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
Properties of Multi-word Expressions

- Syntactically flexible/rigid
- Syntactically normal/idiosyncratic
- Semantically transparent/semi-transparent/opaque
Inflexible Multi-word Expressions

- **Words with spaces**
  - No variation: *ad hoc, in spite of, by and large, in short*
    - *in shorter, in very short*

- **Semi-fixed expressions**
  - Inflectional morphology
    - *kicked the bucket, shot the breeze, parts of speech*
  - Reflexive pronoun variants
    - *kick oneself/herself/ourselves/...* (to regret)
  - Syntactic inflexibility
    - *the breeze was shot, shot the big breeze, shot a breeze*
Flexible Multi-word Expressions

• Head plus dependent
  Verb-particle constructions: look X up, call X up, phone X up
  but *telephone X up
  Determinerless PPs: on top, at eye/ground level
  but *on very top, *at level
  Light verbs: take a bath, give a lecture
  but *do a bath, *make a lecture
• Idioms: keep tabs on X, take note of X
• Proper names:
  International Business Machines, Macy’s Inc.
  the San Francisco 49ers
Words with spaces in the ERG lexicon

One should avoid ad hoc solutions.

ad_hoc_a1 := aj_i_le &
[ ORTH < "ad", "hoc" >,
  SEMPRED ".ad+hoc_a_1.rel" ].

They were by and large just starting to understand.

by_and_large_adv1 := av_i-vp_le &
[ ORTH < "by", "and", "large" >,
  SEMPRED ".by+and+large_a_1.rel" ].
Inflexible idioms in the ERG lexicon

That politician finally kicked the bucket.

```
kick_v1_i := v_np_sg-idm_le &
   [ ORTH < "kick" >,
      SEMPRED "_kick_v_i_rel" ].
```

```
bucket_n1 := n-_c_le &
   [ ORTH < "bucket" >,
      SEMPRED "_bucket_n_1_rel" ].
```

```
kick+the+bucket := v_np_idiom_mtr &
   [ INPUT.RELS < [ PRED "_kick_v_i_rel" ],
      [ PRED "_bucket_n_1_rel" ] > ].
```
Verb-particle constructions

- Usually intransitive or transitive verbs
- Transitives usually allow reordering of NP and particle
  So not [V + PP] but rather [V + P + NP] or [V + NP + P]
- Transitives also allow passives, both verbal and adjectival
- Semantics can be transparent or opaque
Some more verb-particle constructions

- **Verb + Particle + PP**
  
  *get away with X, keep up with X, make up for X.*

- **Verb + Particle + Clause, with expletive *it* subject**
  
  *It turns out that he needs a ride.*

- **Verb + Particle + Clause, with normal subject**
  
  *We pointed out that he needs a ride.*

- **Verb + Particle + PP + Clause**
  
  *We pointed out to her that he needs a ride.*

- **Verb + Particle + Infinitival-VP**
  
  *She turned out to be a genius.*

- **Verb + Particle + VP-ing**
  
  *We kept on making the same mistake.*

- **Verb + Particle + PredPhrase**
  
  *He ended up richer than you.*
One more frequent verb-particle construction

• Verb + NP + Particle + PP

  *We boiled the discussion down to two basic ideas.*
  *She made the check out to her brother.*
  *The wall was made up of old bricks.*
  *You should trade in your old phone for a new one.*
Verb + Particle in the ERG lexicon

We need to look that answer up.

\[
\begin{align*}
\text{look_up_v1} & := v\_p\text{-np}\_le \ & \text{&} \\
& [ \text{ORTH} < "\text{look}" >, \\
& \text{SEMPRED } "\text{\_look_v\_up_rel}"], \\
& \text{COMPKEY } \text{\_up_p\_sel_rel }]. \\
\end{align*}
\]

\[
\begin{align*}
\text{up_prtcl} & := p\_np\_ptcl\_le \ & \text{&} \\
& [ \text{ORTH} < "\text{up}" >, \\
& \text{SEMPRED } \text{\_up_p\_sel_rel }]. \\
\end{align*}
\]

\[
\begin{align*}
\text{NP\_particle\_lr} & := \text{lex\_rule} \ & \text{&} \\
& [ \text{DTR } [ \text{SYNSEM } \text{generic\_NP\_particle\_verb }], \\
& \text{SYNSEM } \text{generic\_particle\_NP\_verb}, \\
& \text{ALTS.\_NPPART } + ]; \\
\end{align*}
\]
Verb + Particle + PP

They made up for their earlier mistakes.

make_up_for_v1 := v_p-pp_le &
[ ORTH < "make" >,
  SEMPRED "_make_v_up-for_rel",
  COMPKEY _up_p_sel_rel,
  OCOMPKEY _for_p_rel ].
Inflexible order for Verb + Particle

We will see you around.
*We will see around you.

