May 3, 2017

Learning socio-indexical features of words

Janet B. Pierrehumbert

University of Oxford e-Research Centre
Northwestern University

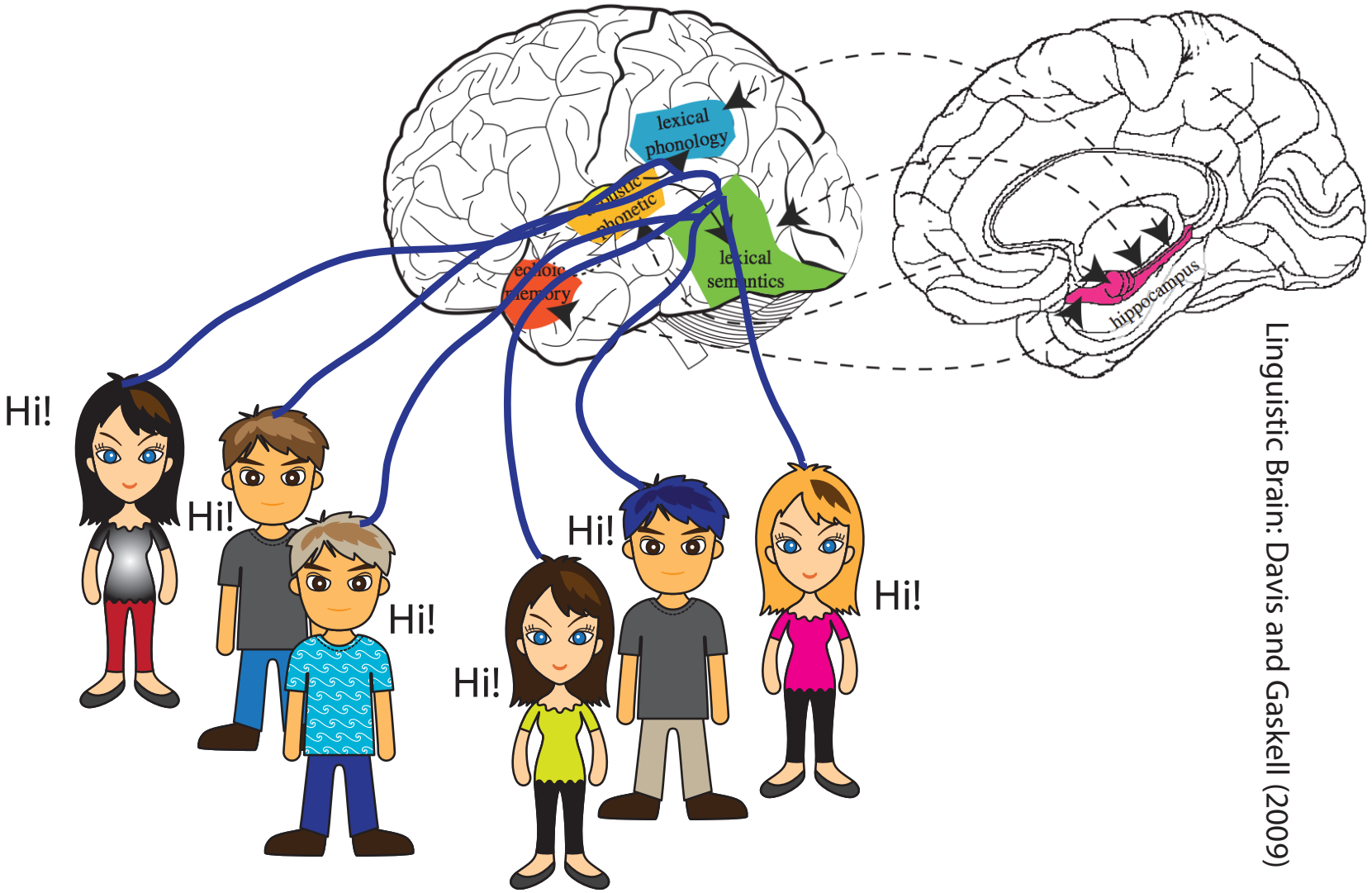
New Zealand Institute of Language Brain and Behaviour, University of Canterbury

Thanks to:

