Universal Dependencies
Dubious Linguistics and Crappy Parsing?

Joakim Nivre
Uppsala University
Department of Linguistics and Philology

Based on collaborative work with Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher Manning, Ryan McDonald, Natalia Silveira, Slav Petrov, Sampo Pyysalo, Sebastian Schuster, Reut Tsarfaty, Francis Tyers, Daniel Zeman and many others
Introduction
Introduction

Growing interest in multilingual and cross-lingual NLP
  • Multilingual evaluation campaigns to test generality of approaches
  • Cross-lingual learning to support low-resource languages
Introduction

Growing interest in multilingual and cross-lingual NLP
  • Multilingual evaluation campaigns to test generality of approaches
  • Cross-lingual learning to support low-resource languages

Growing awareness of methodological problems
  • Current NLP relies heavily on linguistic annotation
  • Annotation guidelines vary across languages
A cat chases rats and mice.
A cat chases rats and mice.

En katt jagar råttor och möss.
A cat chases rats and mice

En katt jagar råttor och möss

En kat jager rotter og mus
A cat chases rats and mice.

En katt jagar råttor och möss.

En kat jager rotter og mus.

Toutefois, les filles adorent les desserts.
Why is this a problem?
Why is this a problem?

- Hard to compare empirical results across languages
Why is this a problem?

• Hard to compare empirical results across languages
• Hard to usefully do cross-lingual structure transfer
Why is this a problem?

- Hard to compare empirical results across languages
- Hard to usefully do cross-lingual structure transfer
- Hard to evaluate cross-lingual learning
Why is this a problem?

- Hard to compare empirical results across languages
- Hard to usefully do cross-lingual structure transfer
- Hard to evaluate cross-lingual learning
- Hard to build and maintain multilingual systems
Why is this a problem?

- Hard to compare empirical results across languages
- Hard to usefully do cross-lingual structure transfer
- Hard to evaluate cross-lingual learning
- Hard to build and maintain multilingual systems
- Hard to make comparative linguistic studies
Why is this a problem?

- Hard to compare empirical results across languages
- Hard to usefully do cross-lingual structure transfer
- Hard to evaluate cross-lingual learning
- Hard to build and maintain multilingual systems
- Hard to make comparative linguistic studies
- Hard to validate linguistic typology
Why is this a problem?

- Hard to compare empirical results across languages
- Hard to usefully do cross-lingual structure transfer
- Hard to evaluate cross-lingual learning
- Hard to build and maintain multilingual systems
- Hard to make comparative linguistic studies
- Hard to validate linguistic typology
- Hard to make progress towards a universal parser
Universal Dependencies

http://universaldependencies.org

Osaka welcomes the participants of COLING.

大阪はCOLINGの参加者を歓迎します。

Le chat chasse les chiens.

The chat chases the dogs.

http://universaldependencies.org
Universal Dependencies

http://universaldependencies.org

Napoli accoglie i partecipanti di il CLiC-it

Osaka welcomes the participants of COLING

Le chat chasse les chiens.

Part-of-speech tags
Universal Dependencies

http://universaldependencies.org

Part-of-speech tags

Morphological features
Universal Dependencies

http://universaldependencies.org

Dependency relations

大阪 is COLING の 参加者 を 欅迎し ます

PROPNN ADPN PROPN NOUN ADP VERB AUX

case nsubj nmod case obj aux

Osaka welcomes the participants of COLING

PROPNN VERBN DET NOUN ADP

nsbj obj det nmod case

Number=Sing Number=Sing Definite=Def PronType=Art Number=Sing Number=Sing

Person=3 Tense=Pres Definite=Def Gender=Masc

Part-of-speech tags

Morphological features
Universal Dependencies

http://universaldependencies.org
Universal Dependencies

http://universaldependencies.org

Stanford Dependencies
Universal Dependencies

http://universaldependencies.org

Stanford Dependencies

Google UD
Universal Dependencies

http://universaldependencies.org

Stanford Dependencies

Stanford UD

Google UD
Universal Dependencies

http://universaldependencies.org

Stanford Dependencies

Stanford UD

Google UD

HamleDT
Universal Dependencies

http://universaldependencies.org
Universal Dependencies

http://universaldependencies.org
Universal Dependencies
http://universaldependencies.org

Milestones:
• Kick-off meeting at EACL in Gothenburg, April 2014
• Release of annotation guidelines, v1, October 2014
• Releases of treebanks every 6 months, v1.0–v1.4
• Release of annotation guidelines, v2, December 2016

Open community effort – anyone can contribute!
December 13, 2016:

- 47 languages
- 64 treebanks
- 145 contributors
- 7000+ downloads
December 13, 2016:
- 47 languages
- 64 treebanks
- 145 contributors
- 7000+ downloads
December 13, 2016:

- 47 languages
- 64 treebanks
- 145 contributors
- 7000+ downloads

UD Japanese
Masayuki Asahara
Hiroshi Kanayama
Yuji Matsumoto
Yusuke Miyao
Shinsuke Mori
Takaaki Tanaka
Sumire Uematsu

Chief Cat Herder
Release and Documentation Task Force

Universal Guidelines Group
A guided tour of the UD framework
A guided tour of the UD framework

Why such weird dependency trees?
A guided tour of the UD framework

Why such weird dependency trees?

