Universal Stanford Dependencies:
A cross-linguistic typology

Ekaterina Panfilova
ekaterina.panfilova@student.uni-tuebingen.de

May 23, 2017
Outline

1 Introduction
   • Stanford Dependencies Representation
2 A Proposed Universal Taxonomy
   • Lexicalist Approach
3 Basic Dependencies
   • Dependents of clausal predicates
   • Non-core dependents of clausal predicates
   • Treatment of copulas
   • Multiword Expressions
   • Treatment of prepositions and case marking
   • Informal text genres
4 Enhanced dependency representation
5 Language-particular relations
6 Mapping of existing schemes
7 Summary
8 References
Stanford Dependencies (SD) Representation

- *a simple description of the grammatical relationships in a sentence* that can easily be understood and effectively used by people without linguistic expertise who want to extract textual relations.

- originally developed as a practical representation of English syntax

**Universal Dependencies v2**

- Basic dependency Representation
- Enhanced dependency Representation
The **basic dependency representation** forms a tree, where exactly one word is the head of the sentence.

Basic dependency representation **is obligatory for all UD treebanks**

SD adopts the **lexicalist hypothesis** in syntax, whereby grammatical relations should be between whole words (or lexemes).
## A Proposed Universal Taxonomy

<table>
<thead>
<tr>
<th></th>
<th>Nominals</th>
<th>Clauses</th>
<th>Modifier words</th>
<th>Function Words</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Core arguments</strong></td>
<td>nsubj</td>
<td>csubj</td>
<td>advmod*</td>
<td>aux</td>
</tr>
<tr>
<td></td>
<td>obj</td>
<td>ccomp</td>
<td>discourse</td>
<td>cop</td>
</tr>
<tr>
<td></td>
<td>iobj</td>
<td>xcomp</td>
<td></td>
<td>mark</td>
</tr>
<tr>
<td><strong>Non-core dependents</strong></td>
<td>obl</td>
<td>advcl</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>vocative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>expl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>dislocated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Nominal dependents</strong></td>
<td>nmod</td>
<td>acl</td>
<td>amod</td>
<td>det</td>
</tr>
<tr>
<td></td>
<td>appos</td>
<td></td>
<td></td>
<td>clf</td>
</tr>
<tr>
<td></td>
<td>nummod</td>
<td></td>
<td></td>
<td>case</td>
</tr>
<tr>
<td><strong>Coordination</strong></td>
<td>conj</td>
<td>fixed</td>
<td>list</td>
<td>orphan</td>
</tr>
<tr>
<td></td>
<td>cc</td>
<td>flat</td>
<td>parataxis</td>
<td>goeswith</td>
</tr>
<tr>
<td></td>
<td></td>
<td>compound</td>
<td></td>
<td>reparandum</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>punct</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>root</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>dep</td>
</tr>
</tbody>
</table>
The system of clausal dependents most closely follows **Lexical Functional Grammar (LFG)**

The taxonomy differs from LFG in several aspects:

- Clear distinction between core dependents and other dependents
- The model reverts to the traditional grammar notion of direct and other objects. In cases of multiple objects, the theme/patient argument is the direct object.
The object of a verb is the second most core argument of a verb after the subject. Typically, it is the noun phrase that denotes the entity acted upon or which undergoes a change of state or motion (the proto-patient).

**She gave me a raise**
**Direct Object**

**obj**

In languages distinguishing morphological cases, the object will often be marked by the **accusative case**.

If there is just one object, it should be labeled **obj**.

If there are two or more objects, one of them should be **obj** and the others should be **iobj**. In such cases it is necessary to decide what is the most directly affected object (patient).
**iobj**
The indirect object of a verb is any nominal phrase that is a core argument of the verb but is not its subject or (direct) object.
iobj

In languages distinguishing morphological cases, the indirect object will often be marked by the **dative case**.

\[
\text{On učí mou đceru matematiku }. \\
\text{He teaches my daughter. Acc maths. Acc .}
\]
A design goal of SD has been *to distinguish in dependency names where a new clause is being introduced*.

