Syntactic Analysis
Daniel Zeman

December 16, 2019
1 Constituents vs. Dependencies

2 Universal Dependencies
   - A Tour through UD Syntax
   - Nonverbal Predicate and Copula
   - Ellipsis and Enhanced UD
   - Core Arguments vs. Oblique Dependents
Constituents vs. Dependencies
Outline

1. Constituents vs. Dependencies

2. Universal Dependencies
   - A Tour through UD Syntax
   - Nonverbal Predicate and Copula
   - Ellipsis and Enhanced UD
   - Core Arguments vs. Oblique Dependents
Syntactic Structure

- Different shapes in different theories
- Typically a tree
  - Constituents (phrase tree structure)
  - Dependencies (dependency tree structure)
Constituent Tree

(S (NP (N Paul)) (VP (V gave) (NP (N Peter)) (NP (C two) (N pears))))
Paul gave Peter two pears.
Dependency Tree with Labels

[#,0] (gave,2) ([Paul,1], [Peter,3], [pears,5] ([two,4])), [,6])

# / AuxS
  
gave / Pred
    
    Paul / Sb
    
    Peter / Obj
    
    pears / Obj
    
    two / Atr
Dependency Tree with Labels

Paul gave Peter two pears

root

nsubj iobj obj

Paul gave Peter
two pears

nummod
Constituents vs. Dependencies

- The two models are interconnected

- Sentence divided to **phrases** (constituents)
  - Recursive: phrases divided to smaller phrases
  - The smallest phrases are words

- There are **dependencies** (relations) between words (constituents)
  - **Head** of phrase = governing node, parent node
  - The other nodes are dependent nodes, children of the head
Phrase vs. Dependency Trees

Syntactic Analysis

Universal Dependencies
Phrase vs. Dependency Trees

- **Phrase trees**
  - Usually do not mark the head
  - May not mark the function of the constituent in the superordinate constituent
Phrase vs. Dependency Trees

- **Phrase trees**
  - Usually do not mark the head
  - May not mark the function of the constituent in the superordinate constituent

- **Dependency trees**
  - Do not show nonterminals (phrase types)
    - Nor any other phrase-level features
  - Do not show “how the sentence is generated” (order, recursion, proximity of constituents)
Example

Syntactic Analysis
Constituents vs. Dependencies

Universal Dependencies

11/111
Discontinuous Phrases

- Classical context-free grammar cannot describe them!
- They cannot be represented by bracketing.
- English example: *I found the best example ever.*
- Czech example: *(Soubor (se nepodařilo) otevřít).* “File couldn’t be opened.”

![Syntactic Analysis Diagram]

**Constituents vs. Dependencies**

**Universal Dependencies**
Nonprojectivity

- Dependency tree including word order (x-coordinate of nodes).
- Projection to the base: the vertical from the node crosses a dependency (nonprojective edge).
- Formally:
  - Dependency \([g, x_g], [d, x_d]\) where \(x_w\) is the order of the word \(w\) in the sentence.
  - There exists a node \([n, x_n]\) that \(x_g < x_n < x_d\) or \(x_d < x_n < x_g\) and \([n, x_n]\) is not in subtree rooted by \([g, x_g]\).
- Informally: The string spanned by the subtree of the governing node is discontinuous, it contains gaps.
Nonprojectivity Can Be Handled by a Dependency Tree!

- `nepodařilo / Pred`
  - `se / AuxT`
  - `otevřít / Sb`
    - `soubor / Obj`
Nonprojectivity Can Be Handled by a Dependency Tree!

soubor se nepodařilo otevřít
file itself did-not-succeed to-open
Universal Dependencies
Outline

1. Constituents vs. Dependencies

2. Universal Dependencies
   - A Tour through UD Syntax
   - Nonverbal Predicate and Copula
   - Ellipsis and Enhanced UD
   - Core Arguments vs. Oblique Dependents
Syntactic Analysis

Constituents vs. Dependencies
Syntactic Analysis

Constituents vs. Dependencies
My daughter bought some bread and cheese.
Universal Dependencies

http://universaldependencies.org/

Milestones:
- 2008-05 Interset (morphological features)
- 2012-05 Google Universal POS tags
- 2012-05 HamleDT (harmonized Prague-style treebanks)
- 2013-08 Google Universal Dependency Treebank
- 2014-05 Universal Stanford Dependencies
- 2014-04 EACL Göteborg, kick-off meeting of UD
- 2014-10 UD guidelines version 1
- 2015-01 released first 10 treebanks
- every ~6 months new release
- 2016-12 UD guidelines version 2
- 2017-05 CoNLL Shared Task in parsing UD
- 2018-06 second Shared Task
  every ~6 months new release
• Same things annotated same way across languages...
• ... while highlighting different coding strategies
Manning’s Law

The secret to understanding UD is to realize that the design is a very subtle compromise between approximately 6 things:

1. UD must be satisfactory on linguistic analysis grounds for individual languages.
2. UD must be good for linguistic typology, i.e., providing a suitable basis for bringing out cross-linguistic parallelism across languages and language families.
3. UD must be suitable for rapid, consistent annotation by a human annotator.
4. UD must be easily comprehended and used by a non-linguist, whether a language learner or an engineer with prosaic needs for language processing. … It leads us to favor traditional grammar notions and terminology.
5. UD must be suitable for computer parsing with high accuracy.
6. UD must support well downstream language understanding tasks (relation extraction, reading comprehension, machine translation, …).