Dubious linguistics?  
Crappy parsing?
Goals and Requirements
Goals and Requirements

Cross-linguistically consistent grammatical annotation
Goals and Requirements

Cross-linguistically consistent grammatical annotation

Support multilingual research in NLP and linguistics

• Meaningful linguistic analysis within and across languages
• Syntactic parsing in monolingual and cross-lingual settings
• Useful information for downstream language understanding tasks
Goals and Requirements

Cross-linguistically consistent grammatical annotation

Support multilingual research in NLP and linguistics
  • Meaningful linguistic analysis within and across languages
  • Syntactic parsing in monolingual and cross-lingual settings
  • Useful information for downstream language understanding tasks

Build on common usage and existing de facto standards
Goals and Requirements

Cross-linguistically consistent grammatical annotation

Support multilingual research in NLP and linguistics
  • Meaningful linguistic analysis within and across languages
  • Syntactic parsing in monolingual and cross-lingual settings
  • Useful information for downstream language understanding tasks

Build on common usage and existing de facto standards

Complement – not replace – language-specific schemes
The UD Philosophy
The UD Philosophy

Maximize parallelism – but don’t overdo it

- Don’t annotate the same thing in different ways
- Don’t make different things look the same
- Don’t annotate things that are not there
Maximize parallelism – but don’t overdo it
  • Don’t annotate the same thing in different ways
  • Don’t make different things look the same
  • Don’t annotate things that are not there

Universal taxonomy with language-specific elaboration
  • Languages select from a universal pool of categories
  • Allow language-specific extensions
Design Principles
Design Principles

Dependency

• Widely used in practical NLP systems
• Available in treebanks for many languages
Design Principles

Dependency
- Widely used in practical NLP systems
- Available in treebanks for many languages

Lexicalism
- Basic annotation units are words – syntactic words
- Words have morphological properties
- Words enter into syntactic relations
Design Principles

Dependency
• Widely used in practical NLP systems
• Available in treebanks for many languages

Lexicalism
• Basic annotation units are words – syntactic words
• Words have morphological properties
• Words enter into syntactic relations

Recoverability
• Transparent mapping from input text to word segmentation
Word Segmentation

What is a word?

- Single part-of-speech tag
- Real syntactic relation

Two-level segmentation

- Represent orthographic tokens in addition to syntactic words
What is a word?

- Single part-of-speech tag
- Real syntactic relation

Two-level segmentation

- Represent orthographic tokens in addition to syntactic words
Word Segmentation

What is a word?

- Single part-of-speech tag
- Real syntactic relation

Two-level segmentation

- Represent orthographic tokens in addition to syntactic words

<table>
<thead>
<tr>
<th>Text</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>del</td>
<td>di</td>
</tr>
<tr>
<td></td>
<td>il</td>
</tr>
</tbody>
</table>
Word Segmentation

What is a word?

- Single part-of-speech tag
- Real syntactic relation

Two-level segmentation

- Represent orthographic tokens in addition to syntactic words

<table>
<thead>
<tr>
<th>Text</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>del</td>
<td>di il</td>
</tr>
<tr>
<td>dámelemo</td>
<td>da me lo</td>
</tr>
</tbody>
</table>
Word Segmentation

What is a word?
- Single part-of-speech tag
- Real syntactic relation

Two-level segmentation
- Represent orthographic tokens in addition to syntactic words

<table>
<thead>
<tr>
<th>Text</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>del</td>
<td>di il</td>
</tr>
<tr>
<td>dámeloo</td>
<td>da me lo</td>
</tr>
<tr>
<td>ושמהשמש</td>
<td>da me lo</td>
</tr>
</tbody>
</table>
What is a word?

- Single part-of-speech tag
- Real syntactic relation

Two-level segmentation

- Represent orthographic tokens in addition to syntactic words

<table>
<thead>
<tr>
<th>Text</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>del</td>
<td>di il</td>
</tr>
<tr>
<td>dámeló</td>
<td>da me lo</td>
</tr>
<tr>
<td>רב מشرط</td>
<td>רב מشرط</td>
</tr>
<tr>
<td>大阪国際会議場</td>
<td>大阪 国際 会議場</td>
</tr>
</tbody>
</table>
Morphology

Le chat chasse les chiens.
Morphology

Le le chat chat chasse chasser les le chiens chien .

• Lemma representing the semantic content of the word
Morphology

- Lemma representing the semantic content of the word
- Part-of-speech tag representing its grammatical class
Morphology

- Lemma representing the semantic content of the word
- Part-of-speech tag representing its grammatical class

<table>
<thead>
<tr>
<th>Open</th>
<th>Closed</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJ</td>
<td>ADP</td>
<td>PUNCT</td>
</tr>
<tr>
<td>ADV</td>
<td>AUX</td>
<td>SYM</td>
</tr>
<tr>
<td>INTJ</td>
<td>CCONJ</td>
<td>X</td>
</tr>
<tr>
<td>NOUN</td>
<td>DET</td>
<td></td>
</tr>
<tr>
<td>PROP</td>
<td>NUM</td>
<td></td>
</tr>
<tr>
<td>VERB</td>
<td>PART</td>
<td></td>
</tr>
<tr>
<td>PRON</td>
<td>SCONJ</td>
<td></td>
</tr>
</tbody>
</table>
## Morphology

