50 Years of Linguistics at MIT, Dec 10, 2011

Representation in Phonology

Janet Pierrehumbert, Northwestern University

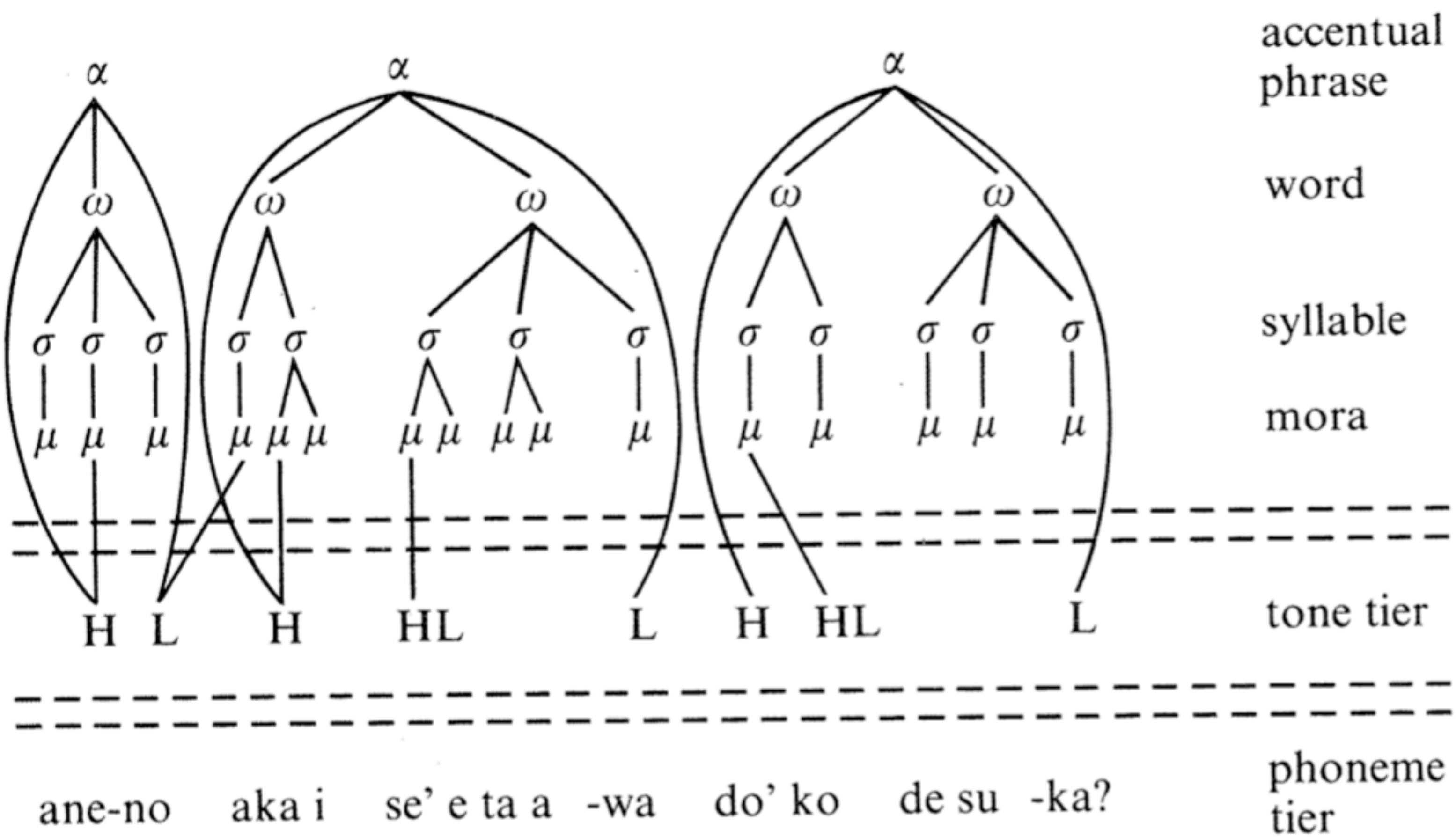
Roadmap

- General properties of autosegmental-metrical representation.
- Intonation as a case study.
- Phonetic realization of segments.
- Current questions.

