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

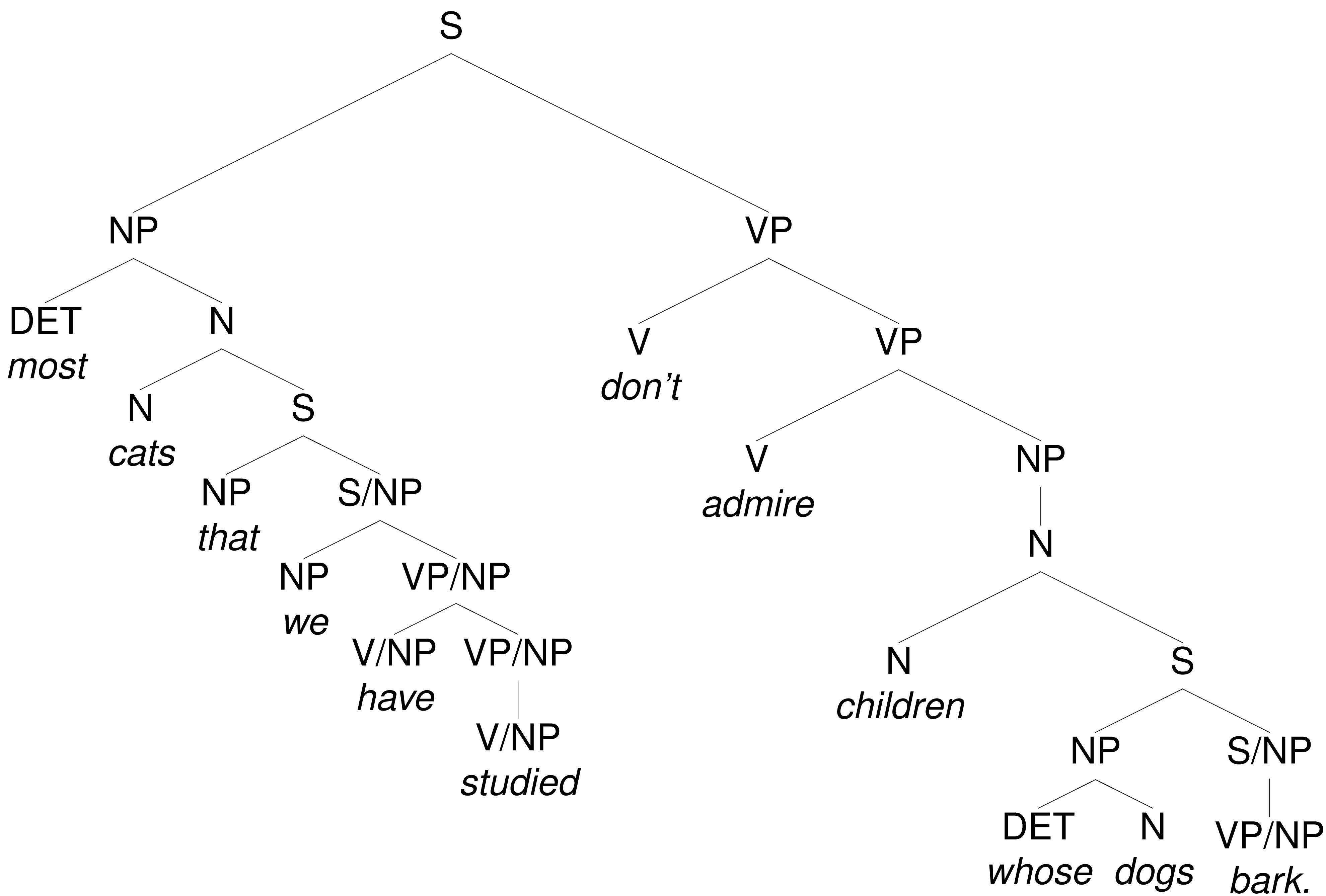
Desiderata for a linguistic theory

- 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



A simple example: MRS – linking

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

<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.

<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: 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

