Syntactic Theory

Lecture 2 (29.10-04.11.2010)

PD Dr.Valia Kordoni

Email: kordoni@coli.uni-sb.de

http://www.coli.uni-saarland.de/~kordoni/

Dependency Grammar (DG)

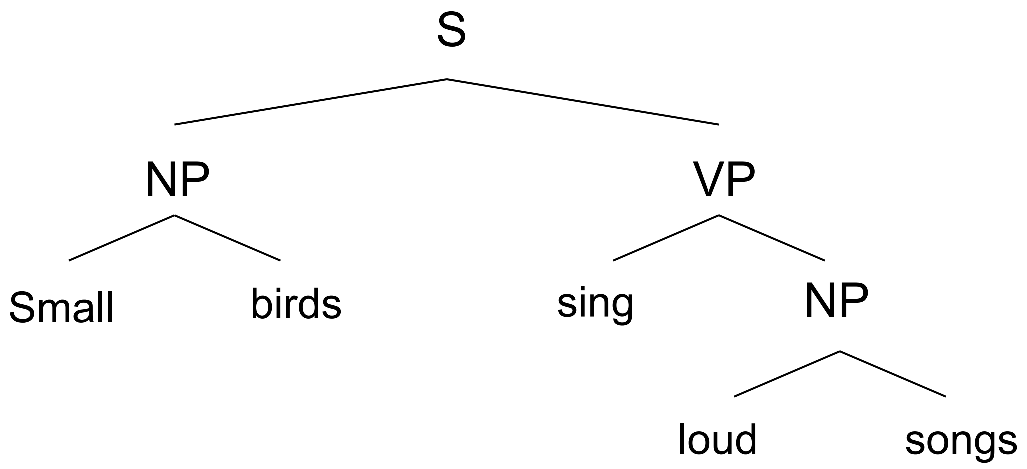
Dependency Grammar

- Not a coherent grammatical framework: wide range of different kinds of DG just as there are wide ranges of ”generative syntax”
- Different core ideas than phrase structure grammar
- We will base a lot of our discussion on Mel’cuk (1988)

Overview

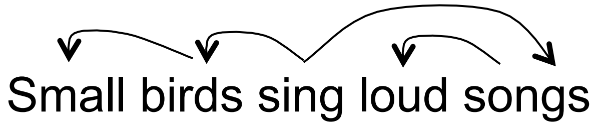
(1) Small birds sing loud songs

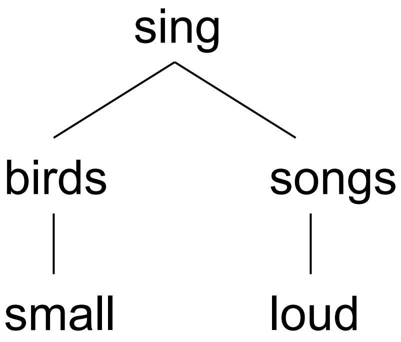
What you might be more used to seeing:



Overview

The corresponding dependency tree representations (Hudson2000):

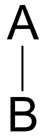
- 
  Small birds sing loud songs

- 
  sing
  birds songs
  small loud

Constituency vs. Relations

- DG is based on relationships between words

A → B means *A governs B or B depends on A* …



- PSG is based on groupings, or constituents

What are these relations?

We'll explore this in more detail, but as a first pass, we're talking about relations like subject, object/complement, (pre-/post-)adjunct, etc.

For example, for the sentence *John loves Mary*, we have:

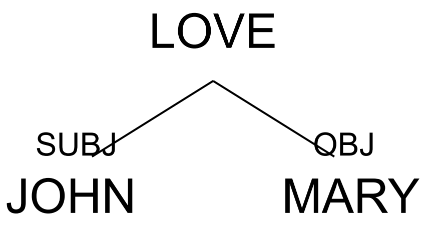
- $\text{LOVE}_{3.sg} \rightarrow_{\text{subj}} \text{JOHN}$
- $\text{LOVE}_{3.sg} \rightarrow_{\text{obj}} \text{MARY}$

Both JOHN and MARY depend on LOVE, which makes LOVE the head of the sentence (i.e., there is no word that governs LOVE)

⇒ The structure of a sentence, then, consists of the set of pairwise relations among words.

In tree form

We can view these dependency relations in tree form:



Adjuncts and Complements

There are two main kinds of dependencies for A → B:

- Head-Complement: if A (the head) has a slot for B, then B is a complement (slots are defined below in the valency section)
- Head-Adjunct: if B has a slot for A (the head), then B is an adjunct

B is dependent on A in either case, but the selector is different

The nature of dependency relations

The relation A → B has certain formal properties (Mel’cuk 1988):

- **antisymmetric**: if A → B, then B ↛ A
  - If A governs B, B does not govern A
  - Consider *box lunch* (LUNCH → BOX) vs. *lunch box* (BOX → LUNCH)
    … can’t have dependency in both directions
  - Eventually, one word is the head of a whole sentence
- **antireflexive**: if A → B, then B ≠ A
  - No word can govern itself.

