



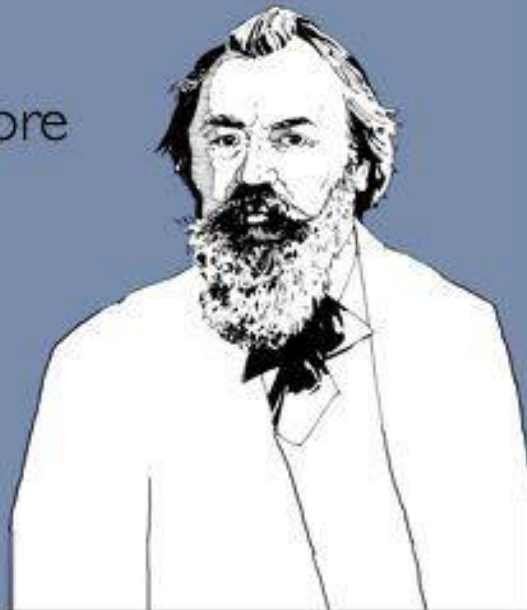
Analyzing Human and Machine Performance In Resolving Ambiguous Spoken Sentences

Hussein Ghaly¹ and Michael Mandel²

¹ Graduate Center, City University of New York, ²Brooklyn College, City
University of New York

I'm a linguist.

I love ambiguity more
than most people.



your  cards
someecards.com



Motivation

When an ambiguous sentence is spoken, what information does speech have which text alone doesn't?

Our goal is to examine this information by analyzing human disambiguation of both text and speech for different types of ambiguities, and developing a model for automatic disambiguation using this information



Research Summary

- Record sentences containing some ambiguity, with the speaker aware of the correct interpretation
- Subjects hear or read sentences, predict the correct interpretation
- Analyze acoustic features of each utterance, including multiple recordings of the same sentence
- Develop a Machine Learning approach to predict the intended reading given the acoustic features

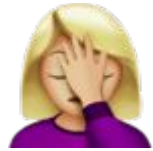
Types of Ambiguity

- Lexical Ambiguity (I forgot my bag at the **bank**)
- Syntactic Ambiguity (old men and women)
 - **Comma Ambiguity**
 - **PP-attachment**
 - NP-ambiguity
 - Coordination ambiguity ... etc



Comma Ambiguity

A woman without her man is nothing.



A woman: without her, man is nothing.





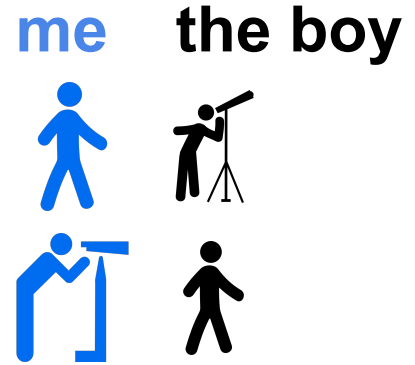
Comma Ambiguity

- Without punctuation (e.g. out of ASR) text can be ambiguous
- Can the written text be disambiguated by humans?
- Can the spoken sentence be disambiguated by humans?

PP-Attachment

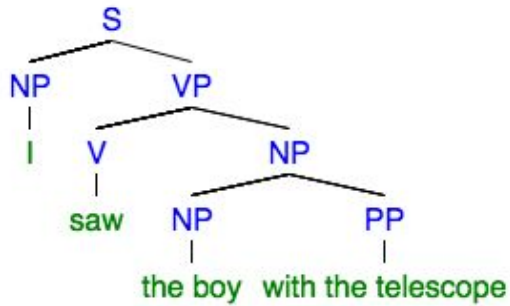
I saw [the boy with the telescope]

I saw the boy [with the telescope]

