A corpus search methodology for focus realization

Jonathan Howell
Department of Linguistics

Mats Rooth
Department of Linguistics
Computing and Information Science

Project Goals

- Study phonetic realization of focus in cases where formal semantic theories make predictions
- Develop and apply methodology for accessing controlled and quantifiable, but diverse and naturally occurring speech data

Focus in Comparative Clauses

- Coherent syntactic-semantic theory about focus placement
- Possibilities constrained: main clause usually antecedent for focus interpretation in comparative clause

Subject focus

(1) He stayed longer than [I]$_F$ did.

-er [[**he** stayed *x* long]$_2$ than [**I** stayed *x* long] ~2

Non-subject focus

(2) I should have like that song more than I [did]$_F$.

more λx[[should λw[I like that song *x* well in **w**]]$_3$

than [I like that song x well in $\mathbf{w_0}$]

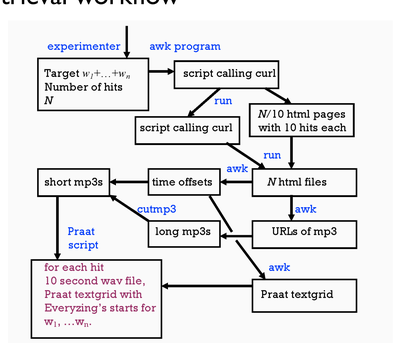
(3) I understand even less than I did [before]$_F$.

even less [[I **pres** understand *x* much]$_4$

then [I understood *x* much **before**]]~4

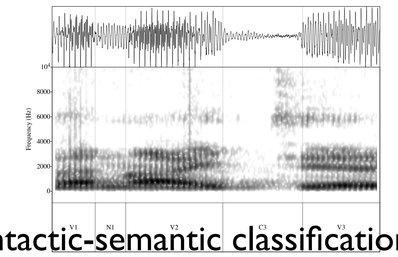
Corpus search methodology

- everyzing.com
  - commercial search engine
  - corpus of podcasts in mp3 format
  - transcription and indexing with BBN speech recognition
- Web harvest
  - Retrieval workflow



Experiment: "than I did"

- Web harvest
  - 179 alleged tokens of "than I did"
  - of which 91 unique correct tokens (56%)
- Subphonemic annotation
  - Experimenter annotation using landmarks

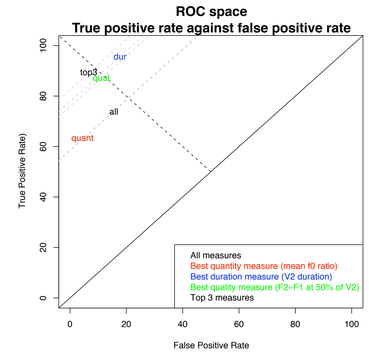


- Syntactic-semantic classification
  - Transcription into English prose
  - Experimenter binary classification into subject focus condition "s" and non-subject focus condition "ns"

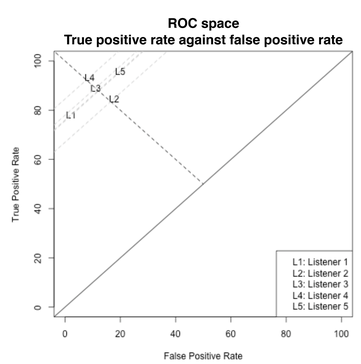
Transcription	Class
[**The aquarium** got more attention] than [**I** did].	s
[I damn sure paid more for **the Apple**] than [I did **this Toshiba**].	ns

- Phonetic feature extraction
  - automated with Praat scripts
  - 308 acoustic measures
    - *Quantity*: F0, amplitude, intensity, power...
    - *Quality*: formants, bandwidth, tilt, jitter...
    - *Duration*: for vowels, stops, syllables
  - Measures taken: mean, extrema, range
  - Locus of measure: regular intervals, time of other extrema (e.g. f0 max, intensity max)
- Supervised machine learning
  - many parameters highly correlated (collinearity condition number κ=127)
  - want a focus detector for classifying novel/ controversial data
  - **SVM** (support vector machine) increasingly common in machine learning
  - **SVM** uses separating hyperplanes (margins) rather than comparison of mean
- Comparing statistical classifiers
  - Classifiers evaluated by one-held-out cross validation (OHOCV)
  - Best measures for each category (using feature selection techniques)

Label	Measures used	Total Accuracy (OHOCV)
All	all 308 measures	89.0%
Duration	duration of vowel "I" [aɪ]	88.9%
Quantity	ratio of f0 mean in "I" [aɪ] and "did" [ɪ]	76.9%
Quality	difference in f1 and f2 at 50% of "I" [aɪ]	87.9%
Top3	Above three measures	91.2%



- Best-performing classifiers use features predictable from formal linguistic theory
- Best-performing quality measure outperforms best-performing quantity measure
- Comparing human classifiers
  - Informal forced-choice discrimination task: "s" condition vs. "ns" condition
  - Stimuli: 91 "than I did" snippets
  - Non-naïve listeners (n=5)



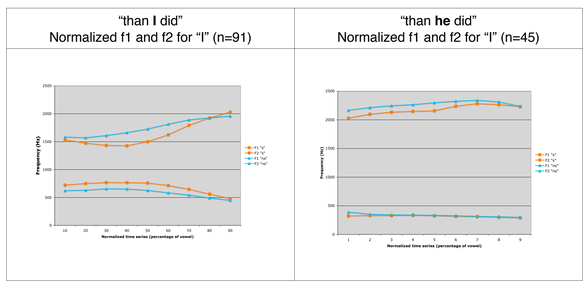
Phonetic models of accent: hyperarticulation vs, sonority expansion

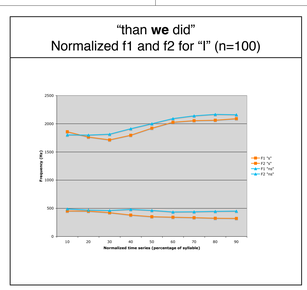
- Hyperarticulation / featural enhancement
  e.g. de Jong (1995), Fowler (1995), Cho (2005)
  - vowel features or gestural targets maximized
- Sonority expansion / tongue or jaw lowering
  e.g. Edwards & Beckman (1988), Beckman *et al.* (1992), Erickson (2002)
  - lower jaw / more open vocal tract

Predictions

	Featural enhancement	Sonority expansion
[a(ɪ)]	[low]: higher f1 [back]: lower f2	higher f1
[(h)i]	[high]: lower f1 [front]: higher f2 ?	higher f1
[(w)i]	[high]: lower f1 [round]: lower f2 ?	higher f1

- Preliminary results consistent with featural enhancement





- References