The nature of dependency relations (cont.)

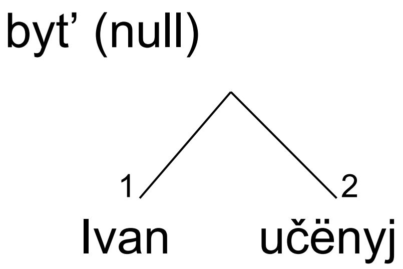
- **antitransitive**: if A → B and B → C, then A ↛ C
  - These are *direct* dependency relations
  - *a usually reliable source*: SOURCE → RELIABLE and RELIABLE → USUALLY, but SOURCE does not govern USUALLY
- **labeled**: ∀ →, → has a label (r)
  - Every dependency relation needs a label
  - Russian žena-vrač ('wife who is a doctor'): WIFE $\rightarrow_1$ DOCTOR vs. žena-vrača WIFE $\rightarrow_2$ DOCTOR ('wife of a doctor')

Unique relations

- uniqueness of A: if A → B, then $\neg \exists$ C s.t. C → B
  - A word can only depend on one other word
  - This is not without controversy ... We'll return to this shortly.

Terminals and Non-terminals

- PS trees contain many non-terminal elements (NP, PP, ...)
- DG trees contain only terminal elements, although there can also be “zero” wordforms, as in the Russian *Ivan učënyj* (’Ivan is scholarly’).



byt’ (null)

1 2

Ivan učënyj

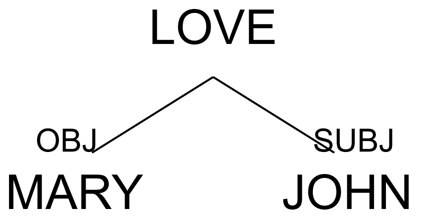
- DG trees also contain definitions of the relations between words (here 1 and 2 are relations roughly corresponding to subject and predicative)

Linear Ordering

- PS trees indicate word order relations along with dominance relations
- Depending on your flavor of DG, the nodes in a DG tree can be unordered

  i.e., the dependency relations are independent of word order ... although, word order may be needed to constrain the dependencies (as we will see later)

  So, the following is a valid tree for *John loves Mary* and is equivalent to our earlier tree:



Syntactic Relations

DG usually maintains a close connection between a tree and the semantics of a sentence

- To do that, the dependency relations need to be labeled
- The labels must correspond to some semantically-relevant entity

⇒ The entities used here are often syntactic roles (e.g., subject, object) which describe the syntactic relations between words.

Dependency Relations

Dependency relations can refer to syntactic properties, semantic properties, or a combination of the two.

- Subject/Agent: *John* fished.
- Object/Patient: *Mary* hit *John*.

→ Some variants of DG separate syntactic and semantic relations by representing different layers of dependency structures (more later)

→ We will discuss the similar notion of grammatical functions in detail in the LFG unit.

Linguistic Analysis

Deciding dependency often comes down to deciding the head of two elements. Roughly, A is the head over B if (Zwicky 1985; Schneider 1998; Hudson 1990):

- A subcategorizes for B (*John runs* → *runs* subcategorizes for a subject)
- A carries the inflection (*red books*, not **reds book*)
- A determines concord/agreement with some other element (*red books read well*, *the red book reads well*)
- A belongs to a category which has the same distribution as A+B (*I like red books/John/books*)
- A is obligatory
- A+B is a hyponym of A (*red book* is a hyponym of *book*)

⇒ Not always a clear-cut issue

Same as PSG?

Are PSG and DG equivalent? Hudson (2000, 1990)

- If a PS tree has heads marked, then you can derive the dependencies
- Likewise, a DG tree can be converted into a PS tree by grouping a word with its dependents

So, are they equivalent representations? ...

Different than PSG

... Not exactly.

- Phrases are only implicit, so they cannot be categorized
- Relations are explicit, so they can be categorized, grouped, put into a hierarchy, whatever
- No unary branches are allowed in DG (why not?)

Valency

An important concept in many variants of DG is that of **valency** = the ability of a verb to take arguments

Each verb takes a specific number of arguments, or valents, and specific types of arguments—this is called a verb's **frame**

Using the PDT-VALLEX notation (Hajič et al. 2003), we would have a lexicon like the following:

	Slot$_1$	Slot$_2$	Slot$_3$
sink$_1$		ACT(NOM)	PAT(ACC)
sink$_2$		PAT(NOM)	
give	ACT(NOM)	PAT(ACC)	ADDR(DAT)

Valency (cont.)

