

“LEXICAL” STATISTICS FOR PROSODY: GOALS AND CHALLENGES

1 October 2015
3rd VARIAMU WORKSHOP
Laboratoire Parole et Langage

About SIRL

- 2+ years funded project
 - 2 postdocs x 2 years
 - 30k for coding, labeling, analysis
 - Workshop
- Explores implicit, socially-driven dynamic processes in production/perception
 - Production: Convergence-like effects that cannot be explained by accommodation (Mendoza-Denton et al., 2003; Delvaux & Soquet, 2007)
 - Perception: Dynamic changes in perception that cannot be explained by speaker normalization (Hay & Drager, 2010; Cutler et al., 2010)

About SIRL

- Exemplar theory
 - Social-contextual information (e.g., ethnicity, class, formality, etc.) does not just interact with cognitive representations of linguistic forms and rules, but is actually intrinsic to those representations (Goldinger, 1998; Pierrehumbert, 2001; Johnson 2006)
- Existing studies emphasize phonetic changes either within or between phoneme categories
 - Shift in boundary between /s/ - /ʃ/
 - Changes in the *phonetic value* of a vowel percept
 - Differences in the degree of diphthongization/monophthongization of a vowel phoneme

About SIRL

- Do the (social) predictions of exemplar-based models bear out for other types of linguistic representations?
 - Phonology, morphophonology, syntax, etc.
 - Esp. prosody and intonation

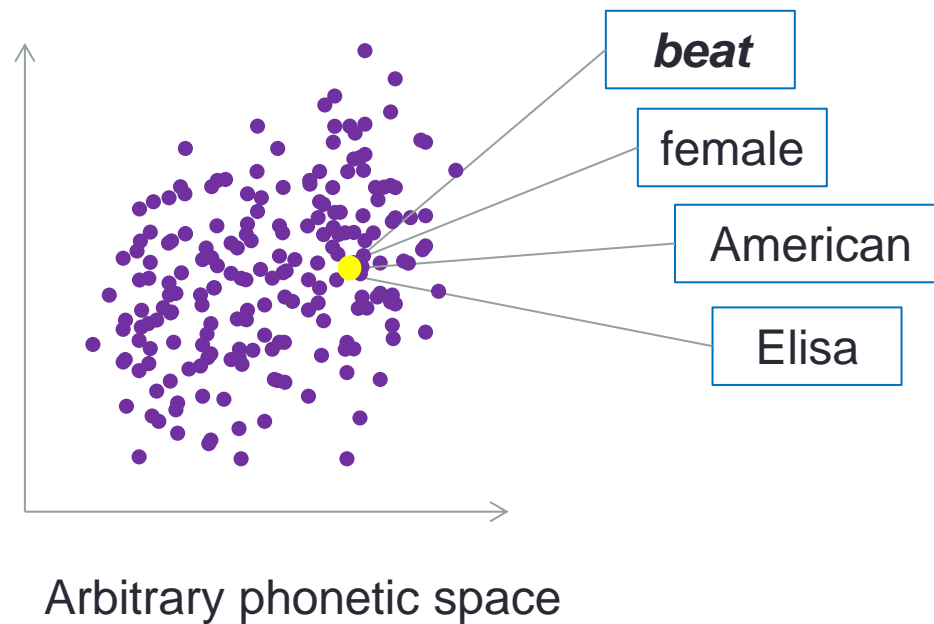
Corsican French Study

- w. Cristel Portes, Amandine Michelas
- A single prosodic contour is interpreted as a confirmation question in Corsican French and as an assertion in standard (continental) French
- A typical Corsican has significant experience with both varieties
- Theory predicts that the presence of an implicit “Corsican” conceptual prime should bias listeners towards a question interpretation, and similarly for an implicit “continental” conceptual prime

Exemplar theory

- Individual experiences of language use are stored in memory
 - Rich phonetic detail (virtual acoustic recording)
 - An index to the lexical wordform
 - Facts about the linguistic context the word occurred in (e.g., surrounding lexical items, phonological parse, syntactic parse)
 - Information about the speaker
 - Index to the speaker's identity (i.e., "Fred Smith")
 - Speaker's gender, age, ethnicity, nationality, socio-economic status, etc.
 - Information about the context
 - Background noise, color of the walls, etc.