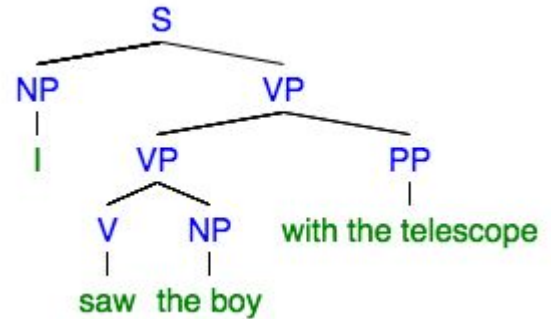
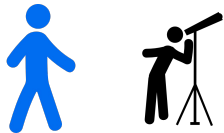


PP-Attachment

This sentence has two possible interpretations, i.e., a structural ambiguity



me **the boy**

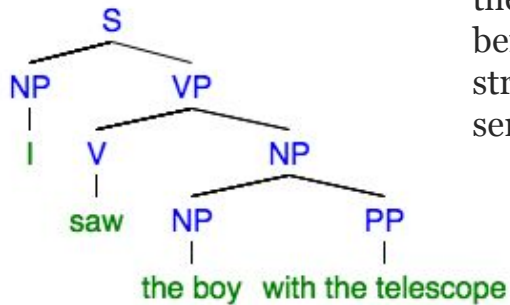


me **the boy**

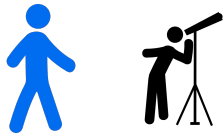


PP-Attachment - Early vs. Late Closure

Late Closure

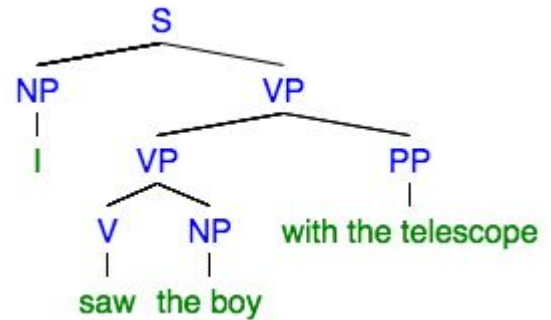


me the boy



late closure is the principle that new words (or "incoming lexical items") tend to be associated with the phrase or clause currently being processed rather than with structures farther back in the sentence. *

Early Closure



me the boy





Hypotheses

- When there is ambiguity in any sentence and the speaker is aware of the correct reading, they will convey their knowledge of the correct reading using certain prosodic cues.
- Listeners will be able to use these cues to identify the correct reading better than readers will
- These prosodic cues can be measured and analyzed and used as features for automatic disambiguation system using machine learning

Previous Research

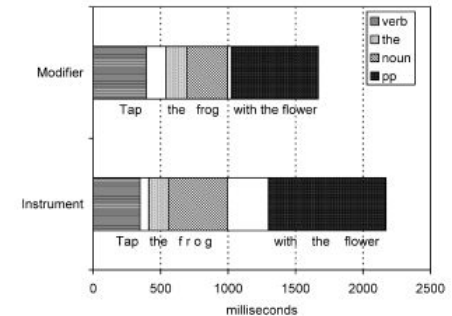
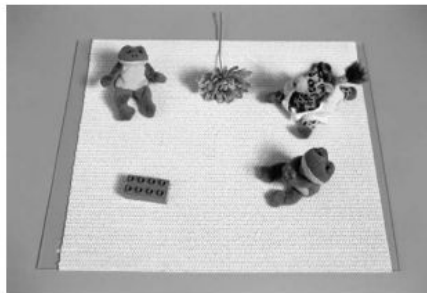
Psychology - Snedeker and Trueswell, 2003 - Using prosody to avoid ambiguity: Effects of speaker awareness and referential context.

“informative prosodic cues depend upon speaker's knowledge of the situation: speakers provide prosodic cues when needed; listeners use these prosodic cues when present.”

Prosodic cues include pauses and word durations, as shown from the utterances of a speaker who is aware of the intended meaning.

