Multi-word Expressions in HPSG

Dan Flickinger
Center for the Study of Language and Information
Stanford University
danf@stanford.edu

PARSEME Training School
Charles University, Prague
January 2015
Overview for the week

• Day One
  Brief introduction to Head-driven Phrase Structure Grammar
  Implementation in the English Resource Grammar (ERG)
  Meaning representation in Minimal Recursion Semantics

• Day Two
  Classification of Multi-word Expressions (MWEs)
  Implementation of MWEs in the ERG
  Strengths and weaknesses of the approach

• Day Three
  Case study of one class of MWEs: idioms with possessives
  Interactions with other linguistic phenomena and processing
  Disambiguation challenges

• Day Four
  Lab session using the ERG to identify and analyse MWEs
Desiderata for a linguistic theory

- General principles that hold true of human language
- Formal descriptive devices to make falsifiable predictions
- Cross-linguistic validity
- Simplicity
Overview of HPSG

- Linguistic objects can be described as attribute-value pairs (phonology), morphology, syntax, semantics, ...
- General principles constrain values for some attributes e.g. a phrase and its head daughter share certain values
- Most constraints are in the lexicon
- Small number of syntactic (phrase structure) rules
A simple example

most cats that we have studied don't admire whose dogs bark.

children have studied whose dogs bark.
A simple example: MRS – linking

*Most cats that we have studied don’t admire children whose dogs bark.*

```xml
<h1,e2:prop:pres:indicative:-:-,
    h3:_most_q(x4:3:pl:+, h5, h6),
    h7:_cat_n_1(x4),
    h8:pron(x9:1:pl),
    h10:pronoun_q(x9, h11, h12),
    h7:_study_v_1(e13:prop:pres:indicative:-:+, x9, x4),
    h14:neg(e16, h15),
    h17:_admire_v_1(e2, x4, x18:3:pl:+),
    h19:udef_q(x18, h20, h21),
    h22:_child_n_1(x18),
    h23:def_explicit_q(x25:3:pl:+, h26, h24),
    h22:poss(e27, x25, x18),
    h28:_dog_n_1(x25),
    h22:_bark_v_1(e29:prop:pres:indicative:-:-, x25),
    h5 qeq h7, h11 qeq h8, h15 qeq h17, h20 qeq h22, h26 qeq h28>
```
A simple example: MRS – scope

Most cats that we have studied don’t admire children whose dogs bark.

<\textit{h1,e2:prop:pres:indicative:-:-, h3:_most_q(x4:3:pl:+, h5, h6), h7:_cat_n_1(x4), h8:pron(x9:1:pl), h10:pronoun_q(x9, h11, h12), h7:_study_v_1(e13:prop:pres:indicative:-:+, x9, x4), h14:neg(e16, h15), h17:_admire_v_1(e2, x4, x18:3:pl:+), h19:undef_q(x18, h20, h21), h22:_child_n_1(x18), h23:def_explicit_q(x25:3:pl:+, h26, h24), h22:poss(e27, x25, x18), h28:_dog_n_1(x25), h22:_bark_v_1(e29:prop:pres:indicative:-:-, x25), h5 qeq h7, h11 qeq h8, h15 qeq h17, h20 qeq h22, h26 qeq h28}>
A simple example: Semantic dependencies

Most cats that we have studied don’t admire children whose dogs bark.

x4: _most_q[]
x9: pronoun_q[]
e13: _study_v_1[ARG1 x9: pron, ARG2 x4: _cat_n_1]
e16: neg[ARG1 e2: _admire_v_1]
e2: _admire_v_1[ARG1 x4: _cat_n_1, ARG2 x18: _child_n_1]
x18: udef_q[]
x26: def_explicit_q[]
e28: poss[ARG1 x26: _dog_n_1, ARG2 x18: _child_n_1]
e30: _bark_v_1[ARG1 x26: _dog_n_1]
Principles of HPSG Encoded in ERG

- Head Feature Principle
  ’The HEAD value of a phrase is identified with that of its (syntactic) head daughter’

- Semantics Principle (MRS version)
  ’The RELS value of a phrase is the result of appending the RELS values of its daughters’
  ’The (semantic) HOOK value of a phrase is identified with the HOOK value of its semantic head’
Sample feature structure
Desiderata for a grammar implementation

- Coverage of linguistic phenomena
- Accuracy of linguistic analyses
- Ambiguity suitably constrained
- Efficiency in processing
- Maintainence and extensibility
- Reversibility (parsing and generation)
English Resource Grammar (ERG)

