



**HAL**  
open science

# Combined low level and high level features for Out-Of-Vocabulary Word detection

Benjamin Lecouteux, Georges Linarès, Benoit Favre

► **To cite this version:**

Benjamin Lecouteux, Georges Linarès, Benoit Favre. Combined low level and high level features for Out-Of- Vocabulary Word detection. Interspeech, 2009, Brighton, United Kingdom. hal-02088875

**HAL Id: hal-02088875**

**<https://hal.science/hal-02088875>**

Submitted on 3 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Combined low level and high level features for Out-Of-Vocabulary Word detection

Benjamin Lecouteux, Georges Linarès, Benoit Favre

## Principle

- Extract low and high level features (acoustic, search graph topology, linguistic) for each word
- Features are classified with a boosting algorithm
- A semantic modul refines the detection process

### Acoustic features :

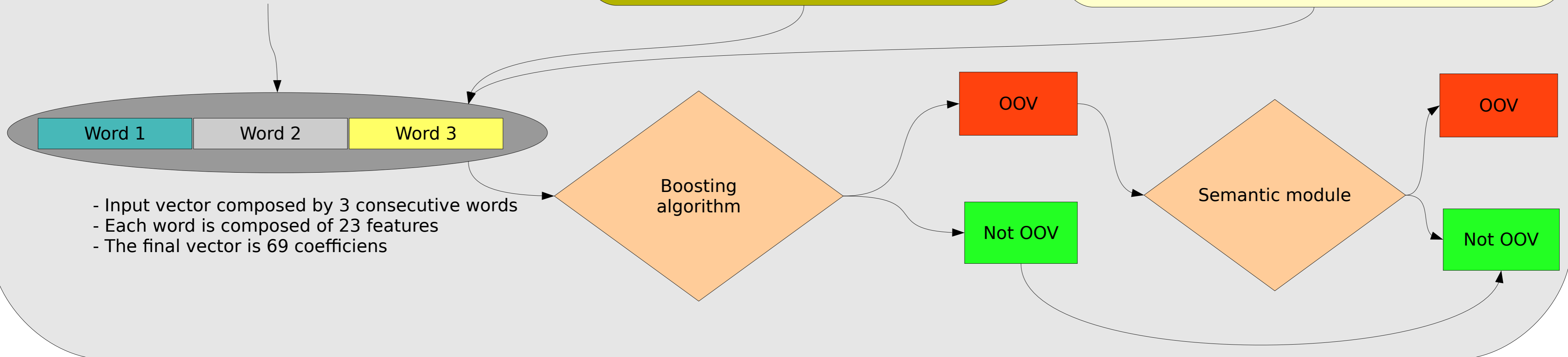
- Log-likelihood
- Average log-likelihood per frame
- Difference between the word log-likelihood and the unconstrained acoustic decoding of the corresponding speech segment.

### Linguistic features :

- 3-gram probabilities
- perplexity of the word in the window
- unigram probability
- Current backoff level of the word

### Graph features :

- Based on word confusion network
- posteriors
- number of alternative paths
- distribution of posteriors (min, max, mean)
- number of null links in the window (500 ms)



