Resources and algorithms for broad-coverage multilingual syntactic analysis

Ryan McDonald

Collaborators: Emily Pitler (Google), Oscar Täckström (Google), Joakim Nivre (Uppsala), and Universal Dependency Treebank group.
Dependency Parsing

(Labeled) head-modifier relations

Used in: Translation, Search, IE, Sentiment, ...
They solved the problem with statistics.

Collins et al. 05, Xu et al. 09
They solved the problem with statistics.

Collins et al. 05, Xu et al. 09
They solved the problem with statistics.

Collins et al. 05, Xu et al. 09
MT Reordering

Collins et al. 05, Xu et al. 09

They solved the problem with statistics

They PRON the DET problem NOUN solved VERB

with ADP statistics NOUN

nsubj dobj prep pobj
They solved the problem with statistics, according to Collins et al. 05, Xu et al. 09.
MT Reordering

Collins et al. 05, Xu et al. 09

They the problem statistics with solved

Translate
**Outline**

- **Algorithms**
  - Supervised multilingual dependency parsing
  - Mildly non-projective transition-based parsing
  - Projection-based dependency parsing
  - Tying parameters via linguistic resources

- **Resources**
  - Universal Dependency Treebanks
Transition-based Parsing

Arc Eager

Configuration = (Stack, Buffer, Tree)

Stack

Buffer

Transitions

Shift

Left-arc

Right-arc

Reduce

[ root John saw Mary ]
Transition-based Parsing

Arc Eager

[ ] [ root John saw Mary ] { }
Transition-based Parsing

Arc Eager

\[
[ \text{root} ] [ \text{John saw Mary} ] \{ \}
\]

Shift \hspace{1cm} \text{Left-arc} \hspace{1cm} \text{Right-arc} \hspace{1cm} \text{Reduce}
Transition-based Parsing

Arc Eager

[ root ] [ John ] [ saw Mary ] { }

Shift Left-arc Right-arc Reduce
Transition-based Parsing

Arc Eager

[ root ]  [ John ]  [ saw Mary ]  {  }
Transition-based Parsing

Arc Eager

[ root ] [ saw Mary ] { saw }
John

Shift Left-arc Right-arc Reduce
Transition-based Parsing

Arc Eager

[ root ] [ saw Mary ]

Shift Left-arc Right-arc Reduce

{ saw }

John

root
Transition-based Parsing

Arc Eager

[ root  saw ]  [ Mary ]
root → saw
John

Shift  Left-arc  Right-arc  Reduce
Transition-based Parsing

Arc Eager

[ root saw ] [ Mary ]

root { saw }

John

Shift  Left-arc  Right-arc  Reduce
Transition-based Parsing

Arc Eager

[ root saw Mary ] [ ]

root

{ saw }

John Mary

Shift Left-arc Right-arc Reduce
Transition-based Parsing

Arc Eager

[ root saw ] [ ]

{ saw } root

John

Mary

Shift Left-arc Right-arc Reduce
TRANSITION-BASED PARSING

Arc Eager

[ root ] [ ]

Shift  Left-arc  Right-arc  Reduce

root  

{  saw  }

John  Mary
Transition-based Parsing

Arc Eager

[ root ] [ ]

root

{ saw }

John

Mary

DONE

Shift

Left-arc

Right-arc

Reduce
Transition-based Parsing

- Arc-eager transition-based parsing
  - Fast: $O(n)$
  - Accurate: Zhang & Nivre 2011, Bohnet et al. 12-14
  - Usually within 0.5% of graph-based
- Arc-decomposable (Golberg & Nivre 2012)
  - Each transition correlates with properties of tree
- Very little spurious ambiguity
Arc-Eager Parsing

Cross-serial dependencies

Verb-projection raising

Remnant Extraposition

Cannot handle non-projectivity

Scrambling

WH-movement

Extraposition

Examples from Pitler et al. 2013
## Non-projective Transition-based Parsing

### Past solutions

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attardi 06</td>
<td>$O(n)$</td>
</tr>
<tr>
<td>Pseudo-projective</td>
<td>$O(L^2)$</td>
</tr>
<tr>
<td>Covington’s</td>
<td>$O(n^2)$</td>
</tr>
<tr>
<td><strong>Swap (Nivre ‘08)</strong></td>
<td>$O(n^2)$</td>
</tr>
<tr>
<td>2-planar (GR&amp;N ‘08)</td>
<td>$O(n^2)$</td>
</tr>
<tr>
<td>Choi &amp; McCallum ’13</td>
<td>$O(n^2)$</td>
</tr>
</tbody>
</table>
Swap Parsing (Nivre 08)