The syllable structure of *plain*

(according to moraic theory).

Tone associations for a sentence in Japanese



(Pierrehumbert & Beckman 1988, building on work by Leben, Goldsmith, Liberman, Mc-Carthy, and others).

Key features

- Hierarchical structure with no recursion.
- Temporal interpretation from left (first) to right (last).
- Integrates lexical and phrasal phonological structures.
- Not a tree. Directed acyclic graph. Terminal nodes can have more than one parent (due to multiple autosegmental links).
- Not isomorphic to the syntactic tree.
  - Prosodic structures can cross-cut syntactic structures e.g. *He'd # like to*.
  - Alternative intonational phrasing e.g. *Don't (%) go!*

Motivation

- Capture and elucidate the interaction of constraints on sound structure at different time scales.
- Licensing:
  - Supersedes the *phonemic principle* with a general theory of contrasts.
  - Each node in the structure licenses a choice (or set of choices) amongst the elements that are contrastive at the temporal scale for the node.
- Contrasts can be lexical, syntactic, semantic, or pragmatic.

Movation, cont’d

- Statistical interpretation of licensing.
  - Items that are adjacent in the representation exhibit strong correlations (whether positive or negative).
  - Items that are not adjacent are (more) independent.
  - Implicit categorization of statistics into ”structural” and ”ignorable” (compare Jackendoff, Yang).
- Phonetic realization: Model how elements licensed at different scales interact to determine the phonetic outcome.
  - Launchpad for laboratory phonology (1980’s onward).
  - Implicit categorization of phonetic phenomena into ”phonological” and ”implementational”.

Intonation as part of the linguistic system

Primary marking of yes-no questions, in contrast to declaratives in various Niger-Congo languages. (Partial data from Rialland, 2007)

Language Group	Register shift	H tone	L tone
Atlantic	2	1	
Kru		1	3
Gur		1	13
Benue-Congo Bantu	6	4	

Intonation patterns are not universal. They are learned.

Phonetic realization: *L*+H L !H %* in Greek (Arvaniti, 2007)

