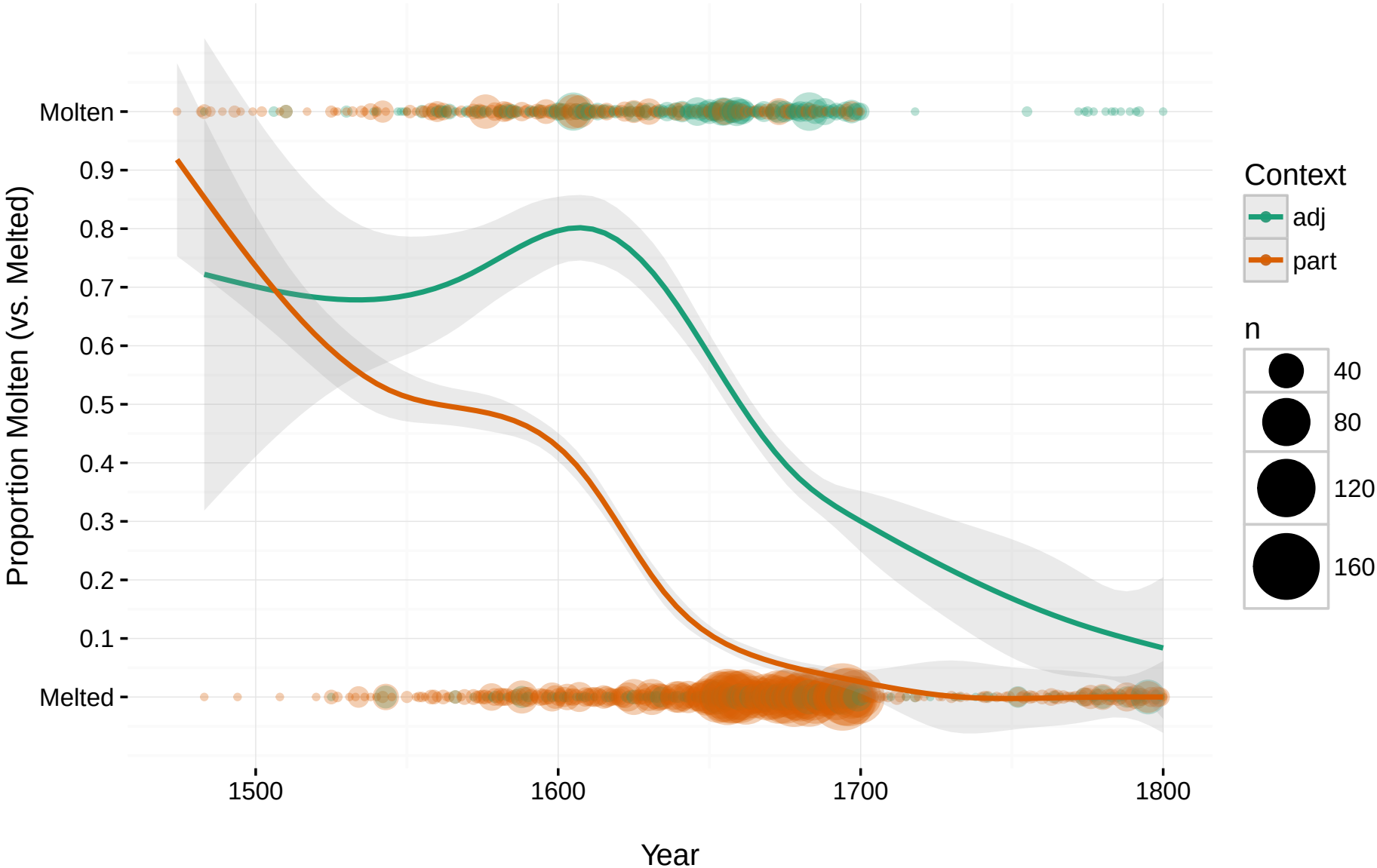
(11) The {?melted / molten} gold flowed down the hill.
(adjectival context)

- 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.