Configuration = (Stack, Buffer, Tree)

root A review came out of Argo

Stack
Buffer
Swap Parsing (Nivre 08)

[ ] [ root A review came out of Argo ]

root A review came out of Argo
Swap Parsing (Nivre 08)

[ root ] [ A review came out of Argo ]

root A review came out of Argo
Swap Parsing (Nivre 08)

[ root A ] [ review came out of Argo ]

root A review came out of Argo
Swap Parsing (Nivre 08)

[ root ] [ review came out of Argo ]

root  A  review  came  out  of  Argo
Swap Parsing (Nivre 08)

[ root ] [ review came out of Argo ]

root A review came out of Argo
swap parsing (nivre 08)

[ root ] [ came review out of Argo ]

root A review came out of Argo
Swap Parsing (Nivre 08)

[ root ] [ came review out of Argo ]
Swap Parsing (Nivre 08)

[ root came ] [ review out of Argo ]

root A review came out of Argo
Swap Parsing (Nivre 08)

[ root came ] [ out review of Argo ]

root A review came out of Argo
Swap Parsing (Nivre 08)

[ root  came ]  [ out  review  of  Argo ]

root  A  review  came  out  of  Argo
Swap Parsing (Nivre 08)

[ root  came  out  ]  [ review  of  Argo ]

root  A  review  came  out  of  Argo
Swap Parsing (Nivre 08)

[ root ] [ came ]

[ review ] [ of Argo ]
Swap Parsing (Nivre 08)

[ root ]

[ came ]

[ review ]

[ of Argo ]

root A review came out of Argo
Swap Parsing (Nivre 08)

[ root came review ] [ of Argo ]

root A review came out of Argo
Swap Parsing (Nivre 08)

[ root came review ] [ of Argo ]

root A review came out of Argo
Swap Parsing (Nivre 08)

[ root came review of ] [ Argo ]
Swap Parsing (Nivre 08)

[ root  came  review  of  ]  [  ]  [  Argo  ]

root  A  review  came  out  of  Argo
Swap Parsing (Nivre 08)

[ root came review of Argo ] [ ]
Swap Parsing (Nivre 08)

[ root    came    review    of ]  [ ]

root  A  review  came  out  of  Argo
Swap Parsing (Nivre 08)

[ root came review ] [ root A review came out of Argo ]
Swap Parsing (Nivre 08)

[ root      came      ] [                     ]

root      A    review      came      out      of      Argo
Swap Parsing (Nivre 08)

[ root ] [ ]

root A review came out of Argo
A review came out of Argo.
Swap Parsing (Nivre 08)

A review came out of Argo

Pros
- Simple
- Parses all non-projective trees

Cons
- O(n^2) + spurious ambiguity
- Parses all non-projective trees

DONE
**Mildly Non-projective Parsing**
Mildly Non-projective Parsing

- A class of dependency trees is mildly non-projective
  - It is a strict *subset* of all non-projective trees
  - It is *characterizable* via local properties
  - It is *linguistically relevant*
Mildly Non-projective Parsing

- A class of dependency trees is mildly non-projective
  - It is a strict **subset** of all non-projective trees
  - It is **characterizable** via local properties
  - It is **linguistically relevant**

Language is subset of non-projectivity
Mildly Non-projective Parsing

- A class of dependency trees is mildly non-projective
  - It is a strict **subset** of all non-projective trees
  - It is **characterizable** via local properties
  - It is **linguistically relevant**

- We care about classes that permit tractable inference algorithms
Mildly Non-projective Parsing

- A class of dependency trees is mildly non-projective
  - It is a strict **subset** of all non-projective trees
  - It is **characterizable** via local properties
  - It is **linguistically relevant**

  ![Language is subset of non-projectivity](image)

- We care about classes that permit tractable inference algorithms
- Popular in graph-based parsing
  - Bodirsky et al. 05, Kuhlmann & Nivre 06, Gomez-Rodriguez et al. 11, Satta & Kuhlmann 13, Pitler et al. 12/13
# Non-projective Transition-based Parsing