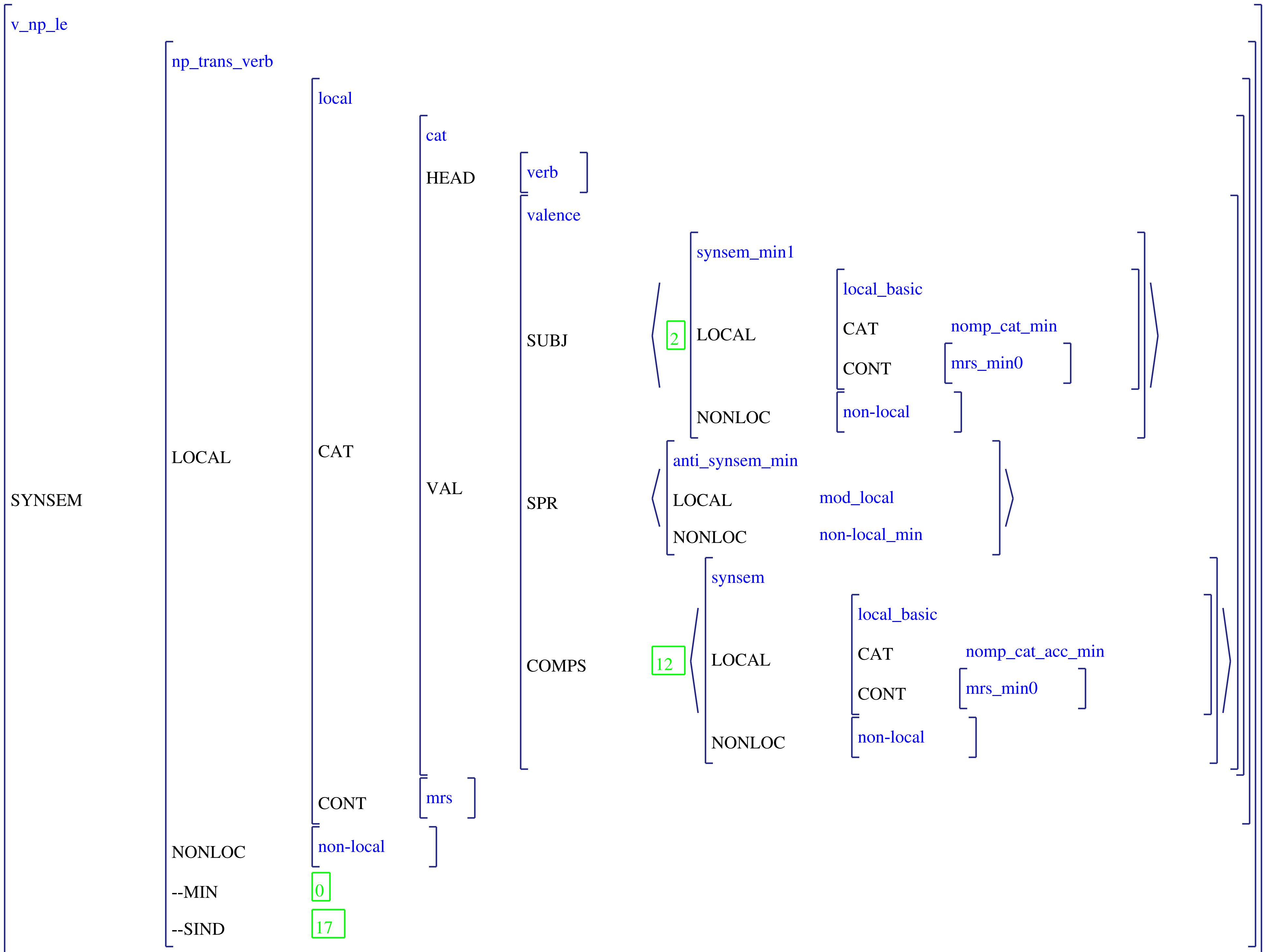
  'The HEAD value of a phrase is identified with that of its (syntactic) head daughter'

- 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


v_np_le
SYNSEM
np_trans_verb
LOCAL
local
CAT
cat
HEAD
verb
VAL
valence
SUBJ
2
synsem_min1
LOCAL
local_basic
CAT
nomp_cat_min
CONT
mrs_min0
NONLOC
non-local
SPR
anti_synsem_min
LOCAL
mod_local
NONLOC
non-local_min
COMPS
12
synsem
LOCAL
local_basic
CAT
nomp_cat_acc_min
CONT
mrs_min0
NONLOC
non-local
CONT
mrs
NONLOC
non-local
--MIN
0
--SIND
17

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)

