Using Prosody to Improve Parsing

Hussein Ghaly[1] and Michael Mandel[2], Linguistics PhD Program, Graduate Center, City University of New York

1) hmghaly@gmail.com; 2) mim@sci.brooklyn.cuny.edu

Background

Goal: Improve automatic syntactic parsing of spontaneous spoken sentences using prosodic cues

Theoretical Motivation:

- Automatic Parsing is negatively affected by syntactic ambiguity (Kummerfeld et al., 2012)
- Prosody can help resolving some syntactic ambiguities (e.g. Cutler et al., 1997)
- Syntactic structure is related to prosodic structure (e.g., Shattuck-Hufnagel and Turk, 1996)
- Automatic Parsing of spontaneous speech is negatively affected by disfluencies and speech repairs (Charniak and Johnson, 2001), which are often marked with prosodic cues (Shriberg et al., 1997)

About Prosody:

- **Definition:** "(1) acoustic patterns of F0, duration, amplitude, spectral tilt, and segmental reduction, and their articulatory correlates, that can be best accounted for by reference to higher-level structures, and (2) the higher-level structures that best account for these patterns." (Shattuck-Hufnagel and Turk, 1996)
- Prosody includes these components: phrasing, stress, intonation and rhythm
- Prosodic phrasing focuses on how spoken utterances are split into groups and on boundaries between such groups (prosodic breaks or prosodic boundaries)
- Prosodic breaks are accompanied with acoustic cues, such as pauses and longer duration of syllables preceding boundaries
- A popular way to represent prosody is ToBI (Tones and Break Indexes) (Silverman et al., 1992), which is a system of annotating prosodic events.
- Break indexes reflect prosodic breaks as the perceived disjuncture between words (0 reflecting no disjuncture and 4 reflecting strong disjuncture)

Parsing Metrics:

In dependency parsing, the main parsing metric used is Unlabelled Attachment Score (UAS), which represents the ratio of how many heads predicted correctly by the parser divided by the number of words in the sentence.

Previous Work

Gregory et al., (2004): using prosody as punctuation. No improvement in parsing metrics (the Switchboard Corpus (SWB) (Godfrey et al., 1992))

Kahn et al., (2005): using prosody as features to a re-ranking system. An improvement of 0.2% absolute in F1-score using prosodic cues and 0.6% absolute using prosodic cues with syntactic cues (SWB)

Dreyer and Shafran (2008): using prosody as latent annotations. An improvement of 0.2% absolute (SWB)

Huang and Harper (2010): attaching prosodic breaks to the syntactic tree in different ways. An improvement of 0.5-0.8% absolute (A combination of SWB and Fisher Corpus)

Tran et al. (2017): using prosody as input to a neural network-based parser. An improvement of 0.5% absolute in F1-score (SWB)

Hypothesis

- There are correspondences between prosody and syntax, which can be extracted from the dependency structure, providing information about the most likely prosodic structure for a given utterance
- Using an ensemble classifier can improve parsing, by using higher-level features to select the most likely hypothesis.
- An ensemble classifier can be improved further by selecting the parse most compatible with the observed prosody.

Dependency Structure and Prosody

- Dependency parsing is now the norm in computational linguistics, as opposed to constituency parsing used in most previous work.
- We know of no study on the interaction of dependency structure and prosody
- Sense Unit Condition (SUC) (Selkirk, 1984) predicts highest likelihood of prosodic boundaries to be when there is least semantic coherence between words.
- Dependency structure reflects semantic relationships between words, elements of dependency structure can represent this semantic coherence

Dependency Configurations

- Dependency structure is defined by the dependent-head relationship for each word, we use it to quantify semantic coherence
- We propose the concept of dependency configurations, defined in terms of dependency offsets of two consecutive words
- We define dependency offset as the distance (measured by number of words) between a word and its head
- For each word, we quantify offset as:
  - 0 if the word is root
  - +1 if it depends on the word immediately to the right
  - +2 if it depends on a word further to the right
  - -1 if it depends on the word immediately to the left
  - -2 if it depends on a word further to the left
- Each pair of consecutive words is characterized by a duple representation (e.g. (+1,-2)) to describe the configurations
- This duple is then used to predict the prosodic break between the words

Analysis

Configuration	Abstract Representation	Example
(-1,-2)	word1 word2	well take it back thirty years
(-1,-1)	word1 word2	you 've found something else
(-1,0)	word1 word2	both of my kids do
(-1,+1)	word1 word2	is that what you 're saying
(0,-1)	word1 word2	you 've found something else
(0,0)	word1 word2	and in the long-term
(0,+1)	word1 word2	and they 're pretty close
(+1,-2)	word1 word2	both of my kids do
(+1,0)	word1 word2	you 've found something else
(+1,+1)	word1 word2	what branch were you in
(+2,-1)	word1 word2	both of my kids do
(+2,+1)	word1 word2	you 've found something else