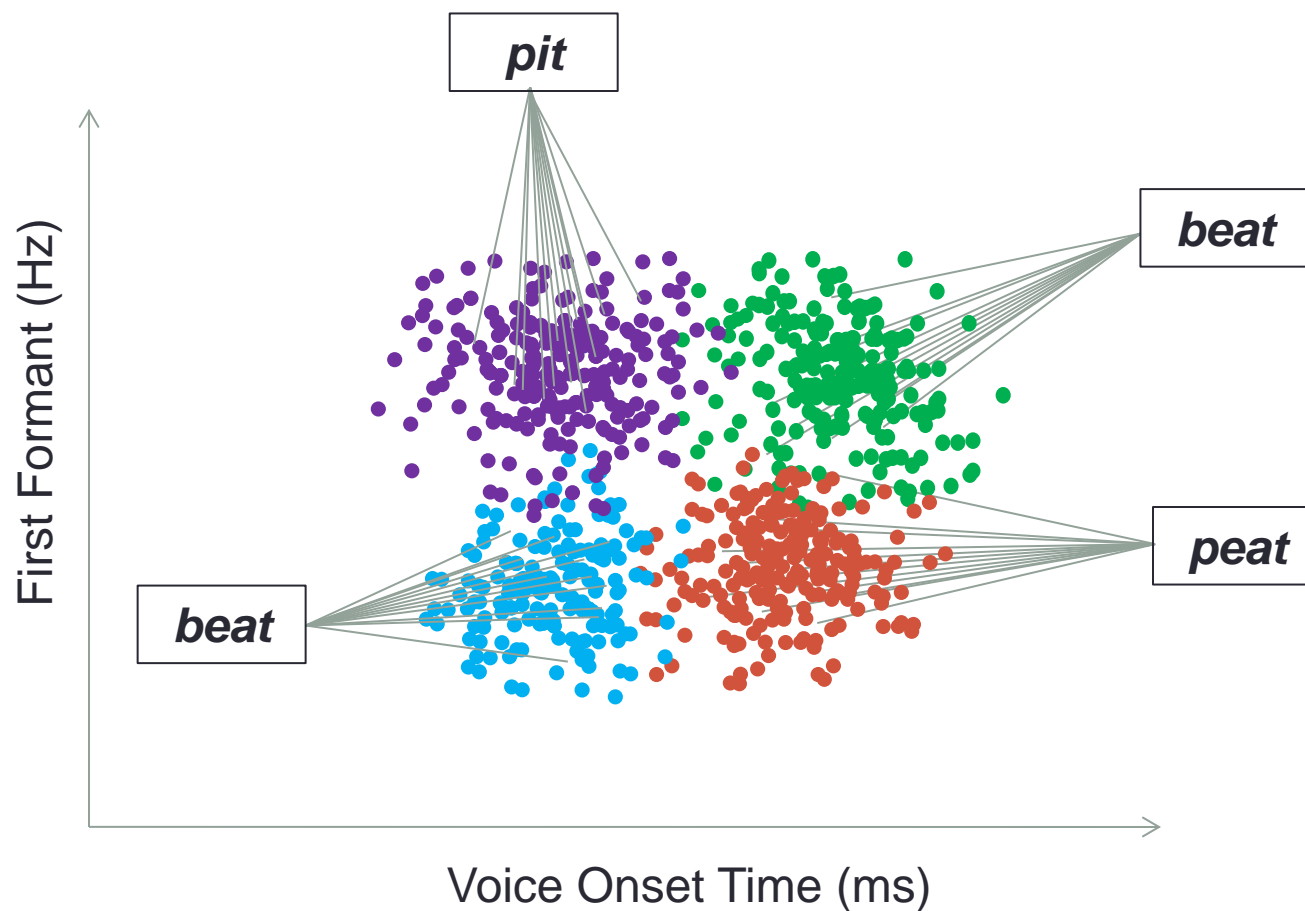
Exemplar/episodic representations



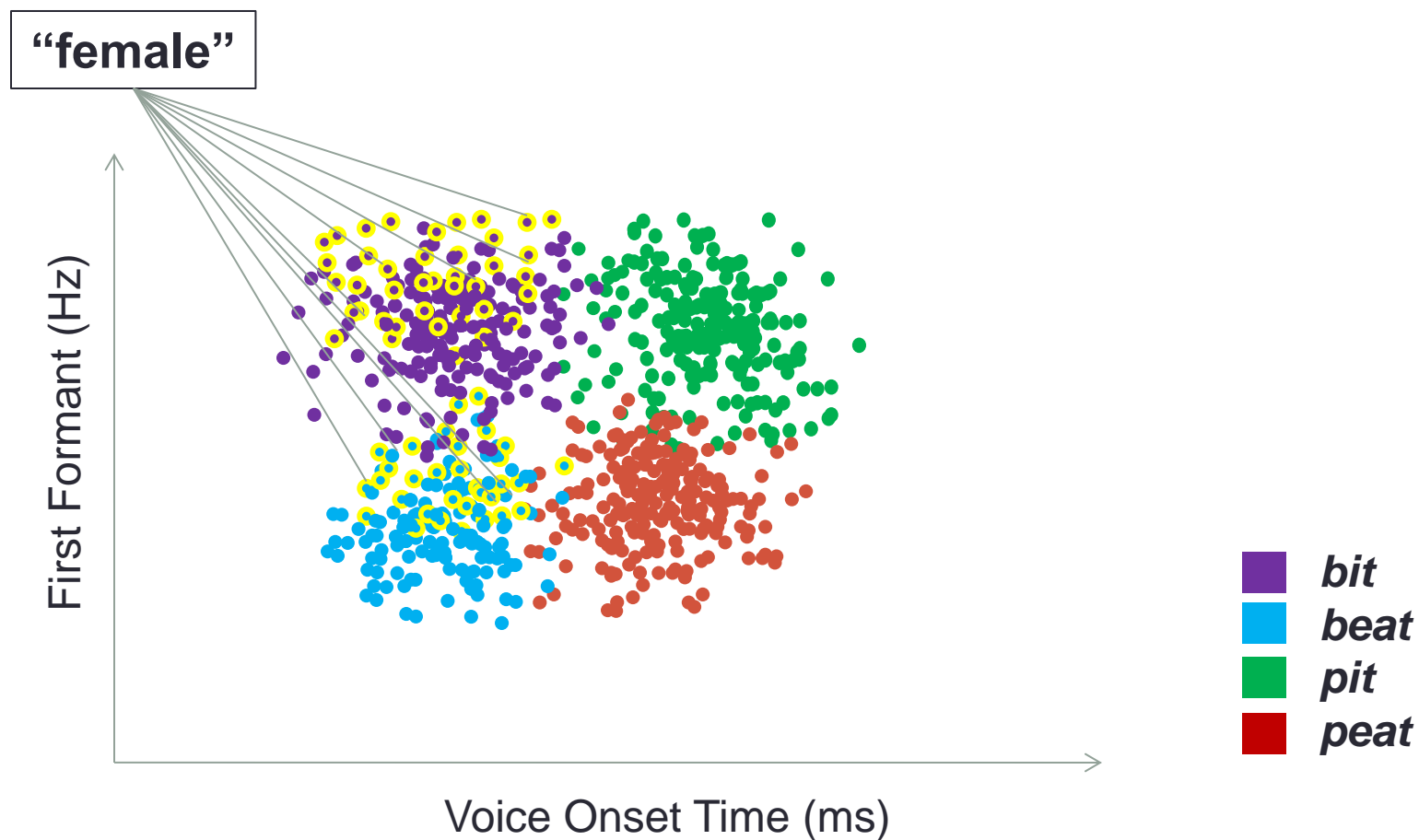
Exemplar-based models

- Dynamic updating of the perceptual space depending on speaker characteristics (e.g., shifts of category boundaries)
- Socially conditioned variation in production which can be distinguished from accommodation
- Spontaneous phonetic imitation of specific talkers on a word-by-word basis (Goldinger 1997, 1998)
- Talker convergence effects (Pardo 2006, 2009; Kim, Horton & Bradlow 2011; i.a.)
- Word-specific phonetics: Homophones do not have exactly the same realization (Pierrehumbert, 2002)
- Prosodically-conditioned variation in phoneme realization (i.e., traditional allophony)

