A theory of variational specialization across domains of the grammar

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

3rd December 2016
International Conference on Formal Linguistics 7
Nankai University
Outline

Specialization and Replacement
   The Principle of Contrast and dimensions of specialization
   Imperfect 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.
- 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)

Parsed Corpora: YCOE, PPCME2, PPCEME, PPCMME
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, 2013b; 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 Yemen, [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.
Hypotheses for the diachronic study

- **Hypothesis 2**: All IE relative clauses derive historically from clause-adjoined relatives (Kiparsky, 1995).
- **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 (Wallenberg, to appear).
Specialization within individuals

- Relative clause extraposition in the Parsed Corpus of Early English Correspondence (PCEEC; Taylor et al. 2006).
- 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 Ingason and MacKenzie, 2011).
Diachronically, Crosslinguistically

- **English:** YCOE (Taylor et al., 2003), PPCME2 (Kroch and Taylor, 2000), PPCEME (Kroch et al., 2005), PPCMME (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).
English, over time (N = 18530 clauses)
cf. Yiddish Tense, N = 1030 clauses

Wallenberg (2013a), building on Santorini (1993)
Pause-fillers, *um* vs. *uh*, $n_{speakers} = 308$

(Fruehwald 2015, using *Philadelphia Neighborhood Corpus*) $n_{tokens} = 25514$
English, over time (N = 18530 clauses)
English, average weight over time (N = 18530)
Icelandic, over time (N = 3486)
Old/Middle French, over time (N = 8207)
Portuguese, over time (N = 2398)
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.
- 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).
The origin of the slow change?

(7) yó mártyaḥ śíśīte áty aktúbhir,
    which mortal sharpen-Mid-Sg overly nights-INSTR,
    mā naḥ sá ripúr īśata
    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 Kiparsky, 1995, 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 Kiparsky (1995) 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: 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):

  (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?
- 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

The diagram illustrates the probability of participle (vs. adjective) forms of the terms 'melted' and 'molten' over different years. The graph shows a downward trend, indicating a decrease in the use of the participle form as time progresses.
Simultaneous Replacement? N = 7946 tokens

Proportion Molten (vs. Melted)

Year

Melted

Molten

Context

adj
part

n

40
80
120
160
Model Comparison: specialization by context

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

<table>
<thead>
<tr>
<th>model</th>
<th>AIC</th>
<th>BIC</th>
<th>p-value (Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant Rate</td>
<td>4777.4</td>
<td>4812.3</td>
<td>–</td>
</tr>
<tr>
<td>with Date*Context</td>
<td>4766.0</td>
<td>4807.8</td>
<td>0.0003</td>
</tr>
</tbody>
</table>
471 identifiable speakers, N = 3601 tokens
Individual Speakers, 1570-1670 midlifes

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

(15)  a. Method of breeding Horses... Molten grease and fatning balls

b. ... which may bring away any melted grease

(16)  a. ... the grease is molten into them

b. ... considering that if grease should be melted

(17)  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.
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:
   a. For categorical variants along categorical dimensions, decouple tracked frequencies of variants for each context, $C_1,\ldots,C_n$, 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 C₁ and C₂.

• Specialization completes in a categorical dimension:
  • Actively, by augmenting the represented frequency of Variant A in C₁ and augmenting Variant B in C₂?
  • 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: \( \mu_{C_A}, \mu_{C_B} \)

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?
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
melted/molten: consider Corollary 2
Specialization in Acquisition: active or passive?

**Passive Hypothesis:** PrinCon is the decoupling of contexts $C_1, \ldots, C_n$ 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 $C_1, \ldots, C_n$.
- Contextual effects within the CRE can be thought of as transformations on a single tracked frequency for Variant A across all contexts.
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).
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 Yang (2000, 2002)’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.
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.
- Can we observe either well enough to test active vs. passive hypotheses?
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 Anton Karl Ingason, Laurel Mackenzie, anonymous reviewers.

https://github.com/joelcw/tyneside/tree/master/extraposition
https://github.com/joelcw/molten
References I


References II


References III

Fruehwald, Josef. 2015. Social meaning and information theory. Poster Presentation.

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


References IV


References V


References VI


References VII


http://www-users.york.ac.uk/~lang22/YcoeHome1.htm.


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


Simultaneous replacement or extreme specialization?

(18) *(molten implies heat in PDE:)*
Is silly putty molten rubber?

(19) *(molten implies liquidy/sludgy state in PDE:)*
melted spatula vs. molten spatula

(20) *(both:)*
melted cheese vs. molten cheese
(J. Fruehwald, p.c., for examples above)

(21) *(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)
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** → 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** → Variant A loses in both C\(_1\) and C\(_2\) but at different rates.
  - **augmentation > 2 x selective pressure** → A wins in C\(_1\) at the same rate B wins in C\(_2\).
  - **selective pressure < augmentation < 2 x selective pressure** → 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)