Verbal Polysemy and Transfer Rules

Sanghoun Song
(sanghoun@uw.edu)

DELPH-IN summit X
2014 July 15

In a nutshell,

Using wordnets for (semi-)automatic creation of transfer rules

Idiomatic Expressions

(1) a. kick the bucket: die

b. kick the ball (object, ...)

c. kick the bottle: $\nparallel$ (a), $\parallel$ (b)

d. *kkangthong-ul cha-ta* 'bucket-ACC kick-DECL' $\nLeftrightarrow$ die

mek- 'eat'

Table : Frequency of the co-occurring objects

rank	noun	meaning	#
1	*pap*	'meal'	28
2	*swul*	'liquor'	22
3	*cenyek*	'dinner'	18
4	*maum*	'mind'	13
5	*umsik*	'food'	12
6	*nai*	'age'	9
6	*yak*	'medicine'	9
8	*kep*	'fear'	8
8	*cemsim*	'lunch'	8
10	*koki*	'meat'	7

cf. The *Sejong* Korean Treebank

mek- 'eat'

Table : Frequency of the co-occurring objects

rank	noun	meaning	#
1	**pap**	**'meal'**	**28**
2	*swul*	'liquor'	22
3	*cenyek*	'dinner'	18
4	*maum*	'mind'	13
5	**umsik**	**'food'**	**12**
6	*nai*	'age'	9
6	*yak*	'medicine'	9
8	*kep*	'fear'	8
8	*cemsim*	'lunch'	8
10	**koki**	**'meat'**	**7**

TYPE I (= eat): *mek-* is translated into just 'eat'.

pap/umsik/koki-(l)ul mek-ta
meal/food/meat-ACC eat-DECL
'eat a meal/some food/meat' [kor]

Selectional Preference Strength

mek- ‘eat’

Table : Frequency of the co-occurring objects

rank	noun	meaning	#
1	*pap*	‘meal’	28
2	*swul*	‘liquor’	22
3	**cenyek**	**‘dinner’**	**18**
4	*maum*	‘mind’	13
5	*umsik*	‘food’	12
6	*nai*	‘age’	9
8	*yak*	‘medicine’	9
8	*kep*	‘fear’	8
8	**cemsim**	**‘lunch’**	**8**
10	*koki*	‘meat’	7

TYPE II ($\approx$ eat): slightly different, but still translated into ‘eat’.

cenyek/cemsim-ul mek-ta
dinner/lunch-ACC eat-DECL
‘eat dinner/lunch’ [kor]

mek- 'eat'

Table : Frequency of the co-occurring objects

rank	noun	meaning	#
1	*pap*	'meal'	28
2	**swul**	**'liquor'**	**22**
3	*cenyek*	'dinner'	18
4	*maum*	'mind'	13
5	*umsik*	'food'	12
6	*nai*	'age'	9
6	**yak**	**'medicine'**	**9**
8	*kep*	'fear'	8
8	*cemsim*	'lunch'	8
10	*koki*	'meat'	7

TYPE III ($\geq$ eat): The corresponding words are different in English.

swul/yak-ul mek-ta
liquor/medicine-ACC eat-DECL
'**drink** liquor' / '**take** a medicine'
[kor]

mek- 'eat'

Table : Frequency of the co-occurring objects

rank	noun	meaning	#
1	*pap*	'meal'	28
2	*swul*	'liquor'	22
3	*cenyek*	'dinner'	18
4	**maum**	**'mind'**	**13**
5	*umsik*	'food'	12
6	**nai**	**'age'**	**9**
6	*yak*	'medicine'	9
8	**kep**	**'fear'**	**8**
8	*cemsim*	'lunch'	8
10	*koki*	'meat'	7

TYPE IV ($\neq$ eat): In this case, *mek-* does not directly denote an action of eating.

maum-ul mek-ta
mind-ACC eat-DECL
'**decide**' / '#eat a mind'

nai-lul mek-ta
age-ACC eat-DECL
'**get old**' / '#eat an age'

kep-ul mek-ta
fear-ACC eat-DECL
'**be frightened**' / '#eat a fear'
[kor]

Type IV

(2) a. maum/#cengsin-ul mek-ta
mind-ACC mek-DECL ⇔ 'decide'

b. nai/#yenlyeng-(l)ul mek-ta
age-ACC mek-DECL ⇔ 'get old'

c. kep/#kongpho-(l)ul mek-ta
fear-ACC mek-DECL ⇔ 'be frightened'

Type IV

maum-ul mek-ta
mind-ACC eat-DECL
'**decide**' / '#eat a mind'

These metaphorical expressions are crucial for facilitating **felicitous translations** in multilingual processing as well as in studying the **semantic restriction** on argument structure.

