Intonation for Conversational AI

Joseph Tyler
josephctyler@gmail.com
ObEN, Inc.
Siri example:

“What is the weather like today?”

OK, I found this on the web for ‘What is the weather today?’:

“Today’s weather forecast for [insert city]:

Partly cloudy with a high of [insert temperature]°F and a low of [insert temperature]°F.”
Siri example:

Joseph: What’s the difference?
Siri example:

Joseph: What’s the difference?
Siri example:

Joseph:  

Siri:  

"Can you find a store that sells chicken sausage and potatoes"

OK, I found this on the web for ‘Can you find a store that sells chicken sausage and potatoes‘:
OK, I found this on the web for ‘Can you find a store that sells chicken sausage and potatoes’:

Chicken Sausage Recipes - Allrecipes.com
Chicken Sausage Recipes Looking for chicken sausage recipes? Allrecipes has more than 40 trusted chicken sausage recipes complete with ratings, reviews and cooking tips.

Apple Chicken Sausage Links - Johnsonville.com
... flavorful chicken sausage. ... Best Of Us C-Store - Give your C-Store some Brattitude Foodservice - Discover what ... Chicken, water, ...

Chicken and Sausage Main Dish Recipes | MyRecipes.com
Pair chicken and sausage to make some absolutely
Defining speech, intonation, and prosody
E.g. [p] vs [b]

Speech

Segments (the letter sounds)

Suprasegmentals (Intonation/Prosody)

E.g. [p] vs [b]
Acoustic features of prosody:

- Intonation and tone (f0, measured in Hz, perceived as pitch)
- Intensity/amplitude (perceived as loudness)
- Pausing
- Speech rate
- Voice quality (voicing, whispering, breathy voice, creaky/vocal fry, more)
- Many more...
Structural components of prosody:

- Boundary tones (H and L)
- Pitch accents (H and L)
I love to just live in dresses.
What can prosody do (in English)?

• Prosody can help with:
  • Disambiguation (within and between sentences)
  • Making language easier to understand
  • Improving naturalness

• Bracketing (indicate word groupings)
• Speech acts (e.g. Q vs A, informing vs. reminding)
• Focus and sets of alternatives (John/JOHN didn’t cheat on the test).
• Social meaning (stereotypes, social groups, dialects)
Prosodic Disambiguation

Ambiguity makes NLP hard:

Violinist Linked to JAL Crash Blossoms
Teacher Strikes Idle Kids
Red Tape Holds Up New Bridges
Hospitals Are Sued by 7 Foot Doctors
Juvenile Court to Try Shooting Defendant
Local High School Dropouts Cut in Half
Prosodic Disambiguation

• Linguistic ambiguity: one set of words, 2+ meanings
  • Prosodic boundaries can disambiguate
  • Great place to see prosody’s work

• “Old men and women”
  • Old men, and women
  • Old (men and women)

• “Paula phoned a friend from Alabama.”
  • Paula phoned (a friend from Alabama).
  • Paula phoned a friend (from Alabama).
Prosodic Disambiguation

"I sat in on a history class. I read about housing prices. And I watched a cool documentary."

(Tyler 2014)
Prosodic Disambiguation

“April brings her beagle, and everyone else stays home.”

• Logical conjunction with accented ‘and’
• Conditional reading with rise-fall-rise
Rises vs. Rise-Plateaus

• Podcast example
• On NPR all the time

Rises vs. Rise-Plateaus

• Rises are for lists that **inform**
  • Listener doesn’t know the items in the list.

• Rise-plateaus are for lists that **remind**
  • Listener does know the items in the list, or at least can finish it based on the examples you gave.
Pitch Accents and Focus Marking

• Pitch accents!
  • JOHN didn’t eat the cake.
  • John didn’t EAT the cake.
  • John didn’t eat the CAKE.

• The meaning difference?
Example 1: Pitch Accent and Focus for Siri?

• Ask: “Are there any western movies playing near me right now?”
• Siri: “Sorry, I don’t see any western movies playing nearby today.”

vs.

“I don’t see any WESTERN movies playing near Oakland, but there are quite a few others.”

Personal, relevant, context-aware.
Example 2: Pitch Accent and Focus for Siri?

• Ask: “What’s traffic look like for getting home?”
• Siri: “The traffic to home is about average, so it should take about 1 minute.”

  “Traffic to HOME is about average, but traffic to Yosemite is terrible right now.”
Prosodic correlates of discourse in production

Figure 2.3  The height of each ‘sentence beginning’ (major tone group onset) in a news report^ (SEC B01). Sections 1 and 12 contain the opening and closing headlines, preceded and followed respectively by metatextual links.