## The semantic module

### Combination of two measures :

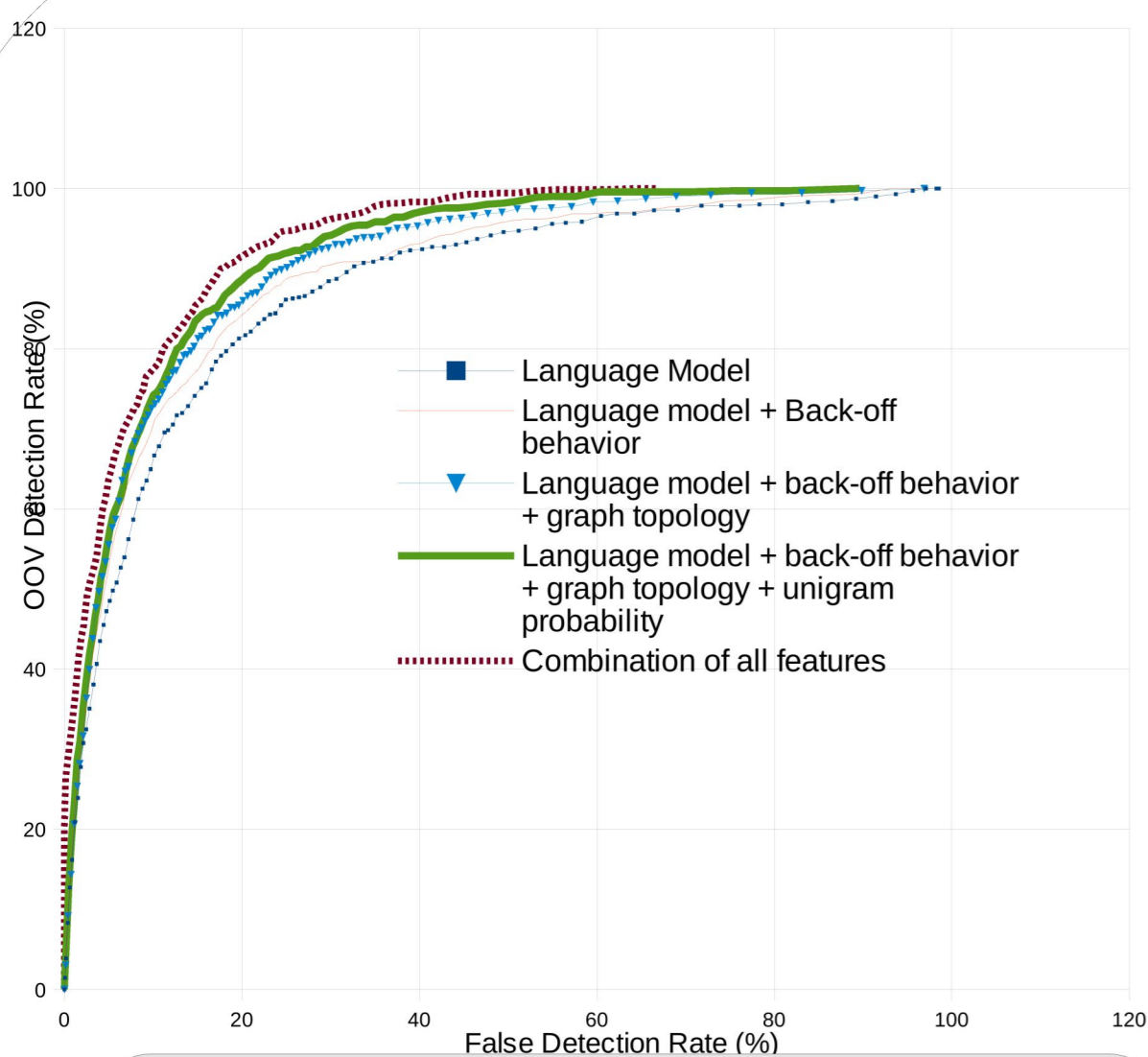
- WEB** : estimates the probability of word co-occurrences as a ratio of Google hits.
- Gigaword Corpus** : based on Latent Semantic Analysis (LSA), to estimate how much the targeted word is semantically close to the current segment.