Properties

- These expressions are interpreted only in a metaphorical manner.
- There are very few or no alternative words.
- These expressions do not follow the principle of semantic compositionality.
- The relationship between the co-occurring items is language-specific.

Selectional Preferences & Verbal Polysemy

	Selectional Preferences	**Verbal Polysemy**
possibility	a class of nouns	a specific noun
distribution	open	fixed
semantics	compositionality	idiomatic expressions
meaning	word-by-word	(near) multiword
cognitive process	compositional	metaphorical
across languages	language-universal	language-specific

Selectional Preferences & Verbal Polysemy (cont'd)

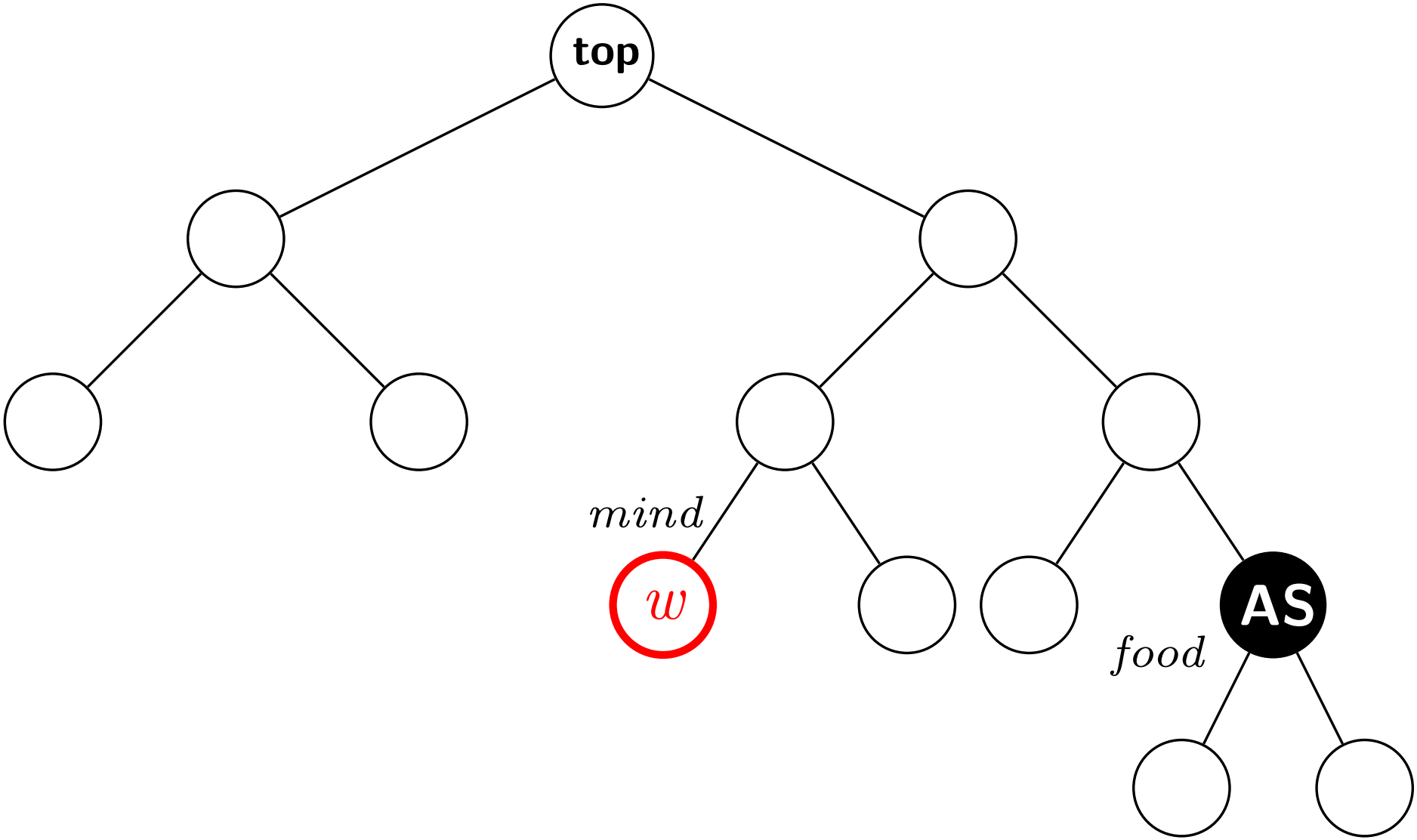
- A list of verbal polysemy can be acquired after identifying selectional preference strength of verbal items wrt lexical hierarchies.
- Finding the locus of a meaningful relation between a verb and its co-occurring objects within WordNets (a.k.a. Association Strength) plays the key role to determine whether each token conveys a metaphorical meaning.