JOHN TEMPLETON FOUNDATION

SUPPORTING SCIENCE ~ INVESTING IN THE BIG QUESTIONS

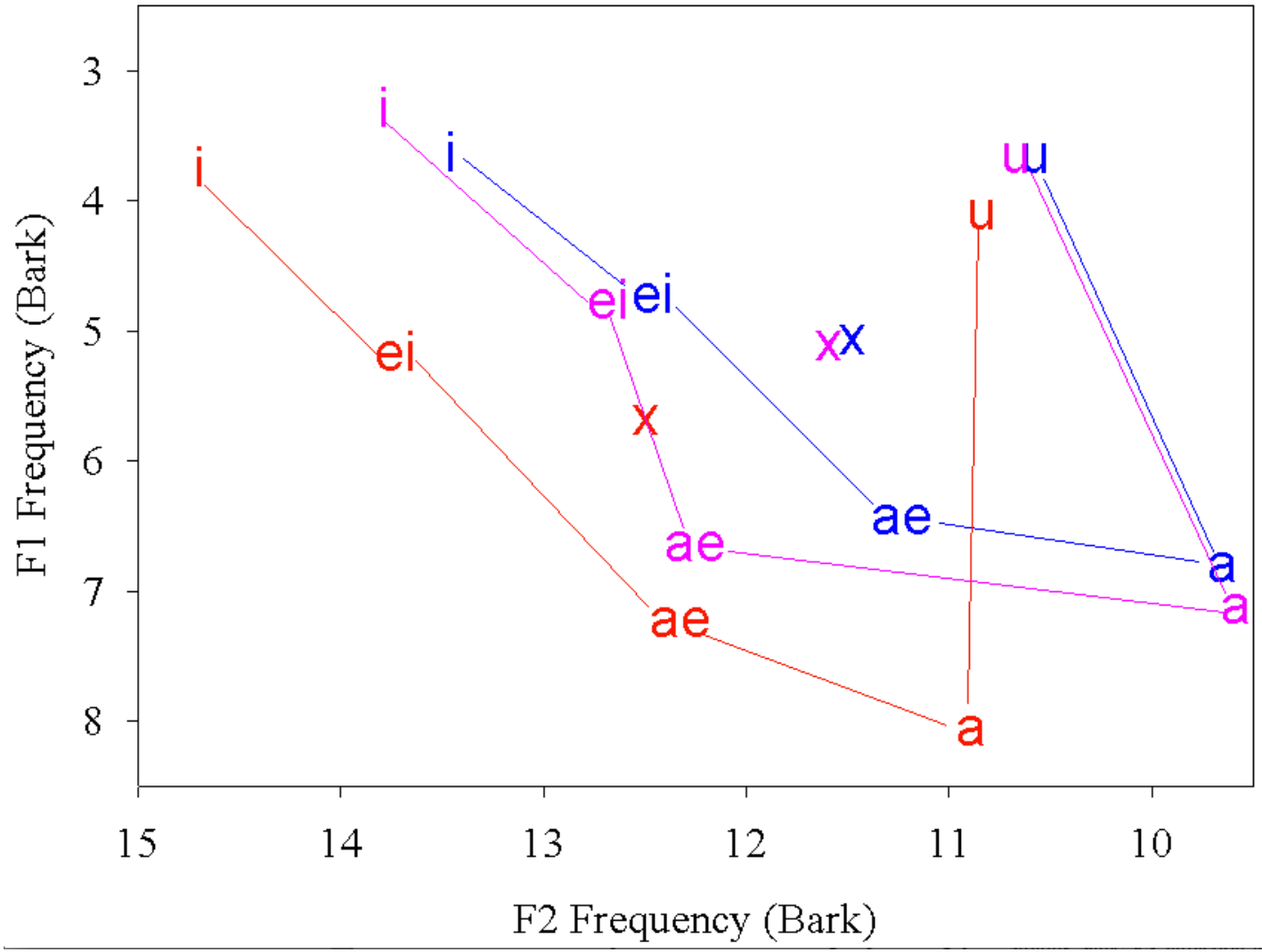
A collective brain



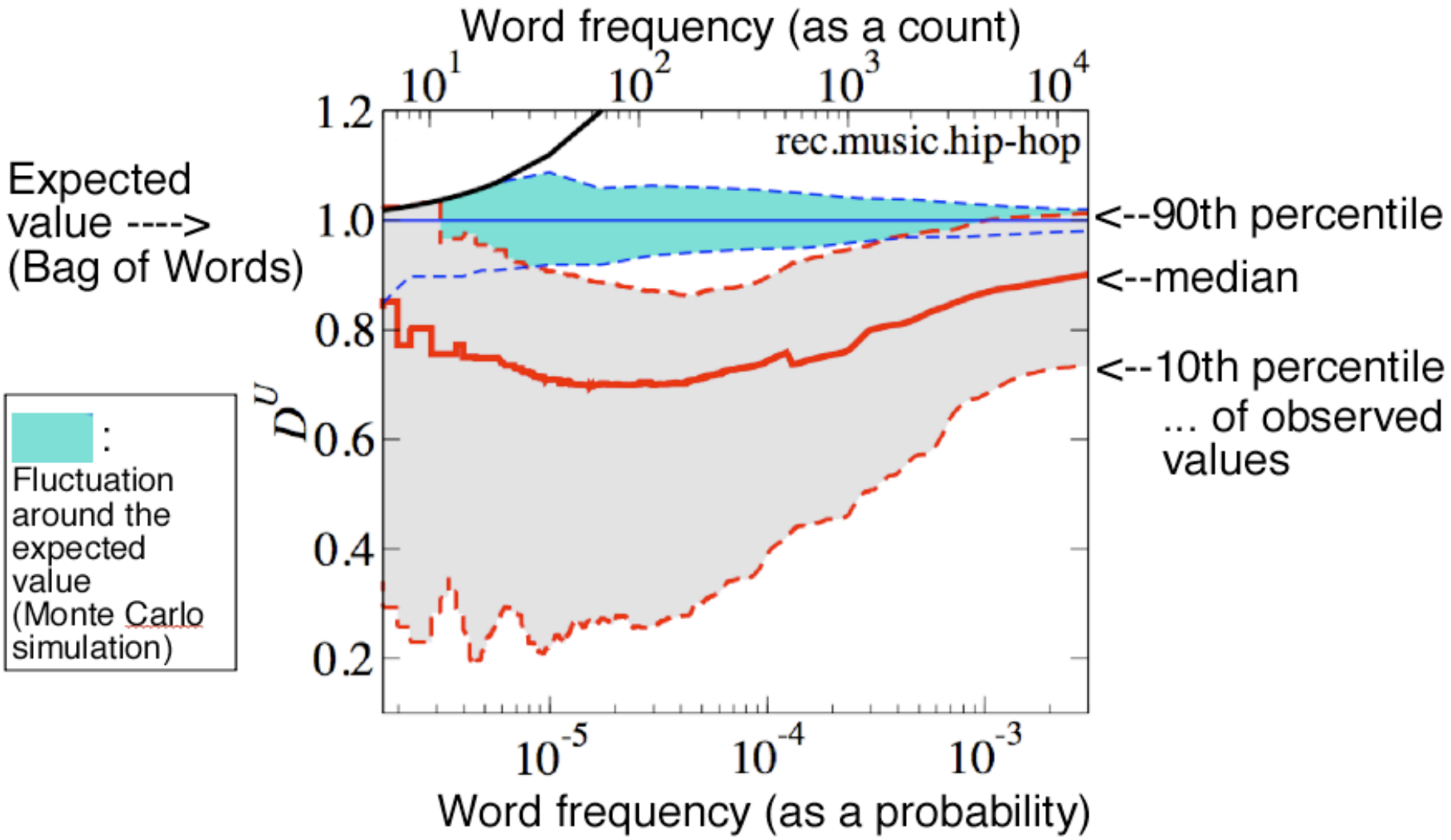
Linguistic Brain: Davis and Gaskell (2009)

This is incorrect! Within speech communities we find many differences.

Phonetic differences : Pierrehumbert et al (2004) on vowels of GLB vs straight speakers. Magenta = self-identified GB. Red, Blue = straight women and men.



Lexical differences : Altmann, Pierrehumbert, Motter (2011) on distributions of words in Usenet discussion groups.



$D^U = 1$ is randomly distributed. $D^U < 1$ means some people use the word more than others.

Questions

- Do individuals have implicit knowledge of the patterns of variation? (... or just of their own variants?)
- Are phonetic differences general, associated with words, or both?
- How are differences learned?
- How are differences generalized?

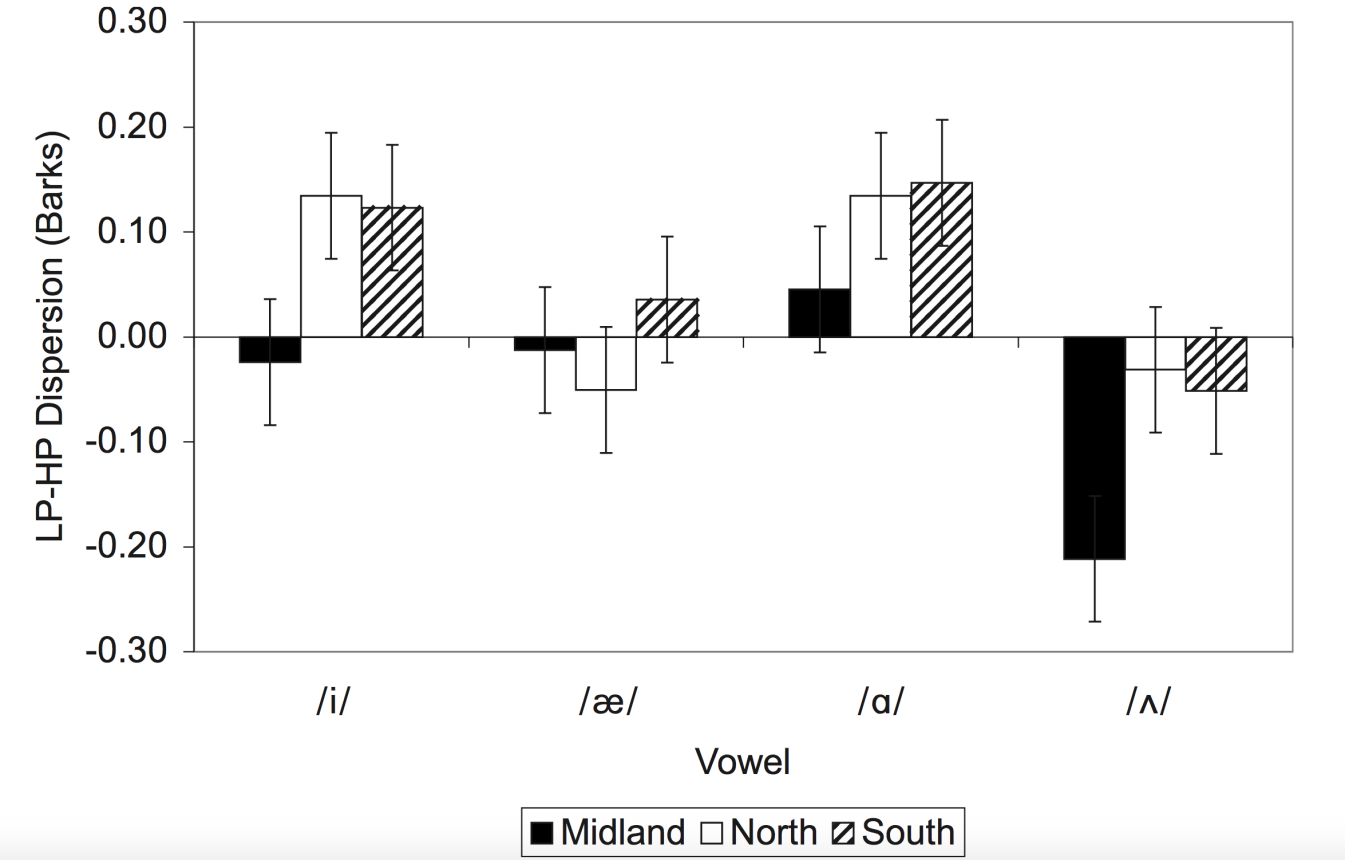
Clopper et al. (2008). Vowel qualities in predictable and unpredictable contexts, in three dialects.