## Past solutions

<table>
<thead>
<tr>
<th></th>
<th>Complexity</th>
<th>Mildly non-projective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attardi 06</td>
<td>$O(n)$</td>
<td>?</td>
</tr>
<tr>
<td>Pseudo-projective</td>
<td>$O(L^2)$</td>
<td>NO</td>
</tr>
<tr>
<td>Covington’s</td>
<td>$O(n^2)$</td>
<td>NO</td>
</tr>
<tr>
<td><strong>Swap (Nivre ‘08)</strong></td>
<td>$O(n^2)$</td>
<td>NO</td>
</tr>
<tr>
<td>2-planar (GR&amp;N ‘08)</td>
<td>$O(n^2)$</td>
<td>YES</td>
</tr>
<tr>
<td>Choi &amp; McCallum ’13</td>
<td>$O(n^2)$</td>
<td>NO</td>
</tr>
</tbody>
</table>
## Non-projective Transition-based Parsing

### Past solutions

<table>
<thead>
<tr>
<th>Past Solutions</th>
<th>Complexity</th>
<th>Mildly non-projective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attardi 06</td>
<td>O(n)</td>
<td>?</td>
</tr>
<tr>
<td>Pseudo-projective</td>
<td>O(L^2)</td>
<td>NO</td>
</tr>
<tr>
<td>Covington’s</td>
<td>O(n^2)</td>
<td>NO</td>
</tr>
<tr>
<td>Swap (Nivre ‘08)</td>
<td>O(n^2)</td>
<td>NO</td>
</tr>
<tr>
<td>2-planar (GR&amp;N ‘08)</td>
<td>O(n^2)</td>
<td>YES</td>
</tr>
<tr>
<td>Choi &amp; McCallum ’13</td>
<td>O(n^2)</td>
<td>NO</td>
</tr>
</tbody>
</table>

### This talk

<table>
<thead>
<tr>
<th>This talk</th>
<th>Complexity</th>
<th>Mildly non-projective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two-registers</td>
<td>O(n)</td>
<td>YES</td>
</tr>
</tbody>
</table>
Two-Interval Crossing Trees

Dependency trees can be divided into arc intervals
Each interval is minimal set of crossing arcs
Two-Interval Crossing Trees

Dependency trees can be divided into arc intervals. Each interval is minimal set of crossing arcs.

Two-Interval Crossing Tree
In each interval, there exists 2 words, x & y, such that every crossing arc is incident to x or y.
Linguistically Motivated

**Cross-serial dependencies**

- das that mer em Hans es the huus helped aastriiche
- $D_AT$ Hans$_{ACC}$ halfed aastricke paint

**Scrambling**

- Kitab Book-Acc Fatma Fatma Esra-Gen read-Ger-3Sg-Acc biliyor know-Prog.

**Verb-projection raising**

- das er em Karajan wil en arie chön-e vorsinge
- that he (to) Karajan wants an aria can sing-for

**WH-movement**

- ROOT When does he think that we left?

**Remnant Extraposition**

- dat Cecilia de reigers beweerde to fotograferen
- that Cecilia the herons claimed to photograph

**Extraposition**

- ROOT A review came out yesterday of this article
Empirically Motivated

2-Interval Crossing
Avg: 95%

Projective
Avg: 73%

Ba   Cz   Nl   En   De   El   Hu   Pt   Sl   Tr

Evalita, December 2014
Can we parse this?
Two-registers Parsing

Configuration = (Stack, Buffer, Register1, Register2, Tree)

Stack
Registers
Buffer

root A review came out of Argo
Two-registers Parsing

[ ] [ ] [ ] [ root A review came out of Argo]

root A review came out of Argo
Two-registers Parsing

\[ \text{[root} \quad \text{][ } \quad \text{][ } \quad \text{][} \quad \text{A review came out of Argo} \]\n
root A review came out of Argo
Two-registers Parsing

