

Syntax 1.2: Lexical categorization

May 4, 2020

Distributions — intuition

A word is characterized by the company it keeps.

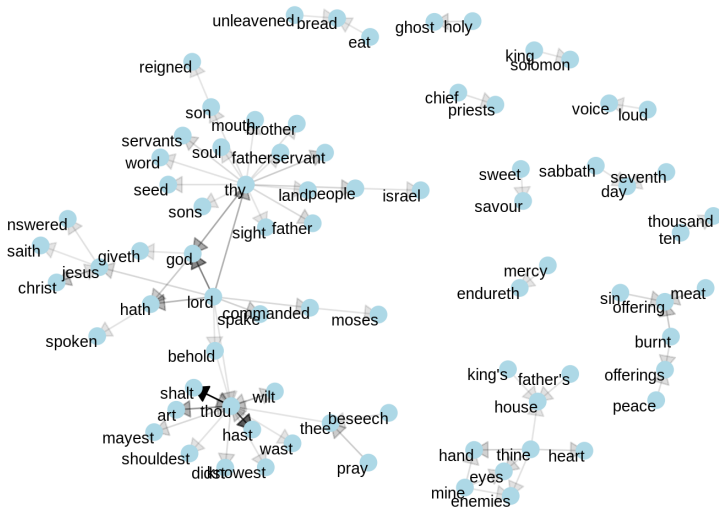
FIRTH 1957

We will use **distributions** to define categories; intuitively, the main hypothesis is ‘similar words will have similar morphological / syntactic distributions’

Upshot: we will arrive at basically the same category labels (noun, verb, preposition, adjective, etc.), but we will do so in a more procedural way.

Why distributions?

Distributions are central to linguistics. A lot of information is contained in the structure of the distribution.



Why distributions?

We use them in phonology and morphology to define phonemes/morphemes and allophones/allomorphs.

They are not subject to interpretation. Consider the definition of noun we learned in primary school: 'a person, place, thing or idea'

- Isn't 'happy' an idea?
- Why is 'happiness' an idea and 'happy' isn't?

We are lucky to have strong, and semi-accurate, intuitions about categorization in English; however, when studying other languages, it is useful to have this tool / method because intuition can easily break down or not exist.

Distributions — more formally

We can think of characterizing a distribution by restrictions on which elements can co-occur.

Elements which are subject to the same **co-occurrence restrictions** are said to be of the same category.

In other words, we define a category by the co-occurrence restrictions of the elements in that category. Think of categories as natural classes of words.

Co-occurrence restrictions

We tend to think of two types of co-occurrence restrictions

- (i) What word level morphology can words have?

MORPHOLOGICAL CO-OCCURRENCE

- (ii) What other words can a word occur with?

SYNTACTIC CO-OCCURRENCE

Morphological co-occurrence

English is not extremely rich in its morphology, so hard to use this to consistently define categories.

- can occur with suffix *-s* NOUN, VERB
- can occur with suffix *-ing* NOUN(?), VERB
- can occur with suffix *-ly* ADJECTIVE
- ...

Syntactic co-occurrence

Instead, consider co-occurrence relations with other words (# indicates end / beginning of sentence; (.) indicates optionality)

- can occur in the context *#the __ must* NOUN
- can occur in the context *must__(it)#* VERB
- can occur in the context *very__* ADJECTIVE, ADVERB
- can occur in the context *is very__#* ADJECTIVE
- can occur in the context *right__* PREPOSITION
 right in the sense of 'completely'
 (e.g. *Go right up the ladder.*)
- can occur in the context *__laptop#* DETERMINER
- can't occur in the determiner or noun context PROPER NOUN

Substitution

We test co-occurrence restrictions by substitution.

Main assumption: substituting (a word from) a different **category** will give different **grammaticality**.

Recall from phonology: substituting (a sound from) a different **phoneme** will give different **meaning**.

A quick example

Only nouns can substitute.

The $\left\{ \begin{array}{l} \text{Animals, planes (?), ...} \\ *In, *up, ... \\ *The, *a, ... \\ *Continue, *walk, ... \\ *Happy, *colorful, ... \\ *Klay, *Steve, ... \end{array} \right\}$ must eat.

Only prepositions can substitute.

(same deal for other categories)

right $\left\{ \begin{array}{l} \text{up, down, through, ...} \\ *Klay, *animals, ... \\ *The, *a, ... \\ *Continue, *walk, ... \\ *Happy, *colorful, ... \end{array} \right\}$ the alley.

A few notes

Note the following:

- (i) A word may belong to multiple categories (e.g. *state*)
- (ii) This is a *rough* partition of the vocabulary... not perfect by any means. People have much finer partitions, but the logic behind the partitioning is similar.
- (iii) Some substitutions may not sound right, but not for grammaticality reasons (e.g. *planes* on last slide)

End of this video's lecture material.