Linguistics 575: MRS in Applications

March 30, 2016

Emily M. Bender

Overview

- Course goals
- DELPH-IN context
- ERG demo
- Course requirements
- Course expectations
- Why use semantics?
- Target tasks

But first…

- https://www.ehs.washington.edu/fsoemerprep/evacinfo.shtm

Course goals

- Explore NLP tasks which can be improved with semantic features
- Understand what information is captured by the ERG’s MRS output that is relevant to those tasks
- Experience with feature design
- Add MRS features to an existing baseline system, and measure the result
- Experience with error analysis
- Experience with academic writing in CL/NLP

DELPH-IN

The DELPH-IN ecology www.delph-in.net

- Head-drive Phrase Structure Grammar (Pollard & Sag 1994)
- Joint reference formalism (Copestake 2002a)
- Shared semantic representation formalism (MRS; Copestake et al 2005)
- Grammars: ERG (Flickinger 2000, 2011), Jacy (Siegel & Bender 2002), NorSource (Hellan & Haugereid 2003), ...
- Grammar generator: Grammar Matrix (Bender et al 2002, 2010)
- Parser generators: LKB (Copestake 2002b), PET (Callmeier 2002), agree, ACE

The DELPH-IN ecology www.delph-in.net

- Parse and realization ranking: (e.g., Toutanova et al 2005, Velldal 2008)
- Robustness measures: (e.g., Zhang & Kordoni 2006, Zhang & Krieger 2011)
- Regression testing: [incr tsdb()] (Oepen 2001)
- Applications: e.g., MT (Oepen et al 2007), QA from structured knowledge sources (Frank et al 2007), Textual entailment (Bergmair 2008), ontology construction (Nichols et al 2006) and grammar checking (Suppes et al 2012)

Multilingual grammar engineering: Other approaches

- The DELPH-IN consortium specializes in large HPSG grammars
- Other broad-coverage precision grammars have been built by/in/with
  - LFG (ParGram: Butt et al 2002)
  - F/XTAG (Doran et al 1994)
  - HPSG: ALE/Controll (Götz & Meurers 1997)
  - SFG (Bateman 1997)
- Proprietary formalisms and Microsoft and Boeing and IBM

Overview

- Course goals
- DELPH-IN context
- ERG demo
- Course requirements
- Course expectations
- Why use semantics?
- Target tasks

Course requirements

- http://faculty.washington.edu/ebender/2016_575/

Term papers v. theses

- Less thorough literature review
- Null results are ok (and don’t need to be made “interesting”)
    - If you get a null result, the paper still has to be well-written :)
- May nonetheless be worth submitting as a conference/workshop paper:
    - EMNLP 2016 (Austin, TX 11/2-6): http://www.emnlp2016.net/ (due: 6/3/16)
    - Coling 2016 (Osaka, Japan 12/11-16): http://coling2016.anlp.jp/ (due: 7/15/16)

Overview

- Course goals
- DELPH-IN context
- ERG demo
- Course requirements
- Course expectations: Why are *you* here?
- Why use semantics?
- Target tasks

Overview

- Course goals
- DELPH-IN context
- ERG demo
- Course requirements
- Course expectations
- Why use semantics?
- Target tasks

Why use semantics?

- “Parsing is a problem in search of a solution” -- Jeremy G. Kahn
- Maybe an overstatement, but it does seem like for many tasks a parsing-based solution doesn’t (easily) improve on a bag-of-words approach
  - Why?

Syntax-semantics mismatches (Bender 2013, Ch 9)

- Valence alternations: passive, dative alternation, middle voice
- Semantically empty elements
- Mediated dependencies: raising/control
- Unrealized arguments
- Coordination and one-to-many/many-to-one dependencies
- Long-distance dependencies

Valence alternations: Passive

- The dog chased the cat./The cat was chased by the dog.
- The cat got chased by the dog.
- The cat chased by the dog ran up the tree.
- Precision and recall were measured using the formulas given above.
- Anyone handed a note will be watched closely.

Valence alternations: Dative alternation, middle voice, causative/inchoative

- Kim gave Sandy the book./Kim gave the book to Sandy.
- Kim threw Sandy a party./Kim threw a party for Sandy.