It’s easy to come up with a proposal that improves UD on one of these dimensions. The interesting and difficult part is to improve UD while remaining sensitive to all these dimensions.
Manning’s Law

The secret to understanding UD is to realize that the design is a very subtle compromise between approximately 6 things:

1. UD must be satisfactory on linguistic analysis grounds for individual languages.
2. UD must be good for linguistic typology, i.e., providing a suitable basis for bringing out cross-linguistic parallelism across languages and language families.

It’s easy to come up with a proposal that improves UD on one of these dimensions. The interesting and difficult part is to improve UD while remaining sensitive to all these dimensions.
Manning’s Law

The secret to understanding UD is to realize that the design is a very subtle compromise between approximately 6 things:

1. UD must be satisfactory on linguistic analysis grounds for individual languages.
2. UD must be good for linguistic typology, i.e., providing a suitable basis for bringing out cross-linguistic parallelism across languages and language families.
3. UD must be suitable for rapid, consistent annotation by a human annotator.

It’s easy to come up with a proposal that improves UD on one of these dimensions. The interesting and difficult part is to improve UD while remaining sensitive to all these dimensions.
Manning’s Law

The secret to understanding UD is to realize that the design is a very subtle compromise between approximately 6 things:

1. UD must be satisfactory on linguistic analysis grounds for individual languages.
2. UD must be good for linguistic typology, i.e., providing a suitable basis for bringing out cross-linguistic parallelism across languages and language families.
3. UD must be suitable for rapid, consistent annotation by a human annotator.
4. UD must be easily comprehended and used by a non-linguist, whether a language learner or an engineer with prosaic needs for language processing. ... it leads us to favor traditional grammar notions and terminology.

It’s easy to come up with a proposal that improves UD on one of these dimensions. The interesting and difficult part is to improve UD while remaining sensitive to all these dimensions.
Manning’s Law

The secret to understanding UD is to realize that the design is a very subtle compromise between approximately 6 things:

1. UD must be satisfactory on linguistic analysis grounds for individual languages.
2. UD must be good for linguistic typology, i.e., providing a suitable basis for bringing out cross-linguistic parallelism across languages and language families.
3. UD must be suitable for rapid, consistent annotation by a human annotator.
4. UD must be easily comprehended and used by a non-linguist, whether a language learner or an engineer with prosaic needs for language processing. ... it leads us to favor traditional grammar notions and terminology.
5. UD must be suitable for computer parsing with high accuracy.

It’s easy to come up with a proposal that improves UD on one of these dimensions. The interesting and difficult part is to improve UD while remaining sensitive to all these dimensions.
Manning’s Law

The secret to understanding UD is to realize that the design is a very subtle compromise between approximately 6 things:

1. UD must be satisfactory on linguistic analysis grounds for individual languages.
2. UD must be good for linguistic typology, i.e., providing a suitable basis for bringing out cross-linguistic parallelism across languages and language families.
3. UD must be suitable for rapid, consistent annotation by a human annotator.
4. UD must be easily comprehended and used by a non-linguist, whether a language learner or an engineer with prosaic needs for language processing. ... it leads us to favor traditional grammar notions and terminology.
5. UD must be suitable for computer parsing with high accuracy.
6. UD must support well downstream language understanding tasks (relation extraction, reading comprehension, machine translation, ...)

It’s easy to come up with a proposal that improves UD on one of these dimensions. The interesting and difficult part is to improve UD while remaining sensitive to all these dimensions.
Same Thing Same Way

George killed the dragon

PROPN VERB DET NOUN
Same Thing Same Way
George killed the dragon
Mharaigh Seoirse an dragan
Jorge mató al dragón
Same Thing Same Way

- **George killed the dragon**
  - **root**
  - **nsubj** George
  - **obj** killed
  - **det** the
  - **det** the
  - **det** the
  - **nsubj** the
  - **obj** dragon

- **Mharaigh Seoirse an dragan**
  - **PROPN**
  - **VERB**
  - **PROPN**
  - **DET**
  - **NOUN**

- **Jorge mató a el dragón**
  - **PROPN**
  - **VERB**
  - **ADP**
  - **DET**
  - **DET**
  - **nsubj** a
  - **case**
  - **det**
  - **nsubj** a
  - **obj** el
  - **obj** el
Same Thing Same Way

Constituents vs. Dependencies
Same Meaning ≠ Same Construction!

He killed the dragon

PRON VERB DET NOUN

The dragon was killed by him

DET NOUN AUX VERB ADP PRON

His killing of the dragon

PRON NOUN ADP DET NOUN

The dragon that was killed

DET NOUN PRON AUX VERB

Syntactic Analysis

Constituents vs. Dependencies
Same Meaning ≠ Same Construction!

He killed the dragon.

The dragon was killed by him.

Syntactic Analysis
Constituents vs. Dependencies
Same Meaning ≠ Same Construction!