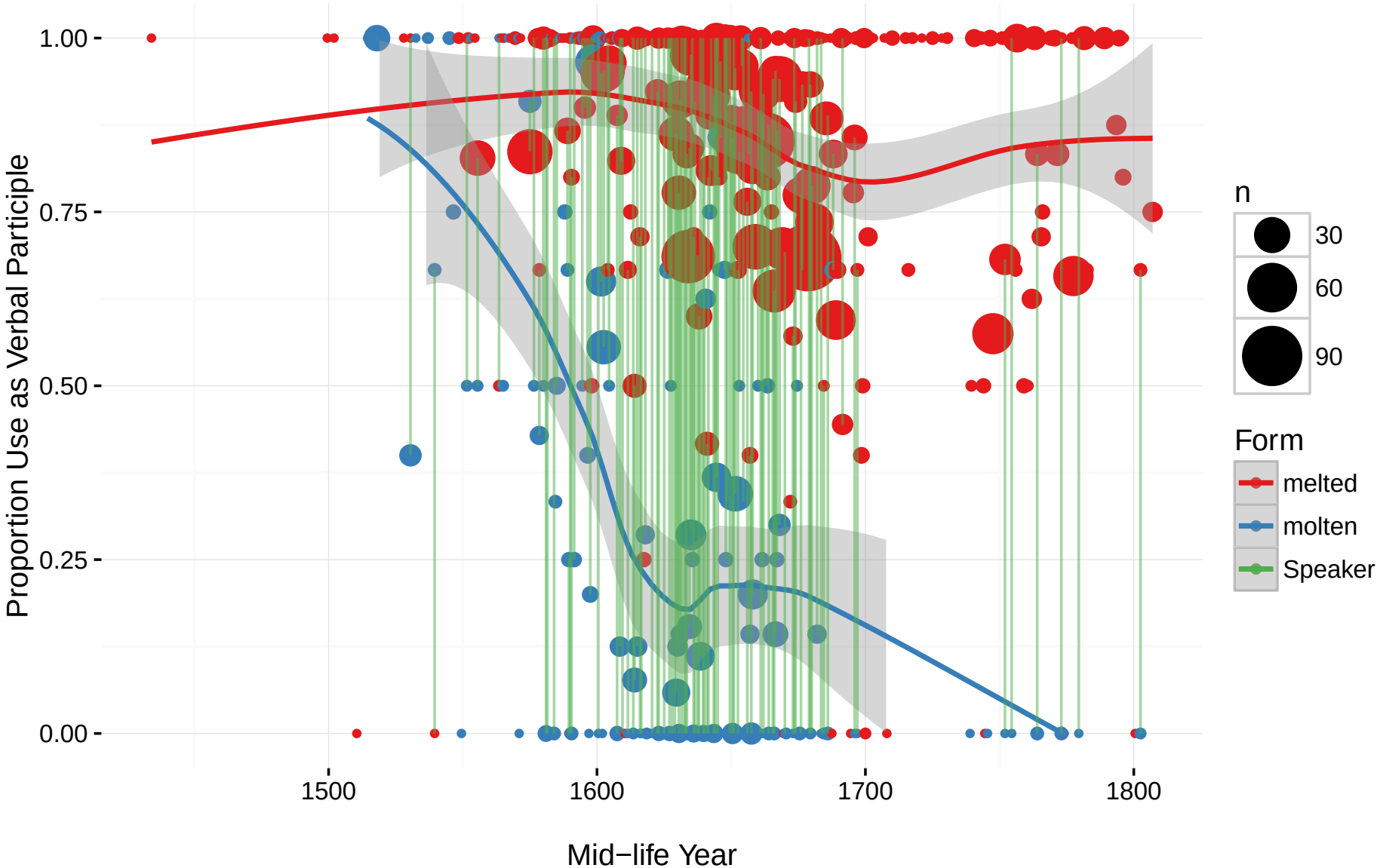
melted/molten specialization N = 7946 tokens