- This truck loads easily.
- *This truck loads easily by movers.

- The vase broke.
- They broke the vase.

Semantically empty elements

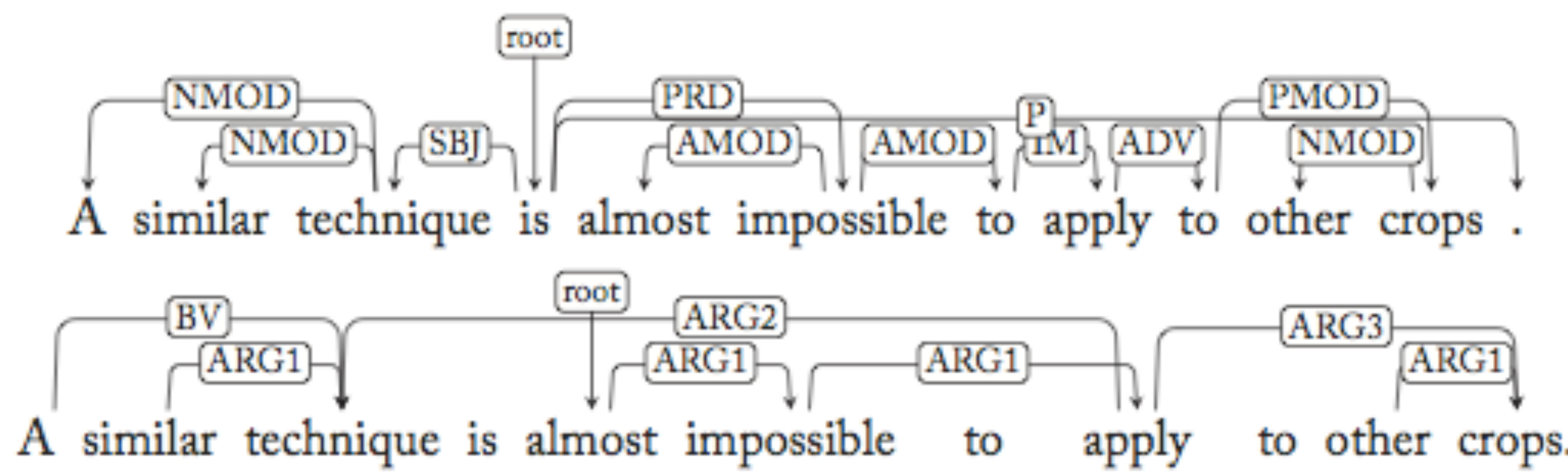

Figure 8.1: Syntactic (CoNLL 2008, top) and semantic (ERG, bottom) dependency structures

- Adapted from Ivanova et al 2012

Mediated dependencies: raising/control

- Kim seems to continue to appear to like sushi.
- Kim tries to like sushi.
- Kim persuaded Sandy to leave.
- Kim appealed to Sandy to leave.
- Kim refrained from laughing.
- Kim will try and find it.
- Kim is anxious to leave.
- It is easy for Kim to leave.

Unrealized arguments

- Mistakes were made.
- Fix those mistakes!
- I ate.
- I watched.
- I’m finished.
- I already told them.

Long-distance dependencies

- What did Sandy claim everyone hoped Lee would believe Kim saw?
- This is the library in which no one believes anyone could imagine Kim read the book.
- I don't think Kim likes eggs. **Bagels**, I seem to recall Sandy saying that Pat had mentioned Kim likes to eat.

Collectively frequent enough to matter

Phenomenon	Frequency	Candidates
barerel	2.12%	546
tough	0.07%	175
rnr	0.69%	1263
itexpl	0.13%	402
vpart	4.07%	765
ned	1.18%	349
absol	0.51%	963
vger	5.16%	679
argadj	3.60%	1346
control	3.78%	124

Table 1: Relative frequencies of phenomena matches in Wikipedia, and number of candidate strings vetted.

