



RESCORING CONFUSION NETWORKS FOR KEYWORD SEARCH

Victor Soto, Erica Cooper, Lidia Mangu, Andrew Rosenberg and Julia Hirschberg

Columbia University, IBM and Queens College/CUNY



Main Findings

Introduce 2-stage cascaded scheme to rescore Confusion Networks (CNs) for Keyword Search in Low-Resource Languages.

1. Rescore CNs to improve error rate of 1-best hypothesis: Using rank-SVM classifier, obtain WER gains between 0.54% and 2.84%.
2. Generate keyword hits from rescored CNs and use logistic regression to detect true hits and false alarms. Gains between 0.45% and 0.9% in MTWV compared to hits generated from unrescored CNs

Keyword Search Task

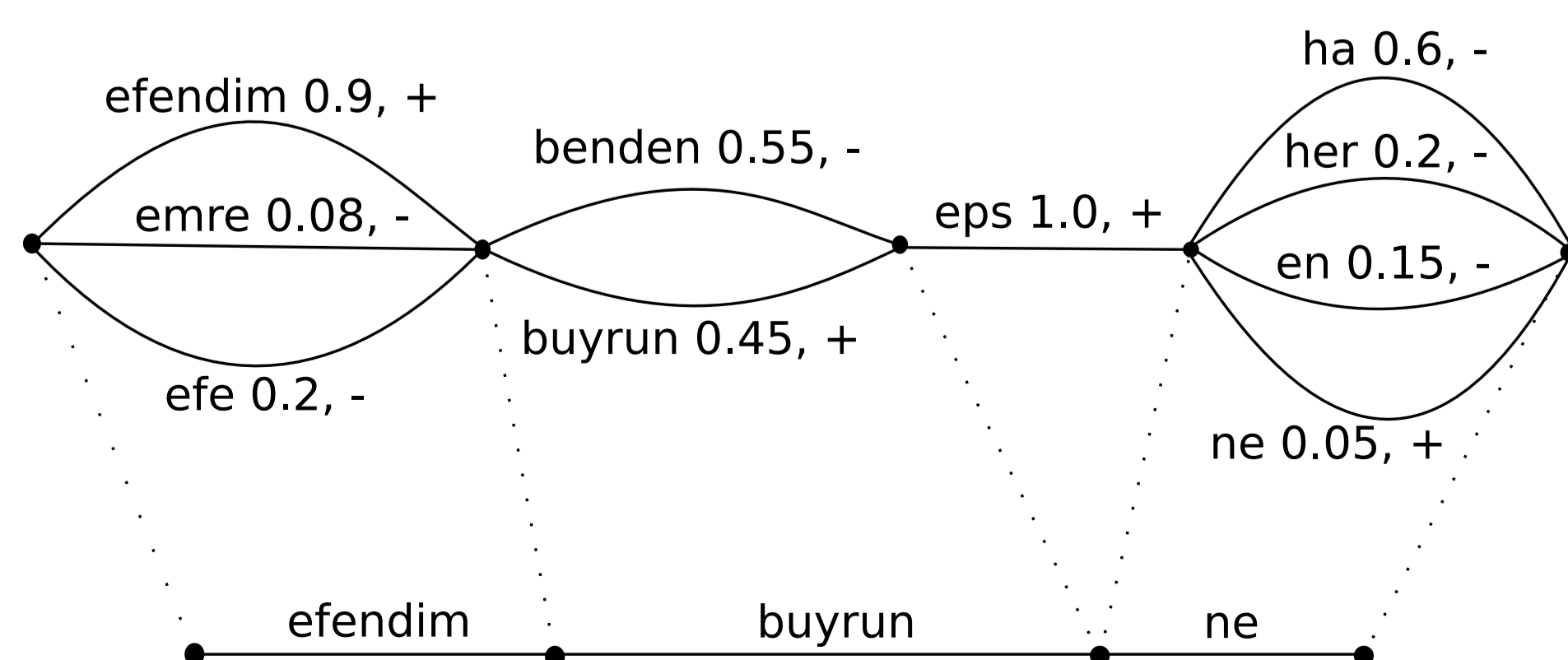
- ▶ IARPA Babel program: rapid creation of speech technology for a diverse set of languages with only a small amount of training data.
- ▶ Spoken keyword search task: identify all exact matches for some set of query terms, provided as text, in a given corpus of speech.
- ▶ Metric: Term-Weighted Value

$$ATWV(\theta) = 1 - [P_{\text{miss}}(\theta) + \beta \cdot P_{\text{FA}}(\theta)] \quad (1)$$

$$MTWV = \max_{\theta} ATWV(\theta) \quad (2)$$

Corpora

- ▶ IARPA Babel conversational speech, Full (FLP) and Limited (LLP) language packs, with 80 and 10 hours of training speech respectively. Dev and Eval partitions are 10 and 5 hours long.
- ▶ Cantonese, Tagalog, Turkish, Pashto, Vietnamese
- ▶ Word Confusion Networks are aligned to transcripts to obtain labels for each arc.