Simultaneous Replacement? N = 7946 tokens

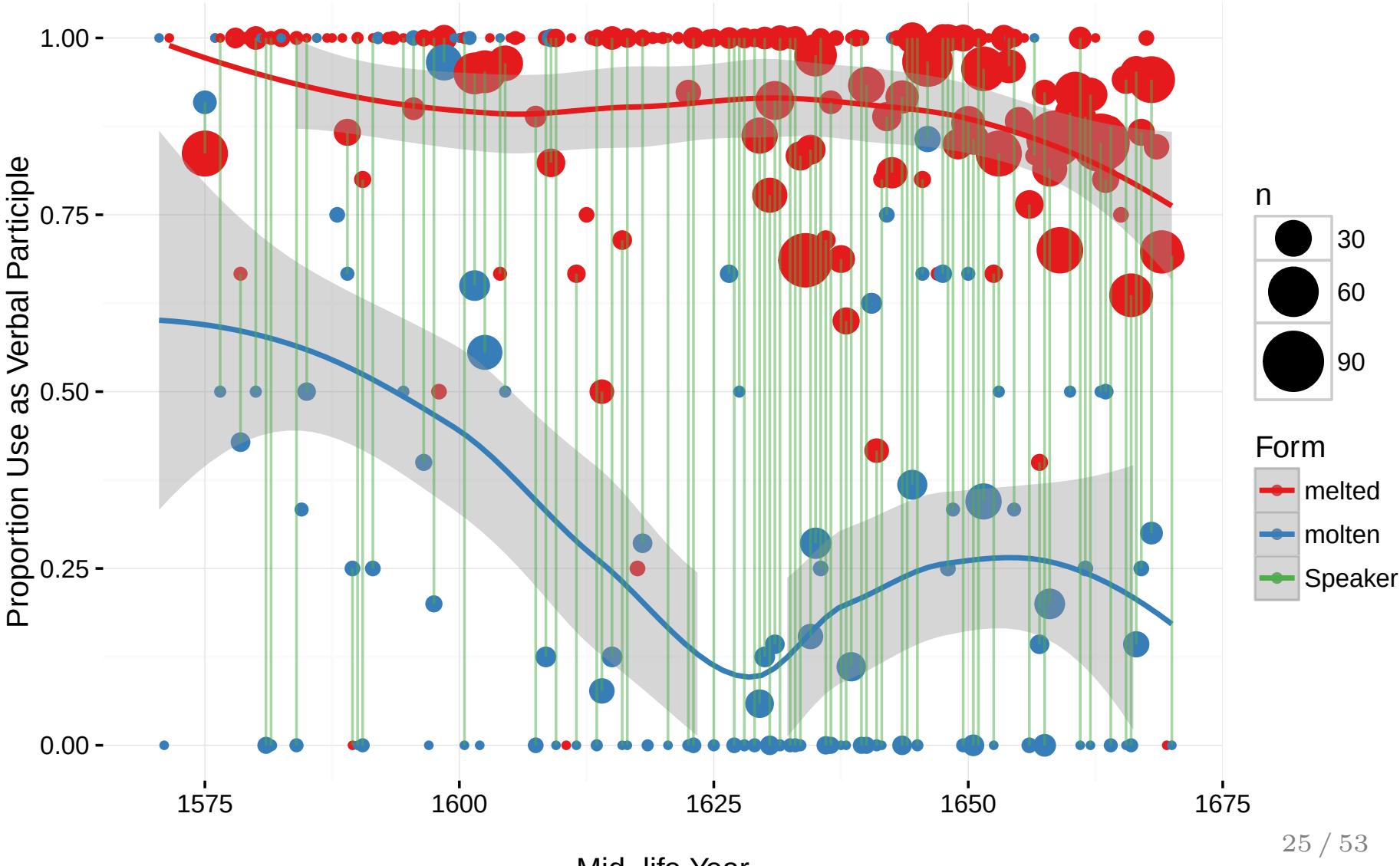


471 identifiable speakers, N = 3601 tokens



Individual Speakers, 1570-1670 midlifes

(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)

Solving Yang's Paradox

- 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**, though 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,...,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 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?

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?

- 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).