- Valency is also a relevant notion for nouns and adjectives

  – noun *picture* requires that is be a picture of something
  – adjective *proud* requires something to be proud of

- Valency is often treated as semantic and thus distinguished from subcategorization, which is a (usually) surface syntactic notion

  – *John eats rice*: two syntactic and two semantic arguments
  – *John eats*: one syntactic argument, but semantically (or "deeply") John still has to eat something (2 valents)

Inventory of valents

PDT-VALLEX (Hajič et al. 2003) distinguishes **inner participants** (selected by the verb) from **free adverbials** (adjuncts)

- Inner participants: actor, patient, addressee, effect, origin
- Free adverbials: when, where, manner, causative, substitution, ...

Valents as syntactic roles

Note that in the PDT the valents are “(deep) syntactic roles”, so, e.g., key is a MEANS in the first sentence and an ACTOR in the second:

(2) The janitor opened the door with a key.

(3) The key opened the door.

The fact that it is an instrumental use in both cases is captured by the lexical semantics.

From valency to dependency

The inventory of valents looks similar to the dependency relations we've seen before ... a verb (noun/adjective) and its frame drive the dependency analysis:

- *sink*$_1$: ACT(NOM), PAT(ACC)
- *You sunk my battleship*

– SINK$_{past}$ $\rightarrow_{\text{act}}$ YOU$_{nom}$
– SINK$_{past}$ $\rightarrow_{\text{pat}}$ BATTLESHIP$_{acc}$
– BATTLESHIP → I$_{gen}$

Putting it all together

How do we put all these pieces together to form an analysis?

1. Words have valency requirements that must be satisfied
2. General rules are applied to the valencies to see if a sentence is valid

Constraining dependency relations: projectivity

One general rule for using valencies to form dependency relations is known as **projectivity**, or **adjacency** (Hudson 1990)

In brief, this states that a head (A) and a dependent (B) must be adjacent;
More technically: A is adjacent to B provided that every word between A and B is a subordinate of A.

⇒ The ordering stipulations can be done separately from the DG trees, which can be order-independent

Projectivity

(4) with great difficulty

(5) *great with difficulty

- WITH → DIFFICULTY
- DIFFICULTY → GREAT

**great with difficulty* is ruled out because branches would have to cross in that case

Different layers of dependencies

- Syntactic and Morphological layers
- Syntactic and Semantic layers

Syntactic and Morphological Layers: “Mutual dependency”

It looks like a subject depends on the verb, but the form of the verb depends on the subject (Mel’cuk 1988):

(6) a. The child is playing.
b. The children are playing.

But the dependence of *child/children* on the verb is syntactic, while the dependence of the verb(form) on the subject is morphological.

Syntactic and Semantic Layers: “Double dependencies”

We said earlier that each word depends on exactly one other word, but it looks like this isn’t true (Mel’cuk 1988):

(7) Wash the dish *clean*.

It seems that *clean* depends both on the verb *wash* and on the noun *dish*

Double dependencies (cont.)

But one can also say that the relation WASH → CLEAN is syntactic and DISH → CLEAN is semantic, cf. the Russian

(8)	My	našli	zal	pust-ym
	We	found	the hall$_{masc}$	empty$_{masc.sg.inst}$

zal (’hall’) provides the gender (semantic), while našli (’found’) dictates instrumental case (syntactic)

Double dependencies: another viewpoint

Most European versions of DG don’t allow for double dependencies, but in theory they’re possible, and Hudson’s Word Grammar (Hudson 2004) explicitly allows for **structure-sharing**

You could, e.g., analyze *Wash the dish clean* as:

- WASH → CLEAN
- DISH → CLEAN

Structure-sharing

Structure-sharing is also how Hudson (1990) accounts for “non-projective” sentences, like *It keeps raining*.

In this case, *keeps* and *raining* both govern *It* because *keeps* structure-shares its subject with the subject of its (in)complement (*raining*).