Feature Extraction and Feature Selection

Lexical :

- ▶ Percentile of word frequency
- ▶ Silence and epsilon flag
- ▶ Number of syllables of token, and syllable index of its primary and secondary stress.

Phonetic:

- ▶ Count of phones
- ▶ Binary features indicating whether word begins/ends in an unvoiced consonant or glottal stop.

Syntactic Proxies: model M class labels.

Structural:

- ▶ Posterior score
- ▶ Arc rank, arc rank - bin size ratio, confusion bin size,...
- ▶ Bin number at the segment and conversation level, distance to previous and next silence and to beginning and end of segment and conversation.

	Cantonese	Pashto	Turkish	Tagalog
CN	CN Posterior P(token + dev) token length chrs model-M # appearances	CN Posterior P(token + dev) pFA # arcs model-M	CN Posterior # ph(?) P(token recognized dev) # apps ends glottal stop	CN Posterior # ph(6w) P(token + dev) # ph(D) # ph(3)
	P(token + tokens in bin) percentile word freq. # ph(kw) segment duration is1English?	ends with glottal stop non-speech tag? reranked Posterior P(token + tokens in bin) # apps	starts glottal stop model-M non-speech tag? P(token + tokens in bin) percentile word freq.	P(token + tokens in bin) # arcs (bin) model-M non-speech tag? percentile word freq.
PL	pFA(kw) prod(r-post) posterior score min(r-post) reranked posterior fraction((eps)) mean (# arcs) min(p) GM(p) GM(r-cn)	pFA(kw) prod(r-post) min(r-post) reranked posterior posterior score min(p) fraction((eps)) GM(p) avg (# arcs) # arcs/# tokens	pFA(kw) min(r-post) prod(r-post) reranked posterior posterior score avg(# arcs) min (p) GM(r-cn) mean(# arcs) GM(p)	pFA(kw) prod(r-post) min(r-post) reranked posterior posterior score avg(# arcs) GM(r-cn) GM(p) min (p) # arcs/token

Table : Most prominent features for CN reranking (LLP) according to QPFS.

Probabilistic: probability of the arc being correct, the probability of the arc being correct, rank-normalized pFA and global re-ranked posterior scores. For PL rescoring include aggregated rescored posteriors from CN.

Acoustic:

- ▶ Median, mean, standard deviation, maximum, minimum and autocorrelation of pitch contour.
- ▶ Number and % of unvoiced cycles in the segment .
- ▶ Harmonics to noise ratio (dB) and noise to harmonics ratio.
- ▶ Number of pulses, number of periods and their mean and standard deviation.
- ▶ Number of voice breaks and their percentage, jitter values and shimmer values.

Prosodic:

- ▶ Intonational phrase boundaries and pitch accent detected by cross-language models.