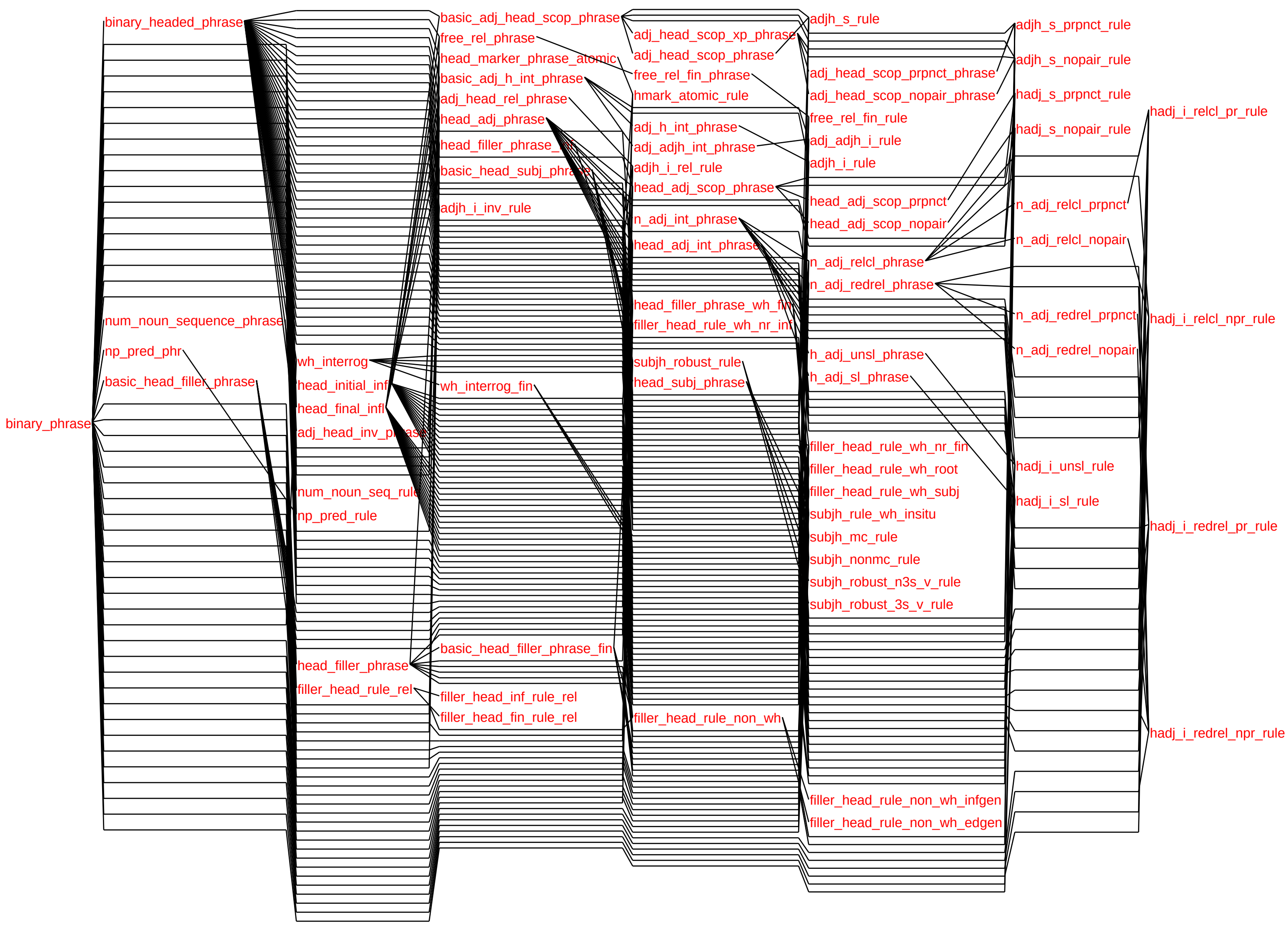
English Resource Grammar (ERG)

- 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

```
v_np_le
SYNSEM      np_trans_verb
            LOCAL     local
                      CAT      cat
                               HEAD   [verb]
                               VAL    valence
                                      SUBJ    < [2] synsem_min1
                                                    LOCAL    local_basic
                                                             CAT   nomp_cat_min
                                                             CONT  [mrs_min0]
                                                    NONLOC   [non-local] >
                                      SPR     < anti_synsem_min
                                                LOCAL    mod_local
                                                NONLOC   non-local_min >
                                      COMPS   [12] < synsem
                                                     LOCAL    local_basic
                                                              CAT   nomp_cat_acc_min
                                                              CONT  [mrs_min0]
                                                     NONLOC   [non-local] >
                      CONT     [mrs]
            NONLOC    [non-local]
            --MIN     [0]
            --SIND    [17]
ARGS        *list*
INFLECTD    -
```

Semantics in feature structures

$$
\begin{bmatrix}
\text{ORTH} & \text{adore} \\
\text{HEAD} & \textit{verb} \\
\text{SUBJ} & \left\langle \begin{bmatrix} \text{HEAD} & \textit{noun} \\ \text{CONT} & \begin{bmatrix} \text{INDEX} & \boxed{2} \end{bmatrix} \end{bmatrix} \right\rangle \\
\text{COMPS} & \left\langle \begin{bmatrix} \text{HEAD} & \textit{noun} \\ \text{CONT} & \begin{bmatrix} \text{INDEX} & \boxed{3} \end{bmatrix} \end{bmatrix} \right\rangle \\
\text{CONT} & \begin{bmatrix} \text{INDEX} & \boxed{1}\ \textit{event} \\ \text{RELS} & \left\langle \begin{bmatrix} \text{PRED} & \text{“adore\_v”} \\ \text{ARG0} & \boxed{1} \\ \text{ARG1} & \boxed{2} \\ \text{ARG2} & \boxed{3} \end{bmatrix} \right\rangle \\ \text{HCONS} & \langle\ \rangle \end{bmatrix}
\end{bmatrix}
$$

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:

$$\begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{SUBJ} & \langle\ \rangle \\ \text{CMPS} & \langle\ \rangle \\ \text{CONT} & \begin{bmatrix} \text{INDX} & \boxed{1}\ \textit{event} \\ \text{RELS} & \left\langle \begin{bmatrix} \text{PRED} & \text{“adore\_v”} \\ \text{ARG0} & \boxed{1} \\ \text{ARG1} & \boxed{2} \\ \text{ARG2} & \boxed{3} \end{bmatrix}, \begin{bmatrix} \text{PRED} & \text{“Kim\_n”} \\ \text{ARG0} & \boxed{2} \end{bmatrix}, \begin{bmatrix} \text{PRED} & \text{“snow\_n”} \\ \text{ARG0} & \boxed{3} \end{bmatrix} \right\rangle \\ \text{HCNS} & \langle\ \rangle \end{bmatrix} \end{bmatrix}$$