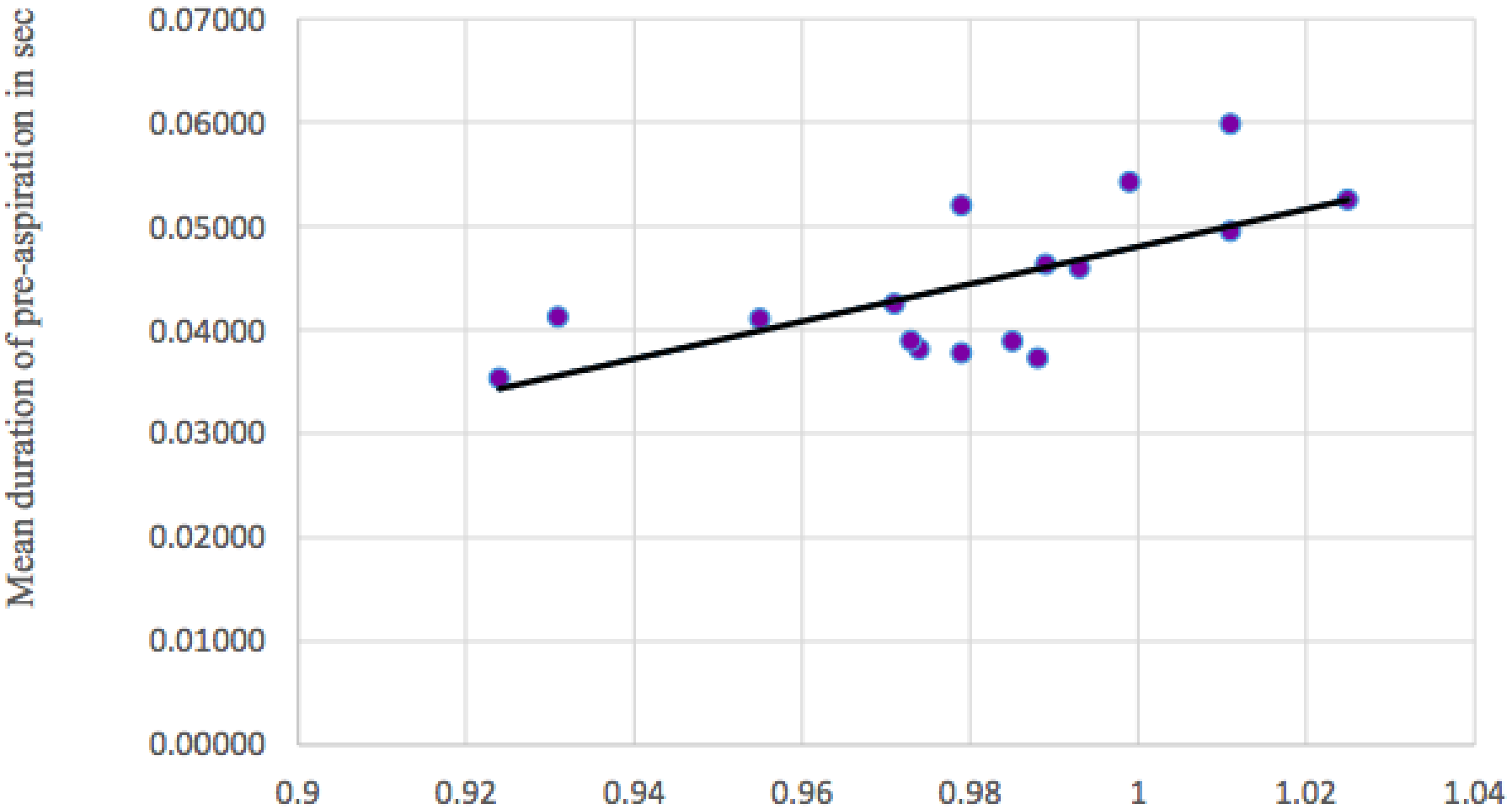
How does it relate to the project?

- 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 (mathematical) theory of this.
  - We need to test quantitative hypotheses.