- Predictable: The wedding banquet was a feast.

- Unpredictable: Tom has been discussing the beads.

Expect: Predictable vowels shorter (true) and more centralized (not exactly true).

Result: Interaction of dialect with predictability



- More predictable contexts lead to more extreme manifestations of dialect.
- Requires contextually activated access to more dialectal and more standard variants of words.

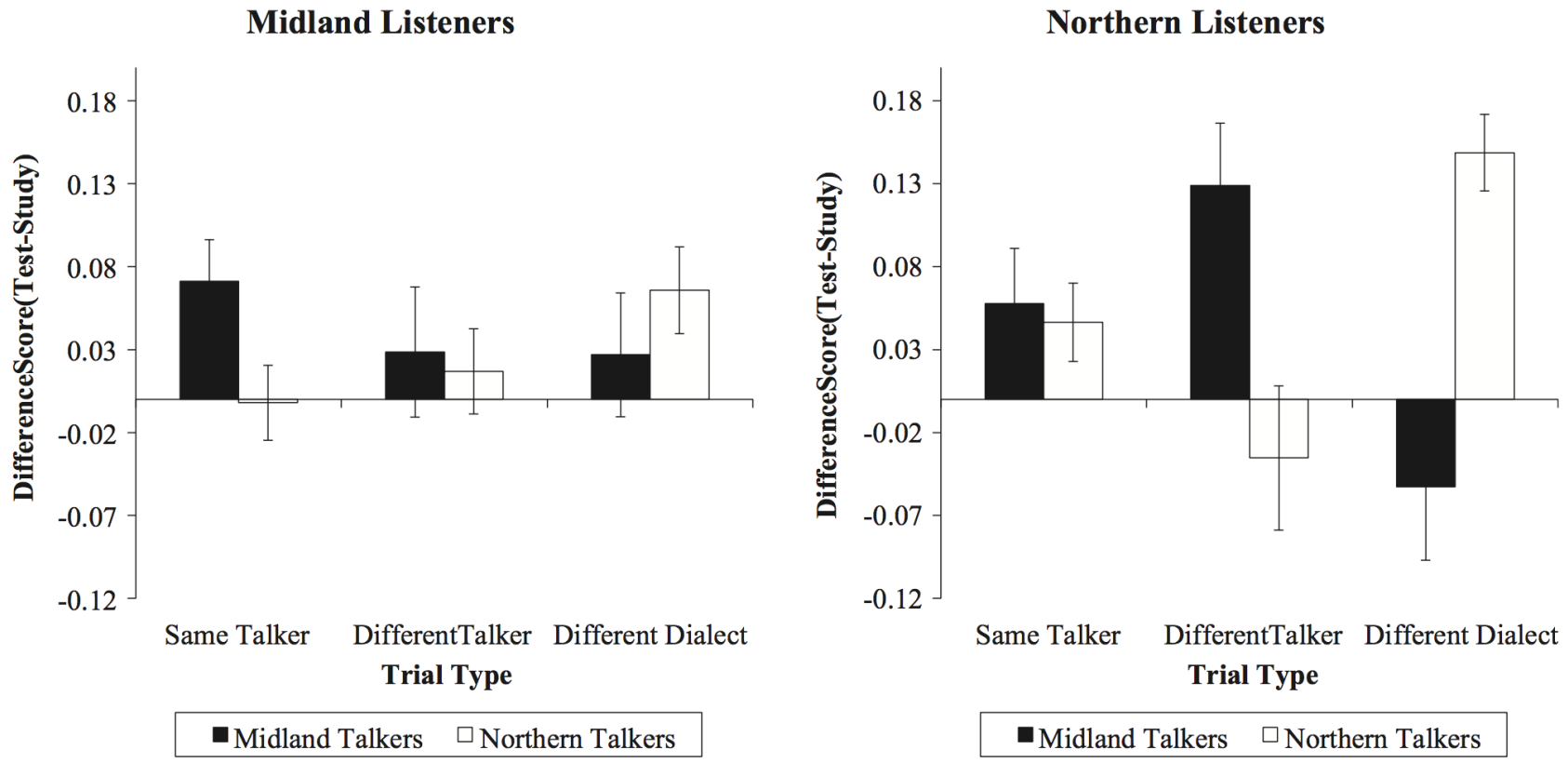
Clopper, Tamati, Pierrehumbert, 2016

- Recognition memory experiment.
  - Explicit recognition memory: Did you hear this before?
  - Implicit recognition memory: Does hearing the word before help word identification?
- Midland vs Northern speakers and listeners. Tested in Midlands (Ohio).
- Issue: How does relevant standard (Midland) vs nonstandard (Northern) dialect affect word recognition, encoding, and memory?

Results

- Northern dialect was slightly more intelligible.
- Explicit recognition accuracy was the same.
- Reaction times for explicit recognition task were SLOWER for Northern dialect for all listeners (whether Midlands or Northern speakers).
- Midlands speech produced an implicit recognition benefit. Northern dialect speech did NOT produce any such benefit.

Results plotted by listener and by talker-type in test word.



Conclusion: Northern words were recognized when produced. But they were not stored as examples of the same word in memory. Why not? Because they were deviant examples of the words, in relation to the contextually relevant standard.

Learning: A dialect imitation study. (German et al. 2013)

Dr. Alistair McGowan, Univ. of Glasgow.

Can Northwestern undergrads learn to talk like Alistair? Will they generalize to novel words? Will they remember how later?