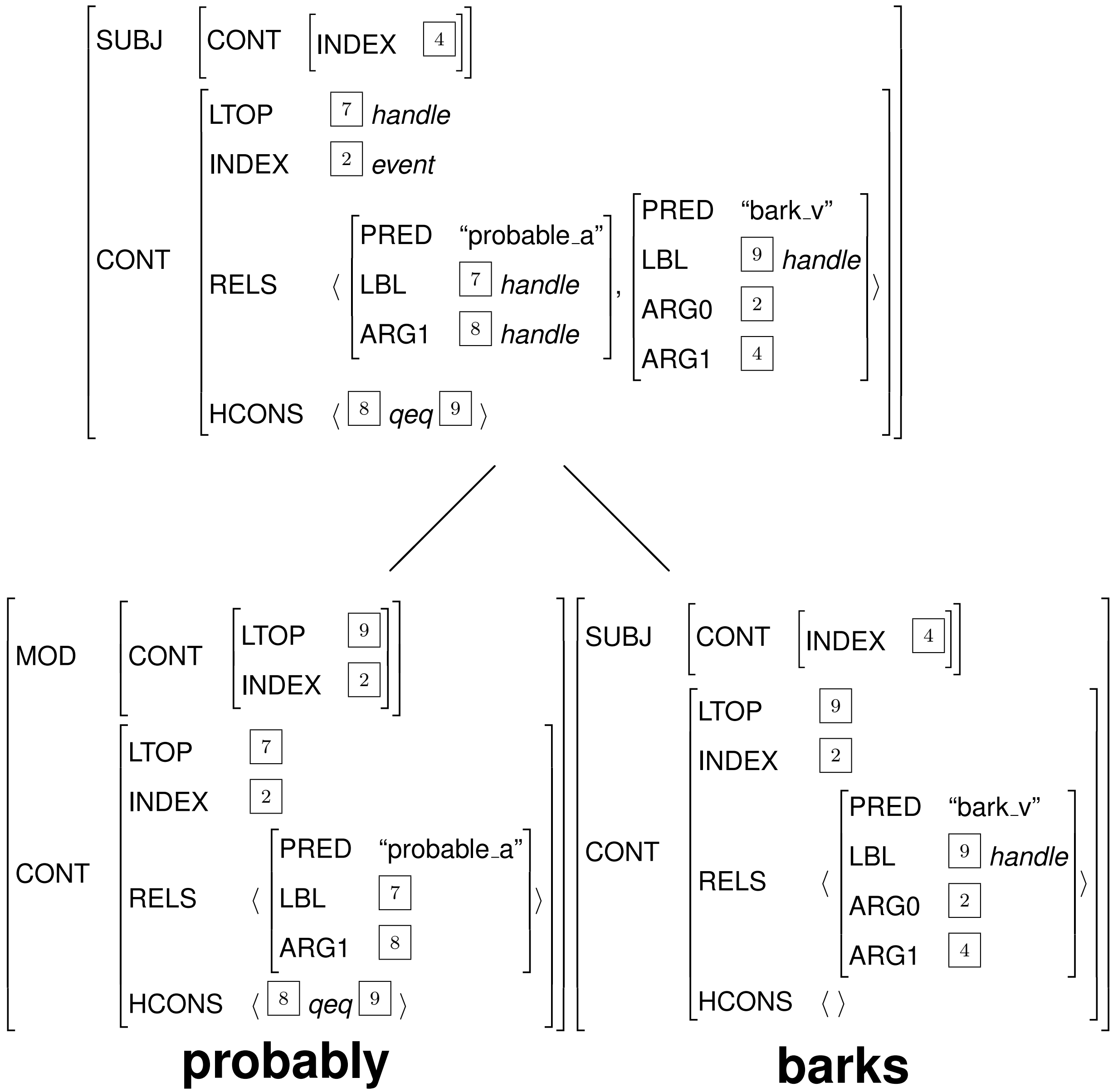
Semantic composition for Minimal Recursion Semantics

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

In addition, to accommodate scope underspecification:

- The **HCONS** value of a phrase is the result of appending the **HCONS** lists of its daughter(s), plus any from the construction itself.
- The **LTOP** value of a phrase is unified with the **LTOP** value of its semantic head daughter.

MRS Composition: “*probably barks*”