(Bender et al 2011)

Bender et al 2015: Central claims

- Sentence meaning, but not speaker meaning, is compositional
- Systems attempting to understand speaker meaning would benefit from reusable, automatically derivable, task-independent representations of sentence meaning
- A compositional approach to creating sentence meaning representations provides
  - Comprehensiveness
  - Consistency
  - Scalability

Sentence meaning, speaker meaning

- Learning correlations between domain-typical surface forms and task-specific representations conflates:
  - timeless/conventional/standing/sentence meaning
  - utterer/occasion/speaker meaning
- Drawbacks:
  - resolving the same problems around grammatical structure for each task
  - unlikely to scale to general-purpose NLU

Importance of grammatical structure

先生によると男の子よりも女の子がポケモンがすきだ。[jpn]

先生 によると 男の子 よりも 女の子 が ポケモン が すき だ。

teacher (.) boy (.) girl (.) Pokemon (.) like (.)

teacher ACCORDING.TO boy THAN girl NOM Pokemon NOM like COP.PRES

according-to(e4, e3, x6)
teachers(x6)
like(e3, x14, x17)

girls(x14)
pokemon(x17)
more-than(e22, x14, x23)
boys(x23)

Leveraging sentence meaning

- Machines don't have access to any direct representation of speaker meaning, only to natural language utterances
- Sentence meaning doesn't determine situated speaker meaning, but is an important cue to it (Quine 1960, Grice 1968, Reddy 1979, Clark 1996)
- A task-independent, comprehensive representation of sentence meaning capturing exactly the information in the linguistic signal itself should benefit NLU systems

A meaning representation system is compositional if (working definition):

- it is grounded in a finite (possibly large) number of atomic symbol-meaning pairings
- it is possible to create larger symbol-meaning pairings by combining the atomic pairings through a finite set of rules;
- the meaning of any non-atomic symbol-meaning pairing is a function of its parts and the way they are combined;
- this function is possibly complex, containing special cases for special types of syntactic combination, but only draws on the immediate constituents and any semantic contribution of the rule combining them; and
- further processing will not need to destructively change a meaning representation created in this way to create another of the same type.

Semantic annotation survey: Compositional layer

- Predicate-argument structure
- Partial constraints on:
    - Scope of negation and other operators
    - Restriction of quantifiers
    - Modality
    - Tense/aspect/mood
    - Information structure
    - Discourse status of referents of NPs
    - Politeness
- Possibly compositional, but not according to sentence grammar:
    - Coherence relations/rhetorical structure

Semantic annotation survey: Sub-compositional layer

- Fine-grained word-sense annotations; entity types; links to ontologies
    - Concern only atoms
    - What is built compositionally is not the choice of atom, but the connections between them
    - Sense underspecification: Only discriminate between supersenses correlating with morphosyntactic differences

Semantic annotation survey: Computation on top of compositional backbone

- Quantifier scope ambiguity resolution (e.g. Higgins & Sadock 2003)
- Coreference resolution (e.g. Hobbs 1979)
- Determination of the focus of negation (e.g. Blanco and Moldovan 2011)
    - Build on partial constraints provided by the grammar
    - Not compositional in that it’s never (strictly) constrained by grammatical structure

Semantic annotation survey: Discourse processing

- Presupposition projection (e.g. van der Sandt 1992, Zaenen & Karttunen 2013, Venhuizen et al 2013)
- Coherence relations/rhetorical structure (e.g. Marcu 1997)
- Discourse moves/adjacency pairs (e.g. Shriberg et al 2004)
  - Again, build on information provided during sentence-level processing
  - Concerns cross-sentential relations, so not compositional according to sentence grammar
  - Open question: compositional processes at higher levels of structure?

