

MRS to DRS

Prescott Klassen
University of Washington
DELPH-IN Summit 2013
Sankt Wendel, Saarland
July 31, 2013

MRS/DRS

- MRS
 - Minimal Recursion Semantics (MRS) (Copestake et al., 2005), as implemented in the English Resource Grammar (ERG) (Flickinger, 2000)
 - ACE 0.9.13¹ and ERG 1111²
- DRS
 - Discourse Representation Structure (DRS), described formally in van Eijck (2006), and based on the Discourse Representation Theory (DRT) of Kamp (1981) and Kamp and Reyle (1990)
 - Groningen Meaning Bank (GMB) 2.0.1³ (Basile et al., 2012)
 - C & C/Boxer 2410 (Curran et al., 2007)⁴

¹<http://sweaglesw.org/linguistics/ace/>

²<http://www.delph-in.net/erg/>

³<http://gmb.let.rug.nl/>

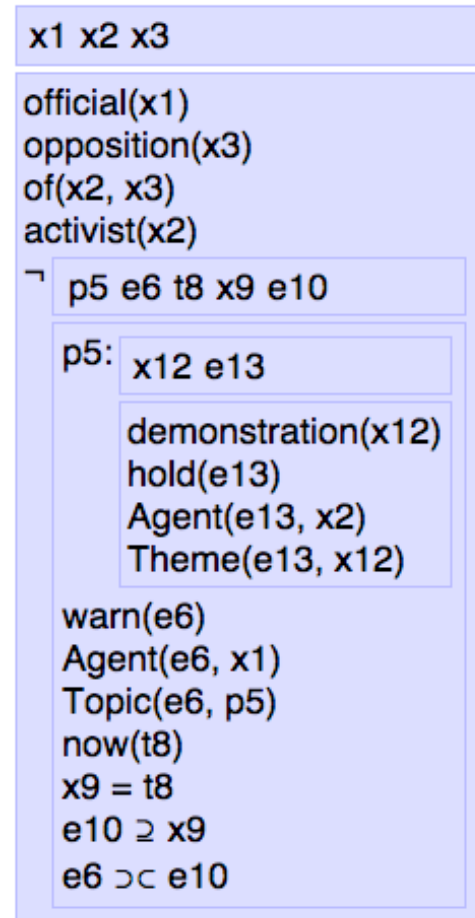
⁴<http://svn.ask.it.usyd.edu.au/trac/candc>

Motivation

- University of Washington LING 575 Seminar Project
 - Explore mapping of MRS to X semantic representation
- MRS as an alternate semantic analysis for building DRS
 - integration with Boxer
 - replacing Boxer with an MRS-Boxer
- Use MRS in place of surface strings for applied tasks (summarization)

DRS Overview

- Discourse Referents and Conditions
 - **basic condition**: one-place or two-place predications for associating the semantics of nouns, verbs, adjectives, and other linguistic elements with discourse referents
 - **complex condition**: an entire discourse structure for example (disjunction, implication, negation, question, proposition, etc)
 - **equality condition**: equality relationships between referents



GMB DRS Details

- **Named Entities (Sekine et al. 2002):**

`named(X, 'John', 'Person')`

- **WordNet senses (Fellbaum, 2010):**

`pred(X, loon, n, 2)`

- **VerbNet thematic roles (Kipper et al., 2008):**

`rel(E, X, 'Agent')`

- **Rhetorical relations (Asher, 1993) :**

`rel(DRS1, DRS2, because)`

Target Corpus Description

- The Groningen Meaning Bank (GMB) (Basile et al., 2012)
 - corpus of DRS-annotated discourse collections available online through a
 - navigation interface
 - download
- Release 2.1.0:
 - 8000 discourse collections
 - 47,000 sentences
 - 1,000,000 tokens—see Table 11.
- Documents are in English extracted from five genres
 - (i) Voice of America (VOA), a US Federal Government newspaper,
 - (ii) the Manually Annotated Sub-Corpus (MASC) from the Open American National Corpus (Ide et al., 2010)
 - (iii) country descriptions from the CIA World Facebook (CIA) (Agency, 2006)
 - (iv) Aesop’s fables
 - (v) a small collection of legal documents from the British Nationality Act 1981, Part I (BNA1),

Corpus Stats

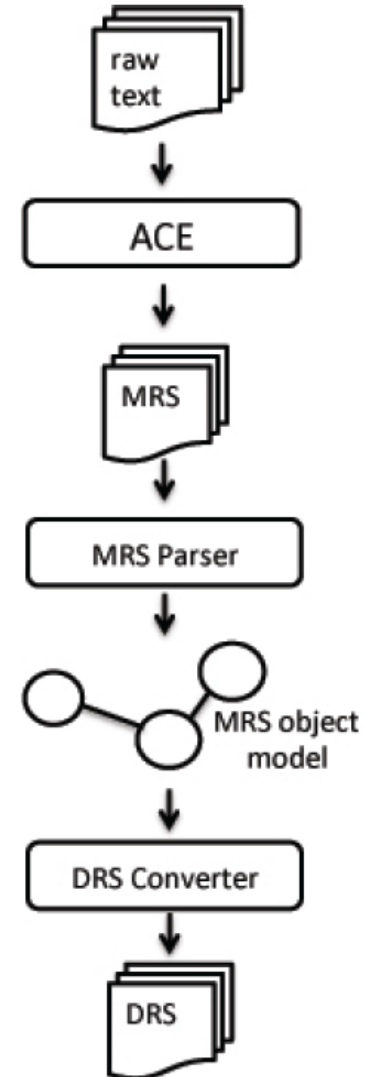
Subcorpus	Documents	Sentences	Tokens	EBOWs	DWFs
Voice of America	7,609	45,610	982,585	20,508	29,395
fables	199	775	19,127	2,180	3,161
CIA World Factbook	178	786	17,131	558	3,238
legal	3	6	287	5	93
MASC Full	11	55	1,237	85	496
all	8,000	47,232	1,020,367	23,336	31,053

Table 1: Groningen Meaning Bank (GMB) corpus statistics for release 2.1.0. An EBOW is an “effective bit of wisdom” and a DWF is a “distinct word form”

Mapping Procedure

The basic task of mapping between MRS and DRS formats requires the conversion of event variables, individual variables, and elementary predicates from MRS into discourse referents and conditions in DRS.

