Mapping between Compositional Semantic Representations and Lexical Semantic Resources: Towards Deep Accurate Semantic Parsing

Sergio Roa, Valia Kordoni and Yi Zhang

46th Annual Meeting of the ACL: HLT
Columbus, OH, USA
June 15-20 2008

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH
UNIVERSITÄT DES SAARLANDES

Main ideas

- Rich compositional semantic representations as parser output.
  - Minimal Recursion Semantics (MRS) gives more fine-grained semantic representations.
- Shallow semantic parsers use lexical semantic resources for evaluation. (PropBank, VerbNet, FrameNet)
- Modelling of a probability distribution for the mapping between MRS and PropBank/VerbNet.

Main ideas

- Features are taken from Robust Minimal Recursion Semantics (RMRS) representation, used by the ERG English Resource Grammar (Head-Driven Phrase Structure formalism).
- Use of WordNet semantic network to reduce complexity of the model (define selectional restrictions).
- Use of the PropBank corpus for evaluation and VerbNet $\leftrightarrow$ PropBank mapping.

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH
UNIVERSITÄT DES SAARLANDES

RMRS representation

Some plans give advertisers discounts

TEXT Some plans give advertisers discounts

TOP h1

RELS {

prop-or-ques_m_rel	
LBL	*h1*
ARG0	*e2* tense=present
MARG	*h3*

_some_q	
LBL	*h6*
ARG0	*x7* num=pl pers=3
RSTR	*h8*
BODY	*h9*

_plan_n	
LBL	*h10*
ARG0	*x7* num=pl pers=3

_give_v	
LBL	*h11*
ARG0	*e2* tense=present
ARG1	*x7* num=pl pers=3
ARG2	*x13* num=pl pers=3
ARG3	*x12* num=pl pers=3

udef_q_rel	
LBL	*h14*
ARG0	*x12* num=pl pers=3
RSTR	*h15*
BODY	*h16*

generic_entity_rel	
LBL	*h17*
ARG0	*x12* num=pl pers=3
CARG	*advertisers*

udef_q_rel	
LBL	*h18*
ARG0	*x13* num=pl pers=3
RSTR	*h19*
BODY	*h20*

_discount_n	
LBL	*h21*
ARG0	*x13* num=pl pers=3

}

HCONS {h3 qeq h11, h8 qeq h10, h15 qeq h17, h19 qeq h21}

ING {}

PropBank semantic roles

- Some plans give advertisers discounts

[$_{\text{Arg}_0}$ Some plans] *give* [$_{\text{Arg}_2}$ advertisers] [$_{\text{Arg}_1}$ discounts]

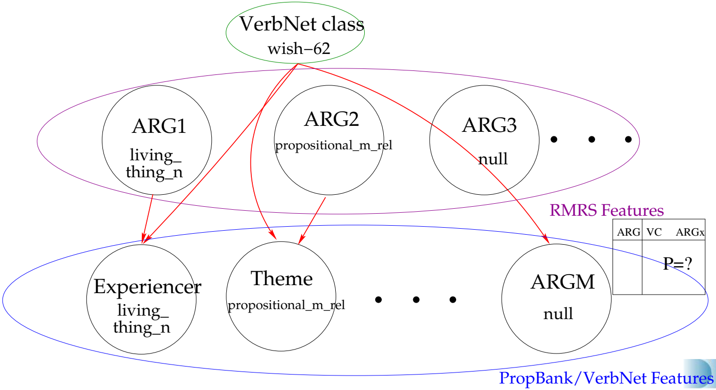
VerbNet thematic roles and classes

31 thematic roles, 237 top-level classes, and 194 new subclasses based on the classification by Levin.

Class: *Give-13.1*		
Parent: –		
Themroles: Agent Theme Recipient		
Selrestr: [+animate or +organization] [+animate or +organization]		
Frame:		
Name	**Example**	**Syntax**
Dative	Some plans give advertisers discounts	Agent V Recipient Theme

Bayesian network based mapping.

- Features: PropBank/VerbNet roles, RMRS, WordNet.



Mapping between RMRS and VerbNet/PropBank

- Other features: adjectives, prepositions (SEM-I classification), verbs, other SEM-I relations.
- Example of SEM-I verb relations:

Predicate			**Synopsis**	
_show_v_1_rel	ARG0 *e*	ARG1 *p*	ARG2 *u*	[ARG3 *h*]
_show_v_1_rel	ARG0 *e*	ARG1 *p*	ARG2 *x*	ARG3 *x*
_show_v_as_rel	ARG0 *e*	ARG1 *p*	ARG2 *h*	
_show_v_up_rel	ARG0 *e*	ARG1 *p*		

Extraction of RMRS and SEM-I features

- *Some plans give advertisers discounts*

SEM-I roles	Features	Words
ARG1	`plan_n`	plans
ARG2	`discount_n`	discounts
ARG3	`generic_entity_rel`	advertisers

- WordNet features (up to 5th level) may replace the nouns features.

