CROSS-LANGUAGE PHRASE BOUNDARY DETECTION

Victor Soto, Erica Cooper, Andrew Rosenberg and Julia Hirschberg

Columbia University and Queens College/CUNY

Main Findings

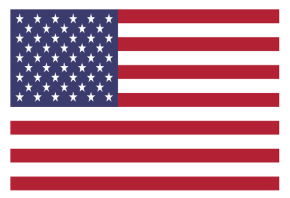
Models of prosodic phrasing trained on **multiple high-resource languages** are used to identify boundaries in an **unseen low-resource language.**

- While **pause** is the most important feature for predicting phrase boundaries in all languages, the annotation of pause varies.
- The relative importance of other features varies by language.
- Different **acoustic correlates** of prosodic boundaries characterize different languages. In some, the relative importance of features is silence > pitch > intensity > duration, while for other languages intensity is more important than pitch.

Motivation

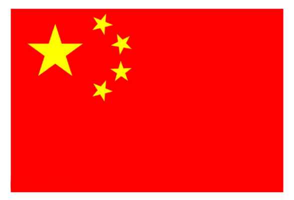
Uses of prosodic event detection:
Part-of-speech tagging, syntactic disambiguation, reducing language model perplexity, salience detection, distinguishing between given and new information, identifying turn-taking behavior and dialogue acts
Typically requires substantial hand-labeled data; not available for most languages.

Corpora



BDC

DIRNDL



DUR

Italian

Phrase Boundary Detection

- **Pause** features: whether the end of word precedes a silence, and duration of that pause.
- **Duration** features: the duration of the word and the difference of the duration of the current and following words.
- **Intensity** (dB) and **Pitch** (log Hz) contour features: raw and speaker-normalized signals at different level of aggregations (mean, maximum, minimum and standard deviation). Speaker normalization is performed by z-score normalization.

Cross-Language Phrase Boundary Detection

	Test Corpus							
	Full				Removed PPW			
Model	BDC	DIRNDL	DUR	Italian	BDC	DIRNDL	DUR	Italian
BDC	(0.79)	**0.89**	0.69	0.64	(0.00)	0.00	0.00	0.00
DIRNDL	**0.79**	(0.91)	**0.71**	**0.65**	0.00	(0,00)	0.00	0.00
DUR	0.74	0.80	(0.88)	0.64	**0.18**	**0.45**	(0.54)	**0.40**
Italian	0.43	0.61	0.61	(0.80)	0.00	0.04	**0.01**	(0.00)

Table: One vs One experiments F-Score results. Left columns show results for the full corpus and right columns show results after having removed pause-preceding words

	Baseline	Full		Remove PPW	
Test Corpus	Acc.	Acc.	F-Score	Acc.	F-Score
BDC	0.84	0.93	0.75	0.84	0.07
DIRNDL	0.88	0.98	0.92	0.89	0.19
DUR	0.63	0.83	0.73	0.63	0.00
Italian	0.89	0.90	0.67	0.90	0.33

Table: Leave-One-Out experiments

Within-Language Feature Analysis

Dataset	All	Silence	All but Silence	Int	F0	Duration
BDC	0.79	0.80	0.64	0.51	0.53	0.49
DIRNDL	0.91	0.91	0.62	0.55	0.20	0.00
DUR	0.88	0.71	0.84	0.81	0.40	0.39
Italian	0.80	0.77	0.69	0.71	0.48	0.40

Table: Within-language F-Scores values using feature subsets.

