

# Roadmap for Adding WordNet Senses to the ERG Lexicon

**Dan Flickinger**

DELPH-IN Summit 2023

Galicia, Spain

27 June 2023

# Motivation

- Addition of word senses
  - Synonymy and other semantic relations
  - Anchoring ERS predicates to standard senses
  - Support for logical reasoning
  - Ideally, improved disambiguation (e.g. PP attachment)
- Larger lexicon (45K growing to 160K)
  - Less dependence on POS tagger for parsing unknown words
  - Improved generation (limited support for unknown words)

# Strategy

- Link to Open English WordNet ([en-word.net](http://en-word.net))

- Use coarse-grained top-level sense types

Nouns (26): *act, animal, artifact, attribute, body, cognition, communication, event, feeling, food, group, location, motive, object, person, phenomenon, plant, possession, process, quantity, relation, shape, state, substance, time, top*

Verbs (17): *body, change, cognition, communication, competition, consumption, contact, creation, emotion, motion, perception, possession, social interaction, state, weather*

Adjectives (3)

- Also record fine-grained synsets in each lexical entry

## Migration path via script (thanks, Francis)

- Produce expanded lexicon for all single-token nouns, verbs, adjs
- Include verb-particle entries such as *look up*
- Postpone MWEs such as *car pool*

## Migration path via script (thanks, Francis)

- Produce expanded lexicon for all single-token nouns, verbs, adjs
  - Include verb-particle entries such as *look up*
  - Postpone MWEs such as *car pool*
- 
- Add coarse-grained senses to words in both ERG, WordNet (50K)  
*inform*
  - Add entries found in WordNet but missing in ERG (75K)  
*paleolithic*
  - Propose additions to WordNet for ERG entries not in WordNet  
*paraglider*

# Old vs. new lexical entries

```
bark_n1 := n_-_mc_le &  
  [ ORTH < "bark" >,  
    SYNSEM.LKEYS.KEYREL.PRED "_bark_n_1_rel" ].
```

```
bark_n1-pln := n_-_mc-pln_le &  
  [ ORTH < "bark" >,  
    SYNSEM.LKEYS.KEYREL.PRED "_bark_n_pln_rel",  
    CONCEPTS < "13183195-n" > ].  
;; DEF tough protective covering of the woody stems and  
;; roots of trees and other woody plants
```

```
bark_n1-evt := n_-_mc-evt_le &  
  [ ORTH < "bark" >,  
    SYNSEM.LKEYS.KEYREL.PRED "_bark_n_evt_rel",  
    CONCEPTS < "07391221-n", "07391331-n" > ].  
;; DEF the sound made by a dog
```

# Enriched lexical types for treebanking

```
v_np-cp_lexent := main_verb &  
  [ SYNSEM np_trans_cp_verb &  
    [ LOCAL.CAT.VAL.COMPS.FIRST.OPT - ] ] .
```

```
v_np-cp_chn_le := v_np-cp_lexent.  
v_np-cp_cog_le := v_np-cp_lexent.  
v_np-cp_com_le := v_np-cp_lexent.  
v_np-cp_cmp_le := v_np-cp_lexent.  
v_np-cp_con_le := v_np-cp_lexent.  
v_np-cp_crt_le := v_np-cp_lexent.  
v_np-cp_emo_le := v_np-cp_lexent.
```

# Manual pruning of expanded lexicon

## Keep:

```
inform_v1-com := v_np-cp_com_le &  
  [ ORTH < "inform" >,  
    SYNSEM.LKEYS.KEYREL.PRED "_inform_v_com_rel" ].  
  CONCEPTS < "00834860-v", "00833312-v" > ].  
;; DEF impart knowledge of some fact, state of affairs, or event to  
;; EX I informed him of his rights
```

## Reject:

```
inform_v1-chn := v_np-cp_chn_le &  
  [ ORTH < "inform" >,  
    SYNSEM.LKEYS.KEYREL.PRED "_inform_v_chn_rel",  
    CONCEPTS < "00523693-v" > ].  
;; DEF give character or essence to  
;; EX The principles that inform modern teaching
```



## Next steps

- Complete manual clean-up of the three lexicon partitions
  - Linking of verb senses to finer-grained subcat frames (230)
  - Consistency for causative/inchoative verbs
  - Mass/count distinction for nouns

## Next steps

- Complete manual clean-up of the three lexicon partitions
  - Linking of verb senses to finer-grained subcat frames (230)
  - Consistency for causative/inchoative verbs
  - Mass/count distinction for nouns
- Update Redwoods treebank adding (coarse-grained) senses
- Train new parsing model which includes sense discriminants

## Next steps

- Complete manual clean-up of the three lexicon partitions
  - Linking of verb senses to finer-grained subcat frames (230)
  - Consistency for causative/inchoative verbs
  - Mass/count distinction for nouns
- Update Redwoods treebank adding (coarse-grained) senses
- Train new parsing model which includes sense discriminants
- Compare with manually sense-tagged data (Sherlock Holmes)
- Evaluate parse selection accuracy
  - PP-attachment
  - Apposition
  - Relative clauses
  - Coordination