Precision Grammar Implementation for Linguistic Hypothesis Testing

Emily M. Bender
University of Washington
ebender@u.washington.edu

Dan Flickinger
CSLI Stanford
danf@csli.stanford.edu

Stephan Oepen
Universitetet i Oslo & CSLI Stanford
oe@csli.stanford.edu

http://lingo.stanford.edu/courses/07/lsa/
Overview

Lecture

- LinGO Grammar Matrix: History and motivations
- LinGO Grammar Matrix: Overview
- Stroll through MRS

Lab

- Run Grammar Matrix customization system
- Test initial coverage
- Plan extension for next lab
Grammar Matrix: Motivations

- Jump-start development of new grammars
- ... by reusing tried and true analyses from existing grammars
- ... while promoting consistency of output for downstream applications
- Explore computational linguistic typology
- (Eventual) use in language documentation
Grammar Matrix: Original construction

- Start with ERG circa Nov 2001
- Strip out what appeared to be English specific
- ... with reference to JACY grammar of Japanese
- Main initial gain: Compositional semantics
- Add in LKB interaction files
- Refine core in light of feedback from other languages
Grammar Matrix: Libraries

- Want to share code beyond cross-linguistic core
- SVO word order seems to work the same way in all SVO languages
- Computational linguistic typology
- Does SVO work the same way in all SVO languages?
- Does our SVO analysis interact properly with our analyses of other phenomena in other languages?
Warning: Different feature/path names

- SUBJ distinct from SPR
- Deeper embedding of feature structures
- Semantic features!
MRS outline

- Semantics in sign-based grammar
- Anatomy of an MRS
- Semantic composition
Semantics in sign-based grammar

- Each sign (lexeme, word, phrase) is a form-meaning pair.
- Lexical items link syntactic arguments to semantic arguments.
- Each sign ‘exposes’ information for further composition.
- Syntactic constructions can also contribute semantic information.

```
S
   /\    
  /  \   
/    \   
NP_i  VP[SUBJ⟨⟩]
       /\    
      /  \   
     /    \  
    SUBJ⟨⟩  COMPS⟨⟩
              /\     
             /  \   
            /    \  
           RELS⟨like(i,j)⟩
                        /\    
                       /  \   
                      /    \  
                     likes bagels
```

LSA.343 Precision Grammar Implementation (7)
Overall strategy

- Represent semantic distinctions corresponding to differences in form.
- Underspecify other semantic distinctions.
- Abstract away from non-semantic information
- Aim for consistency across languages
- Allow for semantic differences between languages.

Minimal Recursion Semantics — Goals

- Adequate representation of natural language semantics
- Grammatical compatibility
- Computational tractability
- Underspecifiability
Working towards MRS

- Every big white horse sleeps.
- $\text{every } (x, \land (\text{big}(x), \land (\text{white}(x), \text{horse}(x))), \text{sleep}(x))$
Flat semantics

\[ h_0: \text{every}(x, h_1, h_2), h_1: \text{big}(x), h_1: \text{white}(x), h_1: \text{horse}(x), \]
\[ h_2: \text{sleep}(x) \]

... with partially specified quantifier scope

\[ h_0: \text{every}(x, h_1, h_2), h_3: \text{big}(x), h_3: \text{white}(x), h_3: \text{horse}(x), \]
\[ h_4: \text{sleep}(x) \]
\[ h_1 \text{ ‘equal modulo quantifiers’ } h_3 \]
Anatomy of an MRS

- An MRS consists of:
  - A top handle
  - A list of relations, each labeled by a handle
  - A list of handle constraints
- An (underspecified) MRS is well-formed iff the constraints can be resolved to form one or more trees.

Linguistic questions

- How do we build MRS representations compositionally?
- Is it linguistically adequate to insist that no process suppress relations?
- Under what circumstances do NLs (partially) constrain scope?
Example MRS (simple MRS view)

```
[ LTOP   h1
 INDEX  e2
 RELS   [ exist_q_rel 
           LBL    h3
           ARG0  x4
           RSTR   h5
           BODY   h6
        ,  _cat_n_rel
        LBL    h7
        ARG0  x4
        ARG1  x4
        _sleep_v_rel
        LBL    h1
        ARG0  e2
        ARG1  x4
     ]
 HCONS  ( h5 qeq h7 )
```
Semantic compositionality

- Phrase structure rules (and lexical rules) gather up RELS and HCONS from daughters.
- Phrase structure rules also (optionally) introduce further RELS and HCONS.

```
basic-unary-phrase := phrase &
    [ SYNSEM.LOCAL.CONT [ RELS [ LIST #first,
                             LAST #last ]],
      C-CONT [ RELS [ LIST #mid,
                     LAST #last ]],
      ARGS < sign & [ SYNSEM.LOCAL
                     [ CONT [ RELS [ LIST #first,
                                   LAST #mid ]]]]>>].
```
Semantic compositionality

- Each sign exposes information for further composition in its HOOK.
- By hypothesis, this includes only:
  - INDEX (the individual or event denoted by the sign, linked to some ARG0)
  - LBL (the local top handle of the sign)
  - XARG (the external argument of the sign)
- Selecting heads find the indices they want in the HOOK of their arguments.
HOOK and LKEYS

intransitive-lex-item := basic-one-arg-no-hcons &
[ ARG-ST < [ LOCAL.CONT.HOOK.INDEX ref-ind & #ind ] >,
  SYNSEM.LKEYS.KEYREL.ARG1 #ind ].

intersective-mod-lex := no-hcons-lex-item &
[ SYNSEM [ LOCAL.CAT.HEAD.MOD < [ ..INDEX #ind ]] >,
  LKEYS.KEYREL.ARG1 #ind ] ].
**Semantic compositionality: Summary**

- Each sign includes a semantic representation.
- The MRS of a phrase (or lexical rule ‘output’) is a function of the MRSs of the daughter(s) and the definition of the rule itself.
- Each sign exposes information for further composition in its HOOK.
- Selecting heads find the indices they want in the HOOK of their arguments.

---

*ABabcdfghiejkl*

*LSA.343 Precision Grammar Implementation (16)*
Labs this week

- Test the Grammar Matrix + libraries for 5 languages
  - Create test suite
  - Customize and download matrix
  - Process test suite with grammar

- Extend the small grammars
  - Examine grammar competence through [incr tsdb()]
  - Explore overgeneration through generation
  - Settle on phenomenon to add
  - Extend test suite to map new territory
  - Implement analysis of new phenomenon