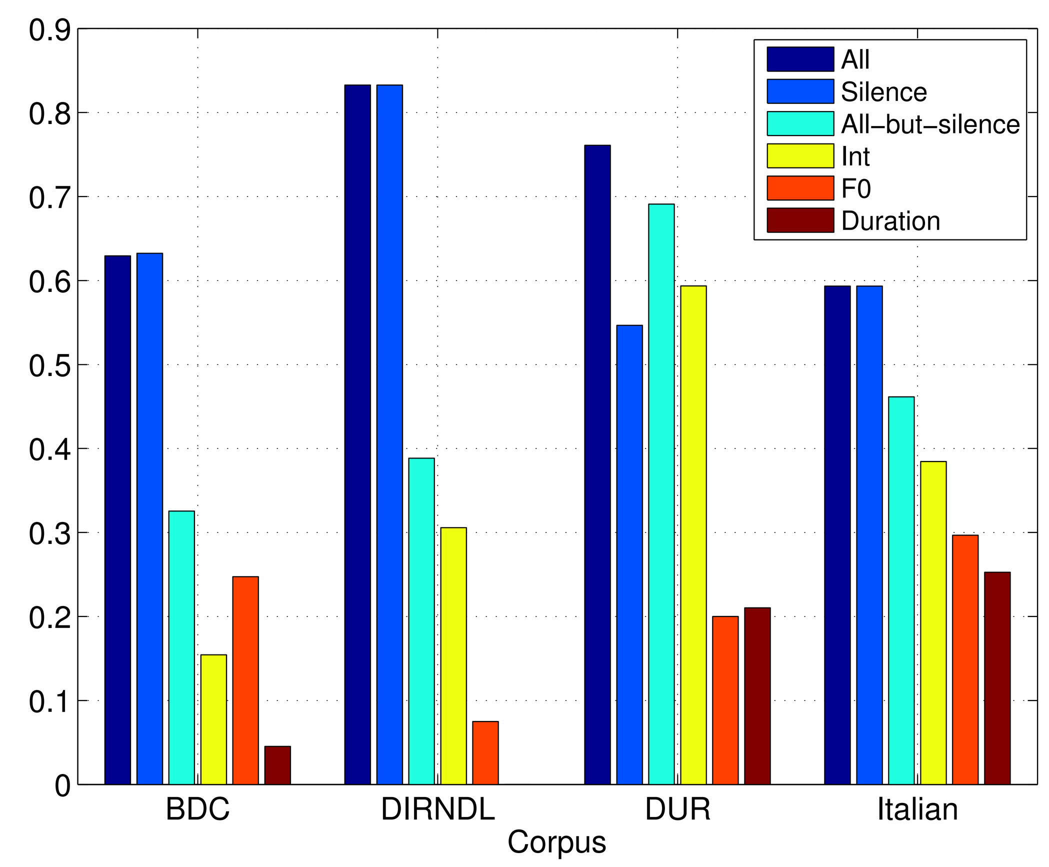
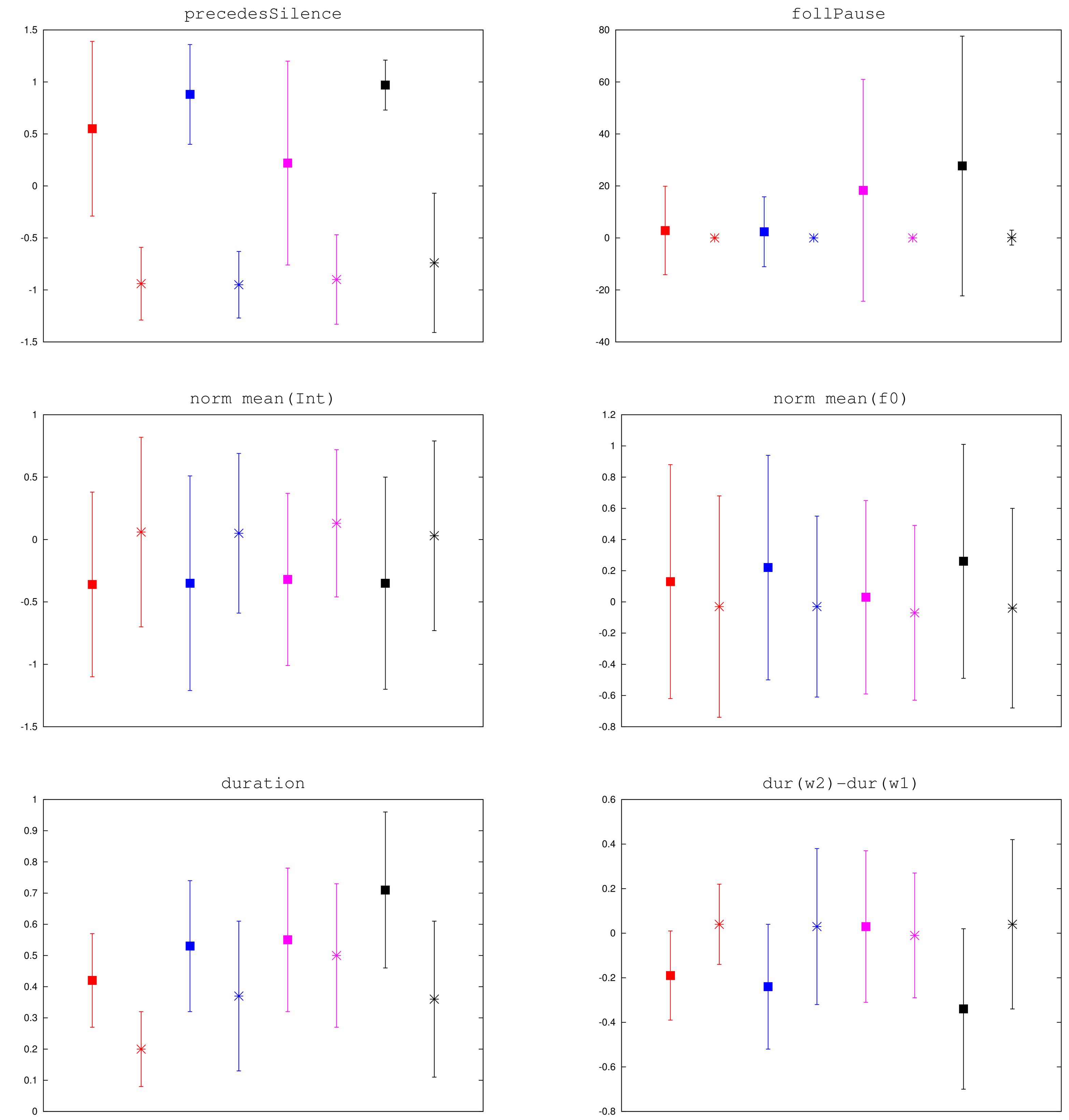


Figure: Relative error reduction using feature subsets.

Comparing Feature Distributions



BDC DIRNDL DUR Italian IPB non-IPB

Mean and std. dev. of example features from the four feature sets.

Corpus	BDC	DIRNDL	DUR	Italian
BDC	0.00	0.13	0.36	0.62
DIRNDL	-	0.00	0.20	0.59
DUR	-	-	0.00	0.42
Italian	-	-	-	0.00

Table: Mean KL-divergence values for each pair of corpus.

Future Work

- Cross-language adaptation
- Additional languages
- Examine which features of a language predict good cross-language performance