## Experimental framework

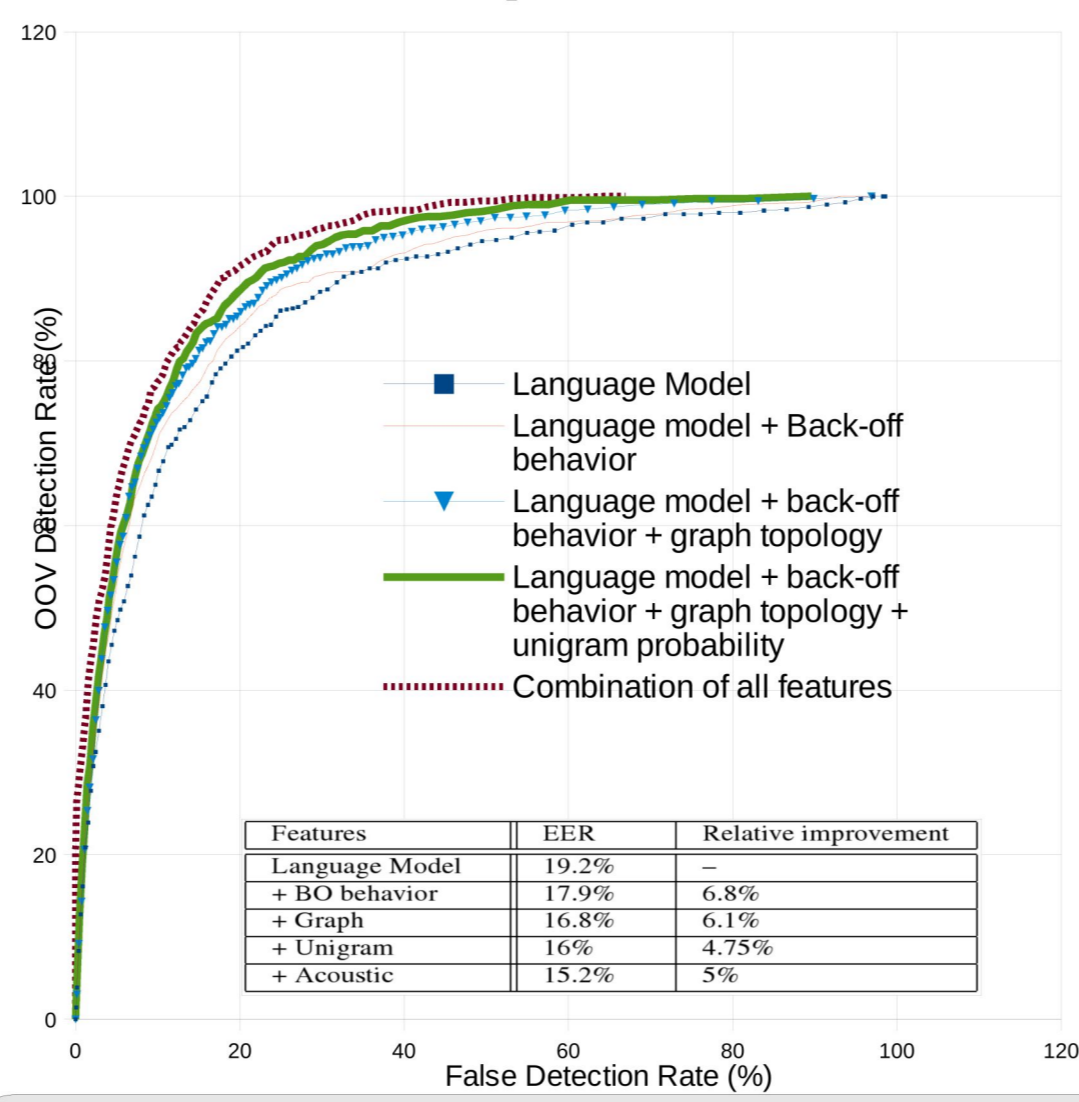
- The Lia broadcast news system used in ESTER campaign
- HMM-based decoder Speeral
  - Asynchronous decoder operating on a phoneme lattice
  - Acoustics models are HMM-based with cross word triphones
  - Language model is 3-gram estimated on 200M words
  - Lexicon is 67 Kwords
  - One pass in 3RT
- The ESTER Corpus :
- French radio broadcasts
  - Training for OOV provided by ESTER-2 train (100 hours)
    - 15K OOV and 1 M of words
  - The test is the ESTER test : 7 hours → 982 OOVs for 70011 word (1.04% OOV)
  - OOV test overlapping with OOV train is 2%

**Detection protocol :** OOV words have been manually specified by selecting all the reference words not available in the lexicon. During the detection, if a marked OOV word overlaps with a true OOV word, the true OOV word is considered as detected. In all other cases we consider a marked word as a false detection.

