

CSE6390 3.0 Special Topics in AI & Interactive Systems II
Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 CSEB – nick@cse.yorku.ca
Tuesdays, Thursdays 10:00-11:30 – South Ross 104
Fall Semester, 2010

Subcategorization

Subcategorization is a natural language phenomenon, which denotes the tendency of verbs to have restrictions on the arguments that they can take. For example, some verbs do not take a noun-phrase object, while some verbs do take an object, or two objects (direct and indirect). The name *subcategorization* comes from the fact that the category of verbs is divided into finer-grained subcategories of different types of verbs based on the arguments they take.

Alternatively, *Subcategorization* is a concept by which differences in syntactic *valency* between words is expressed. For example, a transitive verb has to be followed by a direct object NP contrary to intransitive verbs. We can say that transitive verbs form a *subcategory* of the *category* of verbs, by virtue of the fact that they must be followed by an NP complement. It is the obligatory presence of the object which gives rise to the subcategory of transitive verbs. The object subcategorizes the verb, or the verb is subcategorized by the object.

A *Subcategorization frame* is a formalization of the notion of subcategorization. For example, (i) gives the subcategorization frame of the verb *hit*. It says that the pseudo-transitive verb *hit* optionally (indicated by the parentheses) selects an NP-complement (a sister-node, as indicated by the square brackets) to its right (indicated by the order ' __ NP' rather than 'NP __').

(i) *hit*: [__ (NP)]

Major categories

The main "parts of speech" are called major categories.

- Noun (*teacher*)
- Verb (*teach*)
- Adjective (*stupid*)
- Adverb (*stupidly*)
- Preposition (*with*)

Major categories are typically "open-class". (you can invent new ones easily)

Minor categories

The other parts of speech are called minor categories.

- Determiner (*the, every*)
- Conjunction (*and*)
- Interjection (*oh!*)

Minor categories are typically "closed-class". (you cannot invent new ones easily)

Note: prepositions are a closed-class major category.

Subcategories of Nouns

- Common nouns (*city*)
- Proper nouns (*Helsinki*)
- Pronouns (*she*)

Subcategories have different valency patterns:

- Common nouns (*the city is beautiful*)

- Proper nouns (**the Helsinki is beautiful*)
- Pronouns (**the she is beautiful*)

Common nouns ``subcategorize" for a determiner.

Subcategories of Verbs

- Intransitive verbs (*walk*)
- Transitive verbs (*like*)
- Ditransitive verbs (*give*)

Subcategories have different valency patterns:

- Intransitive verbs (*She walks, *She walks the dog a bone*)
- Transitive verbs (*She likes the dog, *She likes the dog a bone*)
- Ditransitive verbs (*She gives the dog a bone, *She gives*)

Transitive verbs subcategorize for an object NP. Ditransitive verbs subcategorize for two object NPs.

Many subcategories of verbs

(Using Chomsky's early representation of complements).

- walk: V, __ (*She walks*)
- like: V, __ NP (*She likes the dog*)
- give: V, __ NP NP (*She gives the dog a bone*)
- give: V, __ NP PP(to) (*She gives a bone to the dog*)
- pretend: V, __ S(fin) (*He pretended he had gone home*)
- suggest: V, __ S(base) (*He suggested we go home*)
- intend: V, __ VP(to) (*He intended to go home*)
- help: V, __ VP(base) (*He helped clean up*)
- tell: V, __ NP VP(to) (*He told them to clean up*)
- make: V, __ NP VP(base) (*He made them clean up*)
- say: V, __ PP S(fin) (*He said to me he would clean up*)
- bet: V, __ NP NP S(fin) (*He bet me ten pounds he would clean up*)
- become: V, __ AP (*He became unhappy*)
- word: V, __ NP ADVP (*He worded the reply cleverly*)

Difficult to define all possible **valency** patterns of subcategories. Why not allow every word to have its own specific **valency** pattern?