<table>
<thead>
<tr>
<th>Lemma</th>
<th>Part-of-speech</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Le</td>
<td>DET</td>
<td>Gender=Def</td>
</tr>
<tr>
<td>le</td>
<td>DET</td>
<td>Gender=Masc</td>
</tr>
<tr>
<td>chat</td>
<td>NOUN</td>
<td>Number=Sing</td>
</tr>
<tr>
<td>le</td>
<td>DET</td>
<td>Number=Sing</td>
</tr>
<tr>
<td>chat</td>
<td>NOUN</td>
<td>Gender=Masc</td>
</tr>
<tr>
<td>chasse</td>
<td>VERB</td>
<td>Mood=Ind</td>
</tr>
<tr>
<td>les</td>
<td>DET</td>
<td>Gender=Masc</td>
</tr>
<tr>
<td>chiens</td>
<td>NOUN</td>
<td>Number=Plur</td>
</tr>
<tr>
<td>.</td>
<td>DET</td>
<td>Gender=Masc</td>
</tr>
<tr>
<td>.</td>
<td>NOUN</td>
<td>Number=Plur</td>
</tr>
</tbody>
</table>

- Lemma representing the semantic content of the word
- Part-of-speech tag representing its grammatical class
- Features representing lexical and grammatical properties of the lemma or the particular word word form
Morphology

- Lemma representing the semantic content of the word
- Part-of-speech tag representing its grammatical class
- Features representing lexical and grammatical properties of the lemma or the particular word form

<table>
<thead>
<tr>
<th>Lexical</th>
<th>Inflectional Nominal</th>
<th>Inflectional Verbal</th>
</tr>
</thead>
<tbody>
<tr>
<td>PronType</td>
<td>Gender</td>
<td>VerbForm</td>
</tr>
<tr>
<td>NumType</td>
<td>Animacy</td>
<td>Mood</td>
</tr>
<tr>
<td>Poss</td>
<td>Number</td>
<td>Tense</td>
</tr>
<tr>
<td>Reflex</td>
<td>Case</td>
<td>Aspect</td>
</tr>
<tr>
<td>Foreign</td>
<td>Definite</td>
<td>Voice</td>
</tr>
<tr>
<td>Abbr</td>
<td>Degree</td>
<td>Evident</td>
</tr>
</tbody>
</table>

- Definite
- Gender
- Number
- Person
- Polite
- PronType
- Number
- Person
- Polite
- Gender
- Number
- Person
- Polite
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 the content word they modify
• Content words are related by dependency relations
• Function words attach to the content word they modify
• Punctuation attach to head of phrase or clause
The dog was chased by the cat.

Hunden jagades av katten.
The dog was chased by the cat.
The dog was chased by the cat.

Hunden jagades av katten.
The dog was chased by the cat.

Hunden jagades av katten.
Syntactic Relations
Syntactic Relations

Taxonomy of 37 universal syntactic relations

• Three types of structures: nominals, clauses, modifiers
• Core arguments vs. other dependents (not complements vs. adjuncts)
• Language-specific subtypes
Syntactic Relations

Taxonomy of 37 universal syntactic relations

• Three types of structures: nominals, clauses, modifiers
• Core arguments vs. other dependents (not complements vs. adjuncts)
• Language-specific subtypes

Basic and enhanced representations

• Basic dependencies form a (possibly non-projective) tree
• Additional dependencies in the enhanced representation
### Syntactic Relations

<table>
<thead>
<tr>
<th></th>
<th>Nominal Dep</th>
<th>Clause Dep</th>
<th>Modifier Word</th>
<th>Function Word</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Core Predicate Dep</strong></td>
<td>nsubj, obj, iobj</td>
<td>csubj, ccomp, xcomp</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Non-Core Predicate Dep</strong></td>
<td>obl, vocative, expl, dislocated</td>
<td>advcl</td>
<td>advmod*, discourse</td>
<td>aux, cop, mark</td>
</tr>
<tr>
<td><strong>Nominal Dep</strong></td>
<td>nmod, appos, nummodmod</td>
<td>acl</td>
<td>amod</td>
<td>det, clf, case</td>
</tr>
<tr>
<td><strong>Coordination</strong></td>
<td>conj, cc</td>
<td>fixed, flat, compound</td>
<td>parataxis, list</td>
<td>orphan, goeswith, reparandum</td>
</tr>
</tbody>
</table>

* Generalized modifier of predicates and (non-nominal) modifiers
A Two-Level Architecture

Universal relations
  • Broad categories to allow cross-linguistic comparison

Language-specific relations
  • Subtypes to capture language-specific phenomena
A Two-Level Architecture

Universal relations
  • Broad categories to allow cross-linguistic comparison

Language-specific relations
  • Subtypes to capture language-specific phenomena
A Two-Level Architecture

Universal relations
• Broad categories to allow cross-linguistic comparison

Language-specific relations
• Subtypes to capture language-specific phenomena

<table>
<thead>
<tr>
<th>Universal</th>
<th>Subtype</th>
</tr>
</thead>
<tbody>
<tr>
<td>acl</td>
<td>acl:relcl</td>
</tr>
</tbody>
</table>
A Two-Level Architecture