Exemplar Theory



Dynamic variation



Lexical statistics and exemplar representations

- Representations are dynamic
 - New exemplars in a sparsely populated region of a wordform's "cloud" can change the behavior of that wordform in production and perception
- Frequency matters
 - Wordforms with many exemplars are more resistant to the influence of new exemplars
- True test of exemplar-driven vs. accommodation-based processes is whether sensitivity of individual wordforms to new input depends on frequency (e.g., CELEX)

Hybrid Exemplar Models (wordforms)

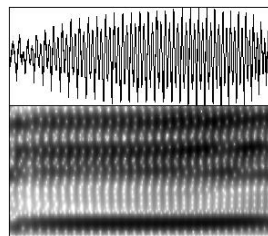
Wordform

city

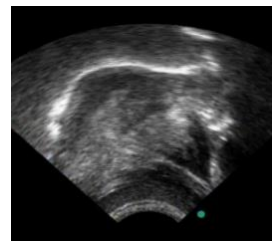
Phonological
Parse

/sɪ.tɪ/

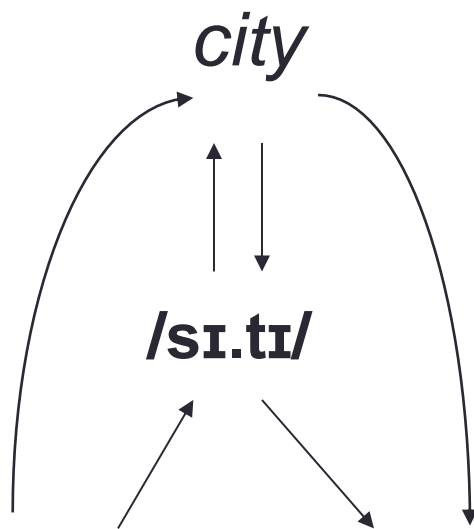
Concrete
Phonetics



Perception



Production



Hybrid Exemplar Models (intonation)

Discourse
meaning

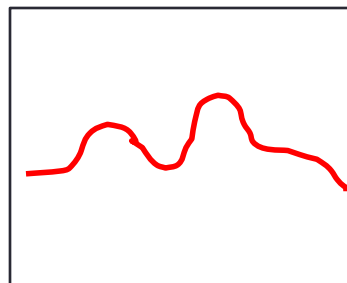
(question, focus-background, etc.)

$$\lambda V_{\langle e, \langle e, t \rangle \rangle} \cdot \lambda O_e \cdot \lambda S_e \cdot \parallel V([o]_F)([s]_T) \parallel^{CT}$$

Phonological
Parse

[L+H* L-H% H* L-L%]

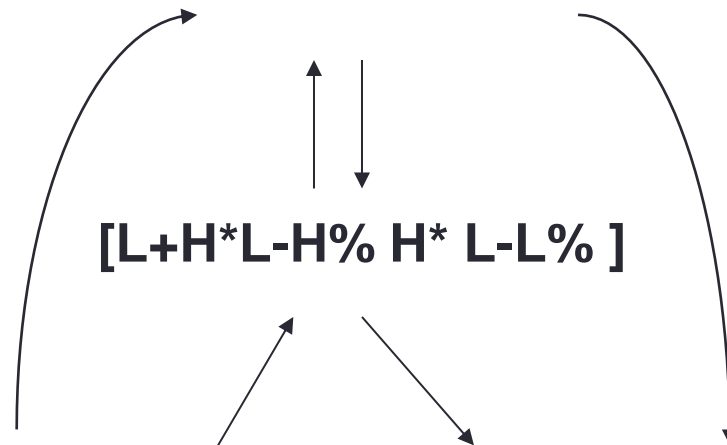
Concrete
Phonetics



Perception

Glottal
gesture
stuff

Production



What is needed?