Four 12-item target blocks. Single target, final position

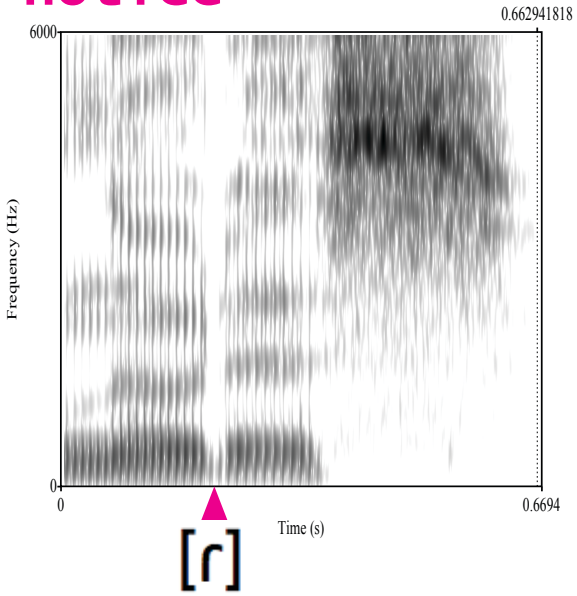
- /t/, strong position: He gave away his only token.
- /t/, weak position: The damp wind made him all sweaty.
- /r/, strong position: All the family's belongings lay beneath the rubble.
- /r/, weak position: The boy swallowed mud because he was curious.

Three 12-item no-target blocks (used for re-familiarization)

- In the piano division, the champion was Michael Hawley.
- The second copy of the code has many bugs.

S4: Baseline

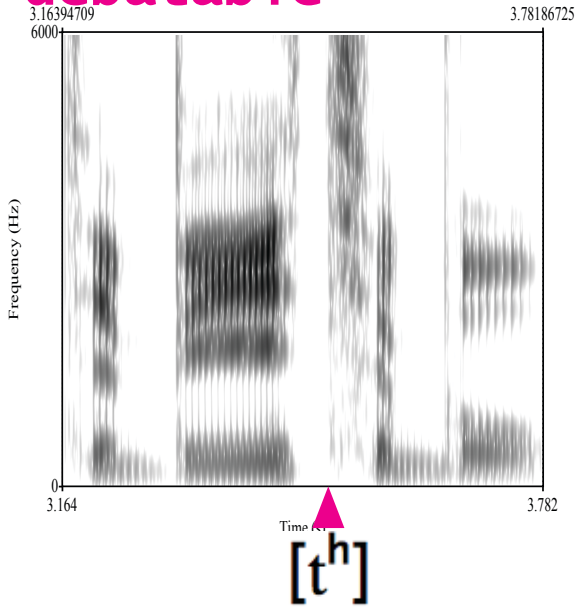
notice



[ɾ]

Alistair:

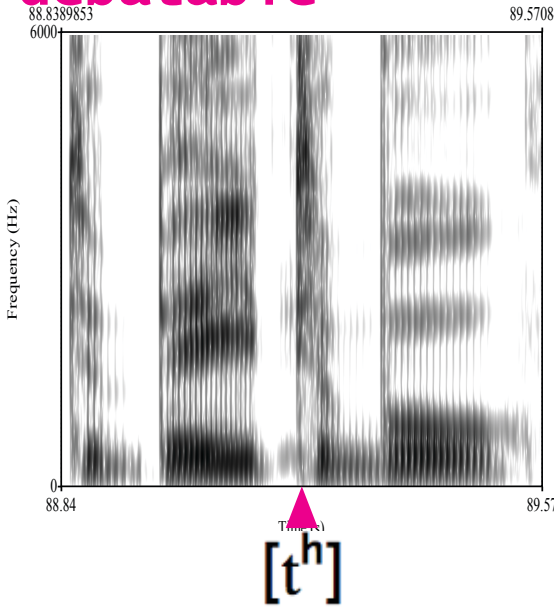
debatable



[tʰ]

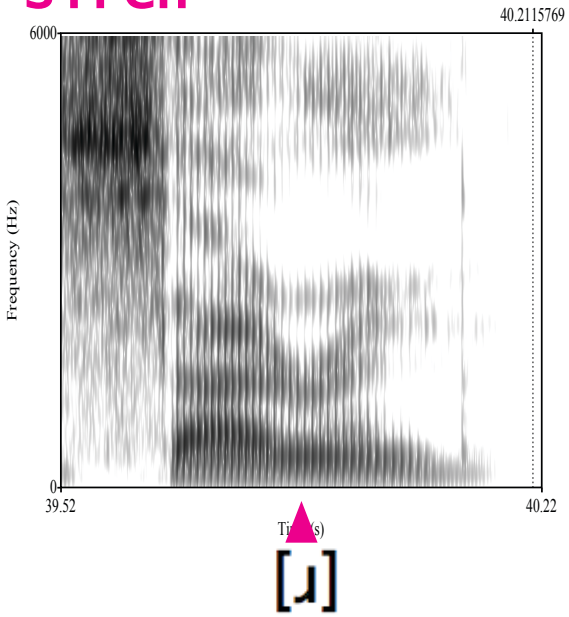
S4: Imitation

debatable



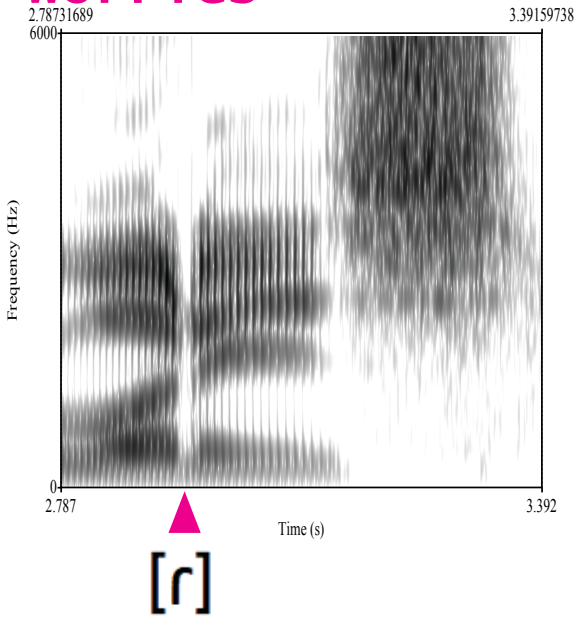
[tʰ]

siren



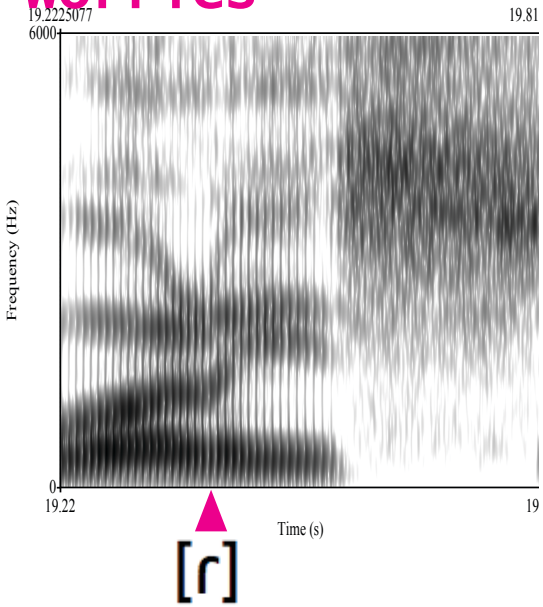
[ɹ]

worries



[ɾ]

worries



[ɾ]