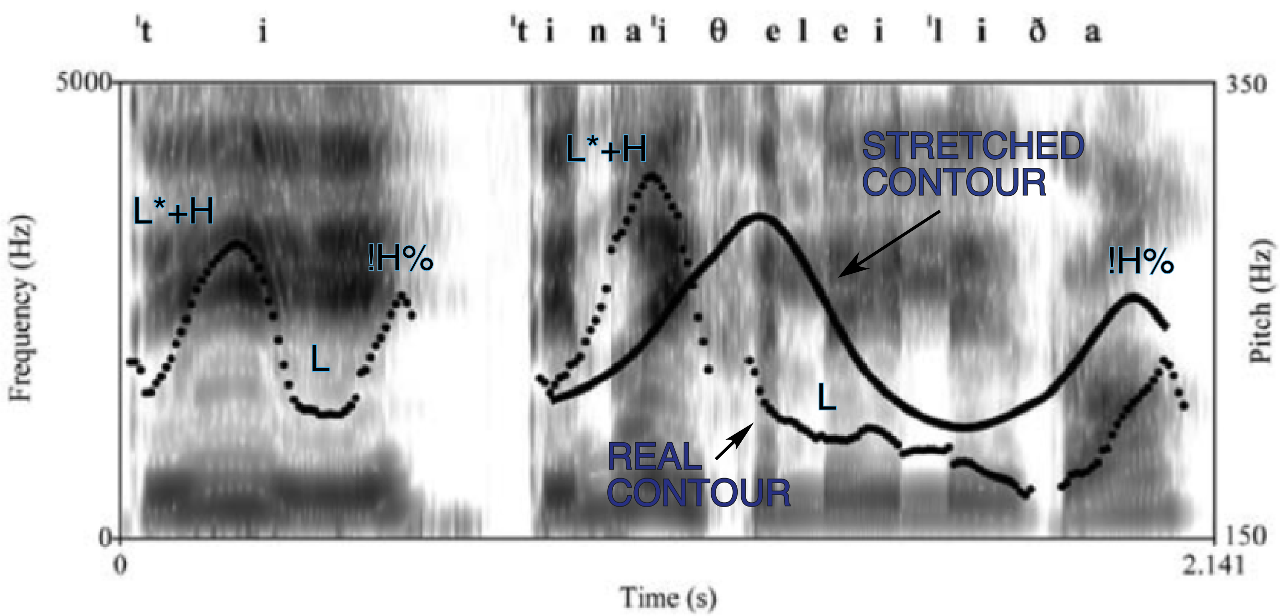
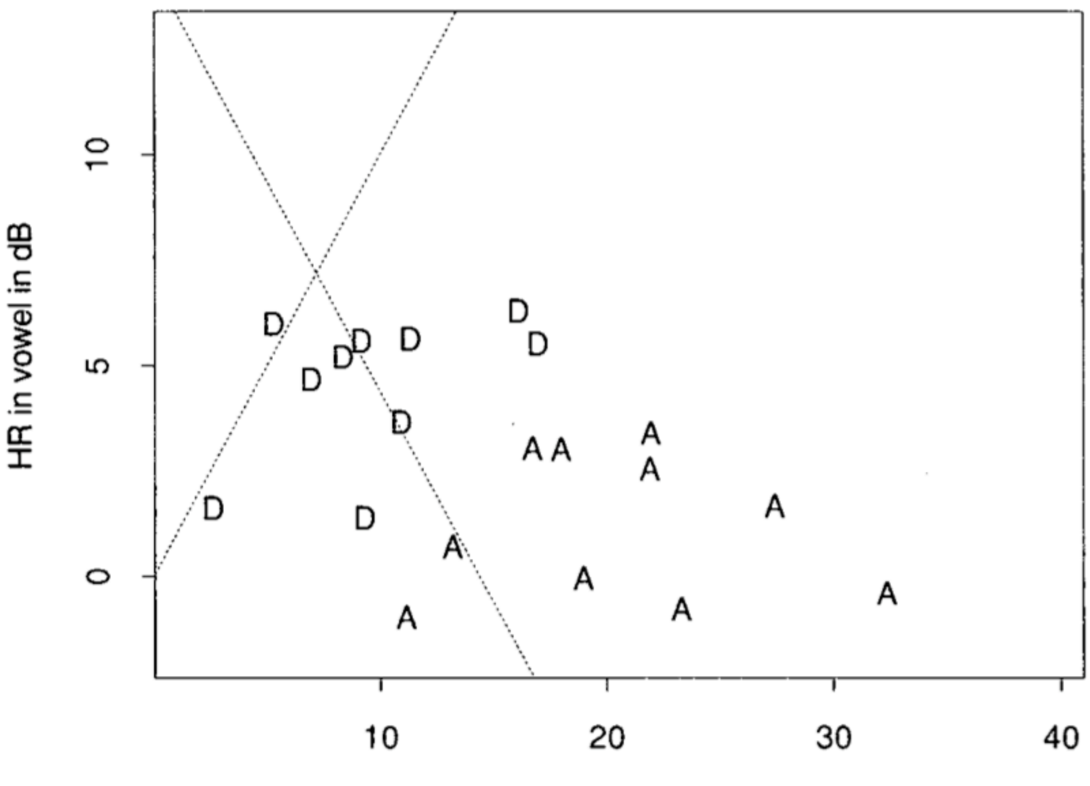


Figure 12. Spectrograms and F0 contours of [ˈti] "what?" and [ˈti na ˈiθele i ˈliða] "what could Lida have wanted?" The thick solid line represents the contour of [ˈti] stretched out to fit the duration of the longer question; as is evident, this contour is not the same as the original contour of the longer question which is represented by the dotted line.