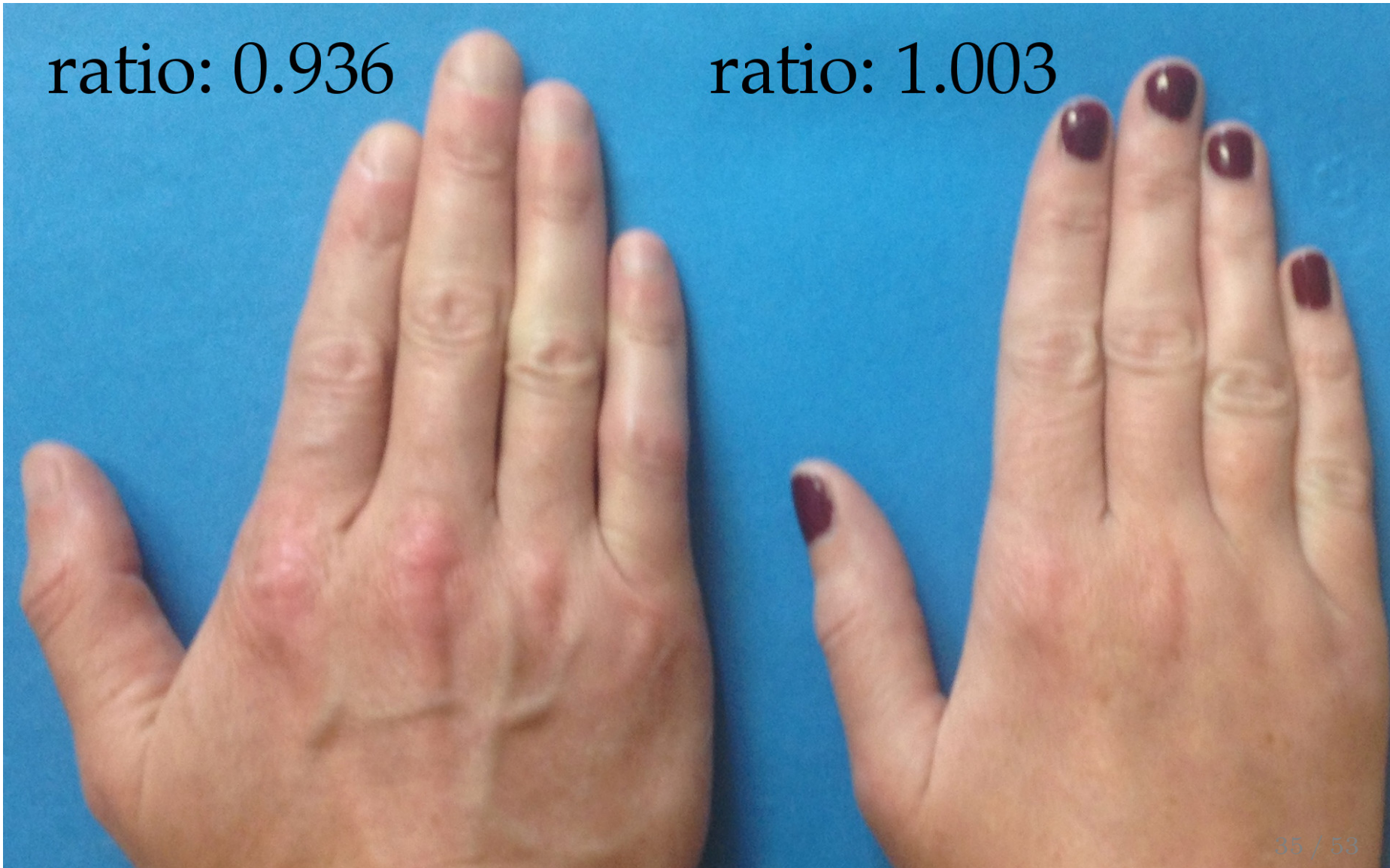
Now for something completely different: gender

- **Linguistic sex/gender effects:** it is well known that speaker-sex has a stochastic effect on the frequency with which linguistic variants are used (Labov, 2001; Eckert, 2011, and many, many more).
- **Hormonal Organising Effects:** the action of sex steroids during sensitive period for sexual differentiation (*in utero*, esp. weeks 8-24 for humans), affecting primary/secondary sex characteristics, and:
  - **Brain morphology:** e.g. the sexually-dimorphic nucleus of the pre-optic area in mammals, including humans, with a correlate in birds (see review in Balthazart, 2011).
  - **Behaviour:** mating behaviours in various mammals, pair-bonding and birdsong in birds, **gender identity** in humans (Hines et al., 2002; Berenbaum and Bailey, 2003; Hines et al., 2004; Cohen-Bendahan et al., 2005; Auyeung et al., 2009), and **gendered social learning in humans** (Hines et al., 2016).

Rise of Pre-aspiration in Tyneside English in 16 women (Cochrane 2016)



2D:4D ratio: smaller → more pre-natal T (or T/E)



How does it relate to the project?

- If there are subtle biases in social learning, based on early life hormone exposure (and correlating with continuous gender identity), this will affect the spread of new variants through a population.
- If we catch new variants in their early spread, we may see these biases more clearly.
- Given the rich sociological information available in Iceland, and the contact situation, we might have a greater opportunity to identify such biases.
- Some measures of hormone exposure (e.g. 2D:4D hand digit ratio) and gender identity (e.g. continuous gender scale) are fairly cheap and quick.