Semantic annotation survey: Doing things with words

- Hedge detection (e.g. Vincze et al 2008)
- Authority claim, alignment move, and other social act annotation (e.g. Morgan et al 2013)
- Pursuit of power in dialogue (e.g. Swayamdipta & Rambow 2012)
    - Not anchored in the structure of sentences
    - Relate to speaker goals

ERG: The English Resource Grammar (Flickinger 2000, 2011)

- Under continuous development since 1993
- Framework: Head-driven Phrase Structure Grammar (Pollard & Sag 1994)
- 1214 release: 225 syntactic rules, 70 lexical rules, 975 leaf lexical types
- Open-source and compatible with open-source DELPH-IN processing engines (www.delph-in.net)
- Broad-coverage: 85-95% on varied domains: newspaper text, Wikipedia, bio-medial research literature (Flickinger et al 2010, 2012; Adolphs et al 2008)
    - ‘Bridging’ rule analyses enable 100% coverage
- Output: derivation trees paired with meaning representations in the Minimal Recursion Semantics framework---English Resource Semantics (ERS)
    - Emerging documentation at moin.delph-in.net/ErgSemantics

Redwoods: ERG-based treebanking (sembanking) (Oepen et al 2004)

- Minimal discriminants (Carter 1997): Properties of derivation trees partitioning parse forest per item
- Allows annotators to swiftly navigate even very large parse forests to select intended analysis or reject all analyses
    - 37,200 words of the Brown corpus annotated in 1400 minutes (1.7 sentences/min)
- All annotation decisions are recorded and can be rerun against updated parse forests produced by updated grammar versions
- Current Redwoods release (7th growth) includes 45,000 sentences of annotated text across genres including Wikipedia, tourism brochures, ...
- Analyses can be viewed as full HPSG analyses, ERS only, or even simpler syntactic or semantic dependency representations

Why a grammar-based compositional approach?

- Argued above for the importance of task-independent, sentence-meaning annotations
- Can be created:
  - Non-compositionally, as in Abstract Meaning Representation (AMR; Langkilde & Knight 1998, Banarescu et al 2013)
  - Compositionally, by hand, as in PropBank (Kingsbury & Palmer 2002) and FrameNet (Baker et al 1998)
  - Compositionally, with a machine-readable grammar, as in Redwoods (Oepen et al 2004), TREPIL (Rosén et al 2005), or the Groningen Meaning Bank (Basile et al 2012)

Benefits of compositionality: Comprehensiveness

- Grammar-based compositional approach ⇒ Every word and syntactic structure must be accounted for, or specifically deemed semantically void

- Narrower paraphrase sets, compare AMR (1), (2) (Banarescu et al 2014) to ERS (3)

(1) a. No one ate.

b. Every person failed to eat.

(2) a. The boy is responsible for the work.

b. The boy is responsible for doing the work.

c. The boy has the responsibility for the work.

Benefits of compositionality: Comprehensiveness

- Grammar-based compositional approach ⇒ Every word and syntactic structure must be accounted for, or specifically deemed semantically void
- Narrower paraphrase sets, compare AMR (1), (2) (Banarescu et al 2014) to ERS (3)

(3)
a. Kim thinks Sandy gave the book to Pat.
b. Kim thinks that Sandy gave the book to Pat.
c. Kim thinks Sandy gave Pat the book.
d. Kim thinks the book was given to Pat by Sandy.
e. The book, Kim thinks Sandy gave to Pat.

Benefits of compositionality: Comprehensiveness

- Task-independent semantic representations can't abstract away from seemingly less relevant nuances of sentence meaning

- Compositional approach facilitates capturing more detail

$$\langle\, h_1, \left| \begin{array}{l} h_4\text{:person}\langle 0\!:\!6\rangle(\text{ARG0 } x_5), \\ h_6\text{:\_no\_q}\langle 0\!:\!6\rangle(\text{ARG0 } x_5,\ \text{RSTR } h_7,\ \text{BODY } h_8), \\ h_2\text{:\_eat\_v\_1}\langle 7\!:\!11\rangle(\text{ARG0 } e_3,\ \text{ARG1 } x_5,\ \text{ARG2 } i_9) \end{array} \right| \{\, h_1 =_q h_2,\ h_7 =_q h_4 \,\}\,\rangle$$

```
(e / eat-01
 :polarity -
 :ARG0 (p / person
   :mod (e / every)))
```

$$\langle\, h_1, \left| \begin{array}{l} h_4\text{:\_every\_q}\langle 0\!:\!5\rangle(\text{ARG0 } x_6,\ \text{RSTR } h_7,\ \text{BODY } h_5), \\ h_8\text{:\_person\_n\_1}\langle 6\!:\!12\rangle(\text{ARG0 } x_6), \\ h_2\text{:\_fail\_v\_1}\langle 13\!:\!19\rangle(\text{ARG0 } e_3,\ \text{ARG1 } h_9), \\ h_{10}\text{:\_eat\_v\_1}\langle 23\!:\!27\rangle(\text{ARG0 } e_{11},\ \text{ARG1 } x_6,\ \text{ARG2 } i_{12}) \end{array} \right| \{\, h_1 =_q h_2,\ h_7 =_q h_8,\ h_9 =_q h_{10} \,\}\,\rangle$$