Universal relations
  • Broad categories to allow cross-linguistic comparison

Language-specific relations
  • Subtypes to capture language-specific phenomena

<table>
<thead>
<tr>
<th>Universal</th>
<th>Subtype</th>
</tr>
</thead>
<tbody>
<tr>
<td>acl</td>
<td>acl:relcl</td>
</tr>
<tr>
<td>compound</td>
<td>compound:pct</td>
</tr>
</tbody>
</table>
A Two-Level Architecture

Universal relations
  • Broad categories to allow cross-linguistic comparison

Language-specific relations
  • Subtypes to capture language-specific phenomena

<table>
<thead>
<tr>
<th>Universal</th>
<th>Subtype</th>
</tr>
</thead>
<tbody>
<tr>
<td>acl</td>
<td>acl:relcl</td>
</tr>
<tr>
<td>compound</td>
<td>compound:prt</td>
</tr>
<tr>
<td>nmod</td>
<td>nmod:poss</td>
</tr>
</tbody>
</table>
Universal Dependencies v2

Executive summary of changes from v1 to v2

- Tokenization and word segmentation
- Morphology
  - General principles
  - Universal POS tags (single document)
  - Universal features (single document)
  - Language-specific features
  - Conversion from other tagsets
- Syntax
  - General principles
  - Basic dependencies
    - Simple clauses
    - Nominals
    - Complex clauses
    - Other constructions
  - Enhanced dependencies
  - Universal dependency relations (single document)
  - Language-specific relations
- CoNLL-U format

This is the online documentation for Universal Dependencies, version 2 (2016-12-01). Note: The treebanks listed below still follow the v1 guidelines available here.
Universal Dependencies

Executive summary of contents

- Tokenization and Normalization
- Morphology
  - General properties
  - Universal properties
  - Universal features
  - Language-specific properties
  - Conversion
- Syntax
  - General properties
  - Basic dependencies
    - Simple
    - Noun phrase dependencies
    - Core argument dependencies
    - Other dependencies
  - Enhanced dependencies
  - Universal features
  - Language-specific features
- CoNLL-U format

Simple Clauses

The UD annotation assumes the clause as one of the basic structures that we expect to find in all languages. A simple clause minimally consists of a predicate together with its core argument dependents, but may be extended with oblique modifiers. Core arguments are typically nominals, while oblique modifiers are either (oblique) nominals or adverbal modifiers. (In complex clauses, both core arguments and oblique modifiers can also be realized as subordinate clauses.) Finally, the predicate may be associated with function words that express different types of grammatical information such as tense, mood, aspect, voice, evidentiality, or type of subordination.

Intransitive and Transitive Clauses

In most clauses, the predicate takes the form of a verb, which may be intransitive or transitive.

An intransitive verb takes a single argument (usually referred to as S in the literature on linguistic typology) with the neub₁ relation. A transitive verb in addition takes an argument with the obj₁ relation. When deciding which relation to use with which argument in a transitive clause, the neub₁ relation should be used with the argument that most resembles the proto-agent (often called A in linguistic typology) and that satisfies additional language-internal criteria for subjecthood based on case-marking, agreement and/or linear position with respect to the predicate. The obj₁ relation should be used for the argument that most resembles the proto-patient (often called O or P in linguistic typology) and that satisfies relevant language-internal criteria. Note that, while case-marking (whether morphological or analytic) can provide important evidence in specific languages, case alignment should not be used to decide the assignment of core argument roles. Thus, in ergative languages, the patient-like argument of a transitive verb (O/P) will take the the obj₁ relation despite the fact that it carries the same case marking as the neub₁ argument (S) of an intransitive verb.

Some languages allow extended transitive clauses, where more than two dependents are realized as core arguments. The additional core arguments then receive the job₁ relation (for "indirect object"), while the obj₁ relation is reserved for the argument most patient-like non-subject argument. The criterion for deciding whether an additional dependent is a core argument is whether it has the typical encoding of a core argument with respect to case-marking, agreement and word order. For example, the English double object construction qualifies as an extended transitive clause because all three nominals appear without prepositions.

This is the online documentation for Universal Dependencies, version 2 (2016-12-01). Note: The treebanks listed below still follow the v1 guidelines available here.
Why such weird dependency trees?
we have come to Osaka

“Content-Head Dependencies”
we have come to Osaka

“Content-Head Dependencies”
“Content-Head Dependencies”

we have come to Osaka
“Content-Head Dependencies”

we have come to Osaka

“Function-Head Dependencies”

we have come to Osaka
we have come to Osaka

“Content-Head Dependencies”

we have come to Osaka

“Function-Head Dependencies”
Dubious Linguistics?

“Such an approach to the syntax of natural languages is contrary to most work in theoretical syntax in the past 35 years, regardless of whether this work is constituency- or dependency-based.” (Groß and Osborne, 2015)
Dubious Linguistics?

“Such an approach to the syntax of natural languages is contrary to most work in theoretical syntax in the past 35 years, regardless of whether this work is constituency- or dependency-based.” (Groß and Osborne, 2015)

Crappy Parsing?