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 to Jacques Balthazart, Claire Cochrane, Anton Karl Ingason, Laurel Mackenzie, anonymous reviewers, and my RAs and participants.

`https://github.com/joelcw/tyneside/tree/master/extraposition`

`https://github.com/joelcw/molten`

References I

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

Auyeung, Bonnie, Simon Baron-Cohen, Emma Ashwin, Rebecca Knickmeyer, Kevin Taylor, Gerald Hackett, and Melissa Hines. 2009. Fetal testosterone predicts sexually differentiated childhood behavior in girls and in boys. *Psychological Science* 20:144–148.

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.

Balthazart, Jacques. 2011. *The Biology of Homosexuality*. Oxford: Oxford University Press.

References II

Berenbaum, Sheri A, and J Michael Bailey. 2003. Effects on gender identity of prenatal androgens and genital appearance: evidence from girls with congenital adrenal hyperplasia. *The Journal of Clinical Endocrinology & Metabolism* 88:1102–1106.

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 30months. *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.

References III

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.

Cohen-Bendahan, Celina CC, Cornelieke van de Beek, and Sheri A Berenbaum. 2005. Prenatal sex hormone effects on child and adult sex-typed behavior: methods and findings. *Neuroscience & Biobehavioral Reviews* 29:353–384.

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 IV

Eckert, Penelope. 2011. language and power in the preadolescent heterosexual market. *American Speech* 86:85–97.

Fruehwald, Josef, and Joel C. Wallenberg. in prep. Optionality is Stable Variation is Competing Grammars.

Galves, Charlotte, and Pablo Faria. 2010. The Tycho Brahe Corpus of Historical Portuguese. Department of Linguistics, University of Campinas (UNICAMP), Online Version, `http://www.tycho.iel.unicamp.br/~tycho/`.

Heycock, Caroline, and Joel C. Wallenberg. 2013. How variational acquisition drives syntactic change: the loss of verb movement in Scandinaviation. *The Journal of Comparative Germanic Linguistics* 16:127–157.

References V

Hines, Melissa, Charles Brook, and Gerard S. Conway. 2004. Androgen and psychosexual development: Core gender identity, sexual orientation, and recalled childhood gender role behavior in women and men with congenital adrenal hyperplasia (CAH). *Journal of Sex Research* 41:75–81.

Hines, Melissa, Susan Golombok, John Rust, Katie J Johnston, Jean Golding, Parents, Children Study Team, et al. 2002. Testosterone during pregnancy and gender role behavior of preschool children: a longitudinal, population study. *Child development* 73:1678–1687.

References VI

Hines, Melissa, Vickie Pasterski, Debra Spencer, Sharon Neufeld, Praveetha Patalay, Peter C Hindmarsh, Ieuan A Hughes, and Carlo L Acerini. 2016. Prenatal androgen exposure alters girls' responses to information indicating gender-appropriate behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371:20150125.

Kauhanen, Henri. 2016. Neutral change. *Journal of Linguistics* 1–32. Draft: `http://henr.in/doc/neutral-change.pdf`.

Kroch, Anthony, Beatrice Santorini, and Ariel Diertani. 2005. Penn-Helsinki Parsed Corpus of Early Modern English. 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 VII

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.

Labov, William. 2001. *Principles of Linguistic Change. Volume 2: Social Factors*. Oxford: Blackwell.

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.

References VIII

Martineau, France, Paul Hirschbühler, Anthony Kroch, and Yves Charles Morin. 2010. Modéliser le changement : les voies du français, Département de français. Size ∼ 994000 words. University of Ottawa, `http://www.arts.uottawa.ca/voies/voies_fr.html`.

Taylor, Ann. 1994. Variation in past tense formation in the history of English. *University of Pennsylvania Working Papers in Linguistics* 1:143–159.

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`.

References IX

Wallenberg, Joel C. 2013. A unified theory of stable variation, syntactic optionality, and syntactic change. In *Presented at the 15th meeting of the Diachronic Generative Syntax (DIGS) Conference*. University of Ottawa.

Wallenberg, Joel C. to appear. Extraposition is disappearing. To Appear in *Language* Historical Syntax Online Section. Draft available upon request.

