

# Increasing harmony between the ERG and Wordnet

Francis **Bond**

Computational Linguistics Lab (凡土研)

Asian Studies

Palacký University, Olomouc

`<bond@ieee.org>`

DELPH-IN 2023



Faculty  
of Arts

# Overview

- Model more lexical semantics
- Add access to word senses in the DELPH-IN representations
- With the goal of richer representations and possibly better disambiguation (syntactic and semantic)
- We have several corpora with both senses and trees
  - ▶ Semcor
  - ▶ Gloss corpus
  - ▶ Sherlock Holmes
- So we can test out mappings



# Major Areas

- ERG missing word: *malcontent*, *tenoroon*, *technocrat*, ...
- WN missing word: *(a)*, *a lot*, *'d better*, *appstore*, ...
- WN sense indistinguishable
- MWEs



# Need to get the right predicate I

- Non-trivial even for single words  
 $\text{cat}_{n:1} \rightarrow \textbf{cat}_{n:1}$
- Harder for MWEs
  - ▶ not always the same choice in the two systems
  - ▶ the coverage is patchy in both
  - ▶ we need principled analysis to choose AND ways to represent efficiently
- It is not just a case of constituent/token size
  - ▶ *a very hot dog* can not be *hot dog*<sub>n:1</sub>
  - ▶ most grammars don't model this



# Need to get the right predicate II

- ERG sometimes decomposes, wordnet doesn't

*here*  $\rightarrow \text{in}_p \text{ DEF}_q \text{ here}_a \text{ place}_n$

*here*  $\rightarrow \text{in}_p \text{ this}_q \text{ place}_n$ ; no need for ADV

- ▶ We have added decomposed pronouns to WN

*here*<sub>n:1</sub> has-hypernym *location*<sub>n:1</sub>; quantified-by *this*<sub>a:1</sub>

*this*<sub>a:1</sub> has-hypernym *proximal*<sub>a:1</sub>



# English Resource Grammar vs Wordnet

Mapping Type	#	%	ERG	WN
unknown no match	48	0.3	comedians/nns	comedian
MWE	114	0.7	a+little	a_little
unknown match	136	0.9	flannel/nn	flannel
morphy	239	1.6	animate	animated
lemma+sense	274	1.8	look_v_like	look_like
ADJ+ly-ADV	405	2.6	usual	usually
mismatch	636	4.1	foul	foul-smelling
exact (ignore sense)	3,603	23.4	story_n_of	story
exact	9,948	64.6	depravity	depravity
Total	15,403	100		

- Not trivial to match lemmas (6% not matched at all)

Matching over SemCor by Pozen:2013



# Mismatches: A long and lovely tail

	ERG		Wordnet
take	v	of-i	take_advantage
rest	v	1	rest_on
step	n	1	steps
join	v	1	join_forces
hold	v	1	hold_out
come	v	1	come_off
well	x	deg	well-kept
troop	n	1	troops
fasten	v	cause-to	unfasten
grey	a	1	gray
moral	n	1	morals
let	v	go-of	let_go_of
late	a	for	later



- *Linus coppers his bets.*  
idiom processing with variables  
need to add a lot more
- *Sherlock knocked her up.*  
verb-particles are being added
- *They are few and far between.*  
not sure what to do here
- *They spilled the government beans.*  
add weird entries for *spill* and *bean*, constrain somehow?
- *I ate a hot dog* very non-compositional  
add as a word-with-spaces?
- *I bought a green tea chest* sort of compositional  
ignore for now, but we are losing something
  - ▶ *black and white television* “白黒テレビ” [white-black TV]
  - ▶ *guard dog*
  - ▶ *hunting dog*





# How to find compositional entries

- Assign MRS to all wordnet entries  
parse (expecting fragments)
- Find the (semantic) head and see if it is a hypernym  
if so — **compositional**
  - ▶ or we didn't find the head correctly
  - ▶ or the WN hierarchy is weird  
*prime minister*<sub>n:1</sub> is not *minister*<sub>n:1</sub> (but a *leader*<sub>n:1</sub>)  
fix WN, it allows multiple inheritance
  - ▶ ideally annotate component senses in wordnet, ...



# Acknowledgements

- Thanks to many, many Wordnet people
- Many students, especially Zina Pozen, Michael Goodman and Le Tuan Anh

Anyone else leaving on Saturday morning at 6:30am?