- Frequency as psycholinguistic parameter is an estimate of the number of times that an individual has encountered a particular *wordform* in his/her life experience with a language
- In order to explore whether prosodic meaning is based on exemplar representations analogous to wordforms, it is necessary to estimate the number of times that a particular (discourse-level) form-meaning pair has occurred in an individual's experience
- Wide range of such pairs needed (c.f., number of wordforms in CELEX)

Prosodic “wordforms”

- Focus-background structure is robust and identifiable w.r.t. nuclear accent placement (English)
- Post-focal deaccenting is obligatory in most dialects

Meaning

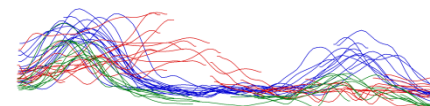
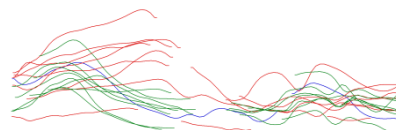
S [V]_F O

[S]_F V O

Pattern

S V O
|
L+H* L-L%

S V O
|
L+H* L-L%



Instance

He WANTS the job.

HE wants the job.

Prosodic “wordforms”

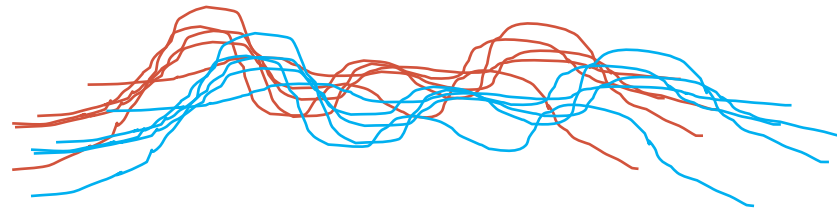
- Contrastive topic

Meaning

[S]_{CT} V [O]_F

Pattern

S V O
| | |
L+H* L-H% H* L-L%



Instance

Melanie studies prosody.

Evokes: Other people study other things.

Prosodic “wordforms”

- Focus/background wordforms
 - [S] V O; S V [O]; S [V] O; S [VO]; [S V O]
- Contrastive topic wordforms
 - [S]_{CT} V [O]_F; [S]_{CT} [V O]_F; [S]_{CT} [V]_F O; S [V]_{CT} [O]_F; [S]_F V [O]_{CT};
[S]_F [V]_{CT} O; [S]_F [V O]_{CT}; S [V]_F [O]_{CT}

Prosodic “wordforms”

Utterance-final patterns:

- English, French, others
 - High rise signaling a yes/no question
 - Low rise signaling a continuation
 - Fall signaling an assertion
 - Fall signaling end of a wh-question
- French
 - Penultimate rise-fall signaling a confirmation question
 - Penultimate rise-fall signaling an assertion

Prosodic “wordforms”

- Certain less common pitch accents carry more specific types of meaning that may potentially be treated in isolation as their own forms
 - Perhaps in combination with focus/background structure
 - “...the L*+H accent is used to convey lack of predication and to evoke a scale...” (Pierrehumbert & Hirschberg, 1990)

Applications

- Test whether “wordforms” from the same comparison class differ in their sensitivity to new input
 - Implicit convergence
 - Direct imitation
 - Passive training
- Explore the extent to which distinct “wordforms” with identical phonological parses (i.e., prosodic homophones) exhibit different degrees of variability
 - Are highly frequent wordforms more likely to exhibit phonetic patterns that depart from what is expected based on the phonological parse?
 - c.f., word-specific phonetics (c.f., Pierrehumbert, 2002)

Hybrid Exemplar Models (wordforms)

