Predicting Phrasing and Accent

Julia Hirschberg

CS 4706

Today

- Motivation for intonation assignment algorithms
- Approaches: hand-built vs. corpus-based rules
- Predicting phrasing
- Predicting accent
- Future research: emotion, personalization, personality

Why worry about accent and phrasing?

A car bomb attack on a police station in the northern Iraqi city of Kirkuk early Monday killed four civilians and wounded 10 others U.S. military officials said. A leading Shiite member of Iraq's Governing Council on Sunday demanded no more "stalling" on arranging for elections to rule this country once the U.S.-led occupation ends June 30. Abdel Aziz al-Hakim a Shiite cleric and Governing Council member said the U.S.-run coalition should have begun planning for elections months ago.

-- Loquendo 🔊

Why predict phrasing and accent?

- TTS and CTS
  - Naturalness
  - Intelligibility
- Recognition
  - Decrease perplexity
  - Modify durational predictions for words at phrase boundaries
  - Identify most ‘salient’ words
- Summarization, information extraction

How do we predict phrasing and accent?

- Default prosodic assignment from simple text analysis
  - Accent content words
  - Deaccdent function words

  The president went to Brussels to make up with Europe.

  - Limitations
    - Doesn’t work all that well, e.g. *particles*
    - Hand-built rule-based systems hard to modify or adapt to new domains
- Corpus-based approaches (Sproat et al ’92)
  - Train prosodic variation on large hand-labeled corpora using machine learning techniques

- Accent and phrasing decisions trained separately – a problem?
  - Binary prediction
  - Feat1, Feat2,…Accent
  - Feat1, Feat2,…Boundary
- Associate prosodic labels with simple features of transcripts that can be automatically computed, e.g.
  - distance from beginning or end of phrase
  - orthography: punctuation, paragraphing
  - part of speech, constituent information
- Apply automatically learned rules when processing text

Reminder: Prosodic Phrasing

- 2 `levels' of phrasing in ToBI
  - intermediate phrase: one or more pitch accents plus a phrase accent (H- or L-)
  - intonational phrase: one or more intermediate phrases + boundary tone (H% or L%)
- ToBI break-index tier
  - 0 no word boundary
  - 1 word boundary
  - 2 strong juncture with no tonal markings
  - 3 intermediate phrase boundary
  - 4 intonational phrase boundary

What are the indicators of phrasing in speech?

- Timing
  - Pause
  - Lengthening
- F0 changes
- Vocal fry/glottalization

What linguistic and contextual features are linked to phrasing?

- Syntactic information
  - Abney ’91 chunking major constituents
  - Steedman ’90, Oehrle ’91 CCGs …
  - Which ‘chunks’ tend to stick together?
  - Which ‘chunks’ tend to be separated intonationally?
    - Largest constituent dominating w(i) but not w(j)

    NP[The man in the moon] |? VP[looks down on you]
    - Smallest constituent dominating w(i),w(j)

    NP[The man PP[in |? moon]]
  - Part-of-speech of words around potential boundary site

    The/DET man/NN |? in/Prep moon/NN
- Sentence-level information
  - Length of sentence

This is a very |? very very long sentence ?| which thus might have a lot of phrase boundaries in?| it ?| don’t you think?

This |? isn’t.

- Orthographic information
  - They live in Butte, ?| Montana, ?| don’t they?
- Word co-occurrence information

  Vampire ?| bat …powerful ?| but benign…
- Are words on each side accented or not?

  The cat in |? the
- Where is the most recent previous phrase boundary?

  He asked for pills | but |?
- What else?

Statistical learning methods

- Classification and regression trees (CART)
- Rule induction (Ripper), Support Vector Machines, HMMs, Neural Nets
- All take vector of independent variables and one dependent (predicted) variable, e.g. ‘there’s a phrase boundary here’ or ‘there’s not’

  Feat1 Feat2 …FeatN DepVar
- Input from hand labeled dependent variable and automatically extracted independent variables
- Result can be integrated into TTS text processor

How do we evaluate the result?

- How to define a Gold Standard?
  - Natural speech corpus
  - Multi-speaker/same text
  - Subjective judgments
- No simple mapping from text to prosody
  - Many variants can be acceptable

The car was driven to the border last spring while its owner an elderly man was taking an extended vacation in the south of France.

Integrating More Syntactic Information

- Incremental improvements continue:
  - Adding higher-accuracy parsing (Koehn et al ‘00)
    - Collins or Charniak parser
    - Different learning algorithms (boosting, co-training)
    - Different syntactic representations: relational? Tree-based?
    - Ranking vs. classification?
- Rules always impoverished
- Where to next?

Predicting Pitch Accent

- Accent: Which items are made intonationally prominent and how?
- Accent type:
  - H* simple high (declarative)
  - L* simple low (ynq)
  - L*+H scooped, late rise (uncertainty/ incredulity)
  - L+H* early rise to stress (contrastive focus)
  - H+!H* fall onto stress (implied familiarity)

What are the indicators of accent?

- F0 excursion
- Durational lengthening
- Voice quality
- Vowel quality
- Loudness

What phenomena are associated with accent?

- Word class: content vs. function words
- Information status:
  - Given/new He likes dogs and dogs like him.
  - Topic/Focus Dogs he likes.
  - Contrast He likes dogs but not cats.
- Grammatical function
  - The dog ate his kibble.
- Surface position in sentence: Today George is hungry.

- Association with focus:
  - John only introduced Mary to Sue.
- Semantic parallelism
  - John likes beer but Mary prefers wine.
- How many of these are easy to compute automatically?

How can we approximate such information?

- POS window
- Position of candidate word in sentence
- Location of prior phrase boundary
- Pseudo-given/new
- Location of word in complex nominal and stress prediction for that nominal

  City hall, parking lot, city hall parking lot
- Word co-occurrence

  Blood vessel, blood orange

Current Research

- Concept-to-Speech (CTS) – Pan&McKeown99
  - systems should be able to specify “better” prosody: the system knows what it wants to say and can specify how
- Information status
  - Given/new
  - Topic/focus

Future Intonation Prediction: Beyond Phrasing and Accent

- Assigning affect (emotion) from text – how?
- Personalizing TTS: modeling individual style in intonation – how?
- Conveying personality, charisma – how?

Next Class

- Information status: focus and given/new information