[root A ] [ ] [ ] [ ] [ review came out of Argo ]

root A review came out of Argo
Two-registers Parsing

[root A ] [ review ] [ ] [ came out of Argo ]

root A review came out of Argo
Two-registers Parsing

[ root ] [ review ] [ ] [ came out of Argo ]

root A review came out of Argo
A review came out of Argo
[root out] [review] [came] [of Argo]

root A review came out of Argo
Two-registers Parsing

[ root ] [ review ] [ came ] [ of Argo ]

root → A review → came → out → of Argo
Two-registers Parsing

A review came out of Argo
Two-registers Parsing

A review came out of Argo
A review came out of Argo.
Two-registers Parsing

[ root ] [ review ] [ came ] [ ]

root A review came out of Argo
A review came out of Argo
Two-registers Parsing

[ root review ] [ ] [ came ] [ ]
Two-registers Parsing

A review came out of Argo
Two-registers Parsing

[ root ] [ [ ] [ came ] [ ] ]

root A review came out of Argo
Two-registers Parsing

[A review came out of Argo]
Two-registers Parsing

A review came out of Argo
Two-registers Parsing

Parsing time is $O(n)$
Theorem

Two-registers transition system parses exactly Two-Interval crossing trees
Even Better

K-registers transition system parses exactly

K-Interval crossing trees

K=2 has ~95% coverage
K=3 has ~97% coverage
K=4 has ~99.5% coverage
Experiments: CoNLL 06/07 Data

Accuracy: Two-registers > Swap > Arc-eager

Speed: Arc-eager > Two-registers > Swap
Mildly Non-projective Transition-Based Parsing

- Two-registers system
  - O(n)
  - Arc-eager in nature (less spurious ambiguity)
- Smaller coverage ...
- ... but more accurate
What if we don’t have training data?

Only ~20 treebanks exist
Cross-Lingual Syntactic Transfer

- Learn parsers for resource-poor languages from resource-rich languages
- Hwa et al. 2005 and earlier

English Treebank

- John likes Mary
  - NOUN
  - VERB
  - NOUN

Syntactic Transfer

...
Delexicalized Transfer

English TB

Parser

John NOUN
likes VERB
Mary NOUN

...
Delexicalized Transfer

English TB

John NOUN
likes VERB
Mary NOUN

...
Delexicalized Transfer

English TB

Delexicalized Parser

Ignore words
Delexicalized Transfer

English TB

NOUN  VERB  NOUN

...
Delexicalized Transfer

English TB

Delexicalized Parser

O Γιαννίς βλέπει την Μαρία
DET NOUN VERB DET NOUN
Delexicalized Transfer

English TB

Delexicalized Parser

O Ελένη βλέπει την Μαρία
DET NOUN VERB DET NOUN
Delexicalized Transfer

English TB

NOUN  VERB  NOUN  ...

Delexicalized Parser

O Γιάννης βλέπει την Μαρία
DET  NOUN  VERB  DET  NOUN
Delexicalized Transfer

English TB

Delexicalized Parser

Zeman & Resnik ‘08: POS-tags
Täckström et al. ‘12: xling clusters
Durrett et al. ‘12: xling dict
Multi-Source Delexicalized Transfer

McDonald et al. 11

Treebanks
Multi-Source Delexicalized Transfer

McDonald et al. 11
World Atlas of Language Structure

• Typological database, e.g., English:

John    hit    the    ball    with  the    bat

subject  object  prepositional-phrase  prep-object

NOUN VERB DET NOUN PREP DET NOUN
Typological database, e.g., English:

- Subjects are to the left of verbs

John hit the ball with the bat
NOUN VERB DET NOUN PREP DET NOUN
Typological database, e.g., English:

- Subjects are to the left of verbs
- Objects are to the right of verbs
Typological database, e.g., English:

- Subjects are to the left of verbs
- Objects are to the right of verbs
- Prepositions are after the noun or verb
World Atlas of Language Structure

- Typological database, e.g., English:
  - Subjects are to the left of verbs
  - Objects are to the right of verbs
  - Prepositions are after the noun or verb
  - Nouns are after the preposition
Typological database, e.g., English:

- Subjects are to the left of verbs
- Objects are to the right of verbs
- Prepositions are after the noun or verb

Naseem et al. 2012
Can guide parameter sharing in syntactic transfer
Delex

- Full features, delexicalized
Delex

Full features, delexicalized

Stack-POS=NOUN
Input-POS=NOUN
Stack-POS=NOUN && Input-POS=NOUN
Stack-Word=John
Input-Word=Mary
Distance=2
etc.
Delex