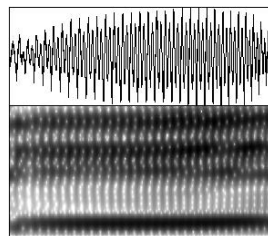
Wordform

city

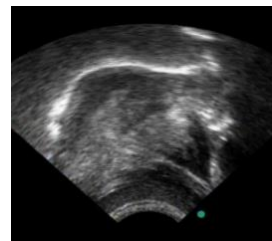
Phonological
Parse

/sɪ.tɪ/

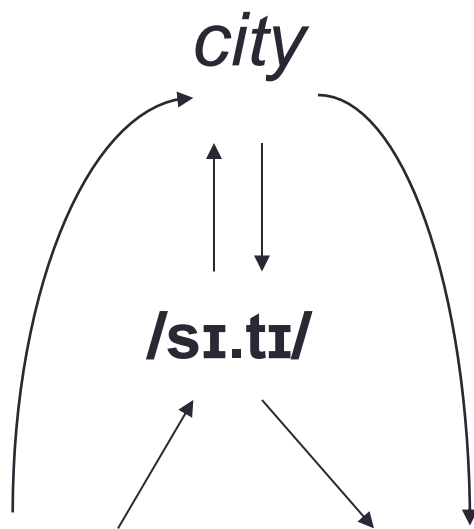
Concrete
Phonetics



Perception



Production



Hybrid Exemplar Models (intonation)

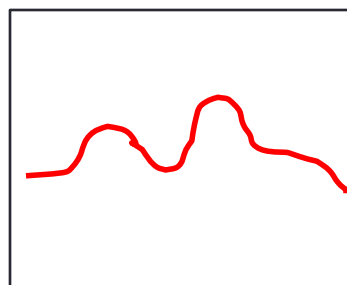
Discourse
meaning

$[S]_{CT} \quad V \quad [O]_F$

Phonological
Parse

$[L+H^*L-H\% \quad H^* \quad L-L\%]$

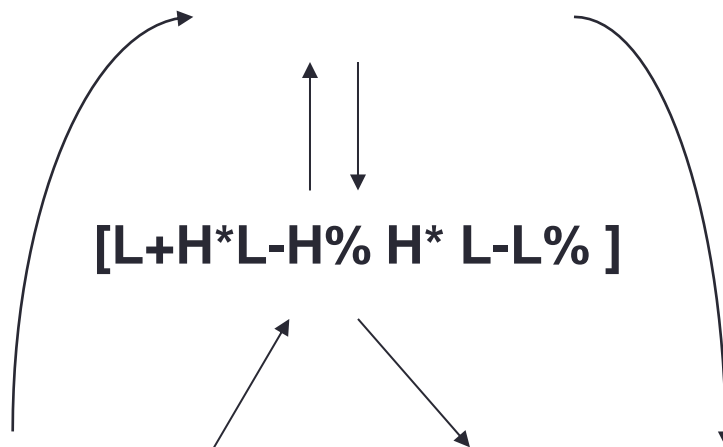
Concrete
Phonetics



Perception

Glottal
gesture
stuff

Production



Applications

- Identify socially-driven variation (e.g., gender)
 - Use this to create specific tests for implicit social priming effects
 - Example: Does the threshold for what counts as a final rise (vs. a fall) depend on which social primes are active during perception?
- Explore how lexico-syntactic context conditions the phonetic realization of prosodic forms
 - Contextual features are indices associated with specific exemplars
- Apply machine-learning to explore how prosodic categories might emerge from noisy inputs
 - Which features are important and why?

Applications

- Move towards a theory of prosodic representation that accounts for...
 - Emergence of categories through learning
 - Late plasticity
 - Episodic effects
 - Dynamic social/contextual conditioning

Open Questions

- Which “wordforms” can be reliably coded?
- Sample size/How many tokens?
- Which corpora represent the best estimates of an individual’s experience with a language?
- Which corpora are likely to include useable types of social variation?

A useable corpus

- Phonetic transcription (hand-corrected force-alignment)
- A simplified ToBI-like transcription (manual)
 - Locations and strength of higher-level boundaries
 - Locations of pitch accents, other tonal events (dep. on language)
 - (Pitch accent categories)
- A reasonably good syntactic parse (automatic)
- Audio is good enough for f0 recovery, segment duration
- Coding of “wordforms” (manual)

Challenges???