“It is now fairly well known that, while dependency representations in which content words are made heads tend to help semantically oriented downstream applications, dependency parsing numbers are higher if you make auxiliary verbs heads […] and if you make prepositions the head of prepositional phrases.” (De Marneffe et al., 2014)
Manning’s Law

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

1. UD needs to be satisfactory on linguistic analysis grounds for individual languages.
2. UD needs to 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 suitable for computer parsing with high accuracy.
5. UD must be easily comprehended and used by a non-linguist, whether a language learner or an engineer with prosaic needs for language processing.
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.
What is a head?

<table>
<thead>
<tr>
<th>Semantic functor</th>
<th>V + NP (V)</th>
<th>P + NP (P)</th>
<th>NP + VP (VP)</th>
<th>Det + N (Det)</th>
<th>Aux + VP (Aux)</th>
<th>Comp + S (Comp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Semantic argument</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>(B) Determinant of concord</td>
<td>(*)</td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>(C) Morphosyntactic locus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>(D) Subcategorizand</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>(E) Governor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(F) Distributional equivalent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>(G) Obligatory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>(H) Ruler</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

Key: = same as entry for ‘Semantic functor’
* different from entry for ‘Semantic functor’

Zwicky (1985), summarised by Hudson (1987)
Why choose one?
Why choose one?

Head properties may be shared by several elements

- So neither content-head nor function-head can be quite right
Why choose one?

Head properties may be shared by several elements
  • So neither content-head nor function-head can be quite right

Linguistic theories capture this in different ways
  • Lexical vs. functional heads (Chomsky, 1995)
  • Surface syntax vs. deep syntax (Sgall et al., 1986; Mel’čuk, 1988)
  • Dissociated nucleus (Tesnière, 1959)
Why choose one?

Head properties may be shared by several elements
• So neither content-head nor function-head can be quite right

Linguistic theories capture this in different ways
• Lexical vs. functional heads (Chomsky, 1995)
• Surface syntax vs. deep syntax (Sgall et al., 1986; Mel’čuk, 1988)
• Dissociated nucleus (Tesnière, 1959)

What about UD?
UD Syntax
UD Syntax

UD representations are mono-stratal – single tree

- Facilitates annotation, parsing and downstream tasks
UD Syntax

UD representations are mono-stratal – single tree
  • Facilitates annotation, parsing and downstream tasks

Tree structure primarily reflects lexical dependencies
  • Brings out parallelism between typologically diverse languages
  • Reveals predicate-argument structure for downstream tasks
UD Syntax

UD representations are mono-stratal – single tree
  • Facilitates annotation, parsing and downstream tasks

Tree structure primarily reflects lexical dependencies
  • Brings out parallelism between typologically diverse languages
  • Reveals predicate-argument structure for downstream tasks
UD Syntax

UD representations are mono-stratal – single tree

- Facilitates annotation, parsing and downstream tasks

Tree structure primarily reflects lexical dependencies

- Brings out parallelism between typologically diverse languages
- Reveals predicate-argument structure for downstream tasks

Reddy et al. (2016) Transforming Dependency Structures to Logical Forms for Semantic Parsing

\[
\lambda z. \exists x y. \text{acquired}(z_e) \land \text{Pixar}(y_a) \land \text{Disney}(x_a) \land \text{arg}_1(z_e, x_a) \land \text{arg}_2(z_e, y_a)
\]
UD Syntax

Other relations encoded in labels – not tree structure

- Functional relations link functional heads to lexical heads
- Coordination relations link equivalent heads/dependents
- Multiword relations link elements of lexicalized expressions
Other relations encoded in **labels** – not tree structure

- Functional relations link functional heads to lexical heads
- Coordination relations link equivalent heads/dependents
- Multiword relations link elements of lexicalized expressions

```
the man in the moon
Jack and Jill left in spite of this
```
Other relations encoded in labels – not tree structure

- Functional relations link functional heads to lexical heads
- Coordination relations link equivalent heads/dependents
- Multiword relations link elements of lexicalized expressions
Other relations encoded in labels – not tree structure

- Functional relations link functional heads to lexical heads
- Coordination relations link equivalent heads/dependents
- Multiword relations link elements of lexicalized expressions
we have come to Osaka
we have come to Osaka
we have come to Osaka
we have come to Osaka

dependency

nucleus
we have come to Osaka
we have come to Osaka
Linguistic Typology

she

nsubj

она

нsubj
	nice

милая

nonverbal predication

Croft et al. (2017) Linguistic Typology Meets Universal Dependencies
Linguistic Typology

Croft et al. (2017) Linguistic Typology Meets Universal Dependencies
Linguistics vs. Parsing

- Mono-stratal but multi-relational representations
- Both lexical and functional heads can be extracted
Linguistics vs. Parsing

• Mono-stratal but multi-relational representations
• Both lexical and functional heads can be extracted

But syntactic parsers don’t know this!? 
we have come to Osaka

“Function-Head Dependencies”

“Content-Head Dependencies”

✔

✘
Schwartz et al. (2012) Learnability-Based Syntactic Annotation Design
“Function-Head Dependencies”
we have come to Osaka