He killed the dragon

PRON VERB DET NOUN

The dragon was killed by him

DET NOUN AUX VERB ADP PRON

His killing of the dragon

PRON NOUN ADP DET NOUN
Same Meaning ≠ Same Construction!
राजा विश्�努शर्माणम् आहूय प्रोवाच
rājā viṣṇuśarmāṇam āhūya provāca
king Vishnusharma having-summoned said

Syntactic Analysis
Constituents vs. Dependencies
राजा विष्णूशर्मणि आहूय प्रौवाच
किंग Vishnusharma having-summoned said

The king summoned Vishnusharma and said.
Universal Dependencies

A Tour through UD Syntax
Outline

1. Constituents vs. Dependencies

2. Universal Dependencies
   - A Tour through UD Syntax
   - Nonverbal Predicate and Copula
   - Ellipsis and Enhanced UD
   - Core Arguments vs. Oblique Dependents
The cat could have chased all the dogs down the street.

DET NOUN AUX AUX VERB DET DET NOUN ADP DET NOUN PUNCT
The cat could have chased all the dogs down the street.

- Content words are related by dependency relations
• Content words are related by dependency relations
• Function words attach to closest content words
Content words are related by dependency relations
Function words attach to closest content words
Punctuation attach to head of phrase or clause
The cat could have chased all the dogs down the street.

**Syntactic Analysis**

Constituents vs. Dependencies

Universal Dependencies

- Content words are related by dependency relations
- Function words attach to closest content words
- Punctuation attach to head of phrase or clause

Not “dependency” in the strictly syntactic sense!
The dog was chased by the cat.
The dog was chased by the cat.

Kučeto se presledvaše ot kotkata.
The dog was chased by the cat.
The dog was chased by the cat.

Pes byl honěn kočkou.
### Dependents of Clauses (Verbal or Not)

<table>
<thead>
<tr>
<th>Core</th>
<th>Nominal</th>
<th>Clausal</th>
<th>Modifier</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nsubj</td>
<td>csubj</td>
<td>advmod</td>
<td>aux</td>
</tr>
<tr>
<td></td>
<td>obl</td>
<td>advcl</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>vocative</td>
<td>discourse</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>dislocated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>expl</td>
<td>cop</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>mark</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Dependents of Verbs, Adjectives and Adverbs

<table>
<thead>
<tr>
<th>Core</th>
<th>Nominal</th>
<th>Clausal</th>
<th>Modifier</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>obj</td>
<td>ccomp</td>
<td>advmod</td>
</tr>
<tr>
<td></td>
<td>iobj</td>
<td>xcomp</td>
<td></td>
</tr>
<tr>
<td>Non-Core</td>
<td>obl</td>
<td>advcl</td>
<td></td>
</tr>
<tr>
<td></td>
<td>expl</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Dependents of Nominals

<table>
<thead>
<tr>
<th>Nominal</th>
<th>Clausal</th>
<th>Modifier</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>nmod</td>
<td>acl</td>
<td>amod</td>
<td>det</td>
</tr>
<tr>
<td>appos</td>
<td></td>
<td>nummod</td>
<td>case</td>
</tr>
</tbody>
</table>
## Noun Phrase

### Dependents of Nominals

<table>
<thead>
<tr>
<th>Nominal</th>
<th>Clausal</th>
<th>Modifier</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>nmod</td>
<td>acl</td>
<td>amod</td>
<td>det</td>
</tr>
<tr>
<td>appos</td>
<td></td>
<td>nummod</td>
<td>case</td>
</tr>
<tr>
<td>compound</td>
<td></td>
<td></td>
<td>clf</td>
</tr>
<tr>
<td>flat</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Syntactic Analysis

#### Constituents vs. Dependencies

```
the American singer Johnny Cash , an icon of country music
```

#### Universal Dependencies

```
det amod flat appos flat punct det nmod case compound
```

34/111
Coordinate structures are headed by the first conjunct

- Subsequent conjuncts depend on it via the `conj` relation
- Conjunctions depend on the next conjunct via the `cc` relation
- Punctuation marks depend on the next conjunct via the `punct` relation
But Some Languages Might Prefer the Opposite

- Coordinate structures would be headed by the last conjunct
  - Preceding conjuncts would depend on it via the `conj` relation
  - Conjunctions would depend on the preceding conjunct
  - Punctuation marks would depend on the preceding conjunct
Multiword Expressions

<table>
<thead>
<tr>
<th>Relation</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>fixed</td>
<td><em>in spite of, as well as, ad hoc</em></td>
</tr>
<tr>
<td>flat</td>
<td><em>president Havel, New York, four thousand</em></td>
</tr>
<tr>
<td>compound</td>
<td><em>phone book, dress up</em></td>
</tr>
<tr>
<td>goeswith</td>
<td><em>notwith standing, with out</em></td>
</tr>
</tbody>
</table>

- UD annotation **almost** does not permit “words with spaces”
  - Multiword expressions are analyzed using special relations
  - The **fixed**, **flat** and **goeswith** relations are always head-initial
  - The **compound** relation reflects the internal structure

- Words with spaces may be allowed in v2:
  - Vietnamese (spaces delimit syllables, not words)
  - Numbers ("1 000 000")
  - Possibly other approved cases, e.g. multi-word abbreviations
### Other Relations

<table>
<thead>
<tr>
<th>Relation</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>parataxis</td>
<td>Loosely linked clauses of same rank</td>
</tr>
<tr>
<td>list</td>
<td>Lists without syntactic structure</td>
</tr>
<tr>
<td>orphan</td>
<td>Orphans in ellipsis linked together</td>
</tr>
<tr>
<td>reparandum</td>
<td>Disfluency linked to (speech) repair</td>
</tr>
<tr>
<td>foreign</td>
<td>Elements within opaque stretches of code switching</td>
</tr>
<tr>
<td>dep</td>
<td>Unspecified dependency</td>
</tr>
<tr>
<td>root</td>
<td>Syntactically independent element of clause/phrase</td>
</tr>
</tbody>
</table>
Language-specific relations are **subtypes** of universal relations added to capture important phenomena.