Some points about phonetic realization to take away

- Tonal elements are in the mind. F0 contours are in the physical world.
- Tonal elements are sparsely specified. F0 contours are continously specified, for all voiced regions of the speech signal.
- The realization of the same element varies as a function of the rest of the structure.
- No prosody and intonation = no speech! Without a well-formed autosegmental-metrical structure, no f0 specifications can be generated!

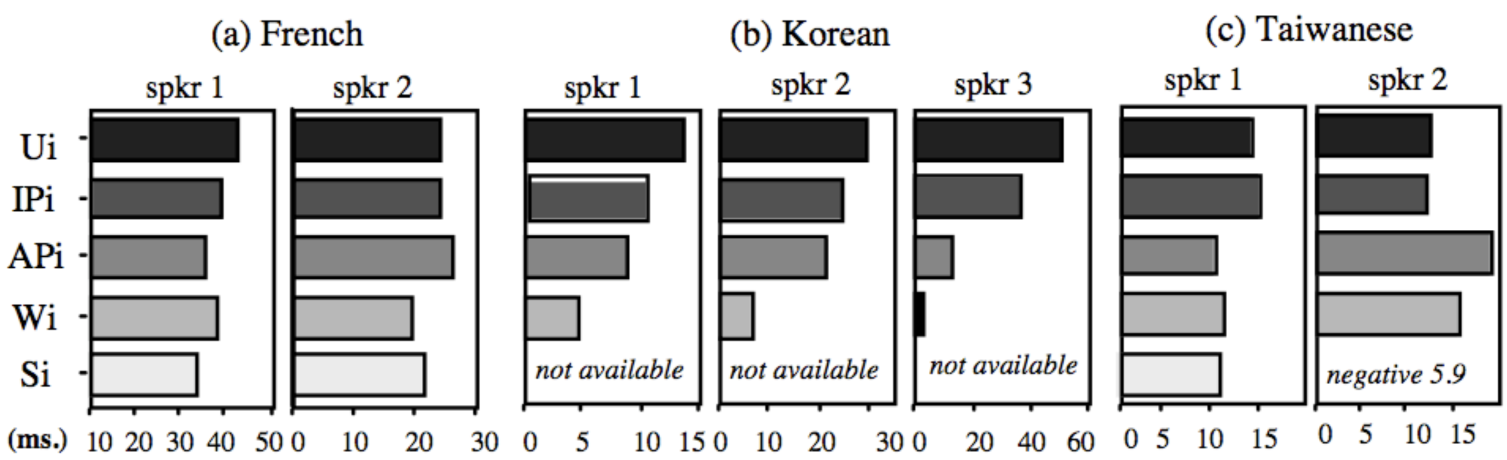
Phrasal prosody also affects segments: Breathiness of /h/ vs V for *hog-farmer*, *hawkweed*. Accented (A). Deaccented (D). Pierrehumbert & Talkin, 1993.



A accented in questions; D deaccented; lines: y = x and y = −x + b

Prosodic effects are pervasive and language-specific

Ex: Voice onset time as a function of strength of prosodic boundary in three languages (Keating et al. 2004).