Rescoring Confusion Networks

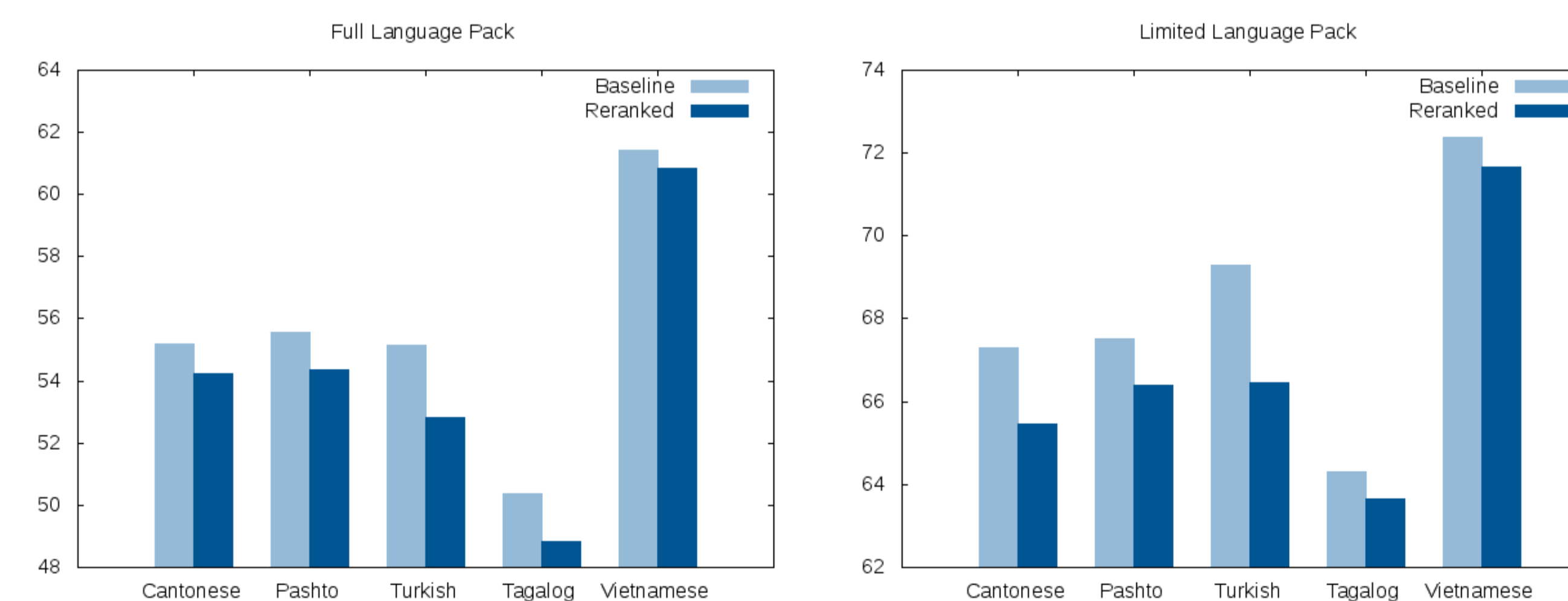


Figure : Token Error Rate results for every language (LLP and FLP).

- ▶ SVM ranks arc in each confusion bin.
- ▶ Positive WER gains from 0.56 to 2.34 for the FLP and 0.71 to 2.84 for the LLP.

Rescoring Posting Lists

Posting List: list of hypothetical hits including keyword id, start time, duration and score.

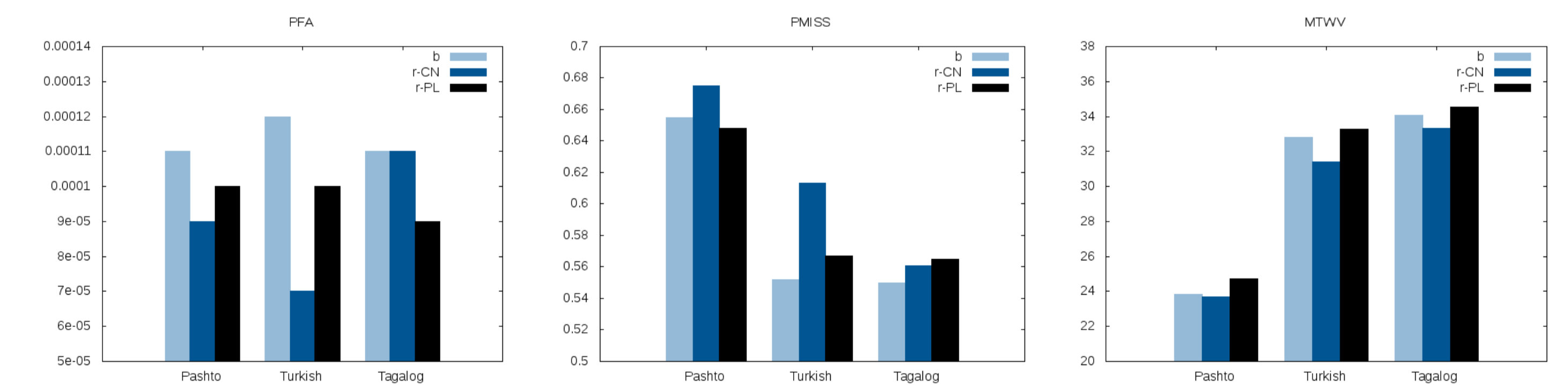


Figure : PFA, PMISS and MTWV results for IV keywords in the LLP conditions.

- ▶ Logistic Regression outputs probability estimate.
- ▶ Weighted F-measure: Miss Penalty > FA Penalty.
- ▶ Improvements between 0.45 and 0.9 MTWV points.

Future Work

- ▶ OOV word handling
- ▶ Cross-language experiments