Subtyping permits us to “back off” to universal relations.

### Language-specific Relation Subtypes

<table>
<thead>
<tr>
<th>Relation</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>acl:relcl</td>
<td>Relative clause</td>
</tr>
<tr>
<td>compound:prt</td>
<td>Verb particle (dress <strong>up</strong>)</td>
</tr>
<tr>
<td>nmod:poss</td>
<td>Possessive nominal (<strong>Mary’s</strong> book)</td>
</tr>
<tr>
<td>obl:agent</td>
<td>Agent in passive (saved <strong>by the bell</strong>)</td>
</tr>
<tr>
<td>cc:preconj</td>
<td>Preconjunction (<strong>both</strong> … and)</td>
</tr>
<tr>
<td>det:predet</td>
<td>Predeterminer (<strong>all</strong> those …)</td>
</tr>
</tbody>
</table>
Universal Dependencies

Nonverbal Predicate and Copula
Outline

1. Constituents vs. Dependencies

2. Universal Dependencies
   - A Tour through UD Syntax
   - Nonverbal Predicate and Copula
   - Ellipsis and Enhanced UD
   - Core Arguments vs. Oblique Dependents
Nonverbal Predicate and Copula

- Some languages use a copula verb:

```
Ivan is the best dancer .
```

- Some languages use a copula pronoun:

```
Ivan – to najlepszy tancerz .
Ivan – it best dancer .
```
Some languages use a copula verb:

Some languages omit the copula:
• Some languages use a copula verb:

- Ivan was the best dancer.

• Some languages use it only in some tenses:

- Ivan byl lučším tancorom.
- Ivan was best dancer.
Copula Verbs: We Are Restrictive!

- *To be* is copula:

  ![Diagram](image)

  Ivan is the best dancer.

- *To become* is not copula:

  ![Diagram](image)

  Ivan became the best dancer.
Once Copula, Always Copula!

- This is parallel with Russian:

```
Ivan is the best dancer.
```

- This is also parallel with Russian:

```
Ivan is today in Moscow.
```
• This is parallel with Russian:

Ivan is today in Moscow.

• But not with this in English:

There is a dancer in Moscow.
Clauses and Copula

- A clause can be the subject:

  The problem is that he is missing.

- But it cannot be annotated as the nonverbal predicate:

  The problem is that he is missing.
1. Constituents vs. Dependencies

2. Universal Dependencies
   - A Tour through UD Syntax
   - Nonverbal Predicate and Copula
   - Ellipsis and Enhanced UD
   - Core Arguments vs. Oblique Dependents
• Some treebanks would use an empty node to represent the second went.
• UD enhanced representation now allows empty nodes!
• But the basic representation sticks with the overt words.
 UD V1: The remnant Relation

Kate went to Florida and Jane (went) to Europe
PDT: The ExD Relation

Kate went to Florida and Jane (went) to Europe

Syntactic Analysis

Constituents vs. Dependencies
Kate went to Florida and Jane (went) to Europe

Syntactic Analysis
Constituents vs. Dependencies
UD V2: The orphan Relation

Kate went to Florida and Jane (went) to Europe

Syntactic Analysis: Constituents vs. Dependencies

Universal Dependencies: 53/111
Enhanced Dependencies: Gapping

Kate wants to go to Florida and Jane (wants) (go) to Europe
Enhanced Dependencies: Gapping and Control

Kate wants to go to Florida and Jane (wants) (go) to Europe

Syntactic Analysis
Constituents vs. Dependencies
Enhanced Dependencies: Coordination

Jane eats sweet apples and oranges

Syntactic Analysis: Constituents vs. Dependencies
A gdzie szukać szamponu, który myje?
And where to-look for-shampoo, that washes?
Universal Dependencies

Core Arguments vs. Oblique Dependents
Outline

1. Constituents vs. Dependencies

2. Universal Dependencies
   - A Tour through UD Syntax
   - Nonverbal Predicate and Copula
   - Ellipsis and Enhanced UD
   - Core Arguments vs. Oblique Dependents
## Dependents of Clauses (Verbal or Not)

<table>
<thead>
<tr>
<th>Core</th>
<th>Nominal</th>
<th>Clausal</th>
<th>Modifier</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>nsubj</td>
<td>csubj</td>
<td>advmod</td>
<td>aux</td>
<td></td>
</tr>
<tr>
<td>obl</td>
<td>advcl</td>
<td>discourse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vocative</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dislocated</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>expl</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non-Core</th>
<th>Nominal</th>
<th>Clausal</th>
<th>Modifier</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>obl</td>
<td></td>
<td>advmod</td>
<td>aux</td>
<td></td>
</tr>
<tr>
<td>expl</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Dependents of Verbs, Adjectives and Adverbs

<table>
<thead>
<tr>
<th>Core</th>
<th>Nominal</th>
<th>Clausal</th>
<th>Modifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>obj</td>
<td>ccomp</td>
<td></td>
<td></td>
</tr>
<tr>
<td>iobj</td>
<td>xcomp</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Core</td>
<td>obl</td>
<td>advcl</td>
<td>advmod</td>
</tr>
<tr>
<td>expl</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Dependents of Nominals

<table>
<thead>
<tr>
<th>Nominal</th>
<th>Clausal</th>
<th>Modifier</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>nmod</td>
<td>acl</td>
<td>amod</td>
<td>det</td>
</tr>
<tr>
<td>appos</td>
<td></td>
<td></td>
<td>case</td>
</tr>
</tbody>
</table>
Information Packaging

Constituents vs. Dependencies
Information Packaging

Constituents vs. Dependencies
He loaded the wagon with hay.
Information Packaging

He loaded the wagon with hay.