- 7000 types in multiple-inheritance monotonic hierarchy
- 975 leaf lexical types
- 39,000 manually constructed lexemes
- 225 syntactic rules
- 70 morphological rules (inflection and derivation)
- Statistical parse selection model trained on 1.5 million word corpus
- Online demo: http://lingo.stanford.edu/erg
Standard HPSG Rules (Pollard & Sag 1994)

- Head-Subject
- Head-Complement
- Head-Specifier
- Head-Modifier
- Head-Marker
- Head-Filler
- Coordination
ERG syntactic rules (binary)
Sample feature structure
Semantics in feature structures

ORTH: adore
HEAD: verb
SUBJ: ⟨HEAD noun, CONT [INDEX 2]⟩
COMPS: ⟨HEAD noun, CONT [INDEX 3]⟩
INDEX: 1 event
CONT: RELS ⟨PRED “adore_v”, ARG0 1, ARG1 2, ARG2 3⟩
HCONS: ⟨⟩
### Semantics of sentences

The value of **CONT** for a sentence is a list of relations in the attribute **RELS**, with the arguments in those relations appropriately linked:

<table>
<thead>
<tr>
<th>HEAD</th>
<th>verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUBJ</td>
<td>⟨⟩</td>
</tr>
<tr>
<td>CMPS</td>
<td>⟨⟩</td>
</tr>
<tr>
<td>CONT</td>
<td>RELS</td>
</tr>
<tr>
<td></td>
<td>⟨ ⟩</td>
</tr>
<tr>
<td></td>
<td>INDX</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ARG0</td>
</tr>
<tr>
<td></td>
<td>ARG1</td>
</tr>
<tr>
<td></td>
<td>ARG2</td>
</tr>
<tr>
<td></td>
<td>HCNS</td>
</tr>
</tbody>
</table>

```
PRAGUE — 19-JAN-15 (danf@stanford.edu)
```
Semantic composition
for Minimal Recursion Semantics

- The \textbf{RELS} value of a phrase is the result of appending the \textbf{RELS} lists of its daughter(s), plus any from the construction itself.
- The \textbf{INDEX} value of a phrase is unified with the \textbf{INDEX} value of its semantic head daughter.

In addition, to accommodate scope underspecification:
- The \textbf{HCONS} value of a phrase is the result of appending the \textbf{HCONS} lists of its daughter(s), plus any from the construction itself.
- The \textbf{LTOP} value of a phrase is unified with the \textbf{LTOP} value of its semantic head daughter.
MRS Composition: “probably barks”
Linking semantic arguments

When heads select a complement or specifier, they constrain its INDEX value – a referential index for nouns, or an event variable for verbs.

trans-verb-lxm

HEAD  verb

SUBJ  \langle [CONT \quad INDEX \quad 1] \rangle

COMPS  \langle [CONT \quad INDEX \quad 2] \rangle

CONT

INDEX  3

RELS  \langle [PRED \quad string] \quad ARG0 \quad 3 \quad ARG1 \quad 1 \quad ARG2 \quad 2] \rangle
Lexical type hierarchy (verbs)
A Wikipedia example

“Computational linguistics” is an [[interdisciplinary]] field dealing with the [[Statistics—statistical]] and/or rule-based modeling of [[natural language]] from a computational perspective.
A Wikipedia example: Syntax

S
  NP
    N
      computational
      linguistics
  VP
    V is
    NP
      DET an
      AP interdisciplinary
      N
        field
        V dealing
        PP with
    VP
    N
      dealing
      PP
      NP
        DET the
        AP statistical
        CONJ and / or
        AP rule-based
        V of
        NP modeling
        PP
        NP
          DET a
          AP computational
          N
            perspective.
A Wikipedia example: MRS

‘Computational linguistics’ is an interdisciplinary field dealing with the statistical and/or rule-based modeling of natural language from a computational perspective.

{h3:undef_q<3:27>(x5, h4, h6),
h7:_computational_a_1<3:15>(e8, x5),
h7:_linguistics_n_1<17:27>(x5),
h9:_be_v_id<32:33>(e2, x5, x10),
h11:_a_q<35:36>(x10, h13, h12),
h14:_interdisciplinary/jj_u_unknown<40:56>(e15, x10),
h14:_field_n_of<60:64>(x10, i16),
h14:_deal_v_with<66:72>(e17, x10, x18),
h19:_the_q<79:81>(x18, h21, h20),
h22:_statistical_a_1<96:106>(e23, x18),
h24:_and-or_c<110:115>(e26, h22, e23, h25, e27),
h25:argument<117:126>(e29, e27, x28),
h33:_rule_n_of<117:126>(x28, i34),
h25:_base_v_1<117:126>(e27, p35, x18),
h24:nominalization<128:135>(x18, h37),
h37:_model_v_1<128:135>(e38, p40, x39:3:SG),
...
## ‘ws13’ Coverage Profile