“tap [the frog with the flower]” - modifier

“tap the frog [with the flower]” - instrument





Previous Research

NLP - Levi et al, 2012 - The effect of pitch, intensity and pause duration in punctuation detection

Predicting punctuation from different prosodic cues of speech using neural networks

Cues included: pitch, intensity and pause duration

Achieved a punctuation detection rate of 54%



Data

- We created a collection of 26 constructed sentences (6 pairs of sentences with comma ambiguity and 7 pairs of sentences with PP-attachment ambiguity)
- We recorded the sentences spoken by a native speaker, each sentence recorded five times (total 130 recording files)



Comma Ambiguity - Speaker Tasks

Record 6 pairs of constructed Comma-ambiguous sentences

Example:

3a: John, said Mary, was the nicest person at the party.

3b: John said Mary was the nicest person at the party.



Comma Ambiguity - Listener Tasks

For each Comma-ambiguous sentence, identify the intended meaning:

Task 1 - Using Text only

Task 2 - Using Audio Only

Example:

Sentence: John, said Mary, was the nicest person at the party.

Question: Who was said to be the nicest person at the party? A- John B- Mary



PP-Attachment Ambiguity - Speaker Tasks

Record 7 pairs of sentences with PP-attachment ambiguity, each pair contains a different preceding context supporting one reading of the sentence

Example:



4a: *One of the boys got a telescope.* I saw the boy with the telescope.

4b:- *I have a new telescope.* I saw the boy with the telescope.

PP-Attachment Ambiguity - Listener Tasks

For the following settings, identify the correct meaning by answering a question. For the last setting, sentences recordings were trimmed from the previous context.

Who has the telescope? A- The boy B- The speaker

Setting	Presentation
Text with context	I have a new telescope. I saw the boy with the telescope.
Audio with context	
Text without context	I saw the boy with the telescope.
Audio without context	



Results - Human Evaluation

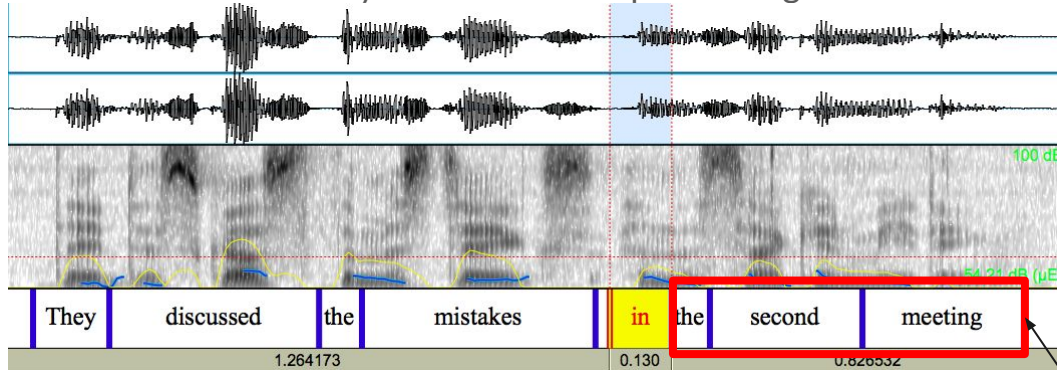
Ambiguity	Modality	Accuracy
Comma	Text	99.3%
Comma	Audio	94.7%
PP-attachment with context	Text	93.1%
PP-attachment with context	Audio	97.1%
PP-attachment without context	Text	52.0%
PP-attachment without context	Audio	74.4%

Results - PP-Attachment - Acoustic Analysis

acoustic feature values averaged over the 20 productions of the following sentences

They discussed the mistakes **in** the second meeting.

The lawyer contested the proceedings **in** the third hearing.



