

Toward an Integrated Semantic Framework for Lexical and Structural Semantics

Francis Bond

and [many more](#)

Linguistics and Multilingual Studies,
Nanyang Technological University

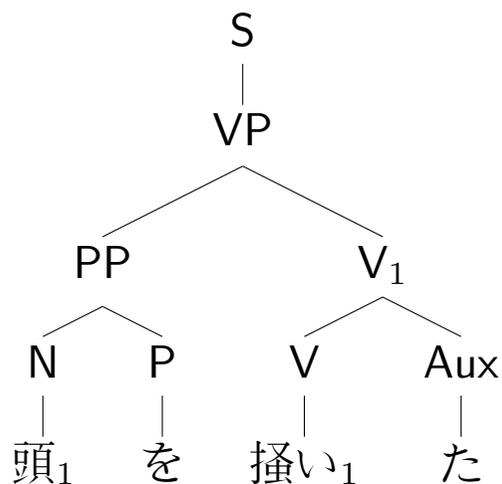
`<bond@ieee.org>`

Overview

- We are developing **rich multilingual meaning representations**
- Currently mainly sense annotation, about to start treebanking
- Goals
 - Scientific inquiry into how languages differ
 - Speeding up development of non-English by comparing analyses to English
 - Reference corpus for our Integrated Semantic Framework (MRS+WN+ α)
- **Spiral model**: annotate; improve model; re-annotate (update);
...
wordnet needs one or two more cycles of revision

Rich Representation

- (1) 頭 を 掻いた
 atama wo kaita
 head ACC scratched
 “I scratched my head.”



Syntax

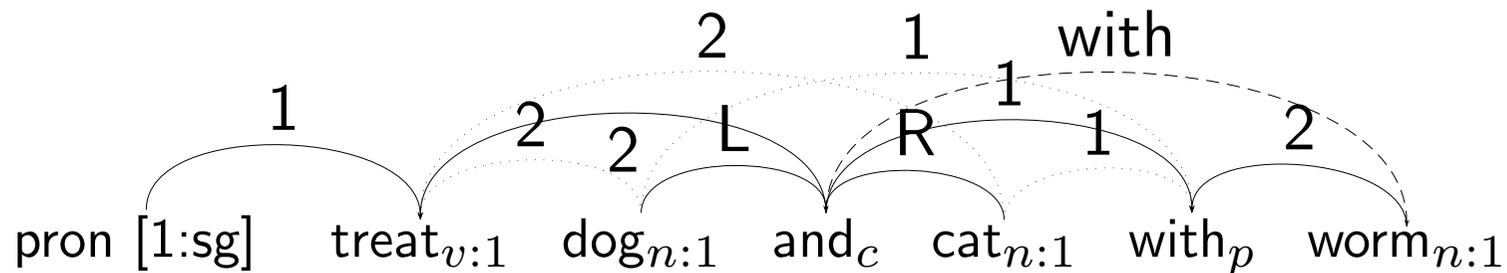
<i>atama</i> _{n:1}	is-a	bodypart
<i>kaku</i> _{v:1}	is-a	change
<i>kaku</i> _{v:1}	ARG1	Speaker
<i>kaku</i> _{v:1}	ARG2	<i>atama</i> _{n:1}
<i>kaku</i> _{v:1}	TENSE	past
Speaker	POSS	<i>atama</i> _{n:1}

Semantics

Extended features on the DMRS

I treat dogs and cats with worms.

HPSG grammars give us dependencies between (contentful) predicates (DMRS) and some information about MWEs.



Dashed lines show Preposition (P) features

Dotted lines show Conjunction (LR) features

Arc labels show the roles: 1 is ARG1, 2 is ARG2,

- Trivial for single words

$\text{cat}_{n:1} \rightarrow \mathbf{cat}_{n:1}$

- Harder for MWEs

not always the same choice in the two systems
and the coverage is patchy in both

- ERG sometimes decomposes, wordnet doesn't

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

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

- We have added decomposed pronouns

$\text{here}_{n:1}$ has-hypernym $\text{location}_{n:1}$; quantified-by $\text{this}_{a:1}$

$\text{this}_{a:1}$ has-hypernym $\text{proximal}_{a:1}$



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)

Mismatches: A long and lovely tail

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
stair	n	1	stairs
fasten	v	cause-to	unfasten
grey	a	1	gray
moral	n	1	morals
let	v	go-of	let_go_of
late	a	for	later