Syntactic Analysis

Constituents vs. Dependencies
UD is NOT about Semantic Roles!
The secret to understanding the design and current success of UD is to realize that the design is a very subtle compromise between approximately 6 things:

1. UD must be satisfactory on linguistic analysis grounds for individual languages.
2. UD must be good for linguistic typology, i.e., providing a suitable basis for bringing out cross-linguistic parallelism across languages and language families.
3. UD must be suitable for rapid, consistent annotation by a human annotator.
4. UD must be easily comprehended and used by a non-linguist, whether a language learner or an engineer with prosaic needs for language processing. ... it leads us to favor traditional grammar notions and terminology.
5. UD must be suitable for computer parsing with high accuracy.
6. UD must support well downstream language understanding tasks (relation extraction, reading comprehension, machine translation, ...)

It's easy to come up with a proposal that improves UD on one of these dimensions. The interesting and difficult part is to improve UD while remaining sensitive to all these dimensions.
UD Avoids Argument-Adjunct Distinction!
Avoiding an Argument-Adjunct Distinction

- From the guidelines:
  - Subtle, unclear, and frequently argued over
  - Questionable as a categorical distinction
  - Best practical solution is to eliminate it

BUT:
- Cannot be eliminated completely
- Some people/data have it and want to keep it
- It aligns well with traditional grammars

AND

I will argue that
- Core-oblique distinction is unclear and argued over too
(Though I will not propose to discard it.)
Avoiding an Argument-Adjunct Distinction

- From the guidelines:
  - Subtle, unclear, and frequently argued over
  - Questionable as a categorical distinction
  - Best practical solution is to eliminate it

- **BUT:**
  - Cannot be eliminated completely
  - Some people/data have it and want to keep it
    - It aligns well with traditional grammars
  - ⇒ there is now a relation subtype \texttt{obl:arg}
Avoiding an Argument-Adjunct Distinction

- From the guidelines:
  - Subtle, unclear, and frequently argued over
  - Questionable as a categorical distinction
  - Best practical solution is to eliminate it

- BUT:
  - Cannot be eliminated completely
  - Some people/data have it and want to keep it
    - It aligns well with traditional grammars
  - \( \implies \) there is now a relation subtype \text{obl:arg}

- AND I will argue that
  - Core-oblique distinction is unclear and argued over too
  - (Though I will not propose to discard it.)
So What Is Core and Why?

Syntactic Analysis

Constituents vs. Dependencies

Universal Dependencies
Community Confusion

- UD v1 guidelines took core-oblique for granted

- English (simplified):
  - Bare noun phrase $\Rightarrow$ core argument (nsubj, obj, iobj)
  - Prepositional phrase $\Rightarrow$ oblique argument or adjunct (obl)

Other languages: not necessarily! (Spanish, Japanese)

But some people simply took the English rule…

Manning’s law: non-linguists!

Clash with traditional terminology

Grammars of German, Czech etc. define prepositional objects

But these are not necessarily core…

Yet some people took their national definition of object…
Community Confusion

- UD v1 guidelines took core-oblique for granted

- English (simplified):
  - Bare noun phrase ⇒ core argument (nsubj, obj, iobj)
  - Prepositional phrase ⇒ oblique argument or adjunct (obl)

- Other languages: not necessarily! (Spanish, Japanese)
  - But some people simply took the English rule...
  - Manning’s law: non-linguists!
Community Confusion

- UD v1 guidelines took core-oblique for granted

- English (simplified):
  - Bare noun phrase $\Rightarrow$ core argument ($\text{nsubj, obj, iobj}$)
  - Prepositional phrase $\Rightarrow$ oblique argument or adjunct ($\text{obl}$)

- Other languages: not necessarily! (Spanish, Japanese)
  - But some people simply took the English rule...
  - Manning’s law: non-linguists!

- Clash with traditional terminology
  - Grammars of German, Czech etc. define prepositional objects
  - But these are not necessarily core...
  - Yet some people took their national definition of object...
Language-specific Coding Strategy

- **Idea:**
  - Oblique arguments are marked *similarly to adjuncts* (prepositions, certain morphological cases...)
  - Core arguments are marked differently
    - ⇒ easy for annotators and non-linguists!

- **Why are core arguments special?**
  - They tend to be *targeted by grammatical rules*
    - Passivization
    - Control verbs
    - Reflexives
    - ...

Syntactic Analysis
Constituents vs. Dependencies
Language-specific Coding Strategy

- Core vs. oblique is not defined in traditional grammar
- How shall we define it?