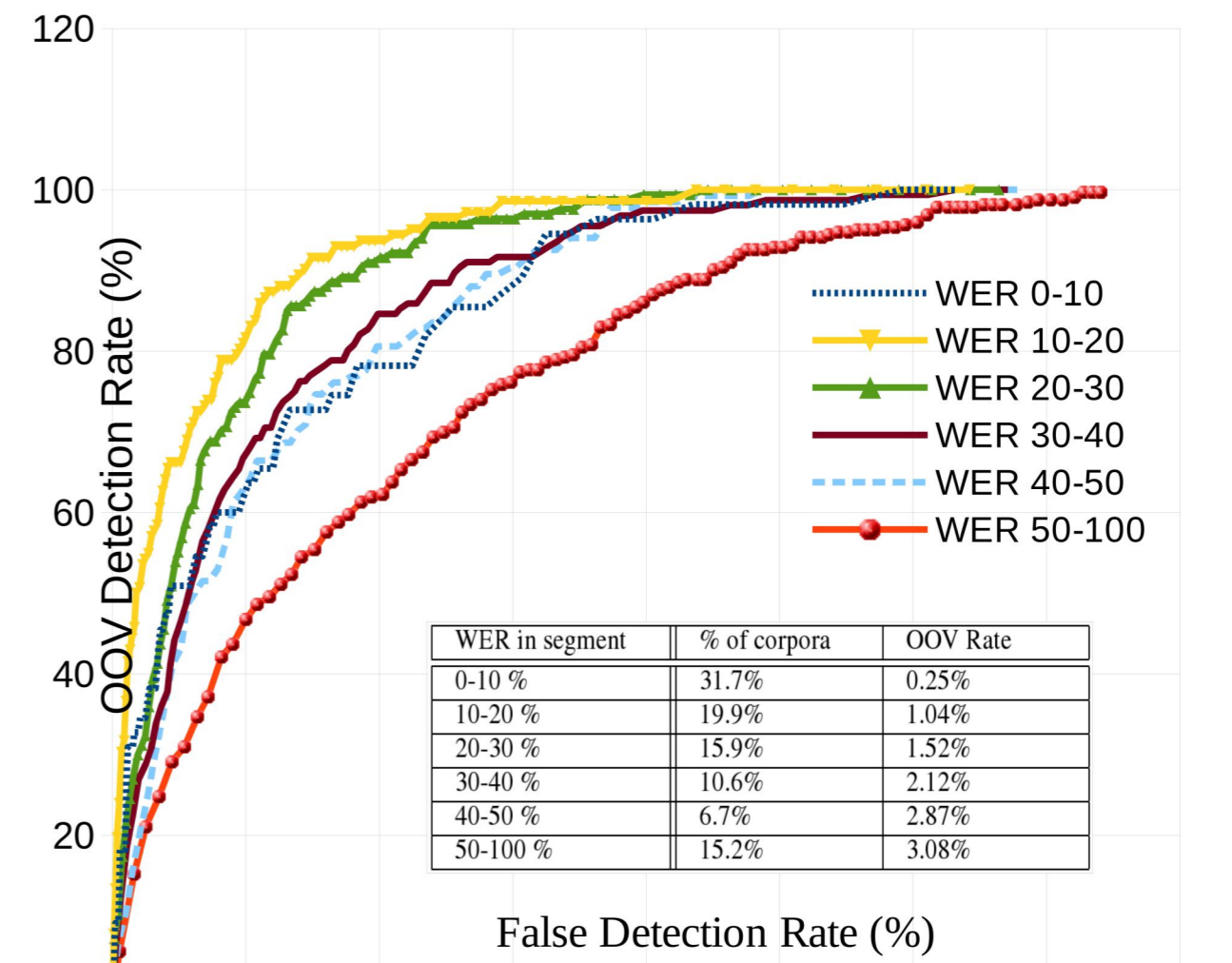
## Experiments



**Relevance of the features :** Results show that the most relevant descriptors are the linguistic and graph-topology features, while acoustic features and posteriors seem weaker in comparison to others.



**Complementarity of feature :** Each set provides some discriminative information. An unexpected result is the prominence of acoustic features. They show a significant improvement, while it exhibits the worst performance when other features are not used.



**OOV and WER :** Beyond 50%, noise decreases the accuracy. Globally, experiments show a relatively good robustness against WER, except when the ASR system dramatically fails.

## Semantic filtering and Conclusions

### Semantic filtering

Semantic module is used only on detected OOV words, with the purpose to refine detection on semantically coherent words. The ROC curves are presented in Figure 5. Results show that the filter reduces the EER by 4% relatively (15.2% to 14.6%), while the false detection rate decreases by about 5% relative.

### Conclusions

Our experiments showed promising results: OOV word detection seems to be homogeneous, despite the inconstant WER. The proposed method allows one to detect 43% of the unknown words with a 2.5% false acceptance rate, or 90% for 17.5% false acceptance. Experiments show that linguistic and graph-based features are the most relevant predictors. However, acoustic features associated to the others make the detection more robust. Finally, semantic filtering provides a slight but significant improvement.