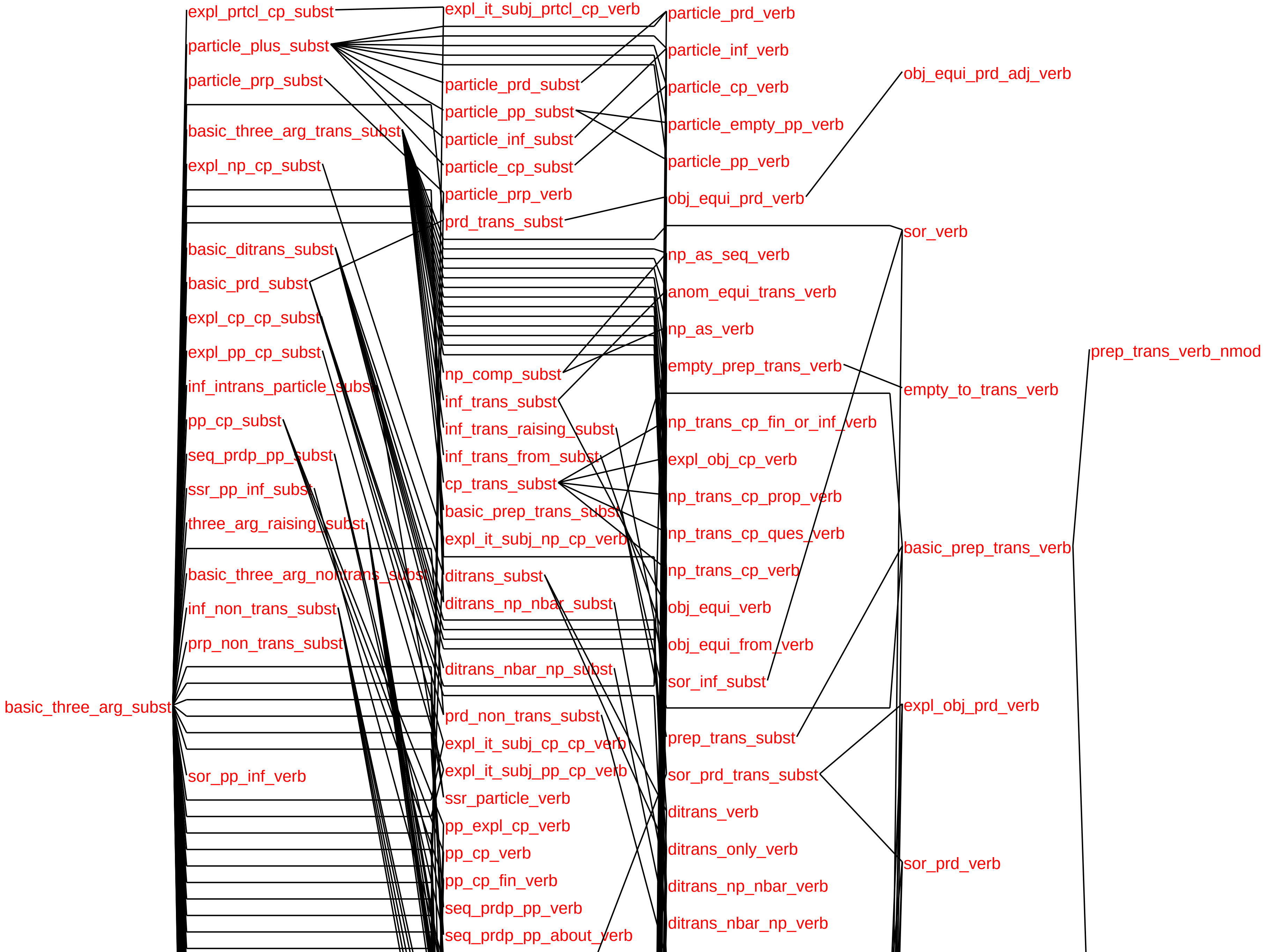


Linking semantic arguments

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

$$
\begin{bmatrix}
\textit{trans-verb-lxm} & \\
\text{HEAD} & \textit{verb} \\
\text{SUBJ} & \left\langle \begin{bmatrix} \text{CONT} & \begin{bmatrix} \text{INDEX} & \boxed{1} \end{bmatrix} \end{bmatrix} \right\rangle \\
\text{COMPS} & \left\langle \begin{bmatrix} \text{CONT} & \begin{bmatrix} \text{INDEX} & \boxed{2} \end{bmatrix} \end{bmatrix} \right\rangle \\
\text{CONT} & \begin{bmatrix} \text{INDEX} & \boxed{3} \\ \text{RELS} & \left\langle \begin{bmatrix} \text{PRED} & \text{string} \\ \text{ARG0} & \boxed{3} \\ \text{ARG1} & \boxed{1} \\ \text{ARG2} & \boxed{2} \end{bmatrix} \right\rangle \end{bmatrix}
\end{bmatrix}
$$

Lexical type hierarchy (verbs)

