Towards Deep Universal Dependencies

Kira Droganova, Daniel Zeman

{droganova,zeman}@ufal.mff.cuni.cz
http://universaldependencies.org/
Multiple Layers of Dependencies

- Form
  - Surface syntax
  - Deep syntax
  - Semantics

Meaning
“The document suggests that this contract affect the persons who make the unemployment lists swell.”
A similar technique is almost impossible to apply to other crops such as cotton, soybean and rice.

similar technique be almost possible #Benef apply #Gen other crop such_as cotton soybean and rice
The thrift holding company said it expects to obtain regulatory approval and complete the transaction by year-end.
"The structural system package should probably be declared unsuccessful."

Le lot gros Œuvre devra probablement être déclaré infructueux.

The package structural works will have to probably be declared unsuccessful.
Universal Dependencies

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

- **MTT:** Russian, English, Spanish, French
- **FGD:** Czech, English
- **PropBank:** English, Arabic, Chinese, Finnish, Hindi, Urdu, Persian, Portuguese, Turkish, German, French
- **AMR:** English, Chinese, Portuguese, Korean, Vietnamese, Spanish, French, German
- **Sequoia:** French
Multilingual Annotation

- **MTT**: Russian, English, Spanish, French
- **FGD**: Czech, English
- **PropBank**: English, Arabic, Chinese, Finnish, Hindi, Urdu, Persian, Portuguese, Turkish, German, French
- **AMR**: English, Chinese, Portuguese, Korean, Vietnamese, Spanish, French, German
- **Sequoia**: French

**Enhanced UD**: Arabic, Bulgarian, Czech, Dutch, English, Estonian, Finnish, Italian, Latvian, Lithuanian, Polish, Russian, Slovak, Swedish, Tamil, Ukrainian
Basic Universal Dependencies: 82 Languages and Growing

I.-E.
- Armenian, Ancient Greek, Greek, Breton, Irish, Welsh
  - Germanic: Afrikaans, Danish, Dutch, English, Faroese, German, Gothic, Norwegian, Swedish
  - 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, Marathi, Sanskrit, Urdu

Uralic: Erzya, Estonian, Finnish, Hungarian, Karelian, Komi, Sámi

Dravidian: Tamil, Telugu

Turkic: Kazakh, Turkish, Uyghur

Af.-As.: Akkadian, Amharic, Arabic, Assyrian, Coptic, Hebrew, Maltese

Sino-Tibetan: 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

Daniel Zeman (ÚFAL MFF UK)
Two-Speed Approach

- Automatic part: derived from basic UD
- Optional manual extras
A Mountain of Work

- Work in progress
  - Only the automatic part now

- WANTED: Feedback
A Mountain of Work

- Work in progress
- WANTED: Feedback
- Automatic part: derived from basic UD
- Optional manual extras

Automatic vs. Manual Annotation

- Automatically derived from UD:
  - Enhanced Universal Dependencies
    - Grammatical coreference (partially)
    - Ellipsis: gapping
  - Normalize syntactic alternations (cf. Candito et al. 2017)
  - Ellipsis: pro-drop

- Manual or with extra resources:
  - Frames, semantic roles
  - Textual coreference
  - Everything else…
  - … and improve the automatic part above
Normalization of Syntactic Alternations

He killed the dragon

ARG1

ARG2

The dragon was killed by him

The dragon that was killed

Daniel Zeman (ÚFAL MFF UK)
Normalization of Syntactic Alternations

He killed the dragon

ARG1

The dragon was killed by him

ARG2

ARG1
Normalization of Syntactic Alternations

He killed the dragon

ARG1

(ARG2)

The dragon was killed by him

ARG2

(ARG1)

She made him kill the dragon

ARG1

(ARG2)

Daniel Zeman (ÚFAL MFF UK)

Towards Deep UD

Paris, 28.8.2019
Normalization of Syntactic Alternations

He killed the dragon.

The dragon was killed by him.

She made him kill the dragon.

The dragon that was killed
Numbered Arguments

- Like in PropBank. Numbers can be mapped to semantic roles if we have a valency dictionary.

