# On word disambiguation features

Maria Boritchev

November 29, 2015

# 1 Integrating Multiple Knowledge Sources to Disambiguate Word Sense: An Exemplar-Based Approach

http://arxiv.org/pdf/cmp-lg/9606032.pdf
Authors: Hwee Tou Ng, Hian Beng Lee.

**Problem:** Word Sense Disambiguation (WSD).
**Input:** set of unrestricted, real-world English sentences. WARNING: the disambiguation is therefore internal to each sentence. Those sentences are pre-tagged with parts-of-speach (POS) tags ($N$, $V$, *Adj*, *Adv*), using existing software. Also: a list of ambiguous words mapped to their different meanings, those meanings being referenced to with numbers.
**Output:** each ambiguous word $w$ tagged with its correct sense number $i$ (context-dependent).

## 1.1 Feature extraction: $L_3, L_2, L_1, R_1, R_2, R_3, M, K_1, \ldots, K_m, C_1, \ldots, C_9, V$

$L_i$: POS of the word $i$-th position to the left of $w$ in the sentence $s$.
$R_i$: POS of the word $i$-th position to the right of $w$ in $s$.
$M$: morphological form of $w$ in $s$ (if $w$ is a noun ($N$), then sing./plur.; if $w$ is a verb ($V$), then prst. $3^d$ pers. sing./past participle/etc).

$K_1, \ldots, K_m$: features corresponding to a set of keywords that frequently co-occur with $w$ in the same sentence. This set is determined by conditional porbabilities (see below). For a sentence $s$, the value of feature $K_i$ is 1 if $K_i$ appears in $s$, 0 otherwise.
$V$: verb-object syntactic relation (useful in disambiguation of nouns).

## 2  Some observations

Other features we might take into consideration: types of prepositions in the sentence (place, time, ...).