Full features, delexicalized

Stack-POS=NOUN
Input-POS=NOUN
Stack-POS=NOUN && Input-POS=NOUN
Stack-Word=John
Input-Word=Mary
Distance=2

e etc.
Parsing + WALS

WALS:

Stack-POS=ADP && Input-POS=NOUN && Direction=right
Parsing + WALS

WALS:

Stack-POS=ADP && Input-POS=NOUN && Direction=right
&& Wals.85A=Postposition

order of adposition
WALS:

Stack-POS=ADP && Input-POS=NOUN && Direction=right && Wals.85A=Postposition
**Parsing + WALS**

**WALS:**

Stack-POS=ADP && Input-POS=NOUN && Direction=right && Wals.85A=Postposition
Parsing + WALS

<table>
<thead>
<tr>
<th>Language</th>
<th>Delex</th>
<th>+WALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ba</td>
<td>47</td>
<td>66</td>
</tr>
<tr>
<td>tr</td>
<td>37</td>
<td>59</td>
</tr>
<tr>
<td>ja</td>
<td>39</td>
<td>47</td>
</tr>
<tr>
<td>cz</td>
<td>41</td>
<td>44</td>
</tr>
<tr>
<td>ar</td>
<td>44</td>
<td>45</td>
</tr>
<tr>
<td>zh</td>
<td>45</td>
<td>53</td>
</tr>
<tr>
<td>hu</td>
<td>53</td>
<td>56</td>
</tr>
<tr>
<td>de</td>
<td>56</td>
<td>57</td>
</tr>
<tr>
<td>nl</td>
<td>57</td>
<td>59</td>
</tr>
<tr>
<td>sv</td>
<td>59</td>
<td>61</td>
</tr>
<tr>
<td>el</td>
<td>61</td>
<td>64</td>
</tr>
<tr>
<td>bg</td>
<td>64</td>
<td>63</td>
</tr>
<tr>
<td>es</td>
<td>63</td>
<td>65</td>
</tr>
<tr>
<td>it</td>
<td>65</td>
<td>67</td>
</tr>
<tr>
<td>ca</td>
<td>67</td>
<td>71</td>
</tr>
<tr>
<td>pt</td>
<td>71</td>
<td>73</td>
</tr>
<tr>
<td>AVG</td>
<td>73</td>
<td>78</td>
</tr>
</tbody>
</table>
Parsing + WALS

Delex + WALS

Indo-European

AVG
**+ Source Semi-Supervised Learning**

AVG-16-langs

- Delex
  - +WALS
- +WALS +SSL

Scores:
- Delex: 55
- AVG-16-langs: 62
- +WALS +SSL: 66
+ Source Semi-Supervised Learning

SOTA Unsupervised
Spitkovsky et al. 2012
16 lang avg: 46.7

AVG-16-langs

Delex

62 +WALS

55

66 +WALS +SSL
+ Source Semi-Supervised Learning

SOTA Unsupervised
Spitkovsky et al. 2012
16 lang avg: **46.7**

SOTA Supervised
16 lang avg: **86.3%**

AVG-16/langs

- Delex
- +WALS
- +WALS +SSL

Evalita, December 2014
SOTA Unsupervised
Spitkovsky et al. 2012
16 lang avg: 46.7

SOTA Supervised
16 lang avg: 86.3%

McDonald et al. 2013 underestimate true performance by ~10%
Resources
Universal Dependency Treebanks

**Goal**: Parsers that produce a common syntactic representation for all the world’s languages
Universal Dependency Treebanks

https://code.google.com/p/uni-dep-tb/
Coordinators: Ryan McDonald, Joakim Nivre, Slav Petrov

Data contributors include Yvonne Quirimbach-Brundage and others at Appen-Butler-Hill; Adam LaMontagne, Milan Soucek, Timo Jarvinen, Alessandra Radici and others at Lionbridge

Joakim Nivre provided the harmonized version of the Swedish Treebank Talbanken portion (http://stp.lingfil.uu.se/~nivre/swedish_treebank/)

Filip Ginter and the group at Turku provided the Finnish data and assisted in the harmonization process (http://bionlp.utu.fi/fintreebank.html)

Maria Simi and other researchers at Pisa provided the harmonized Italian data (http://medialab.di.unipi.it/wiki/ISDT)
### Universal Dependency Treebanks