“Content-Head Dependencies”
we have come to Osaka

Schwartz et al. (2012) Learnability-Based Syntactic Annotation Design

| Function head | Content head |
**Schwartz et al. (2012) Learnability-Based Syntactic Annotation Design**

<table>
<thead>
<tr>
<th>Prep – Noun</th>
<th>Function head</th>
<th>Content head</th>
</tr>
</thead>
<tbody>
<tr>
<td>✔</td>
<td>✔️</td>
<td>✖️</td>
</tr>
</tbody>
</table>
we have come to Osaka

“Function-Head Dependencies”

Schwartz et al. (2012) Learnability-Based Syntactic Annotation Design

<table>
<thead>
<tr>
<th></th>
<th>Function head</th>
<th>Content head</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prep – Noun</td>
<td>✔️</td>
<td>❌</td>
</tr>
<tr>
<td>Det – Noun</td>
<td>❌</td>
<td>✔️</td>
</tr>
</tbody>
</table>
“Function-Head Dependencies”

we have come to Osaka

“Content-Head Dependencies”

we have come to Osaka

Schwartz et al. (2012) Learnability-Based Syntactic Annotation Design

<table>
<thead>
<tr>
<th>Prep – Noun</th>
<th>Function head</th>
<th>Content head</th>
</tr>
</thead>
<tbody>
<tr>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Det – Noun</td>
<td>✘</td>
<td>✔</td>
</tr>
<tr>
<td>CC – Conj</td>
<td>✘</td>
<td>✔</td>
</tr>
</tbody>
</table>
Schwartz et al. (2012) Learnability-Based Syntactic Annotation Design

<table>
<thead>
<tr>
<th></th>
<th>Function head</th>
<th>Content head</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prep – Noun</td>
<td>✔️</td>
<td>✗</td>
</tr>
<tr>
<td>Det – Noun</td>
<td>✗</td>
<td>✔️</td>
</tr>
<tr>
<td>CC – Conj</td>
<td>✗</td>
<td>✔️</td>
</tr>
<tr>
<td>Aux – Verb</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
Schwartz et al. (2012) Learnability-Based Syntactic Annotation Design

```
<table>
<thead>
<tr>
<th></th>
<th>Function head</th>
<th>Content head</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prep – Noun</td>
<td>✔</td>
<td>✘</td>
</tr>
<tr>
<td>Det – Noun</td>
<td>✘</td>
<td>✔</td>
</tr>
<tr>
<td>CC – Conj</td>
<td>✘</td>
<td>✔</td>
</tr>
<tr>
<td>Aux – Verb</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Mark – Infinitive</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
```
UD Parsing
UD Parsing

Silveira and Manning (2015)
Monolingual parsing using transform-dettransform

English  aux case cop  Inconclusive results
## UD Parsing

<table>
<thead>
<tr>
<th>Research</th>
<th>Language</th>
<th>Dependency</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silveira and Manning (2015)</td>
<td>English</td>
<td>aux, case, cop</td>
<td>Inconclusive results</td>
</tr>
<tr>
<td>De Lhoneux and Nivre (2016)</td>
<td>All</td>
<td>aux</td>
<td>Negative results</td>
</tr>
</tbody>
</table>
# UD Parsing

<table>
<thead>
<tr>
<th>Study</th>
<th>Language</th>
<th>Features</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silveira and Manning (2015)</td>
<td>English</td>
<td>aux</td>
<td>Inconclusive results</td>
</tr>
<tr>
<td>Monolingual parsing using transform-detransform</td>
<td></td>
<td>case cop</td>
<td></td>
</tr>
<tr>
<td>De Lhoneux and Nivre (2016)</td>
<td>All</td>
<td>aux</td>
<td>Negative results</td>
</tr>
<tr>
<td>Monolingual parsing using transform-detransform</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attardi et al. (2015)</td>
<td>Italian</td>
<td>case cop</td>
<td>UD &gt; ISDT</td>
</tr>
<tr>
<td>Monolingual parsing using different representations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Language</td>
<td>Feature</td>
<td>Results</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>----------</td>
<td>---------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>Silveira and Manning (2015)</td>
<td>English</td>
<td>aux, case, cop</td>
<td>Inconclusive results</td>
</tr>
<tr>
<td>De Lhoneux and Nivre (2016)</td>
<td>All</td>
<td>aux</td>
<td>Negative results</td>
</tr>
<tr>
<td>Attardi et al. (2015)</td>
<td>Italian</td>
<td>case, cop</td>
<td>UD &gt; ISDT</td>
</tr>
<tr>
<td>Rosa (2015)</td>
<td>All</td>
<td>case</td>
<td>UD &gt; PDT</td>
</tr>
</tbody>
</table>
UD Parsing
UD Parsing

Not so bad after all?

• No clear evidence that “content-head” is harder to parse in general
• In the cross-lingual setting, it even seems to work better
UD Parsing

Not so bad after all?

• No clear evidence that “content-head” is harder to parse in general
• In the cross-lingual setting, it even seems to work better

Can we do better?