- Degree of salience of arguments derived from surface syntax:
  - Subject of active clause ⇒ ARG1
  - Direct object of active clause ⇒ ARG2
  - Indirect object of active clause ⇒ ARG3
  - Subject of passive clause ⇒ ARG2
  - etc…
Predicate Identifiers

- They *could be* sense/frame identifiers
- But now we just take *lemmas*

**Exception:**
- Germanic phrasal verbs: *come_up*
- Inherently reflexive verbs: `[cs] smát_se “laugh”`
- Other compound verbs (incl. light & serial verb constructions)
Enhanced UD: Five Enhancements

- Null nodes for gapping (12 treebanks)
- Dependency propagation in coordination (22 treebanks)
- External subjects of controlled predicates (12 treebanks)
- Cyclic dependencies to/from relative clauses (9 treebanks)
- Enhanced dependency labels (case information) (10 treebanks)

- All 5 types: 6 treebanks, 3 languages
- At least 1 type: 24 treebanks, 16 languages
- Only basic UD: 122 treebanks
Enhanced UD: Five Enhancements

- Null nodes for **gapping** (12 treebanks)
- Dependency propagation in **coordination** (22 treebanks)
- External subjects of **controlled predicates** (12 treebanks)
- Cyclic dependencies to/from **relative clauses** (9 treebanks)
- Enhanced dependency labels (case information) (10 treebanks)

- All 5 types: 6 treebanks, 3 languages
- At least 1 type: 24 treebanks, 16 languages
- Only basic UD: 122 treebanks

**We apply Stanford Enhancer to all UD treebanks.**
Our “Enhanced Plus”

- Enhanced UD help us identify more predicate-argument relations
- But some patterns are still not handled…

- Adverbial infinitives
  - *They will meet to discuss a contract.*
  - *But not ccomp infinitives: He recommended to replace the tyres.*
Our “Enhanced Plus”

- Enhanced UD help us identify more predicate-argument relations
- But some patterns are still not handled…

- Adverbial **infinitives**
  - *They will meet* to discuss a contract.
  - But not **c-comp** infinitives: *He recommended to replace the tyres.*

- Adverbial **converbs (gerunds)**
  - *Terrorists detonated a bomb killing* five people.
Our “Enhanced Plus”

- Enhanced UD help us identify more predicate-argument relations
- But some patterns are still not handled…

- Adverbial **infinitives**
  - They will meet to discuss a contract.
  - But not **ccomp** infinitives: He recommended to replace the tyres.

- Adverbial **converbs (gerunds)**
  - Terrorists detonated a bomb killing five people.

- Attributive **participles**
  - The shares reflected on your statement.
Summary and Next Steps

- **Deep UD 2.4** ([http://hdl.handle.net/11234/1-3022](http://hdl.handle.net/11234/1-3022))
- 121 treebanks of 73 languages
- Enhanced graphs in all treebanks
- (Enhanced Plus: infinitives, gerunds, participles)
- Normalized active-passive

Can be regenerated after each UD release
Summary and Next Steps

- Deep UD 2.4 (http://hdl.handle.net/11234/1-3022)
- 121 treebanks of 73 languages
- Enhanced graphs in all treebanks
  (Enhanced Plus: infinitives, gerunds, participles)
- Normalized active-passive

- Can be regenerated after each UD release
- Evaluate precision and recall (no gold standard yet)
- Test mapping to a valency dictionary
- Oblique arguments?
- Other alternations than passives?
- Non-verbal predicates
- …
Future: Link Arguments to Frame Dictionaries

EngVallex
(http://lindat.mff.cuni.cz/services/EngVallex/EngVallex.html)

- Select the correct frame of the verb
- Map observed arguments to frame slots
  - Use their syntactic function
  - Use their morphological form

The dragon was killed by him

ARG2 PAT

ARG1 ACT

Daniel Zeman (ÚFAL MFF UK)
Thanks!
Merci!
Kate wants to go to Florida and Jane (wants) (go) to Europe.
Jane eats sweet apples and oranges
“And where to look for shampoo that works?”