	ToBI Break Index Values									
	1		2		3		4		Grand Total	
configuration	count	(%)	count	(%)	count	(%)	count	(%)	count	(%)
(+2, +1)	15657	22.97%	1550	2.27%	602	0.88%	860	1.26%	18669	27.39%
(+1, -2)	10328	15.15%	402	0.59%	203	0.30%	135	0.20%	11068	16.24%
(-1, +1)	6125	8.99%	557	0.82%	726	1.07%	1456	2.14%	8864	13.01%
(-1, -1)	6062	8.90%	281	0.41%	308	0.45%	358	0.53%	7009	10.28%
(+1, 0)	6032	8.85%	233	0.34%	114	0.17%	94	0.14%	6473	9.50%
(-1, -2)	3268	4.80%	334	0.49%	465	0.68%	829	1.22%	4896	7.18%
(+1, +1)	3308	4.85%	102	0.15%	42	0.06%	30	0.04%	3482	5.11%
(0, -1)	2802	4.11%	130	0.19%	120	0.18%	115	0.17%	3167	4.65%
(0, +1)	2645	3.88%	130	0.19%	174	0.26%	154	0.23%	3103	4.55%
(+2, -1)	1011	1.48%	25	0.04%	31	0.05%	22	0.03%	1089	1.60%
(-1, 0)	117	0.17%	11	0.02%	27	0.04%	62	0.09%	217	0.32%
(0, 0)	74	0.11%	30	0.04%	2	0.00%	7	0.01%	113	0.17%
Grand Total	**57429**	**84.27%**	**3785**	**5.55%**	**2814**	**4.13%**	**4122**	**6.05%**	**68150**	**100.00%**

- Using the ToBI annotated subset of the Switchboard corpus, we analyzed correspondence between configurations and prosodic breaks
- There is a higher likelihood of breaks (ToBI break indexes 3 and 4) for configurations (-1, +1), followed by (+2,+1) and (-1, -2).
- It is less likely to have a prosodic break for other configurations where one of the two consecutive words depends on the other

System

- An ensemble of three parsers ClearNLP (Choi and McCallum, 2013), Google Syntaxnet (Andor et al, 2016), spaCy (Honnibal et al., to appear), to provide parse hypotheses for each sentence in SWB
- LSTM-based ensemble classifier, using different features to predict the correctness of the head of each word in the parse

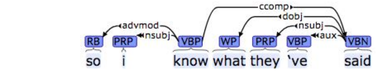
Features and Parameters

Three categories of features:

1- **Lexical features:** what each word and the corresponding head word are (represented by a dense word embedding feature)

2- **Syntactic features:** Parts of Speech (POS), and dependency configurations (configs)

3- **Prosodic features:** pause duration after each word, and normalized duration (dur) (actual duration of the word divided by expected duration, where expected duration is the sum of average duration of each phoneme for each speaker)



word	head word	POS	Config	normalized duration	pause
so	know	RB	(+2,+1)	3.23	0.31
i	know	PRP	(+1,0)	0.45	0
know	-	VBP	(0,+1)	1.13	0
what	said	WP	(+2,+1)	0.97	0
they	said	PRP	(+2,+1)	1.62	0
've	said	VBP	(+1,-2)	0.71	0
said	know	VBN	N/A	2.03	0

Parameters:

- Learning rate: 0.0001; Adam Optimizer; n-layers: 1; n-hidden: 64
- Results reported reflect the best performance before early stopping at 15 epochs

Results

Parser Results

Parser	Dev UAS	Test UAS
clearnlp	**79.76**	**79.59**
spacy	79.06	78.91
syntaxnet	72.54	72.81
Oracle	85.93	85.89

Ensemble Results

Text Features	Prosodic features	UAS Dev	UAS Test
POS, configs		80.69	80.73
POS, configs	Dur, dur log, pause	81.21	81.17
UAS Improvement		**0.51**	**0.44**
Sentences with improved UAS		268	240
Sentences with worse UAS		178	180
Sentences with the Same UAS		4970	5036
P-value (t-test comparing UAS values for all sentences)		< .001	< .001

- Using prosodic features in the ensemble classifier improves UAS over text-only features (configurations and parts of speech), and the top individual parser (clearNLP)
- Further improvement not related to prosody was achieved using lexical features combined with configurations and other features (Dev UAS: 83.47, an improvement of 3.7% absolute in UAS over the best parser)

Conclusions

- This study proposes a new feature, dependency configurations, that can be used in scoring dependency parse hypotheses
- Dependency configurations are related to prosodic breaks, and when used with prosodic features can improve parsing further
- Only pauses and normalized durations have been used as prosodic boundaries, while pitch and intensity can be used in further work with proper normalization