<table>
<thead>
<tr>
<th>Core arguments</th>
<th>Nominals</th>
<th>Clauses</th>
</tr>
</thead>
<tbody>
<tr>
<td>nsubj</td>
<td>obj</td>
<td>csubj</td>
</tr>
<tr>
<td></td>
<td>iobj</td>
<td>ccomp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>xcomp</td>
</tr>
</tbody>
</table>
- **nsubj** (nominal subject)
Dependents of clausal predicates

Examples

- **csusbj** (clausal subject)

  ![Diagram]

  What **she** **said** **makes** **sense**

  Забыть Ваню трудно.

  To-forget Vanya, it-is-hard.
## Dependents of clausal predicates

### Examples

<table>
<thead>
<tr>
<th></th>
<th>Nominals</th>
<th>Clauses</th>
<th>Modifier words</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Core arguments</strong></td>
<td>nsubj</td>
<td>csbj</td>
<td></td>
</tr>
<tr>
<td></td>
<td>obj</td>
<td>ccomp</td>
<td></td>
</tr>
<tr>
<td></td>
<td>iobj</td>
<td>xcomp</td>
<td></td>
</tr>
</tbody>
</table>
| **Non-core dependents** | obl      | advcl   | advmod
|                   | vocative |         | discourse      |
|                   | expl     |         |                |
|                   | dislocated|         |                |
| **Nominal dependents** | nmod    | acl     | amod           |
|                   | appos    |         |                |
|                   | nummod   |         |                |

- **advmod** (adverbial modifier) and **advcl** (adverbial clause modifier)

An adverbial modifier of a word is a (non-clausal) adverb or adverbial phrase that serves to modify a predicate or a modifier word.
Non-core dependents of clausal predicates

**advcl**: adverbial clause modifier
An adverbial clause modifier is a clause which modifies a verb or other predicate (adjective, etc.), as a modifier not as a core complement.
Dependents of clausal predicates

Examples

<table>
<thead>
<tr>
<th>Core arguments</th>
<th>Nominals</th>
<th>Clauses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nsubj, obj, iobj</td>
<td>csubj, ccomp, xcomp</td>
</tr>
</tbody>
</table>

- distinction between **ccomp** (clausal complement of a verb or adjective) and **xcomp** (open clausal complement)
  - The defining difference is a controlled subject.
ccomp: clausal complement
A clausal complement of a verb or adjective is a dependent clause which is a core argument. That is, it functions like an object of the verb, or adjective.
ccomp: clausal complement
Such clausal complements may be finite or non-finite. However, if the subject of the clausal complement is controlled (that is, must be the same as the higher subject or object, with no other possible interpretation) the appropriate relation is xcomp (open clausal complement).
xcomp: open clausal complement
An open clausal complement of a verb or an adjective is a predicative or clausal complement without its own subject. The reference of the subject is necessarily determined by an argument external to the xcomp (normally by the object of the next higher clause, if there is one, or else by the subject of the next higher clause).
## Nominal Dependents

**acl: clausal modifier of noun (adjectival clause)**

<table>
<thead>
<tr>
<th>Core arguments</th>
<th>Nominals</th>
<th>Clauses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nsubj</td>
<td>csubj</td>
</tr>
<tr>
<td></td>
<td>obj</td>
<td>ccomp</td>
</tr>
<tr>
<td></td>
<td>iobj</td>
<td>xcomp</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non-core dependents</th>
<th>Nominals</th>
<th>Clauses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>obl</td>
<td>advcl</td>
</tr>
<tr>
<td></td>
<td>vocative</td>
<td></td>
</tr>
<tr>
<td></td>
<td>expl</td>
<td></td>
</tr>
<tr>
<td></td>
<td>dislocated</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nominal dependents</th>
<th>Nominals</th>
<th>Clauses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nmod</td>
<td>acl</td>
</tr>
<tr>
<td></td>
<td>appos</td>
<td></td>
</tr>
<tr>
<td></td>
<td>nummod</td>
<td></td>
</tr>
</tbody>
</table>
acl: clausal modifier of noun (adjectival clause)