Andrews, 2007 (In Shopen: Language Typology)

Identify primary transitive predicates
- We need semantic roles for this! (One-time only.)

Actor/agent = function $A$
- Undergoer/patient = function $P$

Note the way they are coded
- Note other grammatical rules that target them
- Generalize to other predicates with same coding and rules
- Then define:
  - function $A$ ) nsubj
  - function $P$ ) obj
Language-specific Coding Strategy

- Core vs. oblique is not defined in traditional grammar
- How shall we define it?

- Andrews, 2007 (In Shopen: Language Typology)
  - Identify primary transitive predicates
    - We need semantic roles for this! (One-time only.)
Language-specific Coding Strategy

- Core vs. oblique is not defined in traditional grammar
- How shall we define it?

- Andrews, 2007 (In Shopen: Language Typology)
  - Identify primary transitive predicates
    - We need semantic roles for this! (One-time only.)
  - Actor/agent = function A
  - Undergoer/patient = function P
Language-specific Coding Strategy

- Core vs. oblique is not defined in traditional grammar
- How shall we define it?

- Andrews, 2007 (In Shopen: Language Typology)
  - Identify primary transitive predicates
    - We need semantic roles for this! (One-time only.)
  - Actor/agent = function A
  - Undergoer/patient = function P
  - Note the way they are coded
  - Note other grammatical rules that target them
  - Generalize to other predicates with same coding and rules
Language-specific Coding Strategy

- Core vs. oblique is not defined in traditional grammar
- How shall we define it?

Andrews, 2007 (In Shopen: Language Typology)
- Identify primary transitive predicates
  - We need semantic roles for this! (One-time only.)
- Actor/agent = function A
- Undergoer/patient = function P
- Note the way they are coded
- Note other grammatical rules that target them
- Generalize to other predicates with same coding and rules

- Then define:
  - function A ⇒ nsubj
  - function P ⇒ obj
Transitive Predicates in English

John kills Mary (primary transitive)

PROPN VERB PROPN

Syntactic Analysis
Constituents vs. Dependencies
Transitive Predicates in English

**Primary Transitive**
- **John**
- **kills**
- **Mary**

**Generalized Transitive**
- **John**
- **loves**
- **Mary**
Transitive Predicates in English

- John kills Mary (primary transitive)

- John loves Mary (generalized transitive)
Transitive Predicates in English

Syntactic Analysis

Constituents vs. Dependencies
Passivization in English

Mary is killed by John

Mary is loved by John

Syntactic Analysis

Constituents vs. Dependencies
Subject Control in English

Syntactic Analysis
Constituents vs. Dependencies

John wants to kill Mary
PROPN  VERB  PART  VERB  PROPN

John wants to love Mary
PROPN  VERB  PART  VERB  PROPN
Object Control in English

Syntactic Analysis
Constituents vs. Dependencies
Some Problems

- Some temporal adjuncts are bare noun phrases
  - I work the whole week.
  - I work every Friday.

- At least it cannot passivize:
  - *The whole week is worked by me.
  - *Every Friday is worked by me.

- But...
Some Problems

- Some transitive verbs cannot passivize
  - *John has a new car.*
  - *A new car is had by John.*
  - *Friday does not suit me.*
  - *I am not suited by Friday.*

Syntactic Analysis
Constituents vs. Dependencies
Some Problems

- Some transitive verbs cannot passivize
  - *John has a new car.*
  - *A new car is had by John.*
  - *Friday does not suit me.*
  - *I am not suited by Friday.*

- Some prepositional verbs can passivize
  - *You can rely on Ben.*
  - *Ben can be relied on.*
  - *They will take care of your children.*
  - *Your children will be taken care of.*
Tentative Summary?

- The borderline is inherently fuzzy
- No universally applicable and exact algorithm
- Better described in terms of probability
Tentative Summary?

- The borderline is inherently fuzzy
- No universally applicable and exact algorithm
- Better described in terms of probability
- Core coding not favored by adjuncts
- Oblique coding similar to most adjuncts
- Passivization etc. may help...
- ... but does not work as strict criterion
The borderline is inherently fuzzy
No universally applicable and exact algorithm
Better described in terms of probability

Core coding not favored by adjuncts
Oblique coding similar to most adjuncts
Passivization etc. may help...
... but does not work as strict criterion

Semantic roles needed when starting a new language
Argument-adjunct needed to describe exceptions (*the whole week*)
Intransitive Predicates

- Just one core argument
  - We already “know” how to find out if there are two

- ⇒ function $S$
  - Regardless of semantic role:
    - $John$ runs.
    - $John$ sleeps.
    - $John$ falls.

- Then define:
  - function $S$ ⇒ $nsubj$
Ditransitive Predicates

- Three core arguments
- Is one of them “least core”? \( \Rightarrow \text{iobj} \)
- (Alternatively, we could look at the semantic roles once again.)

```
I gave her a book
```

Passivization:
- *She was given a book by me.*
- *?A book was given her by me.*
Ditransitive Predicates

- Three core arguments
- Is one of them “least core”? \( \Rightarrow iobj \)
- (Alternatively, we could look at the semantic roles once again.)

- Andrews (2007): the status of the notion of ‘indirect object’ is problematic and difficult to sort out. The top priority is to work out what properties recipients and themes do and do not share with P arguments of primary transitive verbs.