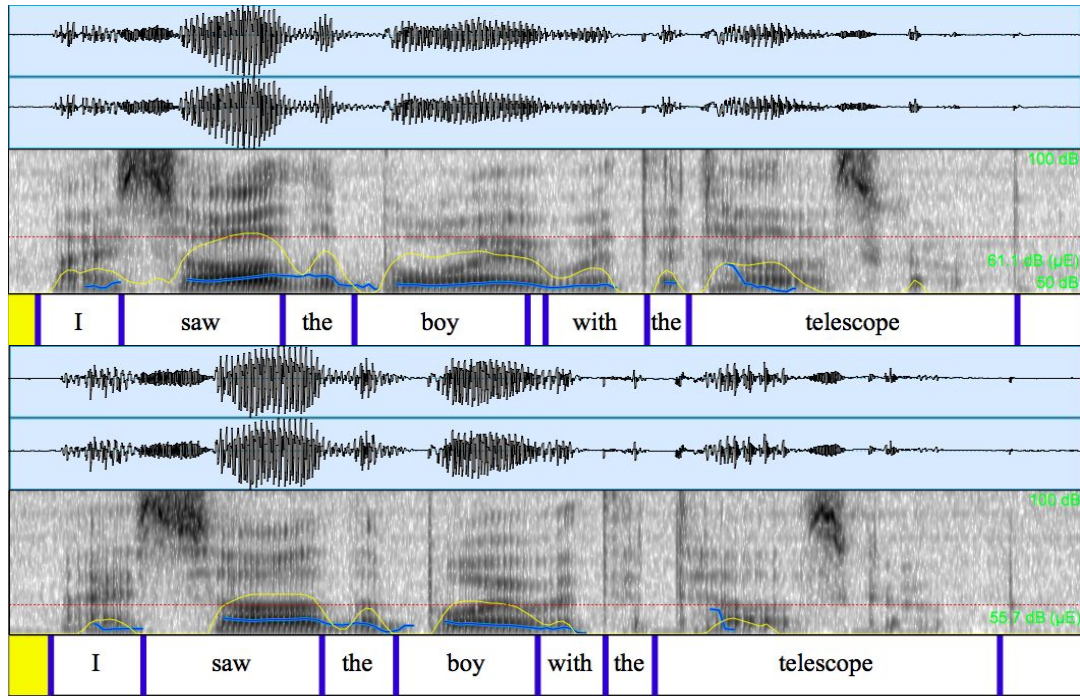
	Late	Early
Preposition Duration (ms)	147	143
Preceding silent pauses (ms)	0	48
Intensity (dB)	57.8	56.4
Following NP duration (ms)	579	640

Preceding Silent Pause

Preposition

Following NP

Acoustic Analysis - Early vs. Late Closure



Early Closure

Late Closure



Results - PP-Attachment - Machine Evaluation

Feature Matrix - Extracted manually from 10 audio files for the sentence “They discussed the mistakes in the second meeting”.

duration of preposition (ms)	preceding silence (ms)	following NP duration (ms)	Preposition Intensity (dB)	Closure Type
160	0	690	56.6	early
175	0	660	59.0	late
120	0	470	56.2	late
140	80	620	55.6	early
145	0	600	58.7	late
140	90	635	57.8	early
135	0	510	61.1	late
150	110	600	57.9	early
130	0	620	61.0	late
140	60	580	58.8	early



Machine Evaluation

Using Decision Trees for 20 data points with 5-fold cross-validation: 80% average accuracy in predicting early vs. late closure

All sentences were using in training and testing each fold



Conclusions

- Humans can disambiguate sentences with comma ambiguity with audio alone almost as well as with text containing punctuation
- Humans can disambiguate spoken sentences with PP-attachment ambiguity without context, but cannot disambiguate the same sentences as text
- When speakers are aware of the intended meaning, they can produce sentences in a way that can
 - Be disambiguated by listeners, even without context
 - Be identified through certain acoustic cues
 - Be disambiguated to some extent by machines, initial results are promising



Thank you!

Questions?