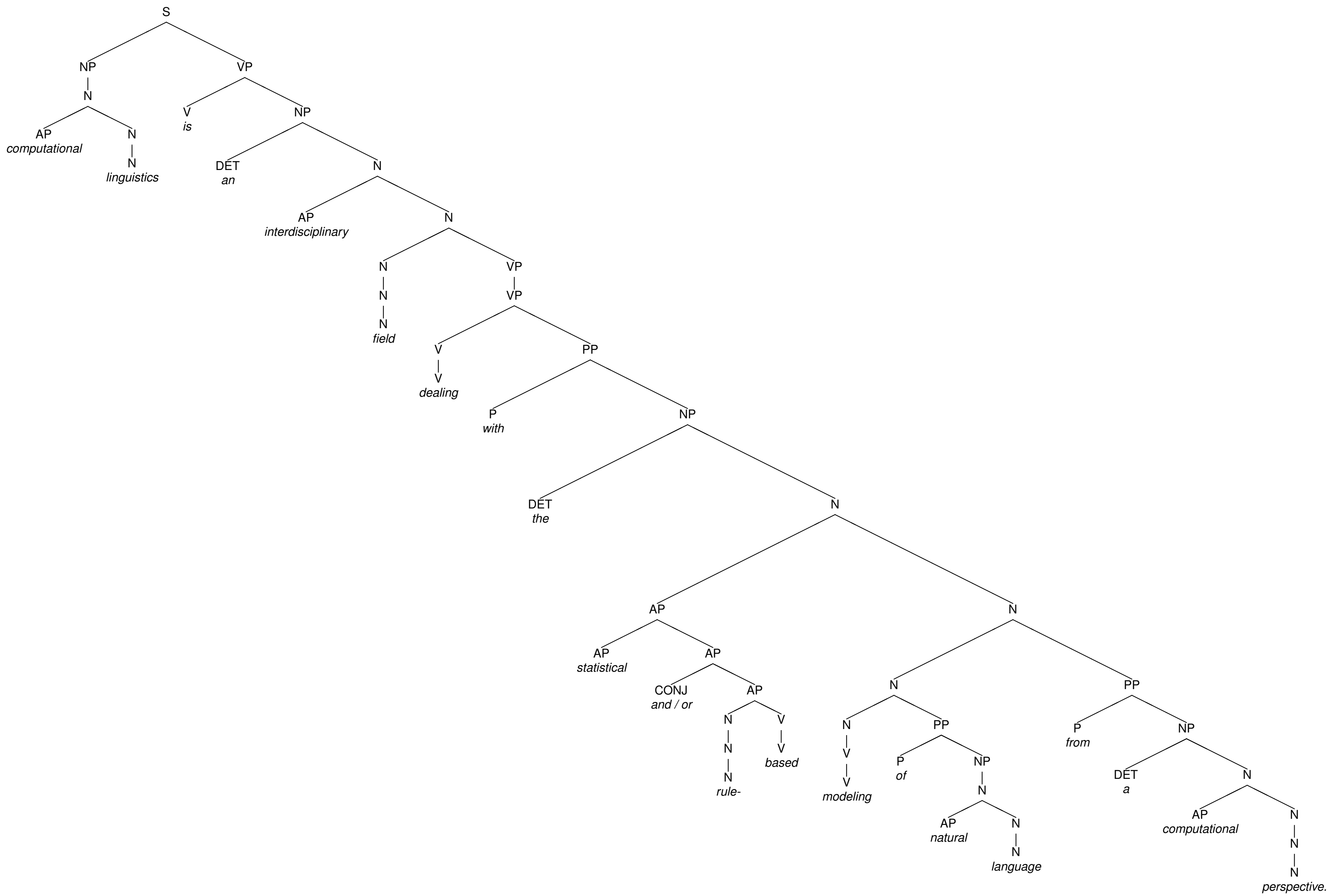


basic_three_arg_subst
expl_prtcl_cp_subst
particle_plus_subst
particle_prp_subst
basic_three_arg_trans_subst
expl_np_cp_subst
basic_ditrans_subst
basic_prd_subst
expl_cp_cp_subst
expl_pp_cp_subst
inf_intrans_particle_subst
pp_cp_subst
seq_prdp_pp_subst
ssr_pp_inf_subst
three_arg_raising_subst
basic_three_arg_nontrans_subst
inf_non_trans_subst
prp_non_trans_subst
sor_pp_inf_verb
expl_it_subj_prtcl_cp_verb
particle_prd_subst
particle_pp_subst
particle_inf_subst
particle_cp_subst
particle_prp_verb
prd_trans_subst
np_comp_subst
inf_trans_subst
inf_trans_raising_subst
inf_trans_from_subst
cp_trans_subst
basic_prep_trans_subst
expl_it_subj_np_cp_verb
ditrans_subst
ditrans_np_nbar_subst
ditrans_nbar_np_subst
prd_non_trans_subst
expl_it_subj_cp_cp_verb
expl_it_subj_pp_cp_verb
ssr_particle_verb
pp_expl_cp_verb
pp_cp_verb
pp_cp_fin_verb
seq_prdp_pp_verb
seq_prdp_pp_about_verb
particle_prd_verb
particle_inf_verb
particle_cp_verb
particle_empty_pp_verb
particle_pp_verb
obj_equi_prd_verb
np_as_seq_verb
anom_equi_trans_verb
np_as_verb
empty_prep_trans_verb
np_trans_cp_fin_or_inf_verb
expl_obj_cp_verb
np_trans_cp_prop_verb
np_trans_cp_ques_verb
np_trans_cp_verb
obj_equi_verb
obj_equi_from_verb
sor_inf_subst
prep_trans_subst
sor_prd_trans_subst
ditrans_verb
ditrans_only_verb
ditrans_np_nbar_verb
ditrans_nbar_np_verb
obj_equi_prd_adj_verb
sor_verb
empty_to_trans_verb
basic_prep_trans_verb
expl_obj_prd_verb
sor_prd_verb
prep_trans_verb_nmod

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



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.