Benefits of Compositionality: Consistency

- Requiring meaning representations to be grounded in both the lexical items and syntactic structure of the strings being annotated significantly reduces the space of possible annotations

- Grammar based approach allows encoding of design decisions for machine application

    - Ex: arguments of *when*

- Human input still required, but choosing among representations is far simpler than authoring them

    - Development of grammar is still a big investment, but with big returns as the same grammar is applied over more and more text

Benefits of Compositionality: Scalability

- In amount of text annotated: Initial development of grammar pays off as it is applied to as much text as desired
- In genre diversity of the resource: One and the same grammar can be applied to texts from multiple different domains
  - Robustness techniques can compensate for lack of grammar coverage
- In the complexity of the annotations themselves: Grammar updates can be efficiently propagated across the treebank by reparsing corpus and rerunning annotation decisions (Oepen et al 2004)
  - Improve analyses of particular phenomena, or add layers of grammar-based annotation (e.g. partial constraints on information structure)

Inter-Annotator Agreement study

- Data source: Sentences sampled from Antoine de Saint Exupéry's *The Little Prince*
- Three expert annotators
- Annotated 50-sentence trial set, then adjudicated, updating annotation guidelines as indicated
- Annotated 150-sentence sample set, then measured IAA, then produced adjudicated gold standard
- Repeat above steps with 'bridging' analyses in

Agreement Metrics

- NB: Chance-corrected IAA measures as yet unavailable for graph-structured annotations
- Exact match: Full ERS identical between annotators
- Elementary Dependency Match (Dridan & Oepen 2011)
  - Computed over sets of triples from reduction of ERS to Elementary Dependency Structures (EDS)
  - EDMa: Argument identification only
  - EDMna: Argument identification + predicate name identification

IAA Results

	Annotator Comparison			
Metric	**A vs. B**	**A vs. C**	**B vs. C**	**Average**
Exact Match	0.73	0.65	0.70	0.70
EDM_a	0.93	0.92	0.94	0.93
EDM_{na}	0.94	0.94	0.95	0.94

- Compare Banarescu et al (2013) triple-based IAA for AMR over web text of 0.71

The best of both worlds

- Grammar-based compositionally created meaning representations can only access information that can be implemented in a grammar

- But having comprehensive, integrated meaning representations is desirable (Basile et al 2012, Banarescu et al 2013), Ide & Suderman 2007)

- Proposal: Start from grammar-based meaning representations and process to produce larger paraphrase sets and/or add layers, such as fine-grained word sense annotations, additional coreference links, cross-sentential discourse relations

Central claims (reprise)

- Sentence meaning, but not speaker meaning, is compositional
- Systems attempting to understand speaker meaning would benefit from reusable, automatically derivable, task-independent representations of sentence meaning
- A compositional approach to creating sentence meaning representations provides
  - Comprehensiveness
  - Consistency
  - Scalability

Bender, Emily M., Dan Flickinger, Stephan Oepen, Woodley Packard and Ann Copestake. 2015. Layers of Interpretation: On Grammar and Compositionality. *IWCS 2015*.

Overview

- Course goals
- DELPH-IN context
- ERG demo
- Course requirements
- Course expectations
- Why use semantics?
- Target tasks

Target tasks

- At least two ways to use MRS:
  - As an interface representation (transfer based MT, ‘deep’ NLU/dialog systems, MRS based abstractive summarization)
  - As an additional source of features for a machine-learner, together with n-grams, syntactic features, etc.
- We’ll focus on the second one this quarter
- What tasks interest people?