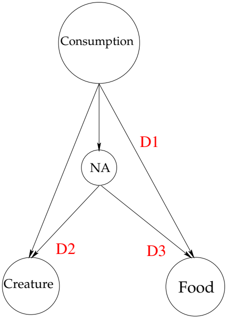
Alignment between RMRS and lexical resources

- PropBank/VerbNet annotation: [$_{\text{Arg}_0\text{(Agent)}}$ Some plans] *give* [$_{\text{Arg}_2\text{(Recipient)}}$ advertisers] [$_{\text{Arg}_1\text{(Theme)}}$ discounts]
- Alignment process:

SEM-I roles	Mapped roles	Features
ARG1	Agent	`plan_n`
ARG2	Theme	`discount_n`
ARG3	Recipient	`generic_entity_rel`

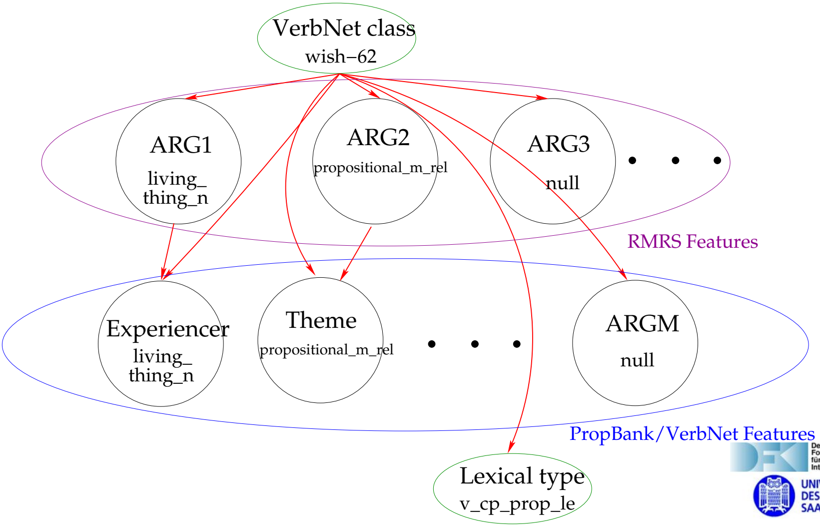
Verb classes inference

A possible realisation of a bayesian network for semantic verb classes (following Sabine Schulte im Walde)



P(Food)=P(Food | Na)=?

A priori bayesian structure for VerbNet classes inference:



Learning and Inference

- **Learning phase**:
  - *Maximum Likelihood (ML)* for the a priori structure, generating conditional probability tables by using multinomial distributions.
  - *Structural Expectation Maximization (SEM)*. Iterative process when parameters and structures are updated based on the best network so far.
- **Inference**: Markov Chain Monte Carlo (MCMC) inference engine (sampling process considering the evidence).

Learning algorithm

- Algorithm for training Bayesian Networks for inference of lexical semantic roles

procedure *Train (Model)*

```
for all Verbs do
    for all Sentences and Parsings which include the current verb do
        Initialize vertices of the network with SEM-I labels and features.
        Initialize vertices with the corresponding VerbNet class.
        Initialize edges connecting corresponding features.
        Append the current features as evidence for the network.
    end for
    Start Training Model for the current Verb, where Model is ML or SEM.
end for
```

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH
UNIVERSITÄT DES SAARLANDES

Results for inference of mapped roles

10370 sentences of the PropBank corpus succesfully parsed using the ERG grammar.

Corpus	Nr. iter. MCMC	Mode	WordNet feature	Verb classes	Accuracy %
PropBank with VerbNet features	1000	ML			78.41
	10000	ML			84.48
	10000	ML		×	87.92
	1000	ML	×		84.74
	10000	ML	×		86.79
	10000	ML	×	×	87.76
	1000	SEM			84.25
	1000	SEM	×		87.26
PropBank with PropBank features	1000	ML			87.46
	1000	ML	×		75.70
	1000	SEM			90.27

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

UNIVERSITÄT DES SAARLANDES

Preliminary results for inference of VerbNet classes

- The inference of a VerbNet class can help disambiguate the parsing.
- Regardless of whether [$_{Theme}$ you] hike from lodge to lodge or **stay**$_{\text{LODGE-46}}$ [$_{Location}$ *in one place*] and take day trips, there are plenty of choices.
- Using ERG, an incorrect mapping for the verb *stay* to the VerbNet class EXIST-47.1-1 with the (falsely) favored parse where the PP *"in one place"* is treated as an adjunct/modifier is derived.

Discussion

- Reliable mapping between MRSs structures and lexical semantic resources.
- Potential improvement of the parse disambiguation task by using verb semantic information.
- Enrichment of shallow semantic information with compositional semantic structures can eventually be helpful for applications like question answering.

The End...

Thanks for your attention!

Results of alignment

Corpus	Nr of parsings	Parsability	Alignment %
PropBank with VerbNet features	10	25.67%	80.65%
PropBank with PropBank features	25	26.54%	84.15%

Pseudocode alignment

- Algorithm for the alignment of SEM-I and PropBank (VerbNet) labels

```
procedure AlignRMRSPropBankLabels (Sentence, Parsing)
for all Verbs at the current Sentence and Parsing do
    for all SEM-I argument role labels do
        for all PropBank (VerbNet) argument role labels do
            if All the words in current SEM-I argument are found in PropBank
            argument then
                Map SEM-I and PropBank (VerbNet) roles.
                Do not allow more mappings for these labels.
            end if
        end for
    end for
    Obtain the VerbNet semantic class for the current verb.
    Obtain the leaf lexical type for the current verb.
end for
```