Basic Data

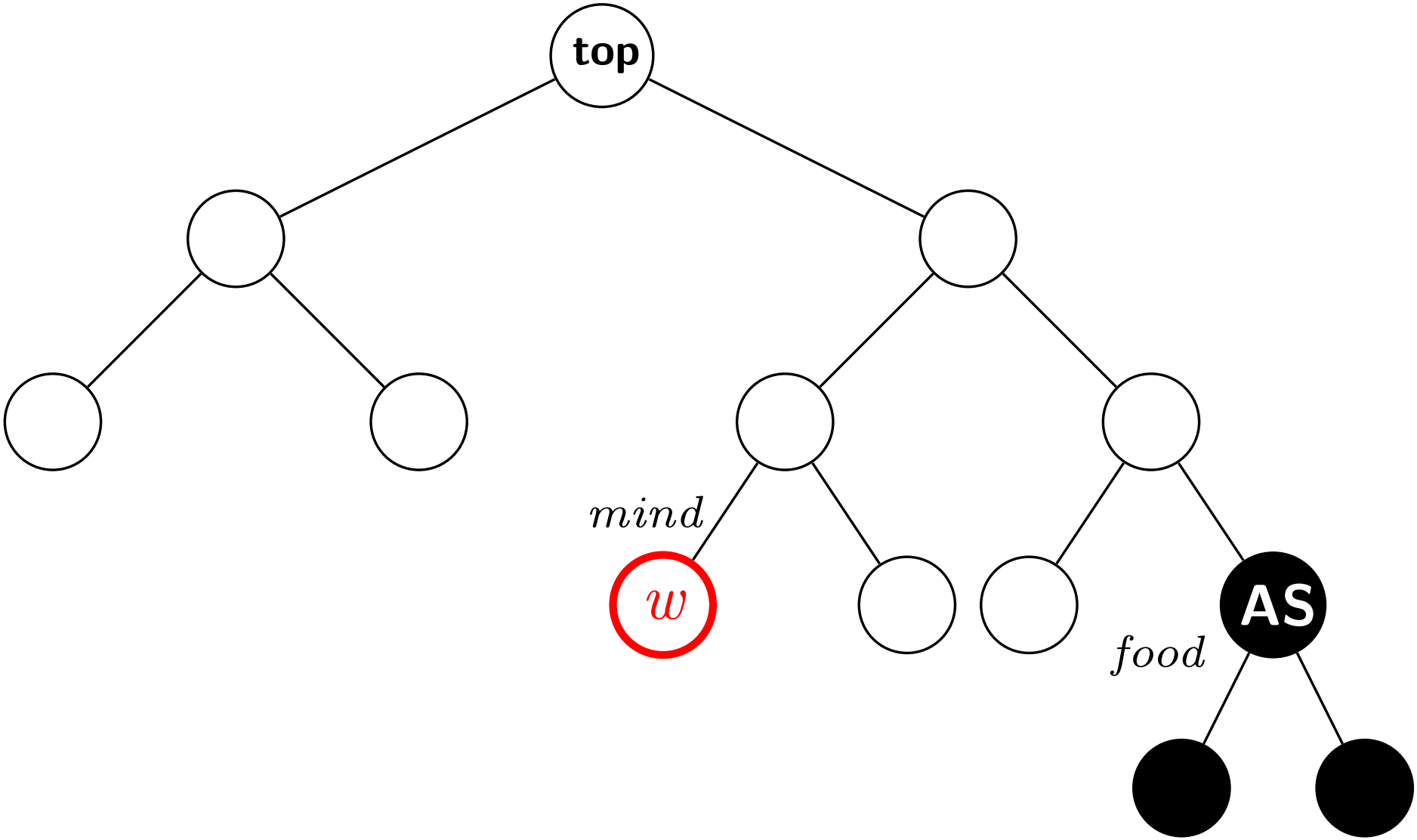
Basically two types of resources are required to calculate Selectional Preference Strength and to acquire a list of verbal polysemy.