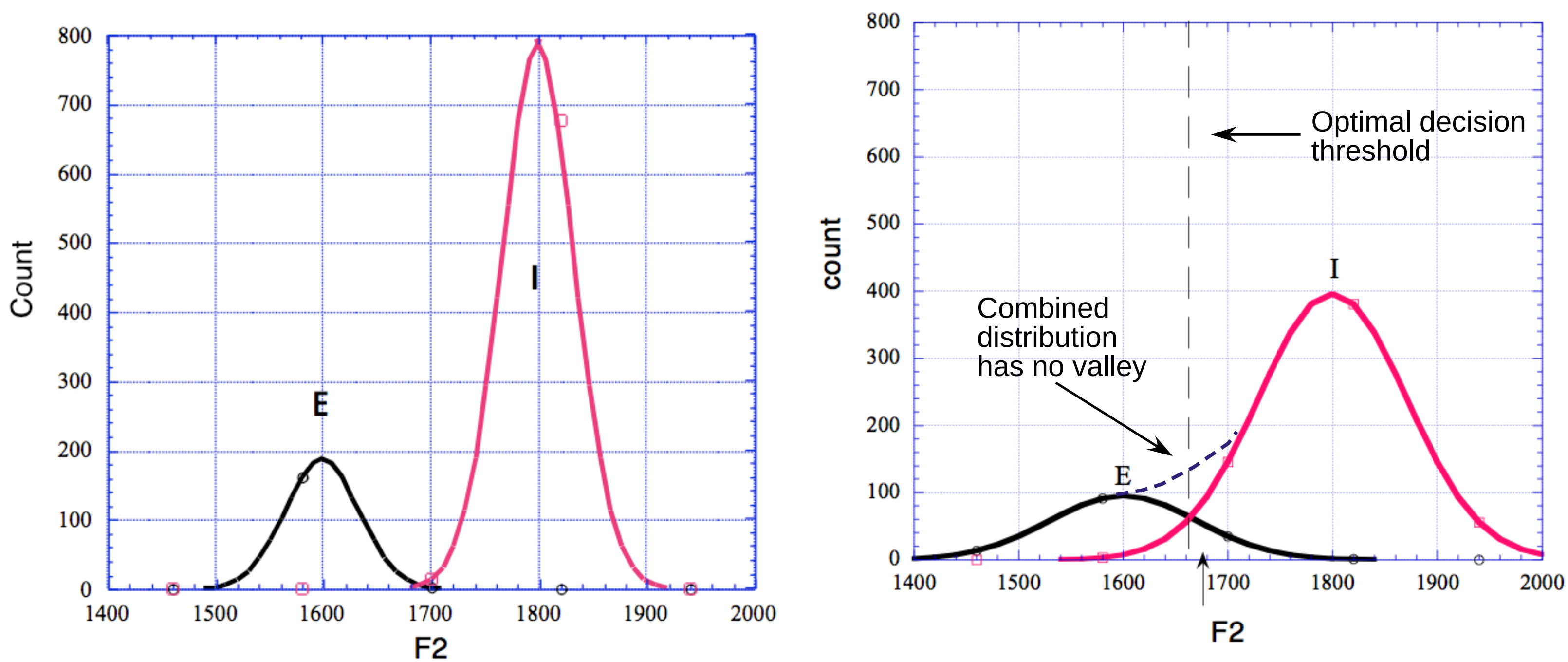
In Korean, like English, VOTs are longer at stronger prosodic boundaries. This pattern is much weaker in French and Taiwanese.