Protocol

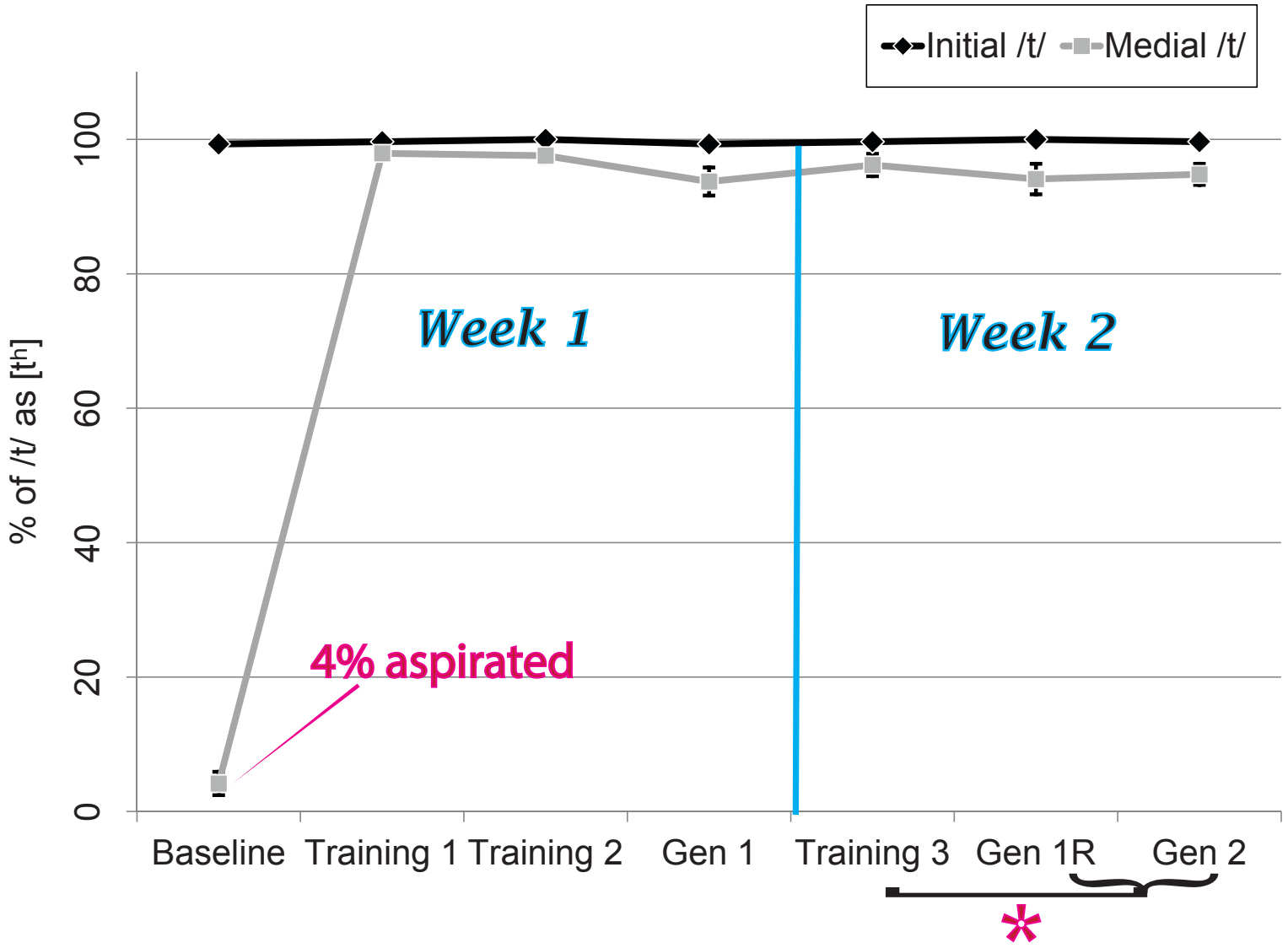
- Week 1: Record Baseline. Training. Repeat training.
  Test generalization.

- Week 2: No-target blocks alternating with test blocks:
  Retest training. Retest generalization. New generalization.

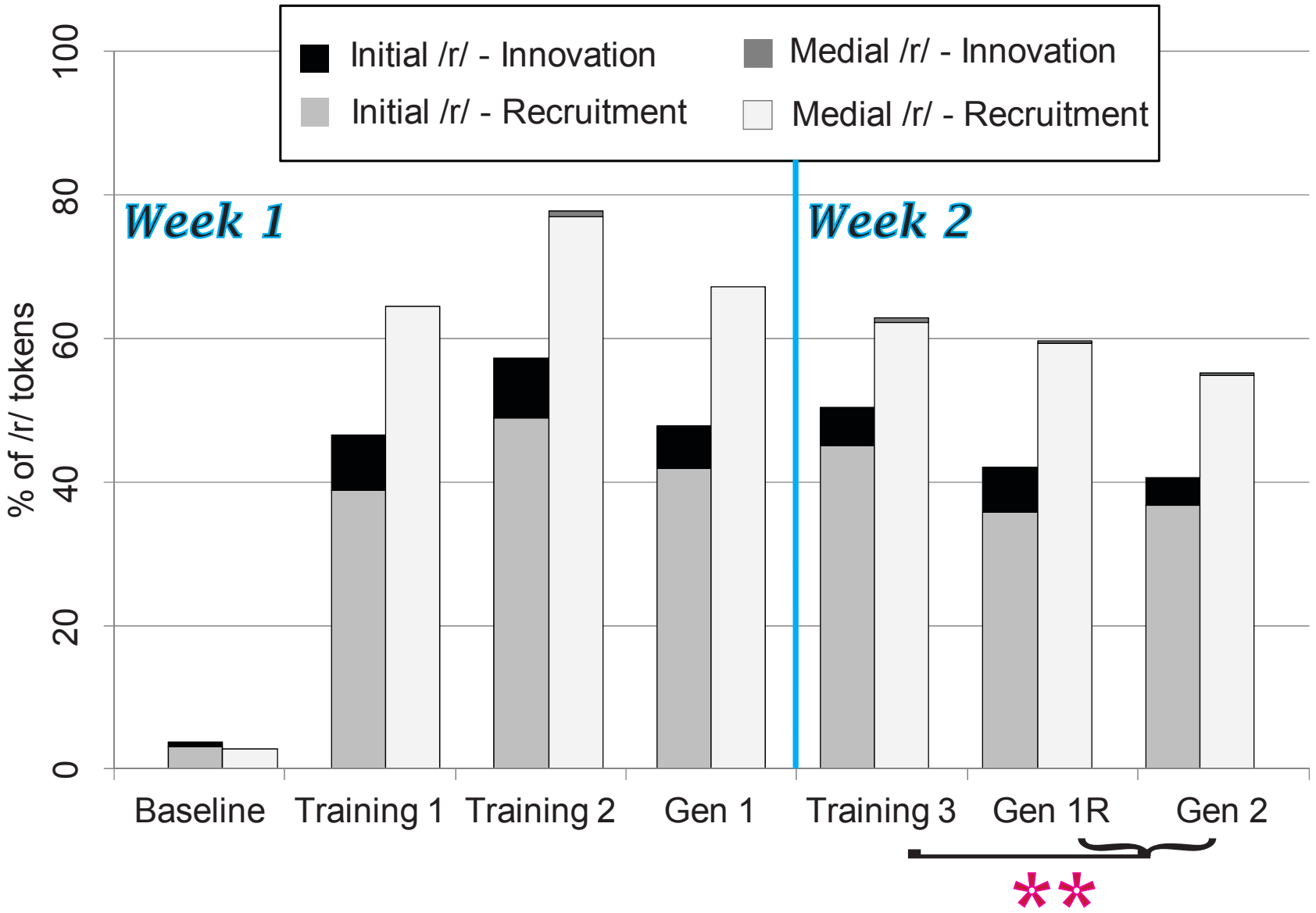
- 24 participants, monolingual NU undergrads. Blocks counterbalanced.