(9) subject of complement of *keep* = subject of *keep*

- KEEP $\rightarrow_{\text{comp}}$ RAIN
- RAIN $\rightarrow_{\text{subj}}$ IT
- KEEP $\rightarrow_{\text{subj}}$ IT

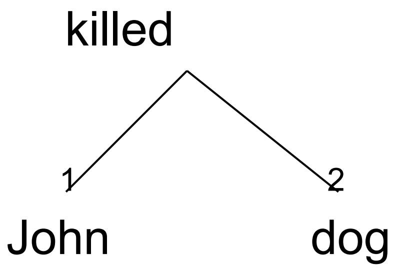
To do this technically, *keep* has to govern both its subject and the verb it shares a subject with (otherwise, there’s nowhere to state the structure-sharing)

Benefits of DG: Connection to semantics

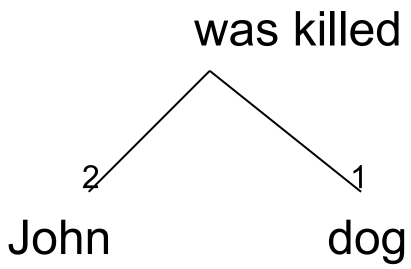
Close connection to semantics allows for

- clean treatment of things like “voice”

John killed the dog



The dog was killed by John



- a representation which allows for (machine) translation between languages

Benefits of DG: More flexible structure

Without the fetters of constituency, certain phenomena are easier to treat (Hudson 1990):

- Can succinctly state that *on* depends on *depend*, whereas in constituency-based accounts, the whole PP has to be marked as *on*.
- In constituency-based accounts, a subject is something like a “second cousin” to the verb, whereas the object is a sister; they are represented parallelly in DG
- Non-constituent coordination is not as much of an issue in DG, e.g., *I drank coffee at eleven and tea at four*.
- The fact that head information percolates up in PSGs indicates that, e.g., N”, N’, and N all share a lot of redundant information.

Benefits of DG: Syntactic Typology

Compare Russian with Hungarian for ’professor’s book’

(10) a. kniga professor+a
book professor
b. professzor könyv+e
professor book

In Russian, BOOK → PROFESSOR, but in Hungarian PROFESSOR → BOOK.
This is claimed to have typological consequences (Mel’cuk 1988)

In general, it is easy to phrase word-order typological rules in terms of heads and dependents

Difficulties for DG

- Coordination
- Modification of groupings (vs. modification of individual words)

Modification of groupings

I lived in Illinois in 1985.

- *in 1985* modifies *lived* → The time when I really lived was in 1985
- *in 1985* modifies rest of sentence → I lived at other places at other times

This latter option is not possible if groupings are not allowed in DG.

Parsing with dependencies

Dependency relations have been used for parsing in different ways.

- To compare parsing output with a gold standard, dependency-based evaluations are more reliable than those comparing bracketings (Carroll et al. 2002)
- Finding the probabilities of bigrams of lexical dependencies has resulted in improved parsing performance (Collins 1996)

References

Carroll, John, Anette Frank, Dekang Lin, Detlef Prescher and Hans Uszkoreit (eds.) (2002). *Proceedings of the Workshop "Beyond PARSEVAL. Towards Improved Evaluation Measures for Parsing Systems" at the 3rd International Conference on Language Resources and Evaluation (LREC 2002)*, Las Palmas, Gran Canaria. http://www.cogs.susx.ac.uk/lab/nlp/carroll/papers/beyond-proceedings.pdf.

Collins, Michael John (1996). A New Statistical Parser Based on Bigram Lexical Dependencies. In Arivind Joshi and Martha Palmer (eds.), *Proceedings of the Thirty-Fourth Annual Meeting of the Association for Computational Linguistics (ACL-96)*. San Francisco, pp. 184–191. http://acl.ldc.upenn.edu/P/P96/P96-1025.pdf.

Hajič, Jan, Jarmila Panevová, Zdeňka Urešová, Alevtina Bémová, Veronika Kolářová and Petr Pajas (2003). PDT-VALLEX: Creating a Large-coverage Valency Lexicon for Treebank Annotation. In *Proceedings of the Second Workshop on Treebanks and Linguistic Theories (TLT 2003)*. Växjö, Sweden, pp. 57–68. http://w3.msi.vxu.se/~rics/TLT2003/doc/hajic et al.pdf.

Hudson, Richard A. (1990). *English Word Grammar*. Oxford, UK: Blackwell. http://www.phon.ucl.ac.uk/home/dick/ewg.htm.

Hudson, Richard A. (2000). Dependency Grammar Course Notes. http://www.cs.bham.ac.uk/research/conferences/esslli/notes/hudson.html.

Hudson, Richard A. (2004). Word Grammar. http://www.phon.ucl.ac.uk/home/dick/intro.htm.

Mel'cuk, Igor A. (1988). *Dependency Syntax: Theory and Practice*. Albany, NY: State University of New York Press.

Schneider, Gerold (1998). A Linguistic Comparison of Constituency, Dependency and Link Grammar. Lizentiatsarbeit, Institut für Informatik der Universität Zürich. http://www.ifi.unizh.ch/cl/study/lizarbeiten/lizgerold.pdf.

Zwicky, Arnold (1985). Heads. *Journal of Linguistics* 12, 1–30.