They should do the problem over.
*They should do over the problem.

doover_v1 := v_np-p_le &
[ ORTH < "do" >,
  SEMPRED "_do_v_over_rel",
  COMPKEY _over_p_sel_rel ].

v_np-p_le := main_verb &
[ SYNSEM np_particle_noalt_verb,
  ALTS.NPPART - ].
Determinerless PPs in the ERG

That book should really be on top (of the stack).
*Top of a tall building is usually a windy place.

top_i_n1 := n_pp_c-brno-o_f_le &
  [ ORTH < "top" >,
    SEMPRED "_top_n_1_rel" ].

on+top := detless_pp_idiom_mtr &
  [ INPUT.RELS < [ PRED _on_p_rel ],
    [ PRED "_top_n_1_rel" ],
    [ PRED idiom_q_i_rel ] > ].

detlesspp_np_phrase := generic_bare_np_phrase &
  [ ARGS < [ SYNSEM.MINOR bare_nom_rel ] > ],
  C-CONT.RELS < [ PRED idiom_q_i_rel ] > ].
Varieties of Detless PPs

- No modification
  - *on top, *on extreme top, *on bookcase top
- Optional modification with adjective
  - in (sharp) contrast, in (full) view
- Optional modification with compound
  - in (lock) step, on (summer) holiday
  - *in equal step, *on long holiday
- Obligatory modification with adjective
  - at close range
  - *at range
- Obligatory modification with compound
  - at eye level, at ground level, at sea level
  - *at level
Lexical types for Detless PPs in the ERG

n_-c-brno_le : No modification
n_-c-brj*_le : Optional adjective modification
n_-c-brj_le : Obligatory adjective modification
n_-c-brn*_le : Optional compound modification
n_-c-brn_le : Obligatory compound modification
### Types of idioms in the ERG

- **V + bare-N + PP**
  - *give rise to, make way for, take note of*
- **V + bare-N**
  - *hit bottom, play catch, take place*
- **Light-V + Adjective**
  - *make sure (of X), make light of X, make good on X*
- **V + NP + XP**
  - *keep X company, keep X a secret, call X quits*
- **V + NP**
  - *kick the bucket, suck eggs*
- **N + P + bare-N**
  - *point of view, sleight of hand*
- **V + NP + P + XP**
  - *take X for granted, take X completely for granted*
Idiom rules in the ERG

That comment gave rise to a big argument.

\[
give + rise + to := v_{nbar\ pp\ idiom\ mtr} \&
\]\[
[ INPUT.RELS < [ PRED "_give_v_to-i_rel" ],
[ PRED "_rise_n_i_rel" ] > ].
\]

Several meetings took place quietly.

\[
take + place := v_{nbar\ idiom\ mtr} \&
\]\[
[ INPUT.RELS < [ PRED "_take_v_of-i_rel" ],
[ PRED "_place_n_i_rel" ] > ].
\]

We shouldn’t make light of his problems.

\[
make + light := v_{light\ adj\ idiom\ mtr} \&
\]\[
[ INPUT.RELS < [ PRED "_make_v_i_rel" ],
[ PRED "_light_a_of_rel" ] > ].
\]
More idiom rules in the ERG

Someone needs to keep him company while he’s ill.

\[
\text{keep+company := v\_np\_xp\_idiom\_mtr &}
\]
\[
[\text{INPUT.RELS < [ PRED "_keep_v_i_rel" ]},
\text{[ PRED "_company_n_of_rel" ] > }].
\]

That politician finally kicked the bucket.

\[
\text{kick+the+bucket := v\_np\_idiom\_mtr &}
\]
\[
[\text{INPUT.RELS < [ PRED "_kick_v_i_rel" ]},
\text{[ PRED "_the_q_rel" ]},
\text{[ PRED "_bucket_n_1_rel" ] > }].
\]
More idiom rules in the ERG

I was sympathetic to that point of view.

point+of+view := noun_detless_pp_idiom_mtr &
    [ INPUT.RELS < [ PRED "_point_n_of_rel" ],
      [ PRED "_view_n_of_rel" ],
      [ PRED idiom_q_i_rel ] > ].

point_n3 := n_pp_c-ns-obl_le &
    [ ORTH < "point" >,
      SEMPRED "_point_n_of_rel",
      COMPKEY _of_p_sel_rel ].

Some syntactically odd MWEs

• We often take him for granted.
• we took ahold of the handle.
• She is the buyer of last resort.
Weaknesses of this approach to MWEs

- Manual effort is required to define lexicon and idiom constraints
- Post-parse filtering on MRS can lead to higher parsing costs
- Some MWE-specific syntactic rules are required (e.g. bare-NP)
Strengths of this approach to MWEs

- Strong limits on overgeneration of interpretation and variation
- Less redundancy in the lexicon and grammar
- Good interaction with syntactic machinery
- No principled obstacle to cross-clausal idioms

Kim resents many tabs we kept on him.