Our main research focus

- create a representation that gets the best of ERG+WN
- tackle hard MWEs
 - *Linus coppers his bets.*
 - *Sherlock knocked her up.*
 - *They are few and far between.*
 - *They spilled the government beans.*
- Have a tool to parse automatically and rank to get the preferred reading: for Chinese, English, Indonesian and Japanese

Concrete next steps

- Compare treebank and sense annotation
 - *Cathedral and the Bazaar*
 - *Speckled Band*
 - *Dancing Man*

- Decide how to annotate senses:
 - Ambiguate MRS?
 - Map MRS to concept table and then tag?

- First attempt:
 - Text → DMRS → synsets (ntumc tagger) → tagged (UKB)

Known problems

- Mapping MWEs
 - compound nouns suddenly become hard :-)
- Solution(s)?
 - Add non-compositional MWEs to the ERG
hot-dog; knock up; copper one's bets
 - Match with MT rules
 - ?

How to do this?

- Assign MRS to all wordnet entries
parse (expecting fragments)
should parse the glosses/examples too, just for fun
- Find the **head** and see if it is a hypernym
- use the parse to make the LHS of the MT rule
- Do this in CEJI (at least)
encourage Bulgarian, Spanish, Norwegian, . . .
automate and cross link

- Consistent Granularity
 - ERG (DELPH-IN) aims to only differentiate when there is a syntactic difference.
 - We want to have the same distinctions made consistently
 - * Argument structure of nominals
 - * Cross-POS links in wordnet

- The ontology/SEM-I is part of the model
 - We don't need to replace *destruction* with *destroy*
It is linked already (derivation)
So we can sidestep the granularity consistency problem of AMR

Some initial applications/tests

- Cross lingual ISF matching (DMRS matching but with synset nodes)
gives us MT rules; multilingual treebanking; MT
same technique can be used for symptom/guideline matching
- Various paraphrases
 - The literalizer: replace idioms/MWEs with literal paraphrases
She knocked Holmes up → *She woke Holmes*
 - The generalizer: replace concepts with hypernyms
She knocked Holmes up → *A woman changed the state of the fictional character* → *An entity changed an entity*
have a slider to change the depth

- The Joke generator
 - *When is a job not a job?*
When it is a nose job.

Differences with AMR

- We have a pipeline to build them automatically although the mapping needs to be refined
- Some expectation that our semantics will be more consistent under the assumption that the ERG/Wordnet are consistent
- We have less money but more existing work in multiple languages
- We don't do (yet)
 - Semantic Role Labeling
 - Named Entities (although if we use Babelnet we can)
 - Co-reference resolution
- We need to make time to do the shared task
annotate the same text with different representations

➤ How to represent idioms?

(2) *bite one's tongue* “refrain from speaking”

(3) *bite the dust* “die”

➤ What is the correct syntactic representation?

pretty much the same as non-idiomatic — different lexical items

➤ What is the correct wordnet representation

bite the dust is in the *die*_{v:1} synset?

bite one's tongue is a hyponym of *silent*_{a:2} (or see-also)?

➤ What about ISF?

one predicate for *bite the dust*

multiple for *bite one's tongue*? = “remain silent”

bite one's foolish tongue need somewhere to link it

Superposition

- Have a packed structure in which both parts are simultaneously true

*bite*_{v:i-bite-ones-tongue} *tongue*_{n:i-bite-ones-tongue}

\wedge *keep*_{v:1} *silent*_{n:i1}

mark them as being in the same group

- For e.g. generation, make sure you only generate from one

- The same predicate may take part in two deep things

Poisson and Gaussian distributions

*Poisson distribution*_{n:1} and *Gaussian distribution*_{n:1}

Terima Kasih

Thanks and more

- We would like to thank:
 - The Creative Commons Catalyst Grant: *Assessing the effect of license choice on the use of lexical resources*
 - The JSPS-NTU grant: *Revealing Meaning Using Multiple Languages*
 - The NTU Tier 1 grant: *Shifted in Translation*
 - The MOE Tier 2 grant: *That's what you meant: A Rich Representation for Representing Meaning*
 - NTU URECA projects
 - HG2002 students



References

Pozen, Z. (2013). Using lexical and compositional semantics to improve HPSG parse selection. Master's thesis, University of Washington.