- a development corpus (e.g. treebanks)
  - The *Sejong* Korean Treebank
- a lexical hierarchy (i.e. WordNet)
  - CoreNet
  - KorLex
  - U-WIN

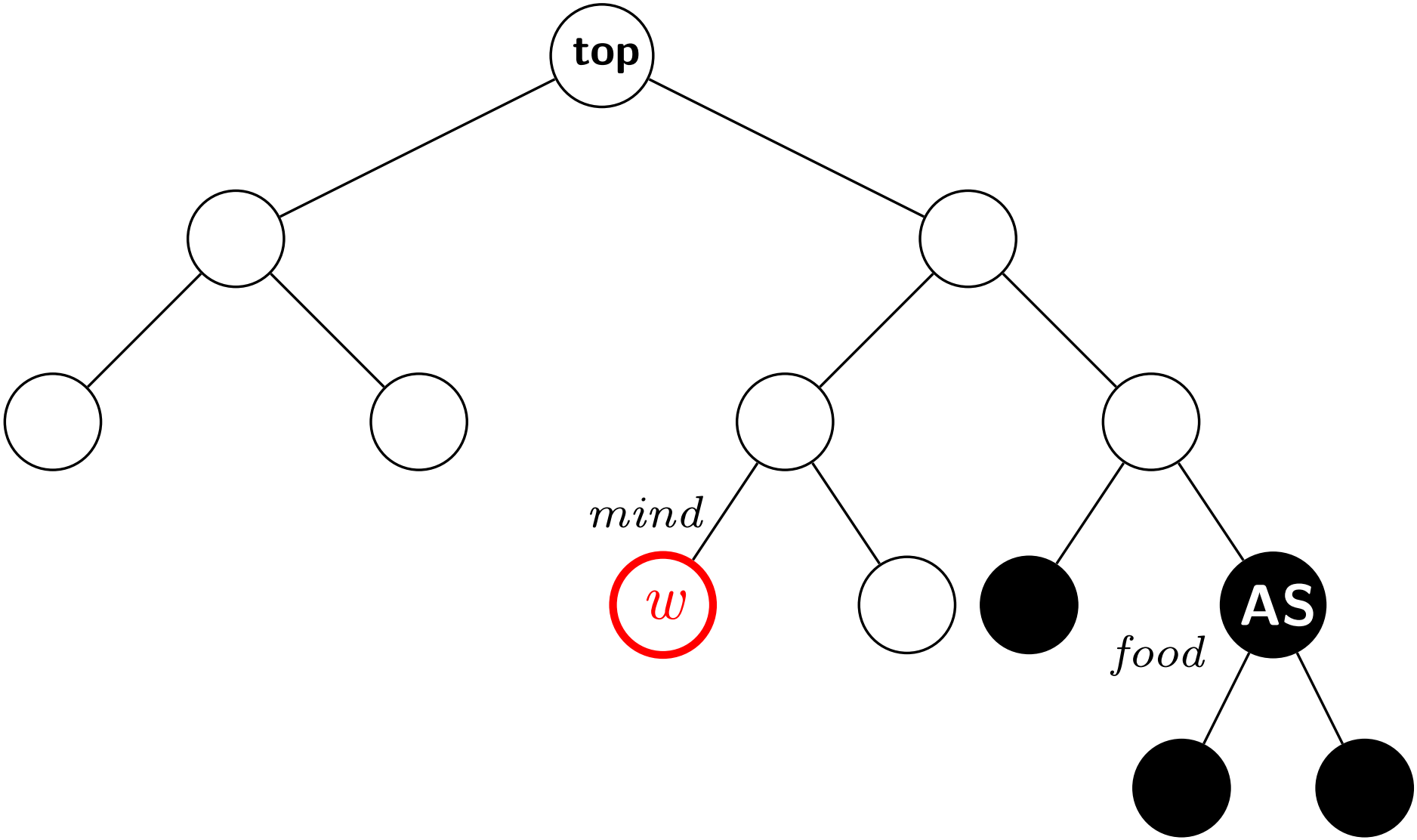
Kinship



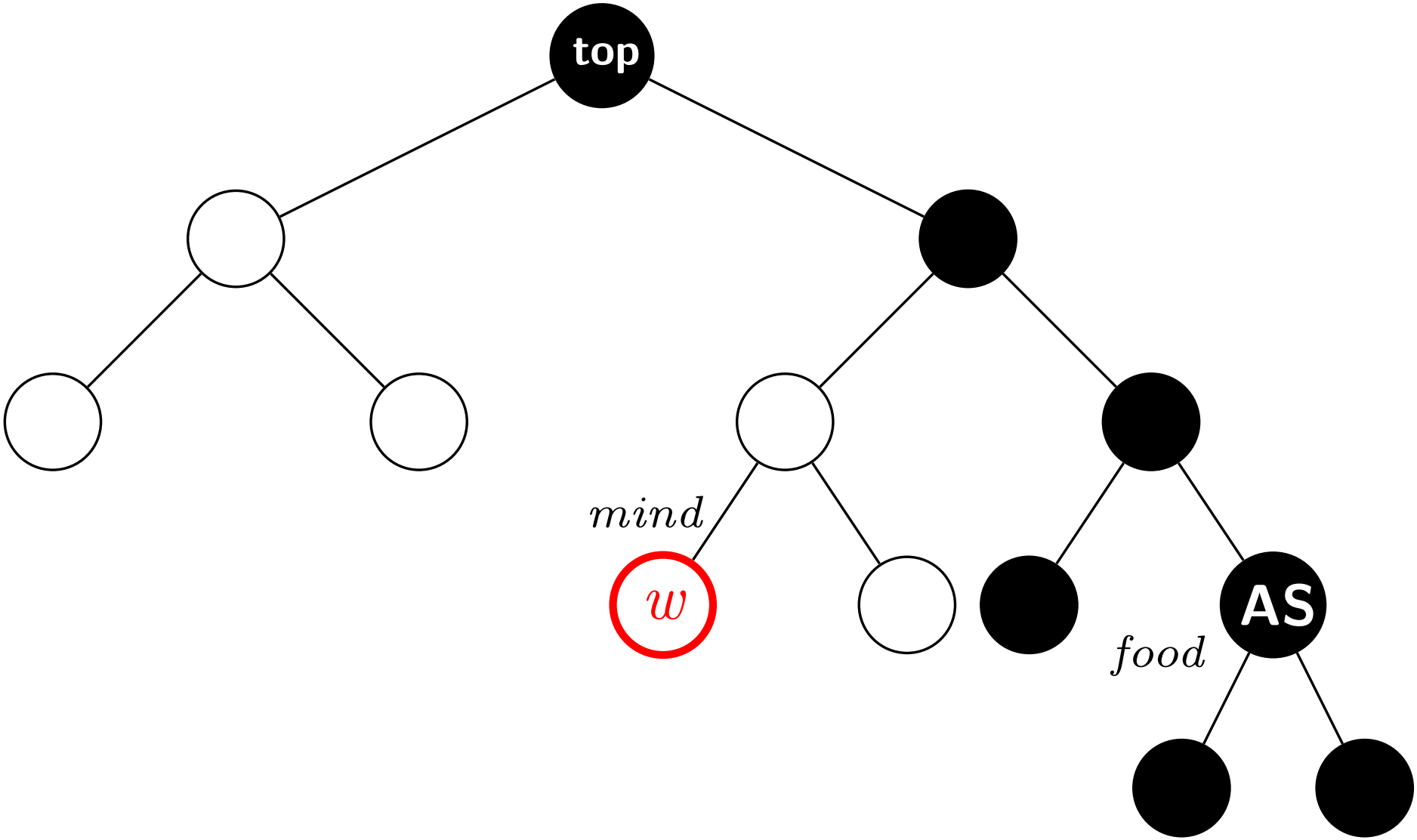
