

# Lexicon of verb alternations

LinGO Project

CSLI, Stanford University

## 1 Overview

The information about lexical entries for high-frequency verbs is stored in a relational database in Microsoft Access format. When you download the database (<http://lingo.stanford.edu/resources/DB10.MDB>), you'll get a single file that can be read by Microsoft Access. Once loaded, you can select "Queries", and then "IBMresults", which will bring up the main table showing the lexical entries and their properties. The structure of the table representing this information is as follows.

Each row represents one lexical entry for a group of word senses which share common subcategorization and thematic role properties, which are encoded in fifteen columns:

Column 1:	Spelling for base form
Columns 2-4:	Wordnet senses
Column 5:	Developer notes
Column 6:	Example sentence for this entry
Columns 7-9:	Syntactic subcategorization for up to three complements
Columns 10-12:	Corresponding semantic constraints for these complements
Columns 13-15:	Corresponding thematic roles for these complements

### 1.1 Senses

The numbers in the three columns immediately following the spelling indicate which of the Wordnet senses are excluded (column 2) or expressed by this entry (column 3). In many cases, an entry expresses more than one Wordnet sense so this value is a list. If a Wordnet sense seems to be just a special usage, we have termed this a subsense, recorded in column 4. For example

accompany : : 2 3 4 : 1 : ...

indicates that this entry for the verb whose base form is spelled "accompany" can be used to express the second, third, and fourth Wordnet senses for this verb. It also expresses the first Wordnet sense as a subsense, and has no excluded Wordnet senses.

### 1.2 Developer notes and example sentences

Column 5 contains any notes provided by the developers. Column 6 provides an example sentence for each entry reflecting the use of the word in the sense specified, and conforming to the argument template.

### 1.3 Syntactic subcategorization

Columns 7-9 provide the syntactic categories for the first, second, and third arguments of the verb of the verb (if they exist) in surface linear order.

### 1.4 Semantic constraints

Columns 10-12 provide sortal restrictions on each of the verb's arguments.

### 1.5 Thematic roles

Columns 13-15 provide the thematic roles assigned to each of the verb's arguments.

For example, the entry for “accompany”

... NP NP sentient sentient AGNT THME ...

indicates that the verb with the above senses

- 1) takes an NP subject and an NP object;
- 2) restricts the subject to being semantically sentient, and the same for the object;
- 3) assigns the role AGNT to the subject, and the role THME to the object.

## 2 Developer guidelines

### 2.1 Templates: Syntax/Semantics/Theta roles

The values of syntax, semantics and theta roles should be limited to a fixed set. There will nearly always be a one-to-one correspondance between syntax, semantics and theta roles: i.e. if syntax positions 1 and 2 are filled, so will semantics 1 and 2 and theta role 1 and 2 (naturally, the slots are filled in order - i.e. the syntax 3 position cannot be filled without 2). The one exception to this is for cases such as the middle construction, e.g. “bread cuts easily” where the adverb is obligatory, but doesn’t have a theta role etc. Where there is no semantic restriction, \* is used in the slot. (The only case where the theta role maybe should be empty is for verbs like “rain” - e.g. “it is raining”. NOTHETA is used here.)

#### 2.1.1 Valid syntax slot fillers

Templates are fully specific - that is, optionality of an argument is indicated by specifying multiple templates. E.g. instead of

NP (PP)

there will be two templates:

```
  syntax1  syntax2
  NP
  NP      PP
```

The format is as follows: major category in caps, modification of this in lower case, and specific words indicated by lower case word in brackets.

```
NP
NPcoll  (collective)
NPrefl  (reflexive)
VP
VPing
VPinf
S
Scomp
Scomp  (that)
PP
PP      (prep)
```

#### 2.1.2 Valid semantic slot fillers

\*

```
sentient  (humans, human organisations etc - animals may occur in frames labelled as sentient
           when emotions etc are being imputed to them. AGNTs will always be sentient)
animate   (weaker than sentient, used for cases like subjects of movement verbs)
location  (this might be a physical object of some sort, provided it can be viewed as a location)
physobj   (physical object - doesn’t exclude animals etc)
subst     (substance - mass-like entity)
bodypart
```

and others, not yet determined.

### 2.1.3 Valid thematic roles

[ See paper by John Sowa for background. ]

If an entity initiates an action, we have labeled it as AGNT even though it may also be affected by the action. For example, Kim is the AGNT in “Kim ran” even though also Moved by the action. In some cases, where predication is symmetric, we have labeled more than one syntactic position with the same role - this is not a generally accepted move, but there is no obvious alternative in examples like “Kim alternated folk songs with Gregory Isaacs tracks”.

EFNT	(We have used this frequently when it is indeterminate whether an action is initiated voluntarily or not.)
AGNT	
EFCT	(very few examples of this - but subj of “sprain” as in “Kim sprained her wrist”)
ORGN	
STRT	??
MTER	
INST	(subject of “carve” in “this chisel carves marble”)
MATR	(object of “carve” in “this chisel carves marble”)
MED	
PATH	
DUR	
TELC	
EXPR	(subject of “like”, etc, object of “scare” etc)
RCPT	
Addressee	
Beneficiary	
RSLT	(object of “carve” in “Kim carves toys”)
DEST	
CMPL	??
FORM	(occasionally used when underspecified whether object is changed by action)
PTNT	
THME	(since THME gets overused, we have used the subroles)
Moved	
Said	
Ascribed	
Experienced	
LOC	
PTIM	??
NOTHETA	?? - may be needed for expletive <i>it</i> etc.

## 2.2 Arguments and adjuncts

When a prepositional phrase fills a given argument position, there are three options as to how this should be recorded:

1. Use PP(preposition) in the frame to specify the particular preposition.
2. Use PP in the frame to specify that some preposition is obligatory, but not specifying what it is. This should only be used when a range of prepositions is possible for the word.
3. Not specify the preposition at all. This is the correct thing to do when the preposition is always optional, and when the choice of preposition is (arguably) semantic rather than fixed for a particular verb. Examples of this include many cases of locative prepositions, benefactive *for* and so on. In these cases the preposition should be classed as an adjunct, not selected for by the verb.

### 2.3 Alternations included and excluded

We have not included all the patterns which would arise from all of the alternations in Levin's classification. In general, we have avoided this when the resulting use seems rather infrequent or marked. We have systematically excluded the Locative inversion "A little white rabbit jumped out of the box" -> "out of the box jumped a little white rabbit", and There-insertion "There jumped out of the box a little white rabbit". We have also excluded the movement sense of sound emission verbs and other similar cases where it appears that a real change of sense is going on. It is likely to be easier to add templates than to remove them, and if we do add frames for cases such as locative inversion, we would indicate these as marked.