<table>
<thead>
<tr>
<th>Language</th>
<th>Sentences</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>De</td>
<td>15k</td>
<td>R,B,N,W</td>
</tr>
<tr>
<td>En</td>
<td>43k</td>
<td>PTB</td>
</tr>
<tr>
<td>Es</td>
<td>16k</td>
<td>R,B,N,W</td>
</tr>
<tr>
<td>Fi</td>
<td>15k</td>
<td>TDT</td>
</tr>
<tr>
<td>Fr</td>
<td>16k</td>
<td>R,B,N,W</td>
</tr>
<tr>
<td>Id</td>
<td>5k</td>
<td>W</td>
</tr>
<tr>
<td>It</td>
<td>7k</td>
<td>SIDT</td>
</tr>
<tr>
<td>Ja</td>
<td>9k</td>
<td>R,B,N,W</td>
</tr>
<tr>
<td>Ko</td>
<td>7k</td>
<td>B,N,W</td>
</tr>
<tr>
<td>Pt-Br</td>
<td>12k</td>
<td>N,W</td>
</tr>
<tr>
<td>Sv</td>
<td>6k</td>
<td>SDT</td>
</tr>
</tbody>
</table>

- **B** = blogs
- **R** = reviews
- **N** = news
- **W** = wiki
Annotation Scheme: Stanford

- Syntactic/semantic in nature
- Primarily content-head
  - Prepositions and copulas are exceptions
- 12 coarse POS tags
Problems: de Marneffe et al. 14

- Not entirely content head
  - Genitive, possessive: adposition parallelism
  - “Son of John” and “John’s son” different
  - Dropped copula
- Poor handling of ellipsis, elision, etc.
- Poor distinction between modification and compounding
- Tag set is impoverished
**Principles: Segmentation**

- Lexicalist view of syntax: dependencies between words
- No segmentation into morphemes
  - Split apart clitics, contractions and even some derivational morphology
  - But we still keep MWEs segmented

**CoNLL-U format**

```
1-2  vamonos  _
 1  vamos  ir
 2  nos  nosotros
3-4  al  _
 3  a  a
 4  el  el
 5  mar  mar
```
Word: lemma + POS + features

- **ADJ**: adjective
- **ADP**: adposition
- **ADV**: adverb
- **AUX**: auxiliary verb
- **CONJ**: coordinating conjunction
- **DET**: determiner
- **INTJ**: interjection
- **NOUN**: noun
- **NUM**: numeral
- **PART**: particle
- **PRON**: pronoun
- **PROPN**: proper noun
- **PUNCT**: punctuation
- **SCONJ**: subordinating conjunction
- **SYM**: symbol
- **VERB**: verb
- **X**: other
**Principles: Morphology/POS**

**Word: lemma + POS + features**

- **ADJ**: adjective
- **ADP**: adposition
- **ADV**: adverb
- **AUX**: auxiliary verb
- **CONJ**: coordinating conjunction
- **DET**: determiner
- **INTJ**: interjection
- **NOUN**: noun
- **NUM**: numeral
- **PART**: particle
- **PRON**: pronoun
- **PROPN**: proper noun
- **PUNCT**: punctuation
- **SCONJ**: subordinating conjunction
- **SYM**: symbol
- **VERB**: verb
- **X**: other

**Refine**

- **Animacy**: animacy
- **Aspect**: aspect
- **Case**: case
- **Definite**: definiteness or state
- **Degree**: degree of comparison
- **Gender**: gender
- **Mood**: mood
- **Negative**: whether the word can be or is negated
- **NumType**: numeral type
- **Number**: number
- **Person**: person
- **Poss**: possessive
- **PronType**: pronominal type
- **Reflex**: reflexive
- **Tense**: tense
- **VerbForm**: form of verb or deverbal
- **Voice**: voice
**Principles: Morphology/POS**