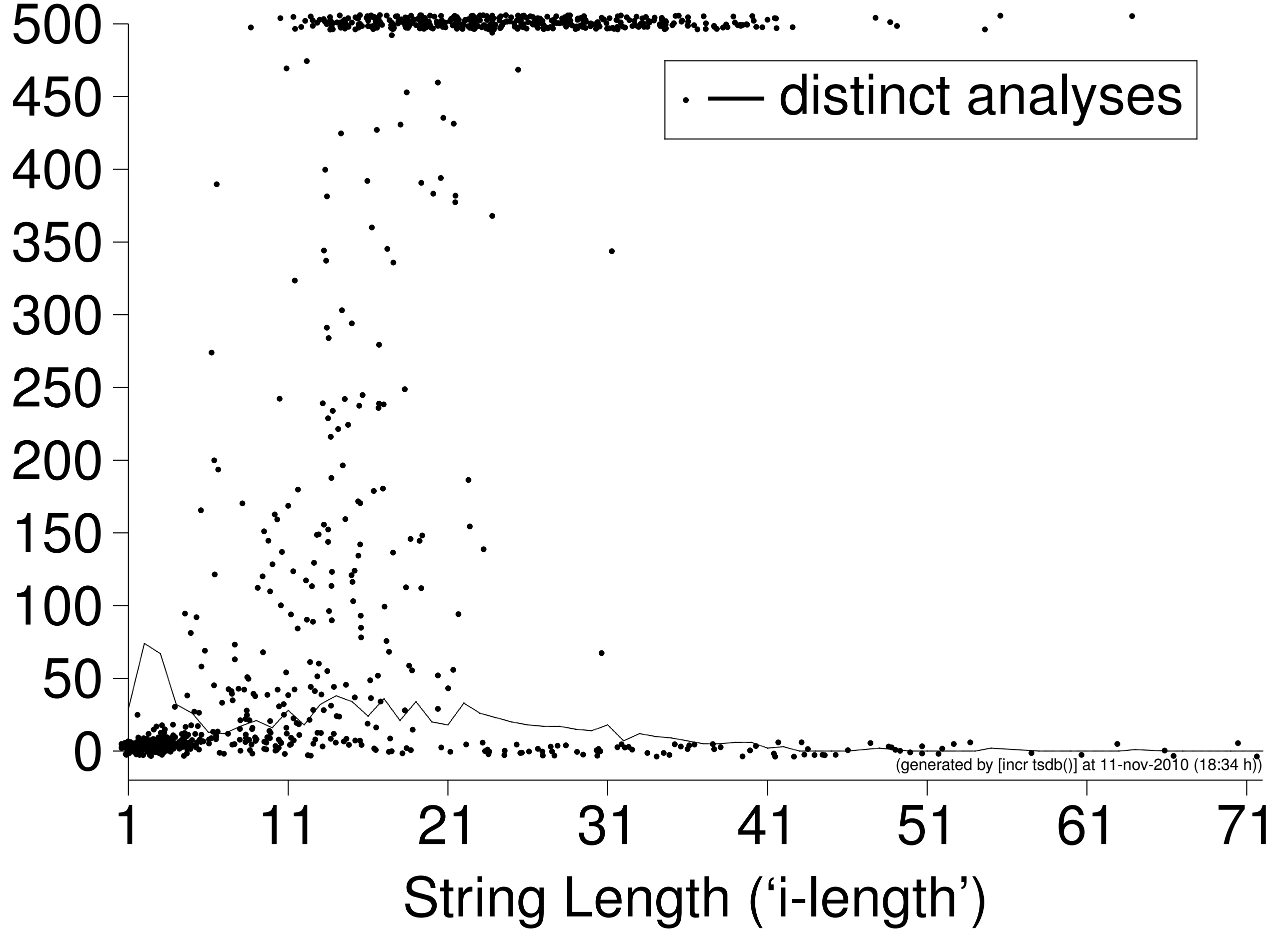
{h3:udef_q<3:27>(x5, h4, h6),
h7:_computational_a_1<3:15>(e8, x5),
h7:_linguistics_n_1<17:27>(x5),
h9:_be_v_id<32:33>(e2, x5, x10),
h11:_a_q<35:36>(x10, h13, h12),
h14:_interdisciplinary/jj_u_unknown<40:56>(e15, x10),
h14:_field_n_of<60:64>(x10, i16),
h14:_**deal_v_with**<66:72>(e17, x10, x18),
h19:_the_q<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),
...