- Data coding: Hand coding plus discriminant analysis on F3 (which is the primary acoustic correlate of the ”American /r/”).

Results for /t/



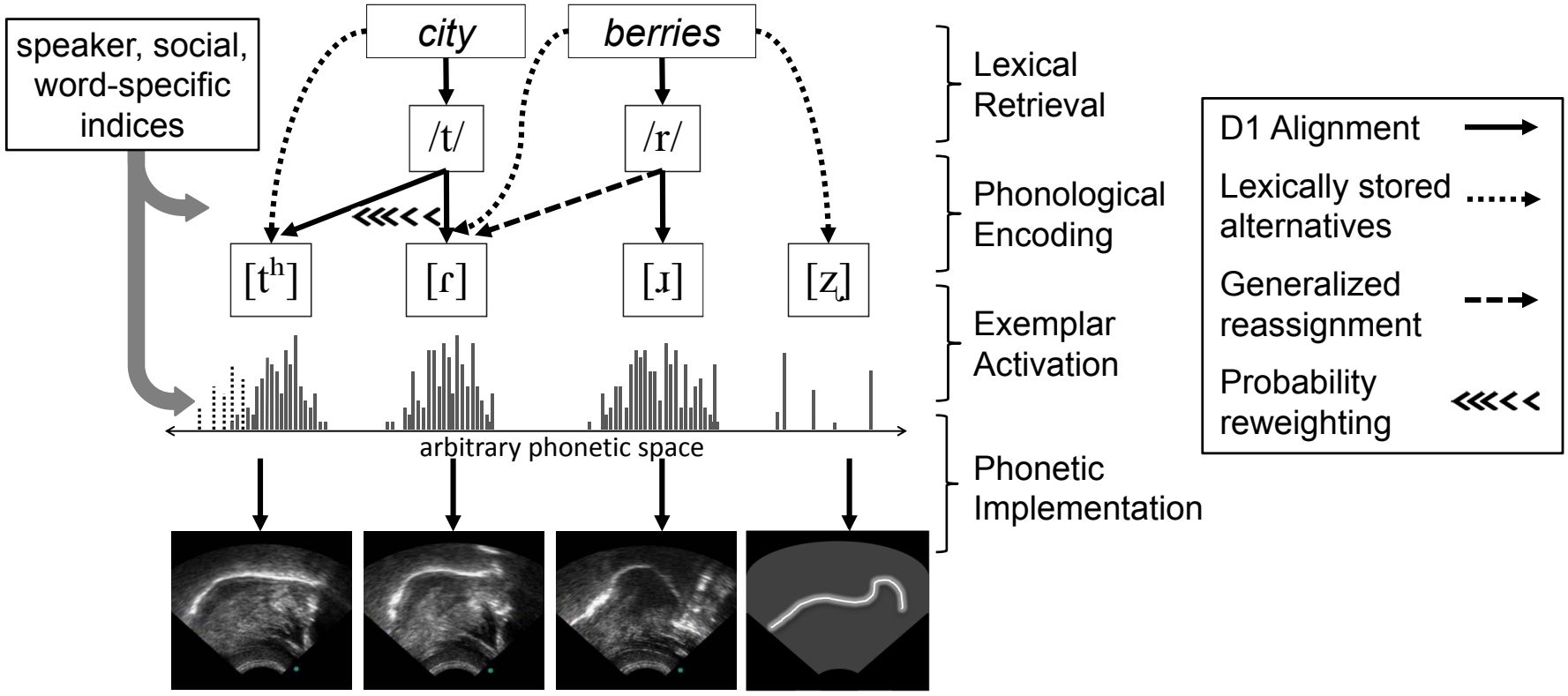
Results for /r/



Summary

- Subjects were extremely successful at raising the probability of an allophone [t^h] that they already knew.
- Transfer of the flap allophone to /r/ occurred rapidly, generalized to new materials, and was retained a week later.
- However, some subjects had difficulty with the initial flap and/or attempted to create a new phoneme.
- Words in the training set were significantly, if slightly, better than generalization words.
- This shows 1-week persistence of memories of specific words, in relation to Alistair's dialect.

Minimal model structure, production side only



Summary.

- Easy: Generalize from examples by increasing the probability of a known allophone.
- Somewhat harder: Associate a known word with an allophone normally used to realize a different phoneme.
- Even harder: Remap the allophonic realization of a phoneme generally.
- Very hard: Learn a new phoneme with a new phonetic pattern.
- Abstract categorical learning is fast and easy, in comparison to phonetic learning.
- Word specific effects are weak in comparison with general learning.

WORDOVATORS

Home | About | People | Publications | Jobs | Games

Welcome to Wordovators

How People Make New Words

Play Games

The Wordovators team. Front row (from left): Christoph Bartneck (NZILBB), Janet B. Pierrehumbert (Northwestern), Jen Hay (NZILBB), Stephanie Stokes (NZILBB). Back row (from left): Chun-Liang Chan (Northwestern), Kayo Takasugi (Visual Voice), Aristotle, Kiwi PaPeRo. Photo credit: University of Canterbury

Announcements

- Postdoctoral Fellow: Northwestern University, Evanston IL

News

- Northwestern press release
- Wordovators featured in Northwestern Annual Research Report 2012
- Kick Off meeting

Sponsors

JOHN TEMPLETON FOUNDATION

Lexical associations w gender : (Needle et al., Workshop on Morphological Typology and Linguistic Cognition, 2017).

More male, more female and neutral words (BNC corpus).

Male	Female	Neutral
platoon	drizzle	cucumber
fieldwork	pillowcase	guidebook
gangster	foundling	droplet

More male, more female and neutral suffixes (BNC corpus).

Male	Female	Neutral
-point , -hold	-woman, -room	-book, -stone
-ship, -ry	-ette, -ish	-ery, -ically

Experimental design

Gender association task for aligned vs non-aligned whole-word and affix cues.

- Male word w male head/affix: checkpoint, internment.
- Male word w female head/affix: gearbox, skewness.
- Female word w male head/affix: fairyland, devilry.
- Female word w female head/affix: pillowcase, wakeful.
- And so forth for neutral items ... 3 x 3 design = 9 types of complex words+ simplex baseline items.

User interface

Provide morphological judgment. Guess whose blog had the word in it.

Results

- People's judgments reflected statistical associations with gender.
- The associations of both the whole word, and the head/affix, were significant in a mixed-effect model.

Learning socio-indexical associations

Guess the right diminutive form for each word in a roof-jumping game.

(Racz, Hay and Pierrehumbert, 2017).