- First sentence manually extracted from GMB Collection
- Parsed by ACE and output as MRS
- Read by Java-based MRS Reader
- Converted to MRS Object Model
- MRS Object Model transformed to DRS
- Output to DRS



Evaluation and Measurability

- Original Plan:
 - 20 development sentences
 - 80 test sentences
 - automated evaluation tool
- Actual result
 - 10 development sentences
 - 10 test sentences
 - manual evaluation (incomplete)
- What to measure?
 - align referents
 - align basic and complex conditions
 - align lemmatization
 - compare box hierarchy?
 - compare labeled box hierarchy?
- Rule-based conversion process incomplete
 - did not present a true evaluation of what was possible, just what had been implemented

Side by Side output

A bad economy can impact people's health as well as their wealth (05.0362).

```
TOP LEVEL DRS BOX: e2
e2 e12
can(e2)
Agent(e2, e12)
```

```
Proposition complex condition box: parentEvent = e2 childEvent = e12
x30 i41 i9 e12 x24 x3 x13
```

```
impact(e12)
Agent(e12, x3)
Patient(e12, x13)
economy(x3)
bad(e8)
Agent(e8, x3)
Patient(e8, i9)
x24 subset-of x13
x30 subset-of x13
health(x24)
```

```
poss_rel complex condition box: parentEvent = e12 childEvent = e28
e28 i22 x18 x24
```

```
poss(e28)
Agent(e28, x24)
Patient(e28, x18)
people(x18)
Agent(x18, i22)
wealth(x30)
Agent(x30, i41)
```

```
poss_rel complex condition box: parentEvent = e12 childEvent = e35
x30 e35 x36
poss(e35)
Agent(e35, x30)
Patient(e35, x36)
pronoun(x36)
```

```
TOP LEVEL DRS BOX: b2
x3, x4, x5, x6, x6, x7, e8,
t11
```

```
bad(x3)
economy(x3)
people(x5)
of(x6, x5)
health(x6)
subset_of(x6, x4)
thing(x6)
of(x7, x6)
wealth(x7)
subset_of(x7, x4)
impact(e8)
Agent(e8, x3)
Theme(e8, x4)
temp_included(e8, t11)
temp_before(t10, t11)
```

```
Box: b9
t10
now(t10)
```

What is and isn't Mapped

- No underspecification – simple scope
 - qeq as equalities
 - No explicit quantification
- No variable properties
- No DRS Temporal logic
- Not all GMB basic and complex conditions realized
- Conversion rules for lexical nouns, verbs, and adpositions
 - generic handlers for elementary predicates with one or more arguments.
 - special cases of abstract unlabeled relations like *named rel*, *pron rel*, *nominalization rel*, *neg rel*, *compound rel*, *compound name rel*, *appos rel*, *subord rel*, and *season rel*.
 - generic handler for unlabeled unknown abstract relations is used.

Conclusion

- No magic discourse analysis as result of conversion
- Creating a rule-based conversion process possible but what is the benefit?
- Need to reevaluate approach
 - adapt select C & C/Boxer processes for MRS
- Integrate C & C/Boxer output into MRS-based applied task

References

- Central Intelligence Agency. 2006. *The CIA World Factbook*. Potomac Books.
- Nicholas Asher. 1993. *Reference to abstract objects in discourse*, volume 50. Kluwer Academic Pub.
- Valerio Basile, Johan Bos, Kilian Evang, and Noortje Venhuizen. 2012. Developing a large semantically annotated corpus. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, pages 3196-3200.
- Ann Copestake, Dan Flickinger, Carl Pollard, and Ivan A Sag. 2005. Minimal recursion semantics: An introduction. *Research on Language and Computation*, 3(2-3):281–332.
- James Curran, Stephen Clark, and Johan Bos. 2007. Linguistically motivated large-scale nlp with c&c and boxer. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 33–36.
- Christiane Fellbaum, 2010. *WordNet*, chapter 10, pages 231–243. Springer.
- Dan Flickinger. 2000. On building a more efficient grammar by exploiting types. *Natural Language Engineering*, 6(1):15–28.
- Nancy Ide, Christiane Fellbaum, Collin Baker, and Rebecca Passonneau. 2010. The manually annotated sub-corpus: a community resource for and by the people. In *Proceedings of the ACL 2010 Conference Short Papers*, pages 68–73.
- Hans Kamp and U. Reyle. 1990. *From discourse to logic: introduction to model theoretic semantics of natural language, formal logic and discourse representation theory*. Kluwer Academic, Dordrecht.
- Hans Kamp. 1981. A theory of truth and semantic representation. *Formal Methods in the Study of Language*, pages 189–222.
- Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2008. A large-scale classification of english verbs. *Language Resources and Evaluation*, 42(1):21–40.
- Satoshi Sekine, Kiyoshi Sudo, and Chikashi Nobata. 2002. Extended named entity hierarchy. In *Proceedings of LREC*, volume 2.
- Jan van Eijck. 2006. Discourse representation theory. *Encyclopedia of Language and Linguistics*, 3:660–669.