• Exploit the full representation – lexical and functional heads
• Use typology of syntactic relations as a bias for learning
A Historical Perspective
A Historical Perspective

Constituency parsing – largely driven by PTB

- Perhaps too much emphasis on English (until recently)
- But deep analysis of categories and representations
A Historical Perspective

Constituency parsing – largely driven by PTB
  • Perhaps too much emphasis on English (until recently)
  • But deep analysis of categories and representations

Dependency parsing – largely driven by CoNLL data
  • More attention to typological diversity from the start
  • But parsers had to remain agnostic about linguistic categories
Dependency Parsing
Dependency Parsing

- Parsers know only one type of syntactic relation
Dependency Parsing

- Parsers know only one type of syntactic relation
- Parsers do not interpret dependency labels
Dependency Parsing

- Parsers know only one type of syntactic relation
- Parsers do not interpret dependency labels
- Parsers represent every construction by its “head”
The dog was chased by the black cat.

- Endocentric construction with cat as head
- Little (syntactic) information is lost by dropping black
**Dependency Parsing**

The dog was chased by the black cat.

- Dissociated nucleus consisting of *was* and *chased*.
- Neither content-head nor function-head is right!
UD and Deep Learning
Figure 5: The representation of a dependency subtree (above) is computed recursively applying composition functions to head, modifier, relation i-triples. In the case of multiple dependents of a single head, the recursive branching order is imposed by the order of the parser’s reduce operations (below).

Parameter optimization was performed using stochastic gradient descent with an initial learning rate of \( \alpha_0 = 0.1 \), and the learning rate was updated on each pass through the training data as \( \alpha_t = \frac{\alpha_0}{1 + \beta t} \), with \( \beta = 0.1 \) and where \( t \) is the number of epochs completed. No momentum was used. To mitigate the effects of “exploding” gradients, we clipped the \( \| \cdot \|_2 \) norm of the gradient to 5 before applying the weight update rule (Sutskever et al., 2014; Graves, 2013). An \( \| \cdot \|_2 \) penalty of \( 1 \times 10^{-6} \) was applied to all weights.

Matrix and vector parameters were initialized with uniform samples in \( \pm \frac{\sqrt{6}}{r + c} \), where \( r \) and \( c \) were the number of rows and columns in the structure (Glorot and Bengio, 2010).

Dimensionality. The full version of our parsing model sets dimensionalities as follows. LSTM hidden states are of size 100, and we use two layers of LSTMs for each stack. Embeddings of the parser actions used in the composition functions have 16 dimensions, and the output embedding size is 20 dimensions. Pretained word embeddings have 100 dimensions (English) and 80 dimensions (Chinese), and the learned word embeddings have 32 dimensions. Part of speech embeddings have 12 dimensions. These dimensions were chosen based on intuitively reasonable values (words should have higher dimensionality than parsing actions, POS tags, and relations; LSTM states should be relatively large), and it was confirmed on development data that they performed well.

Future work might more carefully optimize these parameters; our reported architecture strikes a balance between minimizing computational expense and finding solutions that work.

5 Experiments

We applied our parsing model and several variations of it to two parsing tasks and report results below.

5.1 Data

We used the same data setup as Chen and Manning (2014), namely an English and a Chinese parsing task. This baseline configuration was chosen since they likewise used a neural parameterization to predict actions in an arc-standard transition-based parser.

• For English, we used the Stanford Dependency (SD) treebank (de Marneffe et al., 2006) used in (Chen and Manning, 2014) which is the closest model published, with the same splits. The part-of-speech tags are predicted by using the Stanford Tagger (Toutanova et al., 2003) with an accuracy of 97.3%. This treebank contains a negligible amount of non-projective arcs (Chen and Manning, 2014).

• For Chinese, we use the Penn Chinese Treebank 5.1 (CTB5) following Zhang and Clark (2008), with gold part-of-speech tags which is also the same as in Chen and Manning (2014).

Language model word embeddings were generated, for English, from the AFP portion of the English Gigaword corpus (version 5), and from the complete Chinese Gigaword corpus (version 2), and where

6 Software for replicating the experiments is available from https://github.com/clab/lstm-parser.

3 We did perform preliminary experiments with LSTM states of 32, 50, and 80, but the other dimensions were our initial guesses.

UD and Deep Learning

Stenetorp (2013) Transition-Based Dependency Parsing Using Recursive Neural Networks

Dyer et al. (2015) Transition-Based Dependency Parsing with Stack Long Short-Term Memory

composition functions
+1–2% labeled accuracy

Kuncoro et al. (2016) What Do Recurrent Neural Network Grammars Learn About Syntax?