Syntactic Analysis
Constituents vs. Dependencies
Jorge mató al dragón.

El dragón fue matado por Jorge.
Spanish Transitive Clauses

Syntactic Analysis

Constituents vs. Dependencies

- **nominal**
  - Case=Nom
  - bare NP
  - pre-verb
  - cross-ref on verb

- **VERB**
  - Voice=Act(,Pass)
  - declarative clause
  - agreement

- **(ADP)**
  - nominal
  - Case=Acc
  - (or bare NP)
  - post-verb
Spanish Adjunct Exceptions

Él trabaja toda la semana

He works whole the week

Subiremos a el tren a las cinco

We will board to the train at the five

Syntactic Analysis
Constituents vs. Dependencies
Spanish Ditransitive Clauses

Pedro le dio un libro a Isabel

PROPN PRON VERB DET NOUN ADP PROPN

Pedro le dio un libro a Isabel

PROPN PRON VERB DET NOUN

Syntactic Analysis
Constituents vs. Dependencies
Spanish Ditransitive Clauses

Pedro le dio un libro a Isabel

PROPN PRON VERB DET NOUN ADP PROPN

Un libro fue dado a Isabel por Pedro

DET NOUN AUX VERB ADP PROPN ADP PROPN

Syntactic Analysis
Constituents vs. Dependencies
Jiří zabil draka

Drak byl zabit Jiřím
Czech Transitive Clauses

Constituents vs. Dependencies

Syntactic Analysis

83/111
Czech Adjunct Exceptions

Syntactic Analysis

Constituents vs. Dependencies
Czech Ditransitive Clauses

Syntactic Analysis

Constituents vs. Dependencies

85/111
Dative: Recipient vs. Beneficiary

Petr četl Katce knihu

PROPN VERB PROPN NOUN

Case=Nom Case=Dat Case=Acc

root nsubj obj iobj/obl?

Petr zlomil Katce nohu

PROPN VERB PROPN NOUN

Case=Nom Case=Dat Case=Acc

root nsubj obj iobj/obl?
Monotransitive with Dative?

Zuzka pomohla Martinovi s úkolem

Zuzka helped Martin with homework

Case=Nom Case=Dat Case=Ins

Martinovi bylo pomoženo s úkolem

Martin was helped with homework

Case=Dat Gender=Neut Number=Sing Case=Ins

Syntactic Analysis

Constituents vs. Dependencies
Monotransitive with Genitive?

Novináři musí dbát zásad objektivity.

Journalists must observe principles of objectivity.
Monotransitive with Instrumental?

Karel hýbal nábytkem

PROPN VERB NOUN
Case=Nom Case=Ins

Furniture was moved

NOUN AUX ADJ
Case=Ins Gender=Neut Number=Sing

bylo hýbáno

aux:pass root

root
Spoléhali na ředitelovo rozhodnutí

They-relied on director’s decision

Na ředitelovo rozhodnutí bylo spoléháno

On director’s decision was relied

Syntactic Analysis

Constituents vs. Dependencies
There is a core-oblique scale:
- Nom > Acc > Gen,Dat > Ins > preposition

Where is the borderline?
There is a core-oblique scale:
- Nom > Acc > Gen,Dat > Ins > preposition

Where is the borderline?

UD Czech 1.0: object = argument
- Nom, Acc, Gen, Dat, Ins, ADP > “adverbial”
There is a core-oblique scale:
- Nom > Acc > Gen,Dat > Ins > preposition

Where is the borderline?

UD Czech 1.0: object = argument
- Nom, Acc, Gen, Dat, Ins, ADP > “adverbial”

UD Czech 2.1: bare NP > PP
- Nom, Acc, Gen, Dat, Ins > ADP + adjuncts
There is a core-oblique scale:
\[
\text{Nom} > \text{Acc} > \text{Gen, Dat} > \text{Ins} > \text{preposition}
\]

Where is the borderline?

UD Czech 1.0: \(\text{object} = \text{argument}\)
- Nom, Acc, Gen, Dat, Ins, ADP > “adverbial”

UD Czech 2.1: bare NP > PP
- Nom, Acc, Gen, Dat, Ins > ADP + adjuncts

But perhaps we should go even further?
- Nom, Acc > Gen, Dat, Ins, ADP + adjuncts
There is a core-oblique scale:

- Nom > Acc > Gen, Dat > Ins > preposition

Where is the borderline?

- UD Czech 1.0: object = argument
  - Nom, Acc, Gen, Dat, Ins, ADP > “adverbial”
- UD Czech 2.1: bare NP > PP
  - Nom, Acc, Gen, Dat, Ins > ADP + adjuncts

But perhaps we should go even further?