Game play

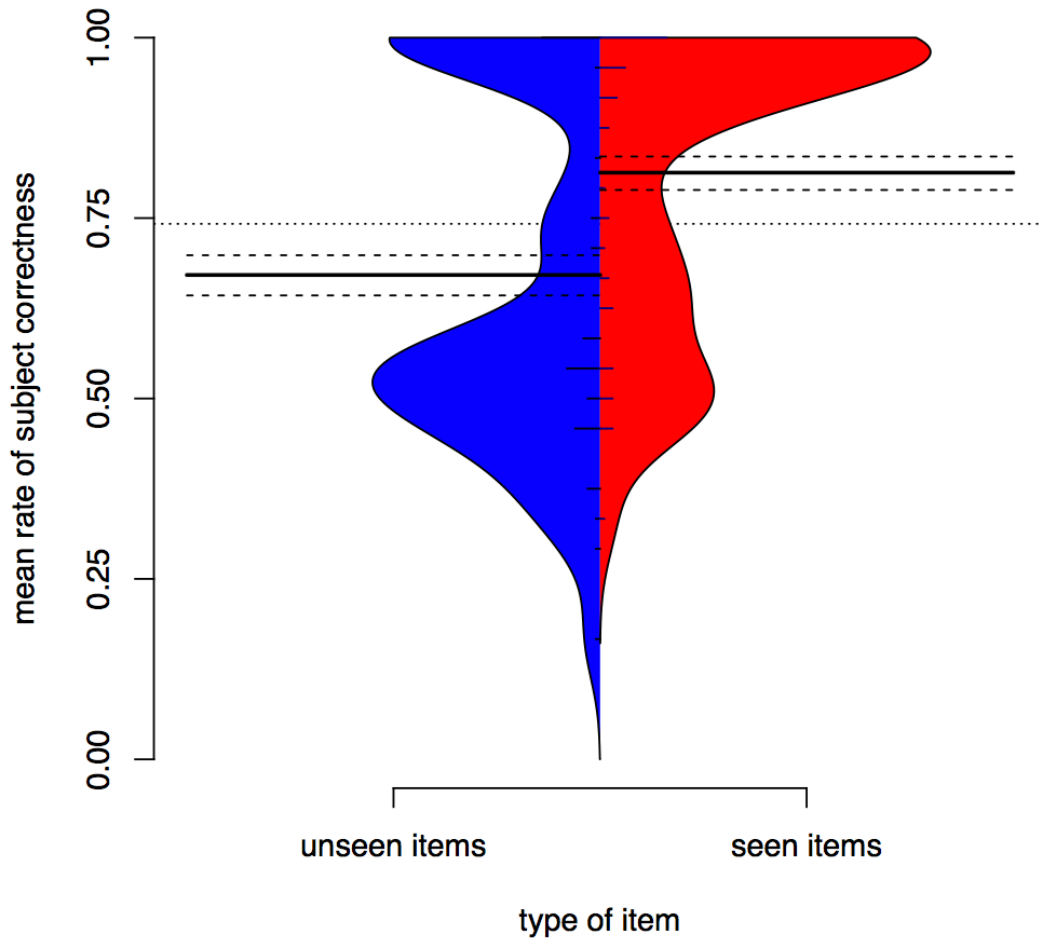
- The correct answer depends on some characteristic of the context
- If you guess the right word, you jump forward to the next roof.
- If you guess wrong, the interlocutor pushes you off your roof, you have to flutter up again.
- Test phase includes previously scene and novel items, no feedback.

Test phase is in the dark

Baseline: Allomorphy rule based on vowel stem

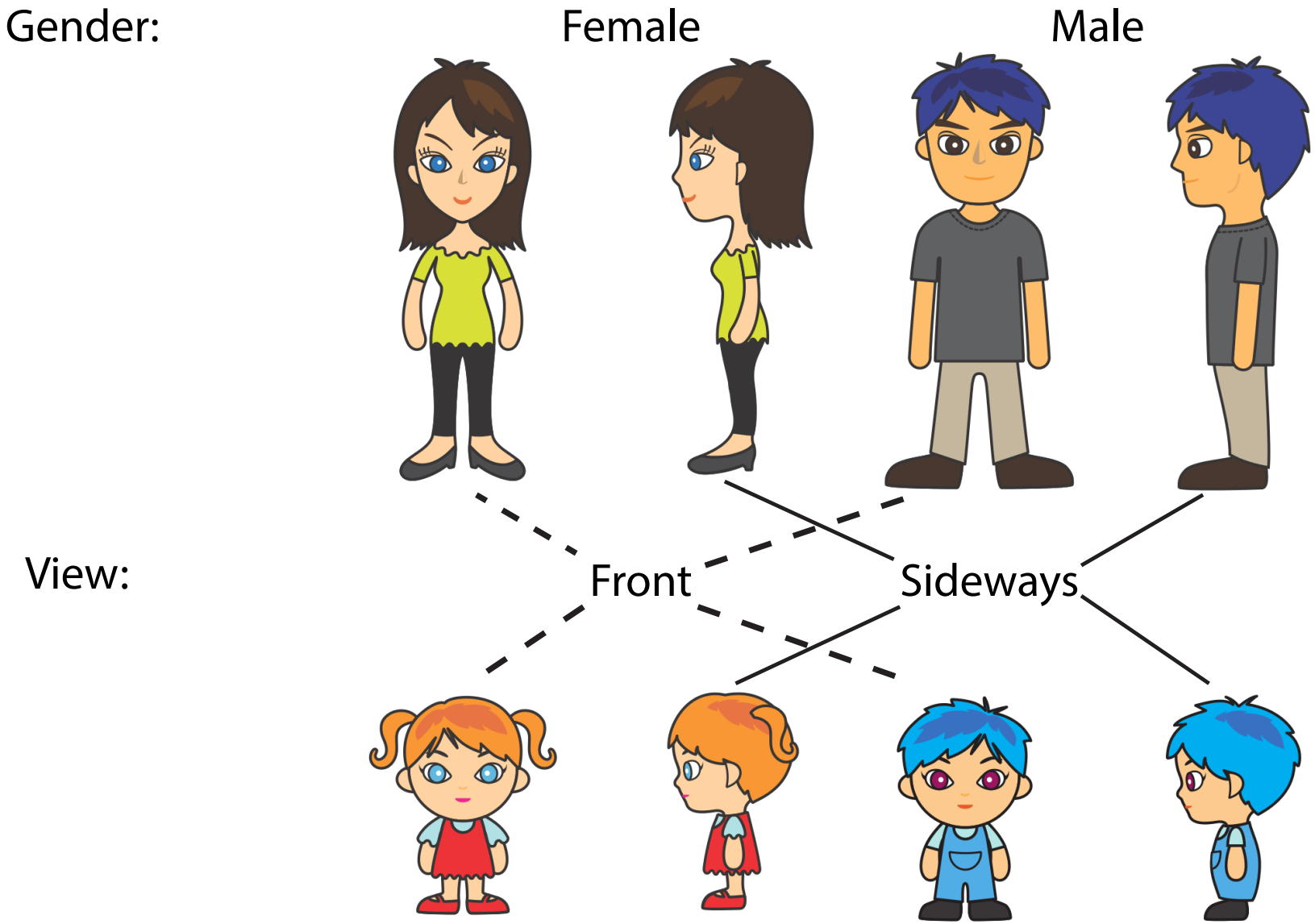
- Half the CVC syllables have the vowel /a/.
- Half the syllables have vowel /e/.
- Correct answer uses the suffix that has the same vowel as the stem.
- This is a reasonably easy example of phonological conditioning.