<table>
<thead>
<tr>
<th>Noun phrases</th>
<th>Verb phrases</th>
<th>Prepositional phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canadian (0.09) Auto (0.31) Workers (0.2) union (0.22) president (0.18) no (0.29) major (0.05) Eurobond (0.32) or (0.01) foreign (0.01) bond (0.1) offerings (0.22) Saatchi (0.12) client (0.14) Philips (0.21) Lighting (0.24) Co. (0.29) nonperforming (0.18) commercial (0.23) real (0.25) estate (0.1) assets (0.25) the (0.1) Jamaica (0.1) Tourist (0.03) Board (0.17) ad (0.20) account (0.40)</td>
<td>buying (0.31) and (0.25) selling (0.21) NP (0.23) ADVP (0.27) show (0.29) PRT (0.23) PP (0.21) pleaded (0.48) ADJP (0.23) PP (0.15) PP (0.08) PP (0.06) received (0.33) PP (0.18) NP (0.32) PP (0.17) cut (0.27) NP (0.37) PP (0.22) PP (0.14)</td>
<td>ADVP (0.14) on (0.72) NP (0.14) ADVP (0.05) for (0.54) NP (0.40) ADVP (0.02) because (0.73) of (0.18) NP (0.07) such (0.31) as (0.65) NP (0.04) from (0.39) NP (0.49) PP (0.12)</td>
</tr>
<tr>
<td>the (0.0) final (0.18) hour (0.81) their (0.0) first (0.23) test (0.77) Apple (0.62) . (0.02) Compaq (0.1) and (0.01) IBM (0.25) both (0.02) stocks (0.03) and (0.06) futures (0.88) NP (0.01) , (0.0) and (0.98) NP (0.01)</td>
<td>to (0.99) VP (0.01) were (0.77) n’t (0.22) VP (0.01) did (0.39) n’t (0.60) VP (0.01) handle (0.09) NP (0.91) VP (0.15) and (0.83) VP (0.02)</td>
<td>of (0.97) NP (0.03) in (0.93) NP (0.07) by (0.96) S (0.04) at (0.99) NP (0.01) NP (0.1) after (0.83) NP (0.06)</td>
</tr>
</tbody>
</table>

attention
graded endocentricity
UD and Deep Learning

Stenetorp (2013) Transition-Based Dependency Parsing Using Recursive Neural Networks

Dyer et al. (2015) Transition-Based Dependency Parsing with Stack Long Short-Term Memory

composition functions
+1–2% labeled accuracy

UD as an inductive bias

graded endocentricity

Kuncoro et al. (2016) What Do Recurrent Neural Network Grammars Learn About Syntax?

<table>
<thead>
<tr>
<th>Noun phrases</th>
<th>Verb phrases</th>
<th>Prepositional phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canadian (0.09) Auto (0.31)</td>
<td>buying (0.31) and (0.25) selling (0.21) NP (0.23)</td>
<td>ADVP (0.14) on (0.72) NP (0.14)</td>
</tr>
<tr>
<td>no (0.29) major (0.05) Eurobond (0.32)</td>
<td>ADVP (0.27) show (0.29) PRT (0.23) PP (0.21)</td>
<td>ADVP (0.05) for (0.54) NP (0.40)</td>
</tr>
<tr>
<td>Saatchi (0.12) client (0.14) Philips (0.21)</td>
<td>pleaded (0.48) ADJP (0.23) PP (0.15) PP (0.08) PP (0.06)</td>
<td>ADVP (0.02) because (0.73) of (0.18) NP (0.07)</td>
</tr>
<tr>
<td>nonperforming (0.18) commercial (0.23)</td>
<td>received (0.33) PP (0.18) NP (0.32) PP (0.17)</td>
<td>such (0.31) as (0.65) NP (0.04)</td>
</tr>
<tr>
<td>real (0.25) estate (0.1)</td>
<td>cut (0.27) NP (0.37) PP (0.22) PP (0.14)</td>
<td>from (0.39) NP (0.49) PP (0.12)</td>
</tr>
<tr>
<td>the (0.1) Jamaica (0.1) Tourist (0.03) Board (0.17)</td>
<td>to (0.99) VP (0.01)</td>
<td>of (0.97) NP (0.03)</td>
</tr>
<tr>
<td>ad (0.20)</td>
<td>were (0.77) n’t (0.22) VP (0.01)</td>
<td>in (0.93) NP (0.07)</td>
</tr>
<tr>
<td>NP (0.01)</td>
<td>did (0.39) n’t (0.60) VP (0.01)</td>
<td>by (0.96) S (0.04)</td>
</tr>
<tr>
<td>and (0.98) NP (0.01)</td>
<td>handle (0.09) NP (0.91)</td>
<td>at (0.99) NP (0.01)</td>
</tr>
<tr>
<td>Apple (0.62)</td>
<td>VP (0.15) and (0.83) VP 0.02</td>
<td>NP (0.1) after (0.83) NP (0.06)</td>
</tr>
<tr>
<td>(0.02) Compaq (0.1) and (0.01) IBM (0.25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>both (0.02) stocks (0.03) and (0.06) futures (0.88)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP (0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusion
Conclusion

Dubious linguistics?

• Lexical dependencies and functional relations encoded in a single tree
• Grounded in linguistic typology and dependency grammar traditions
Conclusion

Dubious linguistics?

• Lexical dependencies and functional relations encoded in a single tree
• Grounded in linguistic typology and dependency grammar traditions

Crappy parsing?

• Not so bad with existing parsers, especially for cross-lingual parsing
• Learn richer parsing models grounded in linguistic typology
UD Events in 2017

CoNLL-2017 Shared Task
http://universaldependencies.org/conll17/

• Multilingual Parsing from Raw Text to Universal Dependencies
• Collocated with ACL, August 3–4, 2017, Vancouver, Canada
• Call for participation in December 2016, data release in March 2017

First Workshop on Universal Dependencies
http://universaldependencies.org/udw17/

• Collocated with NoDaLiDa, May 20, 2017, Gothenburg, Sweden
• Submission deadline: March 20, 2017
Thanks to all UD contributors!