

# Word Sense Disambiguation using Word Sketches

**Vít Baisa**

Lexical Computing

SKEW3, Brno, March 2012

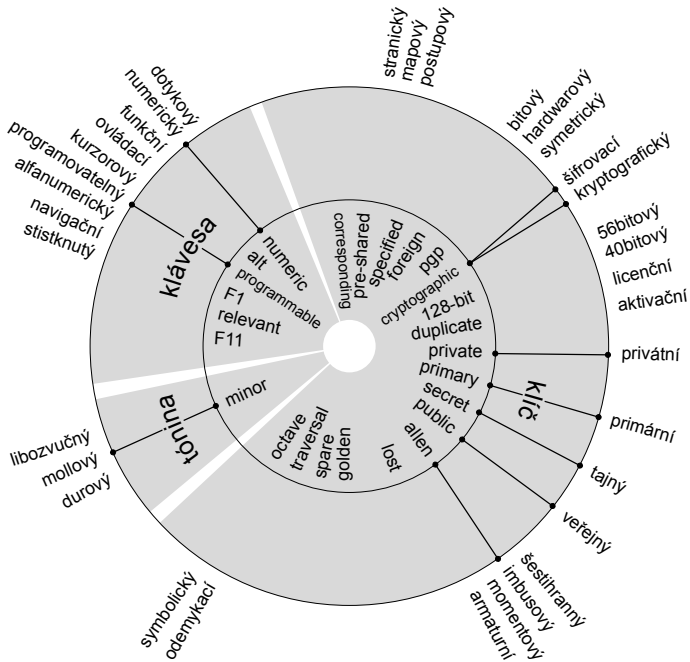
- WSD:
  - a bank of meanings
  - context within a sentence is predominantly used
  - a method for mapping words onto the bank

- WSD:
  - a bank of meanings
  - context within a sentence is predominantly used
  - a method for mapping words onto the bank
- bank of meaning:
  - WSD for MT: choose one proper translation from many
  - pairs of equivalents serve as meanings
  - meanings: *key-klíč*, *key-tónina*, *key-klávesa*
  - pairs may be polysemous

- WSD:
  - a bank of meanings
  - context within a sentence is predominantly used
  - a method for mapping words onto the bank
- bank of meaning:
  - WSD for MT: choose one proper translation from many
  - pairs of equivalents serve as meanings
  - meanings: *key-klíč*, *key-tónina*, *key-klávesa*
  - pairs may be polysemous
- context:
  - word sketches as typical contexts
  - czech, ukWAC, GNU-FDL dictionary

# Example of word sketch for lemma *key*

<b>a_modifier</b>	<b>object_of</b>	<b>n_modifier</b>	<b>modifies</b>
cryptographic	steal	cursor	element
primary	turn	ignition	stakeholder
programmable	remove	shift	point
minor	bend	backspace	area
golden	obtain	activation	aspect
lost	define	hash	principle
F11	enter	F	figure



- general links are rather uninteresting: *small key*
- unique links point at proper translations: *cryptographic key*
  
- original WSGs shared only one gramrel (a\_modifier)
- new WSGs were developed
- 26 compatible grammar rules

# Results by relations

Relation	EN	CS	AL	UL	AL%	UL%
be_adj	38.29	22.67	6.96	4.12	18.17	10.76
n_modifier	45.53	32.05	3.76	2.76	8.26	6.06
subj_be	39.29	31.10	5.17	3.61	13.16	9.19
a_modifier	43.61	<b>38.45</b>	9.71	5.65	22.26	12.96
has_obj	<b>48.39</b>	33.50	8.40	4.56	17.36	9.42
prec_prep	29.99	20.48	<b>15.97</b>	5.66	53.25	18.87
post_prep	<i>29.00</i>	19.78	15.61	5.51	<b>53.83</b>	<b>19.00</b>
modifier	39.17	34.39	15.08	5.97	38.50	15.24
and_other2	33.82	<i>13.62</i>	<i>3.57</i>	2.79	10.55	8.25
is_obj_of	43.40	32.66	11.68	<b>7.46</b>	26.91	17.19



# Overall results

# of retrieved words	44,249
# of retrieved polysemous words	19,316
avg # of Czech eq. per word	2.06
avg # of Czech eq. per polys. word	4.74
avg # of links per word	168.17
avg # of unique links per word	98.73
avg # of links per polysemous word	386.5
avg # of unique links per polys. word	225.84