Wallenberg, Joel C., Anton K. Ingason, Einar F. Sigurðsson, and Eiríkur Rögnvaldsson. 2011. Icelandic Parsed Historical Corpus (IcePaHC). Version 0.9. Size: 1 million words. URL `http://www.linguist.is/icelandic_treebank`.

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?

(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 commes of the snowe waters molten and thawed in those regions. (attr Barnabe Riche, *The famous hystory of Herodotus...*, date: 1584)

Model Comparison: specialization by context

Model 1: Form ~(1 | file) + (1 | author) + zDate + Context

Model 2: Form ~(1 | file) + (1 | author) + zDate * Context

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

Prediction for an Active Hypothesis

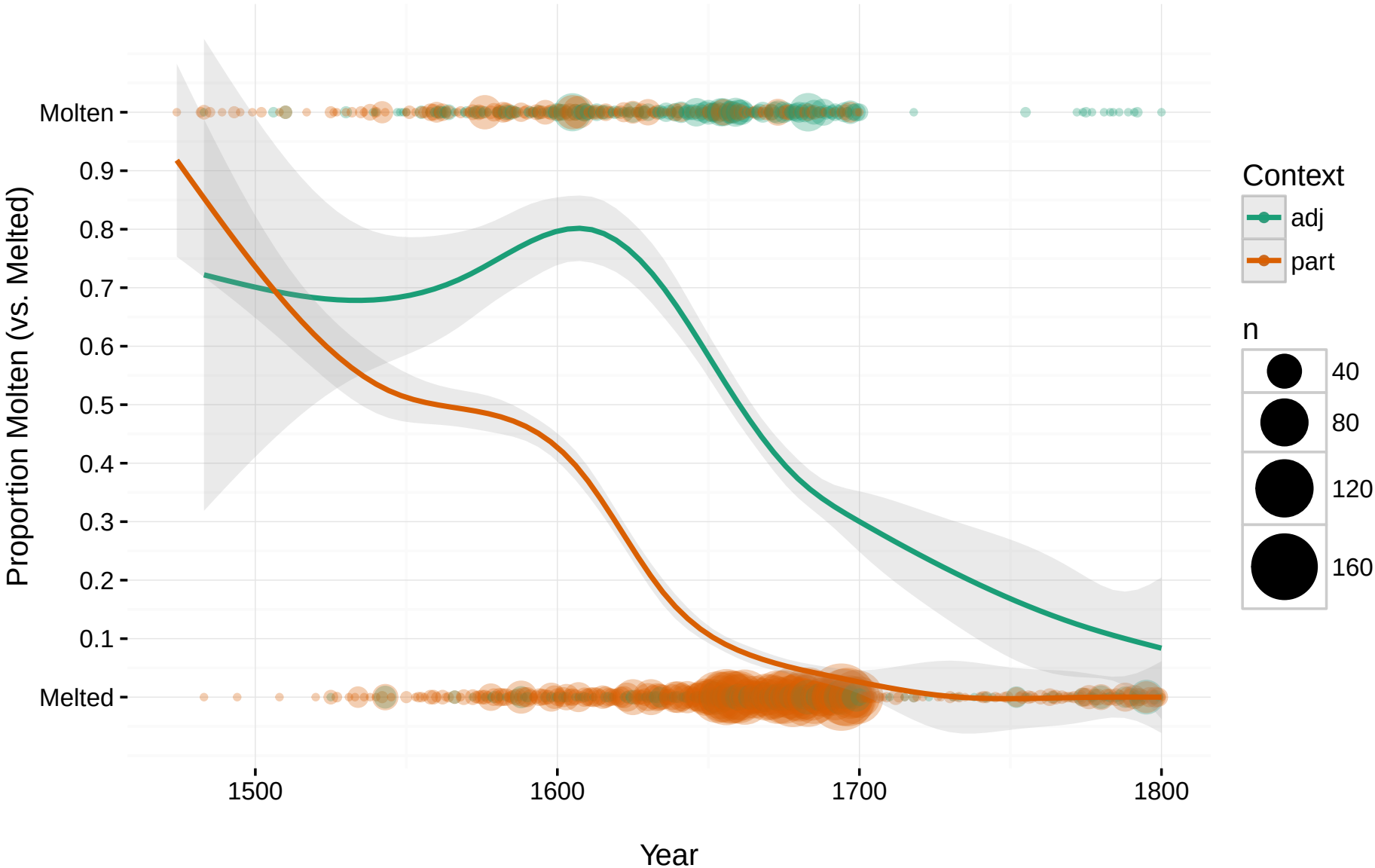
In true neutral change ONLY: once specialization begins to take place, it should be relentless, and **symmetrical**, and both variants should always survive (in the cat-cat case).