Subcategorization lists

(Representing the valency pattern by a list of the complements).

- walk: V, [] (*She walks*)
- like: V, [NP] (*She likes the dog*)
- give: V, [NP, NP] (*She gives the dog a bone*)
- give: V, [NP, PP(to)] (*She gives a bone to the dog*)
- pretend: V, [S(fin)] (*He pretended he had gone home*)
- suggest: V, [S(base)] (*He suggested we go home*)
- intend: V, [VP(to)] (*He intended to go home*)
- help: V, [VP(base)] (*He helped clean up*)
- tell: V, [NP, VP(to)] (*He told them to clean up*)
- make: V, [NP, VP(base)] (*He made them clean up*)
- say: V, [PP, S(fin)] (*He said to me he would clean up*)
- bet: V, [NP, NP, S(fin)] (*He bet me ten pounds he would clean up*)
- become: V, [AP] (*He became unhappy*)
- word: V, [NP, ADVP] (*He worded the reply cleverly*)

Put the subcategorization list in the word's lexicon information. This is called *lexicalism*.

Subcategorization lists with subjects

(Including the subject in the list as well as the complements).

- walk: V, [NP] (*She walks*)
- like: V, [NP, NP] (*She likes the dog*)
- give: V, [NP, NP, NP] (*She gives the dog a bone*)
- give: V, [NP, NP, PP(to)] (*She gives a bone to the dog*)
- pretend: V, [NP, S(fin)] (*He pretended he had gone home*)
- suggest: V, [NP, S(base)] (*He suggested we go home*)
- intend: V, [NP, VP(to)] (*He intended to go home*)
- help: V, [NP, VP(base)] (*He helped clean up*)
- tell: V, [NP, NP, VP(to)] (*He told them to clean up*)
- make: V, [NP, NP, VP(base)] (*He made them clean up*)
- say: V, [NP, PP, S(fin)] (*He said to me he would clean up*)
- bet: V, [NP, NP, NP, S(fin)] (*He bet me ten pounds he would clean up*)
- become: V, [NP, AP] (*He became unhappy*)
- word: V, [NP, NP, ADVP] (*He worded the reply cleverly*)

Phrases and heads

Phrases are constructed from the head word, together with its complements and modifiers.

- Head of NP is N (*a teacher of French*)
- Head of AP is A (*so similar to John*)
- Head of PP is P (*to the dog*)
- Head of VP is V (*would clean up*)

The category of the phrase is the same as the category of its head word.

- Note: Head of S is also V (*he would clean up*)

Context-free grammars and DCGs allow arbitrary rules, for example:

NP → PP V

PP → V AP

These are valid CFG rules, but they are wrong from a linguistic point of view because there is no head word that the phrase is based on.

Which word is the head?

If a phrase consists of X Y or Y X, the head is X if:

1. X is obligatory and Y is optional
2. The phrase is a hyponym of X ("a kind of X")
3. X is subcategorized for Y
4. X is the morphosyntactic locus (where the inflections occur)

Examples:

1. *A teacher of French*
A teacher
**A of French*

2. A teacher of French is a kind of teacher
3. *A teacher of French*
**A teacher at French*
4. *A teacher of French*
Teachers of French
**Teacher ofs French*
?(Teacher of Frenches)

Feature percolation

Features are percolated from the head word to the phrase.

- A noun and its mother NP have the same features
- A preposition and its mother PP have the same features
- An adjective and its mother AP have the same features
- A verb and its mother VP have the same features
- A VP and its mother S have the same features

Head Feature Principle (basic idea):

- A phrase and its head share the same features

One of the shared features is the category.