Wikipedia parsing with ERG/PET: Coverage

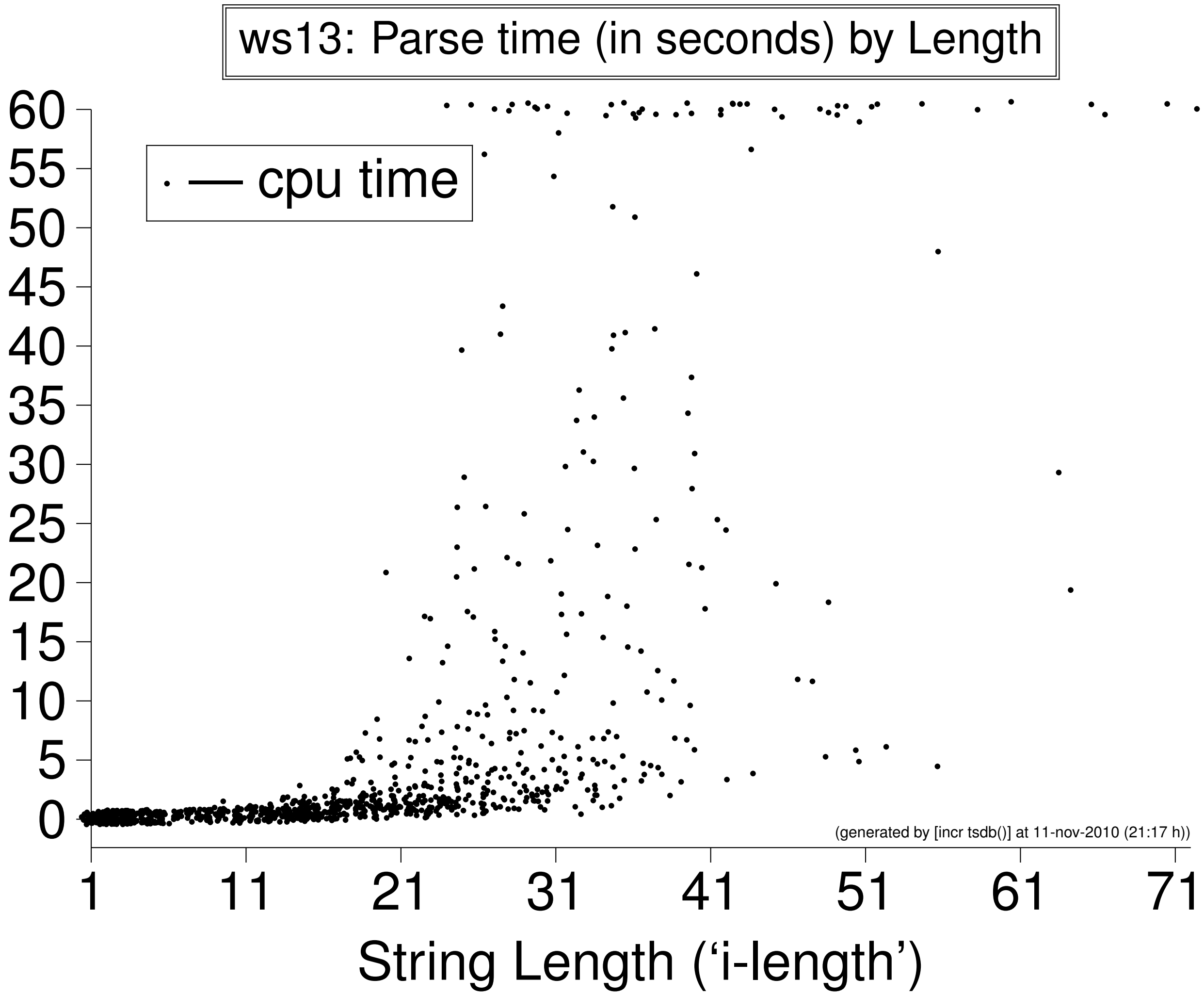
‘ws13’ Coverage Profile						
Length in tokens	**total items**	**word string**	**lexical items**	**distinct analyses**	**total results**	**overall coverage**
65 – 70	2	66.50	755.00	0.00	0	0.0
60 – 65	3	62.33	538.67	500.00	1	33.3
55 – 60	3	56.00	343.33	500.00	2	66.7
50 – 55	8	51.37	563.50	0.00	0	0.0
45 – 50	11	47.27	542.55	500.00	3	27.3
40 – 45	21	41.38	422.90	500.00	11	52.4
35 – 40	42	36.69	433.64	500.00	32	76.2
30 – 35	77	31.78	321.78	490.26	61	79.2
25 – 30	101	26.89	329.47	499.68	87	86.1
20 – 25	130	22.17	220.02	465.31	120	92.3
15 – 20	158	16.98	177.37	374.31	149	94.3
10 – 15	137	12.32	126.58	197.77	132	96.4
5 – 10	97	6.86	67.01	42.46	89	91.8
0 – 5	209	2.47	11.02	3.42	202	96.7
Total	**1001**	**17.55**	**184.78**	**270.04**	**889**	**88.8**

(generated by [incr tsdb()] at 11-nov-2010 (21:09 h))

Wikipedia parsing with ERG/PET: Ambiguity



Wikipedia parsing with ERG/PET: Efficiency



Wikipedia parsing with ERG/PET: Accuracy

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

Frequency of Linguistic Phenomena in Wikipedia 100

Phenomenon	♯Items	%Corpus
Measure NPs	44	0.3
Appositives	1048	9.1
NP Fragments	2126	18.4
NP Coordination	1960	17.0
Multi-NP Coord	558	4.8
VP Coordination	491	4.2
S Coordination	381	3.3
Relative Clauses	2239	19.4
Unbounded Deps	2273	19.7
Yes-No Questions	11	0.1
WH Questions	10	0.1
Imperatives	222	1.9
Free relatives	107	0.9
Passives	3534	30.6

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`
  - 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)