- The frequency of Variant A in C_1 is always being augmented in 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

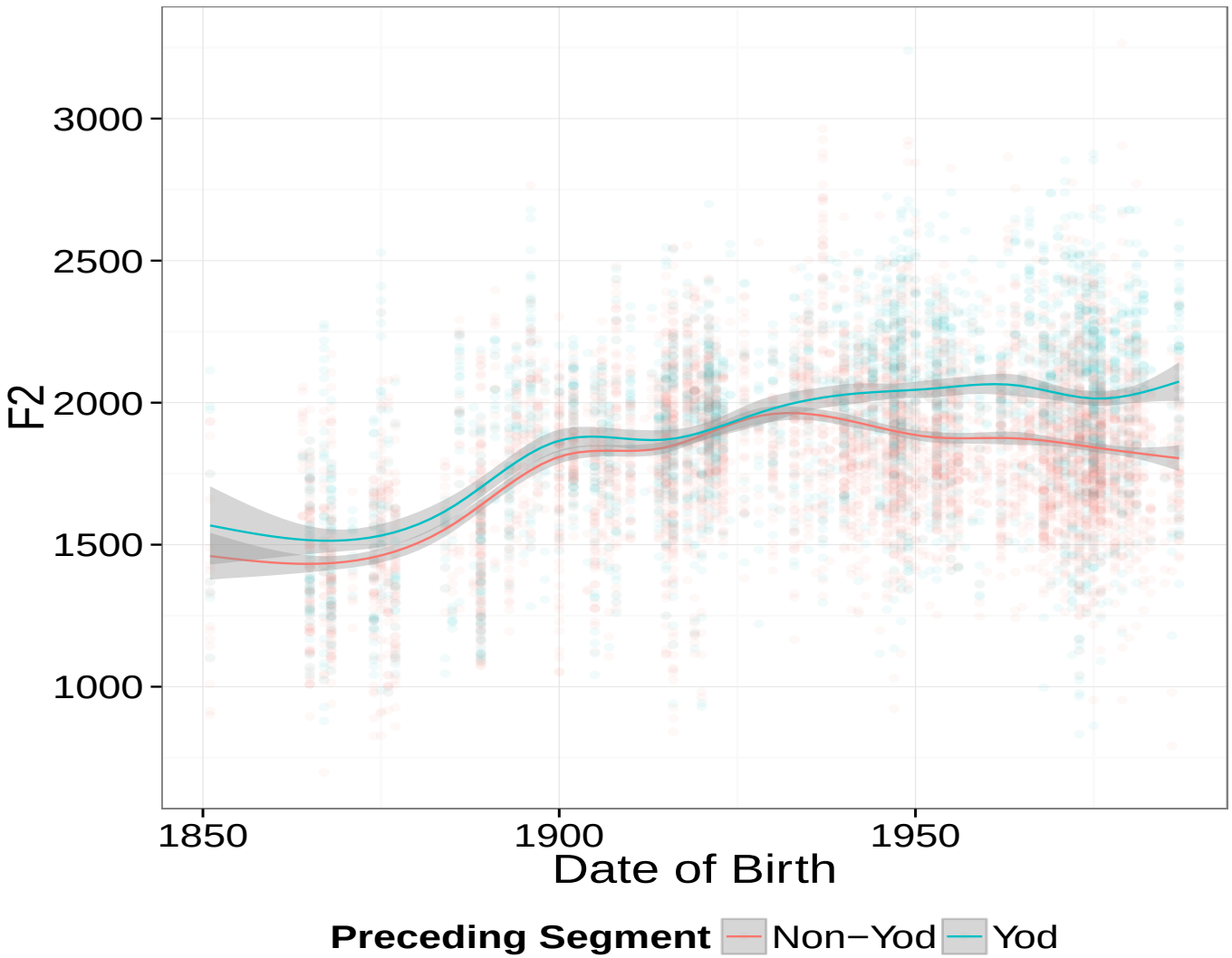


But what if there's a global selective pressure for B?

- 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)