<table>
<thead>
<tr>
<th>Length in tokens</th>
<th>total items</th>
<th>word string</th>
<th>lexical items</th>
<th>distinct analyses</th>
<th>total results</th>
<th>overall coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>65 – 70</td>
<td>2</td>
<td>66.50</td>
<td>755.00</td>
<td>0.00</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>60 – 65</td>
<td>3</td>
<td>62.33</td>
<td>538.67</td>
<td>500.00</td>
<td>1</td>
<td>33.3</td>
</tr>
<tr>
<td>55 – 60</td>
<td>3</td>
<td>56.00</td>
<td>343.33</td>
<td>500.00</td>
<td>2</td>
<td>66.7</td>
</tr>
<tr>
<td>50 – 55</td>
<td>8</td>
<td>51.37</td>
<td>563.50</td>
<td>0.00</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>45 – 50</td>
<td>11</td>
<td>47.27</td>
<td>542.55</td>
<td>500.00</td>
<td>3</td>
<td>27.3</td>
</tr>
<tr>
<td>40 – 45</td>
<td>21</td>
<td>41.38</td>
<td>422.90</td>
<td>500.00</td>
<td>11</td>
<td>52.4</td>
</tr>
<tr>
<td>35 – 40</td>
<td>42</td>
<td>36.69</td>
<td>433.64</td>
<td>500.00</td>
<td>32</td>
<td>76.2</td>
</tr>
<tr>
<td>30 – 35</td>
<td>77</td>
<td>31.78</td>
<td>321.78</td>
<td>490.26</td>
<td>61</td>
<td>79.2</td>
</tr>
<tr>
<td>25 – 30</td>
<td>101</td>
<td>26.89</td>
<td>329.47</td>
<td>499.68</td>
<td>87</td>
<td>86.1</td>
</tr>
<tr>
<td>20 – 25</td>
<td>130</td>
<td>22.17</td>
<td>220.02</td>
<td>465.31</td>
<td>120</td>
<td>92.3</td>
</tr>
<tr>
<td>15 – 20</td>
<td>158</td>
<td>16.98</td>
<td>177.37</td>
<td>374.31</td>
<td>149</td>
<td>94.3</td>
</tr>
<tr>
<td>10 – 15</td>
<td>137</td>
<td>12.32</td>
<td>126.58</td>
<td>197.77</td>
<td>132</td>
<td>96.4</td>
</tr>
<tr>
<td>5 – 10</td>
<td>97</td>
<td>6.86</td>
<td>67.01</td>
<td>42.46</td>
<td>89</td>
<td>91.8</td>
</tr>
<tr>
<td>0 – 5</td>
<td>209</td>
<td>2.47</td>
<td>11.02</td>
<td>3.42</td>
<td>202</td>
<td>96.7</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1001</strong></td>
<td><strong>17.55</strong></td>
<td><strong>184.78</strong></td>
<td><strong>270.04</strong></td>
<td><strong>889</strong></td>
<td><strong>88.8</strong></td>
</tr>
</tbody>
</table>
Wikipedia parsing with ERG/PET: Ambiguity

String Length ('i-length')

--- distinct analyses

(generated by [incr tsdb()] at 11-nov-2010 (18:34 h))
Wikipedia parsing with ERG/PET: Efficiency