Successes of autosegmental-metrical phonology

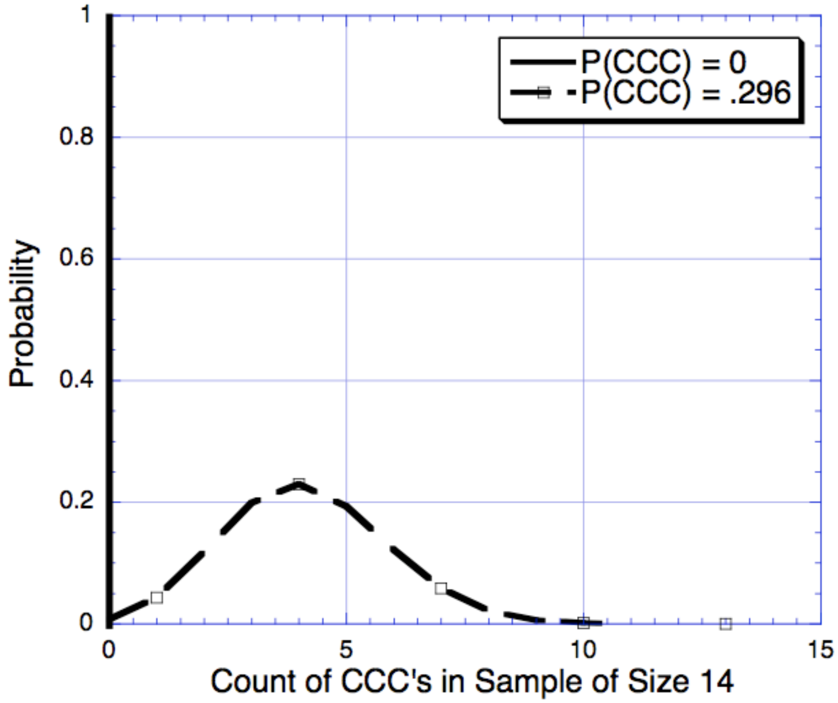
- Vast and illuminating typological literature.
- Explicit generative models of the relation of mental structures to physical reality.
- Cross-disciplinary impact in speech engineering, psycholinguistics, speech and hearing sciences.

Questions around bad categories.
(Bad = significantly different but poorly discriminable).

Left: Learnable with a very general statistical prior. Right: Learnable only with specific prior assumptions. Optimal classification is errorful.



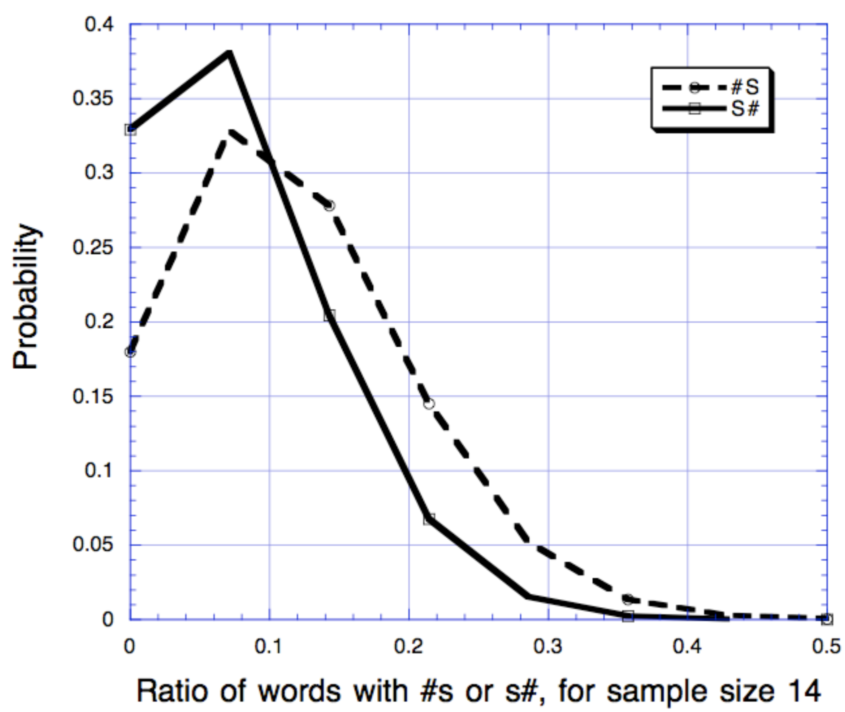
Constraints as statistical categories:
Learnability of *CCC constraint on English wordforms.