Bean plot of rate correct in test: seen vs unseen items



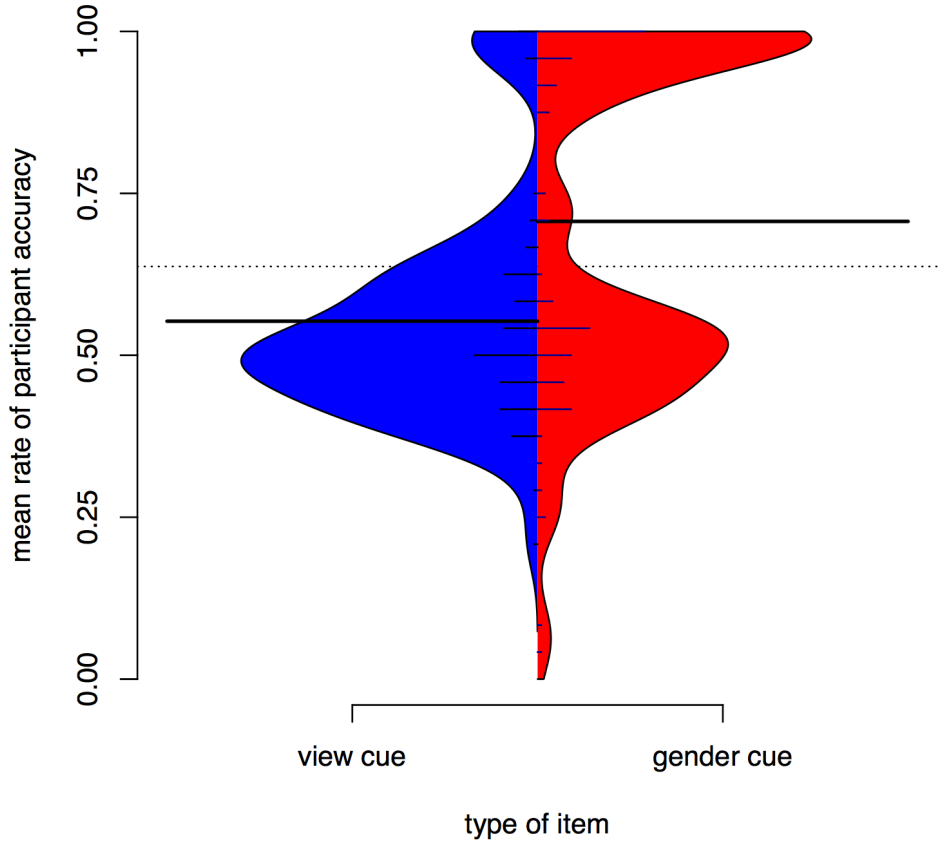
Bimodal distributions: Good vs poor learners. Moderate advantage for previously seen items over unseen items.

What about social context? Same statistical pattern, correct answer depends on interlocutor

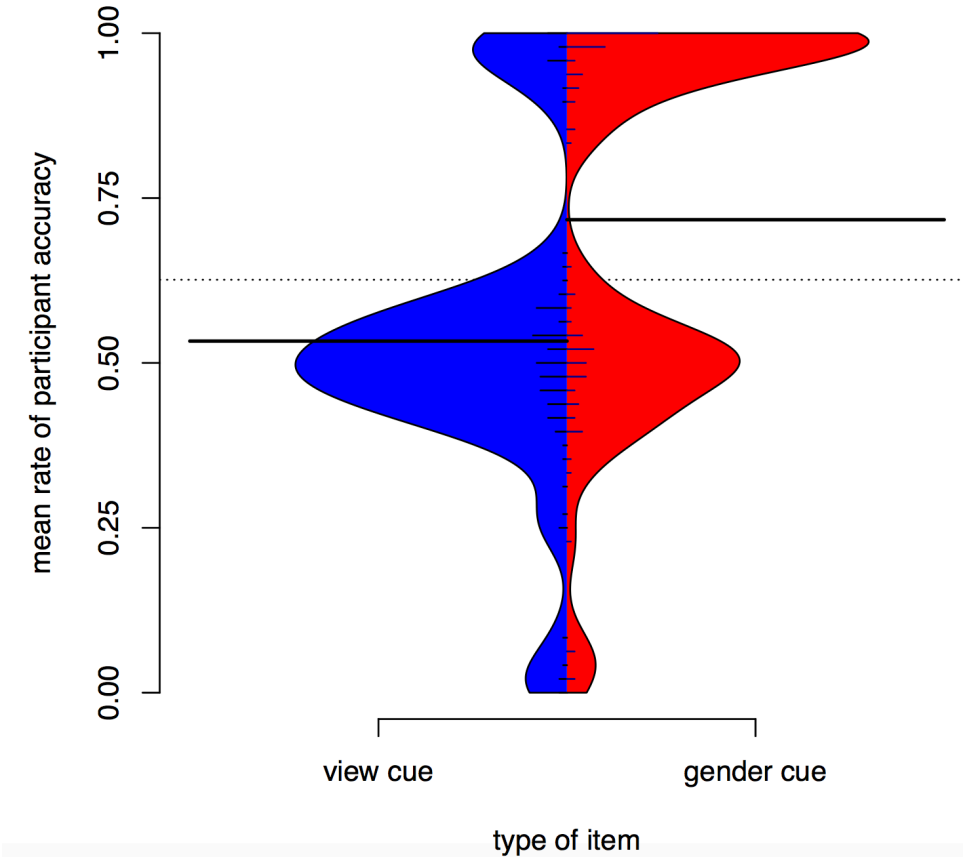


Contextual cues: Gender vs View

Results for one female vs one male interlocutor



Results generalizing to a new female vs male.



To summarize

- Gender cue is about as learnable as the phonological cue.
- No advantage for seen items (advantage for seen people not fully explored).
- The view cue is not very learnable at all. Why? It is not normally relevant for language.
- Outcomes are quite bimodal, something we are seeing repeatedly. (see CogSci 2017 for modelling).
- Points to individual differences in people's ability to focus on the relevant cue and form a generalization.

Additional statistical patterns

- Training length also reflects relative difficulty of conditions.
- Older people do significantly better. Relates to findings by Metcalfe et al that older adults are somewhat better at assimilating feedback.
- General effects replicated for an experiment on a novel Plural.
- Followup shows ethnicity and age indexical associations also quite learnable.

Conclusions

- Indexical associations are quite learnable.
- Reusing known categories is the easiest.
- Statistical learning is influenced by social relevance and by individual cognitive characteristics.

Thank you.

References

Altmann, E.G, J.B. Pierrehumbert, and A.E. Motter, Niche as a determinant of word fate in online groups, PLoS One 6(5), e19009 (2011). doi:10.1371/journal.pone.0019009

Clopper, C, Tamati, T.N. and Pierrehumbert, J. B. (2016) Variation in the strength of lexical encoding across dialects. J. Phonetics 58, 87–103.

Clopper, C. and J. Pierrehumbert (2008) Effects of semantic predictability and regional dialect on vowel space reduction. J. Acoust. Soc. Am. 124(3) 1682-1688.

German, J.S, Carlson, K. and Pierrehumbert, J.B. (2013) Reassignment of consonant allophones in rapid dialect acquisition. J. Phonetics 41 (3), 228-248.

Needle, J, Pierrehumbert, JB and Hay, JB (2017) Gendered associations of English morphology. Workshop on Morphological Typology and Linguistic Cognition. LSA Summer Institute.

Pierrehumbert, J.B., T. Bent, B. Munson, A. Bradlow, J. M. Bailey (2004) The influence of sexual orientation on vowel production. J. Acoust. Soc. Am. 116(4), 1905-1908.

Racz, P., Hay, J.B. and Pierrehumbert, J.B. (2017) Social Salience Discriminates Learnability of Contextual Cues in an Artificial Language Frontiers in Psychology (8) DOI: 10.3389/fpsyg.2017.00051.

Schumacher, R.A and Pierrehumbert, J.B. (2017) Prior Expectations in Language Learning: A Stochastic Model of Individual Differences, CogSci 2017, London