It stands for finite and non-finite clauses that modify a nominal.
acl: clausal modifier of noun (adjectival clause)
A relative clause is an instance of acl, characterized by finiteness and usually omission of the modified noun in the embedded clause.
### Non-Core Dependents

<table>
<thead>
<tr>
<th></th>
<th>Nominals</th>
<th>Clauses</th>
<th>Modifier words</th>
<th>Function Words</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Core arguments</strong></td>
<td>nsubj</td>
<td>csubj</td>
<td>advmod*</td>
<td>aux</td>
</tr>
<tr>
<td></td>
<td>obj</td>
<td>ccomp</td>
<td></td>
<td>cop</td>
</tr>
<tr>
<td></td>
<td>obj</td>
<td>xcomp</td>
<td></td>
<td>mark</td>
</tr>
<tr>
<td><strong>Non-core dependents</strong></td>
<td>obl</td>
<td>advcl</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>vocative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>expl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>dislocated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Nominal dependents</strong></td>
<td>nmod</td>
<td>acl</td>
<td>amod</td>
<td>det</td>
</tr>
<tr>
<td></td>
<td>appos</td>
<td></td>
<td></td>
<td>clf</td>
</tr>
<tr>
<td></td>
<td>nummod</td>
<td></td>
<td></td>
<td>case</td>
</tr>
<tr>
<td><strong>Coordination</strong></td>
<td>conj</td>
<td>fixed</td>
<td>list</td>
<td>orphan</td>
</tr>
<tr>
<td></td>
<td>cc</td>
<td>flat</td>
<td>parataxis</td>
<td>goeswith</td>
</tr>
<tr>
<td></td>
<td></td>
<td>compound</td>
<td></td>
<td>reparandum</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>punct</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>root</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>dep</td>
</tr>
</tbody>
</table>
obl: oblique nominal

This relation is used for a nominal (noun, pronoun, noun phrase) functioning as a non-core (oblique) argument or adjunct.

This means that it functionally corresponds to an adverbial attaching to a verb, adjective or other adverb.
The *obl* relation can be further *specified by the case*. In conjunction with the case relation, it provides a uniform analysis for:

- variant forms with case, a preposition or a postposition, as in Finnish for example:
Non-Core Dependents
Oblique Nominal

- the dative alternation where the prepositional construction gets a similar analysis to the double object construction:

```
give the children the toys
```
```
give the toys to the children
```
```
donne les jouets aux enfants
```
obl is also used for temporal and locational nominal modifiers:
for the agent of a passive verb (with the optional subtype obl:agent):
RECAP: Copula

Copula (plural: copulas or copulae) is a word used to link the subject of a sentence with a predicate (a subject complement), such as the word *is* in the sentence "The sky *is* blue." The word copula derives from the Latin noun for a "link" or "tie" that connects two different things.

In SD the copula "be" is not the head of a clause, but rather the dependent of a lexical predicate.

(1) a. Ivan *is* the best dancer
b. Russian:

```
Ivan
```
lučšij
tancor

```
Ivan
best
dancer
```

c. I judge Ivan the best dancer
Multiword expressions (MWEs) are combinations of words that behave as lexical units rather than compositional syntactic phrases.

Three special relations for analyzing MWEs:

- fixed
- flat
- compound
**fixed**: fixed multiword expression

This relation is used for certain fixed grammaticized expressions that behave like function words or short adverbials.
**Flat Multiword Expression**

*flat* : **flat multiword expression** is used for exocentric (headless) semi-fixed MWEs

**names**

- **Hilary Rodham Clinton**

**titles/honorifics**

- **Mr. Smith**
  
- **President Obama**

- **French actor Gaspard Ulliel**
Other types of names

The names that have a regular syntactic structure should be annotated with regular syntactic relations.
Compounding

Compound is used for any type of $X^0$ compounding:

- **noun compounds** (e.g., phone book)
- **verb and adjective compounds** that are more common in other languages (such as Persian or Japanese light verb constructions)
- **particles of phrasal verbs** (e.g., put up: $compound(put,up)$)
A major proposed change from the extant versions of SD is a **new treatment of prepositions** to provide a uniform analysis of prepositions and case in morphologically rich languages.