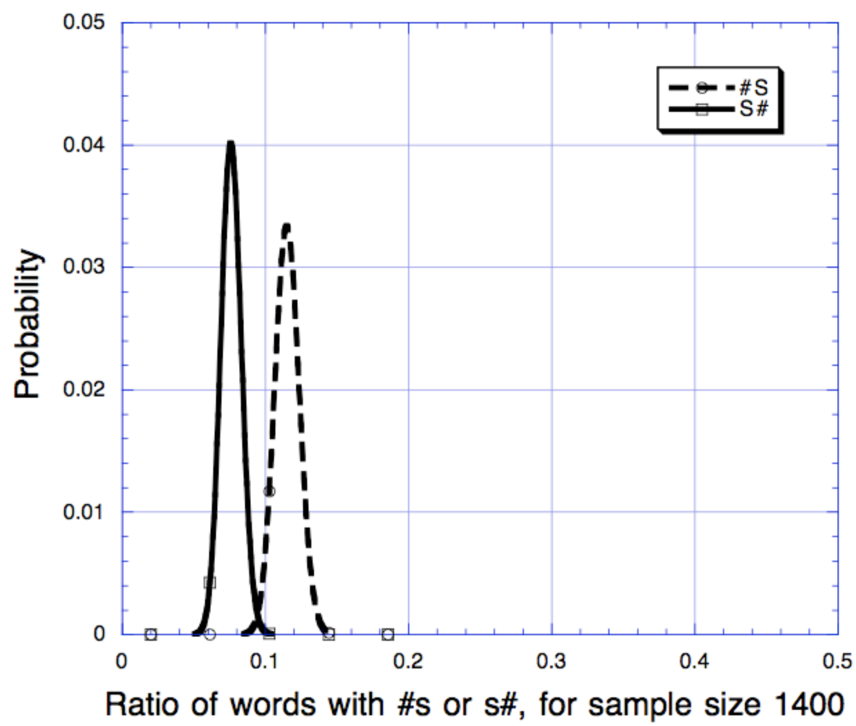


Knowing 14 triphonemic words excludes CCC words at $P < 0.01$.

A marginal English constraint: /s/ is better in word-initial position ($P = 0.11$) than word final position ($P = 0.077$).

Distributions for fraction of words with initial and final /s/ in random samples of various sizes.





Examples of bad categories. Phrasal effects on segments.

- /h/ is just a devoiced vowel.
- Vowels in some phrasal contexts are as breathy as /h/ in other phrasal contexts.
- Parsing the speech stream involves a joint decision about descriptors at different scales.
  - X is an /h/ and it is medial in a deaccented word (*NEAR Omaha, not IN Omaha*).
  - XOR X is a vowel in a phrase-initial unstressed syllable (*Toledo is ...*)

People accomplish this factoring both in learning and in processing. How, exactly? Overlapping effects at different time scales remain a challenge for statistical learning theory.

Second occurrence focus (Beaver et al., 2007).

- *... the defense attorney only named SID in court today. Even the STATE PROSECUTOR only named Sid in court today*.
- The second focus on *Sid* is marked by the phrasal stress pattern but has no pitch accent.
- The focus creates slight phonetic differences. But these are not enough to discriminate focused from non-focused tokens.
- Second occurrence focus is not learnable per se. It could be learned by generalizing applicable features of first-occurrence focus.
- AM representations are challenged by gradient phenomena. But they are also needed to explain how some patterns generalize to others.

Heterogeneity amongst individuals in production and/or perception

- Vowel changes in progress: Older speakers may produce vowel quality contrasts that they cannot reliably perceive (Labov). Younger speakers may perceive vowel contrasts that they don't produce (Hay et al).
- $\approx 10\%$ of English speakers are deaf to lexical stress (Nakatani et al.)
- Similar individual variation in phonological marking of focus and perception of focus (Coburn).
- Expect the linguistic system to stabilize at the population level on representations that are produced and perceived with low error rates. (Wiener, *Cybernetics*).
- How do people factor phonological and socio-indexical information? How do limits on this ability influence the equilibrium states of linguistic systems?

Thank you!