- Nom, Acc > Gen, Dat, Ins, ADP + adjuncts
- ⇒ No ditransitives in Czech!
Basque Transitive Clauses

Syntactic Analysis

Constituents vs. Dependencies

92/111
Basque Transitive Clauses

Syntactic Analysis
Constituents vs. Dependencies

(Storm) sunk it-has-it
Ekaitzak itsasontzia hondoratu du

(Niri) pleasing me-is-it
(To-me) wine ardoa gustatzen zait
Basque Intransitive Clauses

The-man died

Case=Abs

Water boiled it-has-it

Case=Erg

Syntactic Analysis
Constituents vs. Dependencies
Basque Ditransitive Clauses

Syntactic Analysis

Constituents vs. Dependencies

95/111
Basque Ditransitive Clauses

Syntactic Analysis

Constituents vs. Dependencies

Zezenak Bull saihtsa rib pitzatu zion Iñakiri
NOUN NOUN VERB AUX PROPN
Case=Erg Case=Abs

Iñakik liburua eman zion Arantxari
PROPN NOUN VERB AUX PROPN
Case=Erg Case=Abs Case=Dat

Iñakiri book given it-has-it-it to-Arantxa
PROPN NOUN VERB AUX PROPN
Case=Erg Case=Abs Case=Dat
Yidiṅ Transitive Clauses

Syntactic Analysis

Constituents vs. Dependencies
“I, (who) was slapped by the woman, laughed”
Yidiŋ “Dative” Adnominal Clauses

“Ị, (who) was slapped by the woman, laughed”

The coreferential (and elidable) NP must have S or P function.
“I, (who) was laughing, was slapped by the woman”

The coreferential (and elidable) NP must have S or P function.
"I, (who) was slapping the woman, laughed"

Original P is now oblique and original A is now S.
Tagalog Transitive Clauses

Syntactic Analysis
Constituents vs. Dependencies
Magaalis
Will-take
VERB
ang
the
DET
babae
woman
NOUN
ng
bigas
rice
NOUN
sa
from
sako
sack
NOUN
Voice=Act? Case=Nom Case=Acc Case=Loc

Aalisin
Will-take
VERB
ng
babae
woman
NOUN
ang
bigas
rice
NOUN
sa
from
sako
sack
NOUN
Voice=Pass? Case=Acc Case=Nom Case=Loc

Syntactic Analysis
Constituents vs. Dependencies
Tagalog Locative Voice ⇒ Ditransitive!

Syntactic Analysis

Constituents vs. Dependencies

Aalisan
Will-take
VERB
Voice=Locf
Case=Acc

ng
DET

babae
woman
NOUN
Case=Acc

ng
DET

bigas
rice
NOUN
Case=Acc

ang
from-the
DET

sako
sack
NOUN
Case=Nom
Tagalog Benefactive Voice ⇒ Ditransitive!

Ipagaalis ng babae ng bigas ang bata

VERB DET NOUN DET NOUN DET NOUN

Voice=Benf Case=Acc Case=Acc Case=Nom

Syntactic Analysis
Constituents vs. Dependencies
Plains Cree Transitive Clauses

Syntactic Analysis
Constituents vs. Dependencies

nominal
bare NP
Case
pre-verb
cross-ref on verb
}

VERB

obj
cross-ref on verb
agreement

declarative-clause

nominal
bare NP
Case
post-verb

Direct-Inverse Voice in Plains Cree

Animacy hierarchy: 1st person > 3rd person
Direct-Inverse Voice in Plains Cree

Animacy hierarchy: 1st person > 3rd person
Should we set nsubj > obj?
Direct-Inverse Voice in Plains Cree

Johnny helped Mary.

Obviation=Prx  Dir  Obviation=Obv

Johnny helped Mary.

Obviation=Obv  Dir  Obviation=Prx

Johnny helped Mary.

Obviation=Prx  Inv  Obviation=Obv

Johnny helped Mary.

Obviation=Obv  Inv  Obviation=Prx
The theme (not the recipient) is indirect object because it is not cross-referenced on the verb (it is *inanimate*, while the verb references an animate object).
Basic Universal Dependencies: 89 Languages and Growing

- **I.-E.:** Armenian, Greek (+Ancient), Breton, Irish, Scottish, Welsh
  - Germanic: Afrikaans, Danish, Dutch, English, Faroese, German, Gothic, Norwegian, Swedish, Swiss German
  - Romance: Catalan, French, Galician, Italian, Latin, Old French, Portuguese, Romanian, Spanish
  - Balto-Slavic: Belarusian, Bulgarian, Croatian, Czech, Church Slavonic, Old Russian, Polish, Russian, Serbian, Slovak, Slovenian, Ukrainian, Upper Sorbian, Latvian, Lithuanian
  - Indo-Iranian: Kurmanji, Persian, Hindi, Bhojpuri, Marathi, Sanskrit, Urdu
- **Uralic:** Erzya, Estonian, Finnish, Hungarian, Karelian, Livvi, Komi Permyak+Zyrian, Moksha, Sámi North+Skolt
- **Dravidian:** Tamil, Telugu; Turkic: Kazakh, Turkish, Uyghur
- **Af.-As.:** Akkad., Amhar., Arab., Assyr., Coptic, Hebrew, Malt.
- **Sino-Tib.:** Cantonese, Classical Chinese, Chinese; Aus.-As.: Vietnamese
- **Tai-Kadai:** Thai; Austronesian: Indonesian, Tagalog
- **Other:** Buryat, Japanese, Korean, Basque, Sw. Sign, Naija, Bambara, Wolof, Yoruba, Warlpiri, Mbyá Guaraní
Syntactic Analysis

Summary

- Constituent (phrase) trees ... context-free grammar
- Dependency trees (or graphs)
  - Nonprojective dependencies
- Universal Dependencies
  - Unified annotation for all languages
    - Language-specific extensions
  - Content words higher than function words ... better parallelism
  - Clauses – nominals – modifier words
  - Core arguments vs. oblique dependents

https://ufal.cz/courses/npfl094