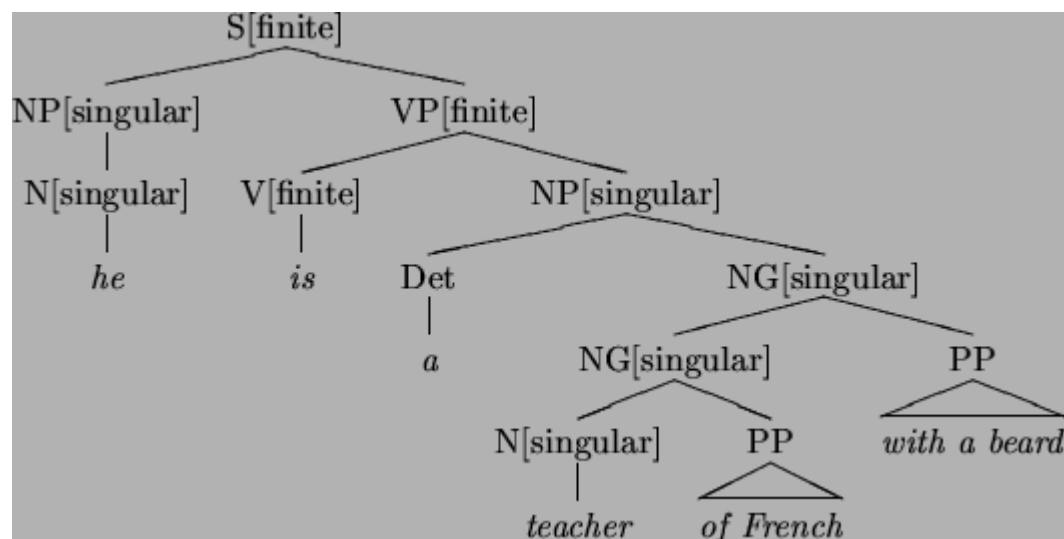
Some NP/NG/N shared features: Number, Case.

Some S/VP/V shared features: Finite, Tense.

Vertical and horizontal mechanisms

Both subcategorization and feature percolation are important mechanisms to make a grammar work properly.

- Feature percolation works ``vertically''
- Subcategorization works ``horizontally''



In linguistics, **verb valency** or **valence** refers to the number of arguments controlled by a verbal predicate. It is related, though not identical, to verb transitivity, which counts only object arguments of the verbal predicate. Verb valency, on the other hand, includes all arguments, including the subject of the verb.

The linguistic meaning of valence is derived from the definition of valency in chemistry. This metaphor is due to Lucien Tesnière.

There are several types of valency:

- An **avalent** verb takes no arguments, e.g. *It rains*. (Though *it* is technically the subject of the verb, it is only a **dummy subject**, that is a syntactic placeholder with no true meaning. No other subject can replace *it*.)
- A **monovalent** verb takes one argument, e.g. *He sleeps*.
- A **divalent** verb takes two, e.g. *He kicks the ball*.
- A **trivalent** verb takes three, e.g. *He gives her a flower*.
- A **tetravalent** verb takes four. They are uncommon, perhaps non-existent in English. Maybe e.g. *He sold you a shirt for 9 bucks*.

The verb requires all of these arguments in a well-formed sentence, although they can sometimes undergo valency reduction or expansion.

For instance, *to eat* is naturally divalent, as in *he eats an apple*, but may be reduced to monovalency in *he eats*. This is called **valency reduction**.

Verbs that are usually monovalent, like *to sleep*, cannot take a direct object. However, there are cases where the valency of such verbs can be expanded, for instance in *He sleeps the sleep of death*. This is called **valency expansion**.

Verb valence can also be described in terms of syntactic versus semantic criteria. The syntactic valency of a verb refers to the number of dependent arguments that the verb can have, while semantic valence describes the thematic relations associated with a verb.

The term valence has a related technical meaning in lexical semantics that elaborates on the role of argument structure - it refers to the capacity of other lexical units to combine with the given word. For instance, valence is one of the elements defining a construction in some Construction Grammars. This sense of the term, sometimes called Lexical Valency, is related to the above, but is far richer than the numerical notion inherited from chemistry.