any case-marking element (including prepositions, postpositions, and clitic case markers) will be treated as a *dependent of the noun it attaches to or introduces*
The **nmod** (nominal modifier) relation is used for nominal dependents of another noun or noun phrase and functionally corresponds to an attribute, or genitive complement. The preposition is now viewed as a case depending on its complement.
(3) a. Hebrew:

\[ \text{wkfraiti} \quad \text{at} \quad \text{hsrj} \]
\[ \text{and when I saw} \quad \text{ACC} \quad \text{the movie} \]

b. \( \text{dobj}(\text{wkfraiti/VERB}, \text{hsrj/NOUN}) \)
\( \text{case}(\text{hsrj/NOUN}, \text{at/PRT-ACC}) \)

*at* in Hebrew is a separate token indicating an accusative object: the case marker depends on the object.
Treatment of case marking

Examples

Morphemes as case markers

c. Russian:

\[
\begin{array}{rrrr}
& nsubj & dobj & nmod \\
Ya & napisal & pis’mo & perom \\
\end{array}
\]

\[I \text{ wrote the letter with a quill}\]

d. \(nsubj(napisal/\text{VERB}, \text{Ya/NOUN-NOM})\)

\(dobj(napisal/\text{VERB}, \text{pis’mo/NOUN-ACC})\)

\(nmod(napisal/\text{VERB}, \text{perom/NOUN-INSTR})\)

The case morpheme is not divided off the noun as a separate case dependent, but \textit{the noun as a whole} is analyzed as a \textit{nmod} of the verb.
The above discussed treatment provides **parallelism between different constructions across and within languages**

- There is now a *greater parallelism between prepositional phrases and subordinate clauses*, which are in practice often introduced by a preposition.

\[(4) \quad \text{a. Sue left after the rehearsal}\]

\[(4) \quad \text{b. Sue left after we did}\]
parallel constructions for the possessive alternation

(2) a. the Chair’s office

b. the office of the Chair
Treatment of prepositions and case marking

The advantages of the proposed analysis

- treatment of prepositional phrases that are predicative complements of "be"

(7) Sue is in shape
**Informal text genres**

**goeswith** is used to connect multiple tokens that correspond to a single standard word, as a result of reanalysis of words as compounds (”hand some” for ”handsome”) or input error (”othe r” for ”other”)

**reparandum** is used to indicate disfluencies overridden in a speech repair

(10) Go to the righ- to the left.

![Diagram showing Universal Stanford Dependencies](image-url)
**orphan** relation is used to provide a treatment of ellipsis (in the case of gapping or stripping, where predicational or verbal material gets elided)
Enhanced dependency representation adds (and in a few cases changes) relations in order to give a more complete basis for semantic interpretation. The enhanced representation is in general not a tree but a general graph structure, as shown below (enhanced dependencies in blue).
Elipsis

In the enhanced representation, special null nodes are added in clauses in which a predicate is elided.
In addition to a universal dependency taxonomy, it is desirable to recognize grammatical relations that are particular to one language or a small group of related languages.

Such language-particular relations are necessary to accurately capture the genius of a particular language but will not involve concepts that generalize broadly.

These relations should always be regarded as a subtype of an existing Universal SD relation.
The authors proclaim that the existing dependency sets for other languages can be fairly straightforwardly mapped onto the new proposal.

The mapping is less difficult because USD adopts some of the ideas and relations that were first developed for other treebanks.

- Finnish
- Italian
- Chinese
- Persian
taxonomy of grammatical relations applicable to a variety of languages, developing the implications of a lexicalist approach for the treatment of morphology and compounding

- **Advantages:**
  - convenient for a practical computational model

- **Disadvantages:**
  - certain clitics
  - phonologically bound words which behave like syntactic words
  - words that are split in unedited writing

- mapping of existing dependency resources for different languages to the taxonomy


http://universaldependencies.org/language-u