Kinship



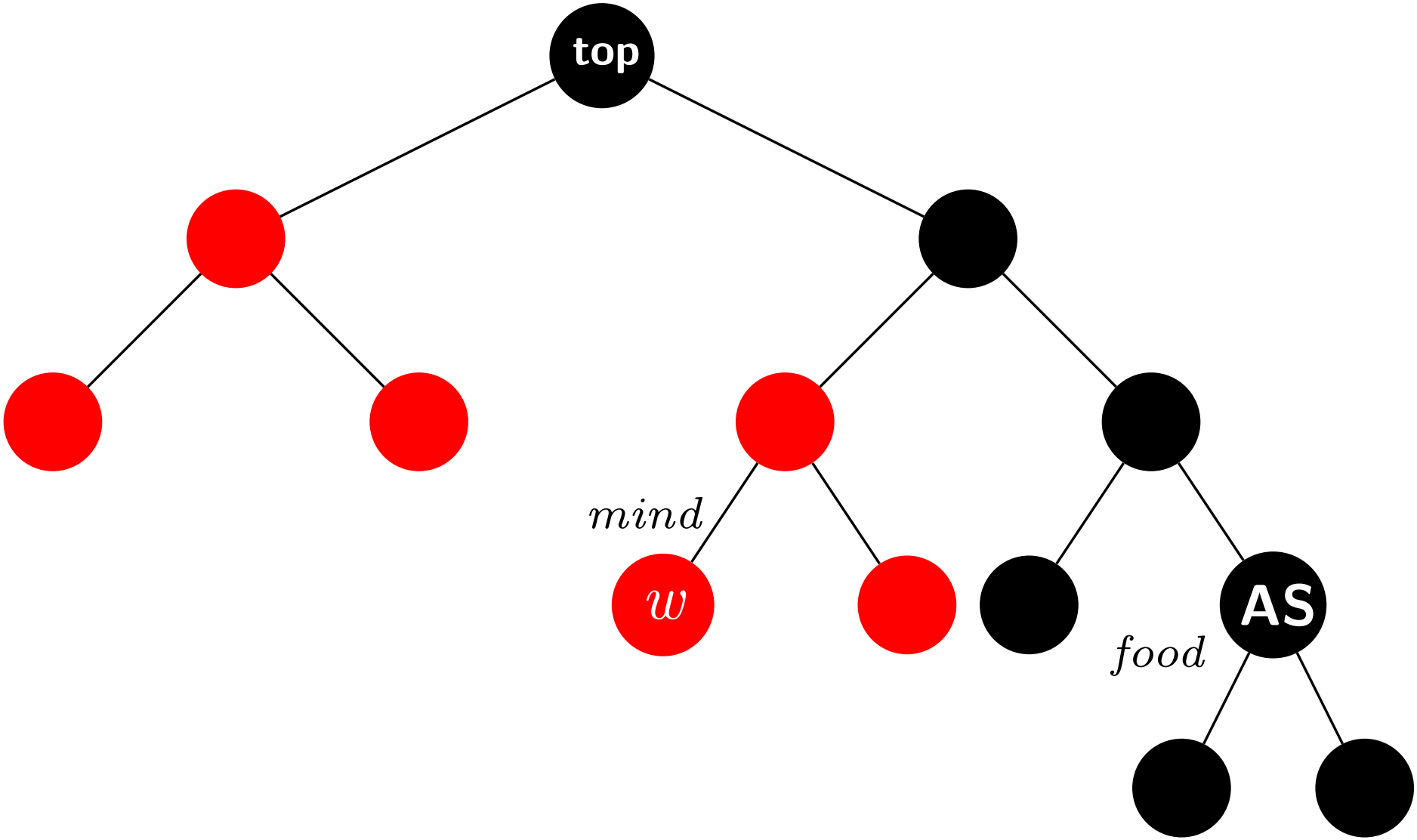
Kinship



Kinship



Kinship



mek- with CoreNet

rank	noun	meaning	SPS	type
1	cenyek	dinner	0.004399507	II
2	**maum**	**mind**	**0.003068423**	**IV**
3	**nai**	**age**	**0.002039022**	**IV**
4	**kep**	**fear**	**0.001788187**	**IV**
5	achim	breakfast	0.00154058	II
6	achimpap	breakfast	0.000822657	II
7	**yok**	**abuse**	**0.000594756**	**IV**
8	**ay**	**difficulty**	**0.000594756**	**IV**
9	**ton**	**money**	**0.000594756**	**IV**