BBC News segments tend to start with high onset pitch
Prosodic correlates of discourse in production

• Monologues have structure beyond the sentence. Aspects of this structure is visible in prosody.
• Reading newspaper article: segmented, relations build full structure
Prosodic correlates of discourse in production

- Larger discourse boundaries $\rightarrow$ louder, higher, longer pauses
- Relationship to prior sentence affects prosody, but only after small boundaries.
Prosody and Social Meaning

• Dialect differences
• Jewish English has lots of low-high pitch accents (L+H*), distinct listing intonation

Why Linguists are Fascinated by the American Jewish Accent
The linguistic field of prosody, the story of melody, pitch, and other hard-to-study verbal traits, is suddenly hot.

(Burbin 2016)
Prosody and Social Meaning

• Dialect differences
• Black English tends to have more frequent pitch accents

Sandra Bland: Talking While Black
August 15, 2015 @ 5:41 pm · Filed by Mark Liberman under Language and politics, Language and society

« previous post | next post »

Below is a guest post by Nicole Holliday, Rachel Burdin, and Joseph Tyler:

Sandra Bland’s traffic stop and the tragic series of events that occurred afterwards have been the subject of many recent think pieces, but few authors have examined why the initial traffic stop went wrong in the first place. The most obvious explanation might be simple racial profiling, which almost certainly played a role, but the dash cam video of the event also shows an interaction that escalated at an alarmingly rapid pace. The
Prosody and Social Meaning

Who do you visualize?
Prosody and Social Meaning

- Uptalk and the uptalk stereotype
I've done everything I can to stop it. Whatever modest sphere of influence I have, I've used. Teaching large undergraduate classes, writing newspaper articles, giving interviews - all to no avail. I'm fighting a steamroller here or, in the more colorful language of Evolutionary Psychology, a very powerful meme.

This is the meme from Hell. The kind of cultural thing Richard Dawkins must have had in mind when he introduced the term in *The Selfish Gene* in 1976. This was, he argued, the way culture spreads - longitudinally as a virus spreads within a population. The meme is the basic unit of culture. As Dawkins argued, memes "travel horizontally, like viruses in an epidemic." They compete with other memes and the winners take up residence in our minds, defining what our culture looks and sounds like. When Susan Blackmore wrote *The Meme Machine* in 1999, she didn't have the topic of this column as an example to draw upon. That's unfortunate. This one is the equivalent of a viral video. About all you can do is stand back and watch it spread. In this case, of course, you'd have to listen to it spread, since it has become part of speech.
Uptalk anxiety

September 7, 2008 @ 6:31 am · Filed by Mark Liberman under Language and gender, Psychology of language

For additional background, here are some of the earlier Language Log posts that deal with related questions:

This is, like, such total crap? (5/15/2005)
Uptalk uptick (12/15/2005)
Angry Rises (2/11/2006)
Further thoughts on "the Affect" (3/22/2006)
Uptalk is not HRT (3/28/2006)
Poem in the key of what (10/9/2006)
Satirical cartoon uptalk is not HRT either (11/14/2006)
Intonation contours and polonium poisoning (12/16/2006)
Uptalk anxiety (9/7/2008)
The phonetics of uptalk (9/13/2008)
Word (in)constancy (9/16/2008)
Prosody and Social Meaning

• The stereotype might not equal reality, meanings are more varied than the popular discussion of uptalk suggests.
  • Positive (emphasis, excitement, normal, happy)
  • Negative (dumb, like a tic)
  • Linguistic (unfinished, conditional)
  • Social (young, female, urban, California)

• Correlation analysis of these perceptions shows two clusters:
  • Unfinished (“continuation rise”, holding the floor)
  • Everything else (social, emotional, regional)
Prosody and Social Meaning

- Vocal fry (aka creaky voice)
- https://www.youtube.com/watch?v=YEqVgtLQ7qM
Prosody and Social Meaning

- Vocal fry (aka creaky voice)
- Men do it too, but tend not to be criticized for it:
Summary:

• Prosodic boundaries == linguistic bracketing
• Speech acts (Q vs. A, informing vs. reminding)
• Pitch accents == sets of alternatives
• Social meaning
  • Construct characters, convey attitudes, define groups
• Lots more I didn’t have time for!

THANKS FOR LISTENING
Siri Problems (as brainstormed at Meetup)

- Punctuation norms, other Siri specific norms
  - Human adapting to Siri, vs other way around
- Proper Nouns, esp. not native to English
- Codeswitching between lgs
- Accents, individual variation (learning)
- Filtering out background noise
- Lack of follow up, just one-and-done
  - Lack of context and memory
- Limited domain knowledge
Rises vs. Rise-Plateaus

• Mark is lost on the east side of town! He calls Stacie and asks how to get home. She says:
  • "Well, what you need to do is get on Hudson and turn on 11th avenue. Then you get on 71, get on 670, take a right on 27…"

• When Mark’s cell phone dies. Does Stacie think that Mark knows how to get home?