**Word: lemma + POS + features**

<table>
<thead>
<tr>
<th>POS</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJ</td>
<td>adjective</td>
</tr>
<tr>
<td>ADP</td>
<td>adposition</td>
</tr>
<tr>
<td>ADV</td>
<td>adverb</td>
</tr>
<tr>
<td>AUX</td>
<td>auxiliary verb</td>
</tr>
<tr>
<td>CONJ</td>
<td>coordinating conjunction</td>
</tr>
<tr>
<td>DET</td>
<td>determiner</td>
</tr>
<tr>
<td>INTJ</td>
<td>interjection</td>
</tr>
<tr>
<td>NOUN</td>
<td>noun</td>
</tr>
<tr>
<td>NUM</td>
<td>numeral</td>
</tr>
<tr>
<td>PART</td>
<td>particle</td>
</tr>
<tr>
<td>PRON</td>
<td>pronoun</td>
</tr>
<tr>
<td>PROPN</td>
<td>proper noun</td>
</tr>
<tr>
<td>PUNCT</td>
<td>punctuation</td>
</tr>
<tr>
<td>SCONJ</td>
<td>subordinating conjunction</td>
</tr>
<tr>
<td>SYM</td>
<td>symbol</td>
</tr>
<tr>
<td>VERB</td>
<td>verb</td>
</tr>
<tr>
<td>X</td>
<td>other</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animacy</td>
<td>animacy</td>
</tr>
<tr>
<td>Aspect</td>
<td>aspect</td>
</tr>
<tr>
<td>Case</td>
<td>case</td>
</tr>
<tr>
<td>Definite</td>
<td>definiteness or state</td>
</tr>
<tr>
<td>Degree</td>
<td>degree of comparison</td>
</tr>
<tr>
<td>Gender</td>
<td>gender</td>
</tr>
<tr>
<td>Mood</td>
<td>mood</td>
</tr>
<tr>
<td>Negative</td>
<td>whether the word can be or is negated</td>
</tr>
<tr>
<td>NumType</td>
<td>numeral type</td>
</tr>
<tr>
<td>Number</td>
<td>number</td>
</tr>
<tr>
<td>Person</td>
<td>person</td>
</tr>
<tr>
<td>Poss</td>
<td>possessive</td>
</tr>
<tr>
<td>PronType</td>
<td>pronominal type</td>
</tr>
<tr>
<td>Reflex</td>
<td>reflexive</td>
</tr>
<tr>
<td>Tense</td>
<td>tense</td>
</tr>
<tr>
<td>VerbForm</td>
<td>form of verb or deverbative</td>
</tr>
<tr>
<td>Voice</td>
<td>voice</td>
</tr>
</tbody>
</table>

άντρα: ἀνδρᾶς + NOUN + Anim:Acc:Masc:Sing
Principles: Dependencies

- Content words reign supreme -- parallelism
  - Dependencies are between content words
  - Not indirectly mediated via function words
- Function words are demoted
  - Function words rarely have modifiers
- Some exceptions
## Dependencies

### Core vs. oblique

<table>
<thead>
<tr>
<th>Core dependents of clausal predicates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal dep</td>
</tr>
<tr>
<td>noun</td>
</tr>
<tr>
<td>noun</td>
</tr>
</tbody>
</table>

| Nominal dep | Predicate dep | Modifier word |
| noun        | verb          | adverb        |
| noun        | verb          | adjective     |

<table>
<thead>
<tr>
<th>Non-core dependents of clausal predicates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal dep</td>
</tr>
<tr>
<td>noun</td>
</tr>
<tr>
<td>noun</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Special clausal dependents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal dep</td>
</tr>
<tr>
<td>noun</td>
</tr>
<tr>
<td>noun</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Compounding and unanalyzed</th>
</tr>
</thead>
<tbody>
<tr>
<td>compound</td>
</tr>
<tr>
<td>word</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coordination</th>
</tr>
</thead>
<tbody>
<tr>
<td>noun1</td>
</tr>
<tr>
<td>preposition</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Loose joining relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>list</td>
</tr>
<tr>
<td>adverb</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sentence head</th>
<th>Unspecified dependency</th>
</tr>
</thead>
<tbody>
<tr>
<td>noun</td>
<td></td>
</tr>
</tbody>
</table>
Dependencies - Examples

“Case”

Possessives

Pre/Postpositional variants

Case markers

Evalita, December 2014
Dependencies - Examples

“Cop”

Parallel with dropped copula
Function words with modifiers

MWE

Conjunctions

Negation & adverbial modifiers

Elision
**Dependencies - Examples**

**Dislocated**

E.g., topic intro

**Compound**

**Name**

**MWE**

**Remnant**

E.g., ellipsis
Community Effort

• By early 2015 want guidelines and data for ~20 languages
• More than a dozen groups involved
Thanks