ws13: Parse time (in seconds) by Length

--- cpu time

String Length ('i-length')
## Wikipedia parsing with ERG/PET: Accuracy

<table>
<thead>
<tr>
<th>Corpus type</th>
<th>Number of items</th>
<th>Av. item length</th>
<th>Observed coverage</th>
<th>Verified coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meeting scheduling</td>
<td>11660</td>
<td>7.5</td>
<td>96.8%</td>
<td>93.8%</td>
</tr>
<tr>
<td>E-commerce</td>
<td>5392</td>
<td>8.0</td>
<td>96.1%</td>
<td>93.0%</td>
</tr>
<tr>
<td>Norwegian tourism</td>
<td>10834</td>
<td>15.0</td>
<td>94.3%</td>
<td>88.5%</td>
</tr>
<tr>
<td>SemCor (partial)</td>
<td>2501</td>
<td>15.0</td>
<td>94.3%</td>
<td>88.5%</td>
</tr>
<tr>
<td>Newspaper (WSJ)</td>
<td>31441</td>
<td>20.4</td>
<td>93.4%</td>
<td>84.9%</td>
</tr>
<tr>
<td><strong>Wikipedia (CmpLng)</strong></td>
<td><strong>11558</strong></td>
<td><strong>19.5</strong></td>
<td><strong>92.9%</strong></td>
<td><strong>81.7%</strong></td>
</tr>
<tr>
<td>Online user forum</td>
<td>578</td>
<td>12.5</td>
<td>85.5%</td>
<td>77.5%</td>
</tr>
<tr>
<td>Dictionary defs.</td>
<td>10000</td>
<td>6.0</td>
<td>81.2%</td>
<td>75.5%</td>
</tr>
<tr>
<td>Essay</td>
<td>769</td>
<td>21.6</td>
<td>83.2%</td>
<td>69.4%</td>
</tr>
<tr>
<td>Chemistry papers</td>
<td>637</td>
<td>27.0</td>
<td>87.8%</td>
<td>65.3%</td>
</tr>
<tr>
<td>Technical manuals</td>
<td>4000</td>
<td>12.5</td>
<td>86.8%</td>
<td>61.9%</td>
</tr>
</tbody>
</table>
## Frequency of Linguistic Phenomena in Wikipedia 100

<table>
<thead>
<tr>
<th>Phenomenon</th>
<th>#Items</th>
<th>%Corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure NPs</td>
<td>44</td>
<td>0.3</td>
</tr>
<tr>
<td>Appositives</td>
<td>1048</td>
<td>9.1</td>
</tr>
<tr>
<td>NP Fragments</td>
<td>2126</td>
<td>18.4</td>
</tr>
<tr>
<td>NP Coordination</td>
<td>1960</td>
<td>17.0</td>
</tr>
<tr>
<td>Multi-NP Coord</td>
<td>558</td>
<td>4.8</td>
</tr>
<tr>
<td>VP Coordination</td>
<td>491</td>
<td>4.2</td>
</tr>
<tr>
<td>S Coordination</td>
<td>381</td>
<td>3.3</td>
</tr>
<tr>
<td>Relative Clauses</td>
<td>2239</td>
<td>19.4</td>
</tr>
<tr>
<td>Unbounded Deps</td>
<td>2273</td>
<td>19.7</td>
</tr>
<tr>
<td>Yes-No Questions</td>
<td>11</td>
<td>0.1</td>
</tr>
<tr>
<td>WH Questions</td>
<td>10</td>
<td>0.1</td>
</tr>
<tr>
<td>Imperatives</td>
<td>222</td>
<td>1.9</td>
</tr>
<tr>
<td>Free relatives</td>
<td>107</td>
<td>0.9</td>
</tr>
<tr>
<td>Passives</td>
<td>3534</td>
<td>30.6</td>
</tr>
</tbody>
</table>
One central challenge: Ambiguity

- Machines lack common sense or real-world knowledge
- So they can propose many unwanted candidate interpretations
- Applications want just the one intended interpretation, out of many:

  Have her report on my desk by Friday.

  Cause her to deliver a report about my desk by Friday.
  Cause her to deliver a report while standing on my desk by Friday.
  Cause her to deliver a report about my desk next to Mr. Friday.
  Take possession of her report which is on my desk, by Friday.
  Cause the report she wrote to be on my desk not later than Friday.
Two possible approaches to resolving ambiguity

- Direct manual coding of preferences/heuristics
  - no underlying theory, so unpredictable behavior
  - brittle: dependence on domain and context
  - expensive to maintain: highly skilled labor
- Statistical modeling trained on manual annotations
  - useful precision with modest amount of training data
  - annotations also useful for grammar development
  - possibly also domain-specific
Treebank construction using Redwoods approach

- Parse corpus using DELPH-IN resources [www.delph-in.net](http://www.delph-in.net)
  - English Resource Grammar (Flickinger 2000)
  - PET parser (Callmeier 2000)
- Store up to 500 top-ranked trees for each sentence
  - Initial ranking using pre-existing stochastic model
  - After first 2000 items treebanked, retrain model
- Manually disambiguate using *discriminants* (Carter 1997)
  - Cf. Redwoods, Alpino, LFG Parsebanker
- Record complete syntax/semantics derivation for each item
- Tools to extract varied output representations
  - Labeled syntax trees
  - Dependency structures
  - Logical form meaning representations (Minimal Recursion Semantics)
Relevance to Multi-word Expressions

- Finding instances of known multi-word expressions by parsing
- Discovering new MWEs by parsing and treebanking
- Using statistical parse selection for disambiguation
- Expressing constraints on MWEs via semantics (MRS)