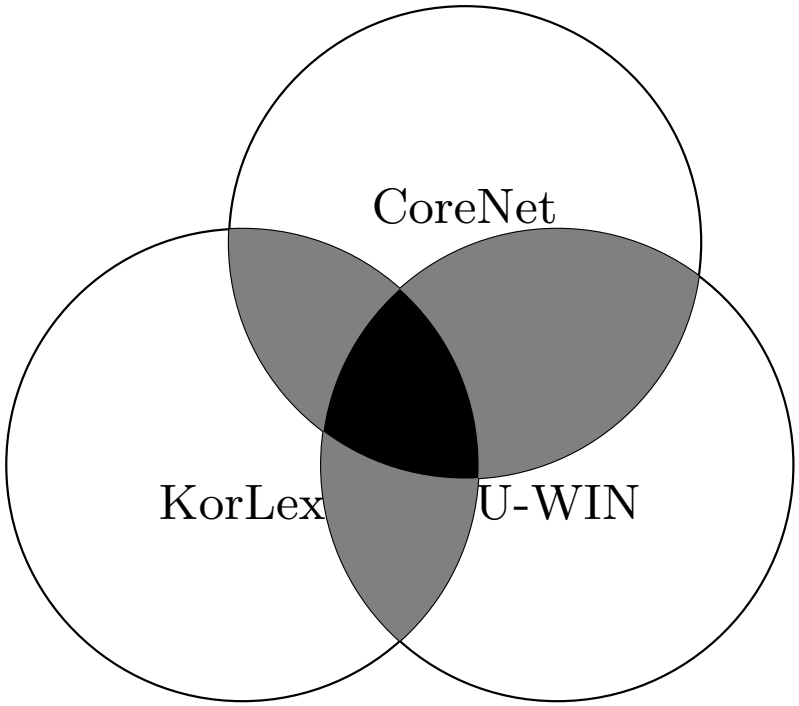
mek- with KorLex

rank	noun	meaning	SPS	type
1	**maum**	**mind**	**0.00439457**	**IV**
2	**kep**	**fear**	**0.002623486**	**IV**
3	**sayngkak**	**thought**	**0.002160117**	**IV**
4	**ton**	**money**	**0.000865791**	**IV**
5	**yok**	**abuse**	**0.000865791**	**IV**
6	**ay**	**difficulty**	**0.000865791**	**IV**
7	ppwuli	root	0.000631654	I
8	pan	half	0.00054467	etc.
9	ocinge	squid	0.00054467	I

mek- with U-WIN

rank	noun	meaning	SPS	type
1	swul	liquor	0.011460771	III
2	**maum**	**mind**	**0.007227785**	**IV**
3	**sayngkak**	**thought**	**0.005256684**	**IV**
4	so	ingredient	0.004335257	I
5	yak	medicine	0.00421386	III
6	**kep**	**fear**	**0.003690057**	**IV**
7	cengto	degree	0.003397661	etc.
8	**nai**	**age**	**0.003359211**	**IV**
9	mwul	water	0.002167629	III

Intersection



maum ‘mind’	⇔ ‘decide’
sayngkak ‘thought’	⇔ ‘think’
nai ‘age’	⇔ ‘get old’
kep ‘fear’	⇔ ‘be frightened’
ton ‘money’	⇔ ‘be bribed’
yok ‘abuse’	⇔ ‘be blamed’
ay ‘difficulty’	⇔ ‘be troubled’

Basic Measures

	CoreNet	KorLex	U-WIN
# of verbal entries	2,760		
# of verbs	1,447		
# of verbal nouns	1,313		
# of tokens of objects	27,044	27,365	26,899
# of types of objects	18,189	18,609	18,144
# of collected LCSs	46,052	32,787	22,259
# of polysemic verbs	236	360	305
# of polysemic relations	762	1,360	894
# of polysemic verbs ($\cap$)	254 (more than 2) / 137 (3)		
# of polysemic relations ($\cap$)	724 (more than 2) / 548 (3)		

A Sample Statement

```
eat-mind-mtr := monotonic_mtr &
[ INPUT [ RELS <! [ PRED "_eat_v_rel",
                    LBL #lbl,
                    ARG0 #arg0,
                    ARG1 #arg1,
                    ARG2 #arg2 ],
                  [ PRED "_mind_n_rel",
                    ARG0 #arg1,
                    LBL #larg ],
                  [ PRED "exist_q_rel",
                    ARG0 #arg1 ]!>,
          HCONS <! [ LARG #larg ] !> ],
  OUTPUT [ RELS <! [ PRED "_decide_v_rel",
                     LBL #lbl,
                     ARG0 #arg0,
                     ARG1 #arg2 ] !>,
          HCONS <! !> ]].
```

Future Work