

Towards Deep Universal Dependencies



FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

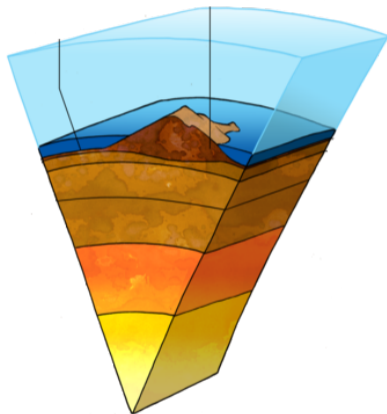


Kira Droганova, **Daniel Zeman**

{droganova,zeman}@ufal.mff.cuni.cz

<http://universaldependencies.org/>

Multiple Layers of Dependencies

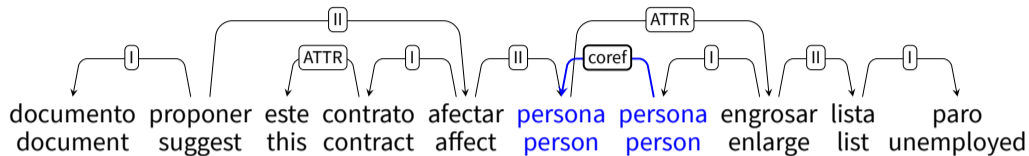


Form

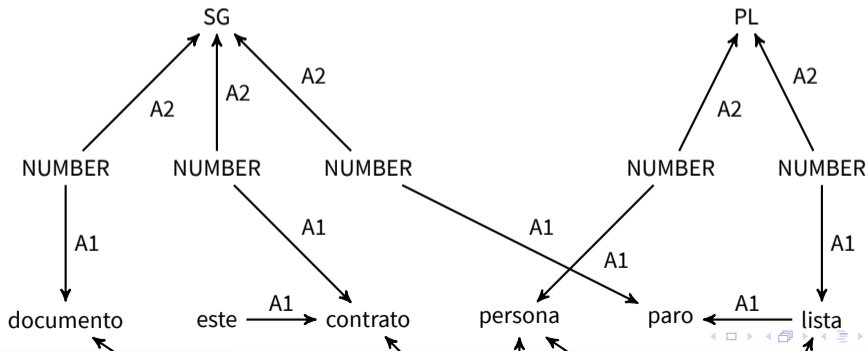
- Surface syntax
- Deep syntax
- Semantics

Meaning

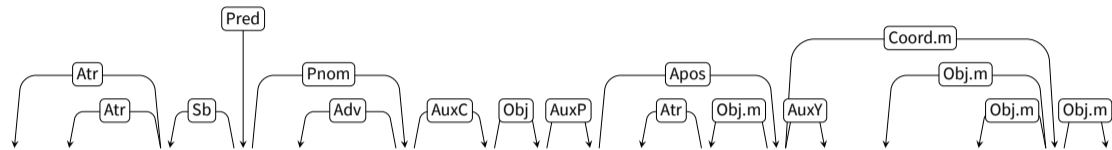
Meaning-Text Theory



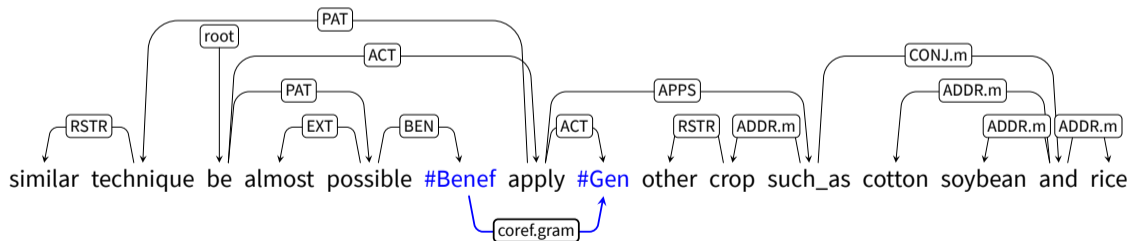
“The document suggests that this contract affect the persons who make the unemployment lists swell.”



Functional Generative Description (Prague Tectogrammatics)

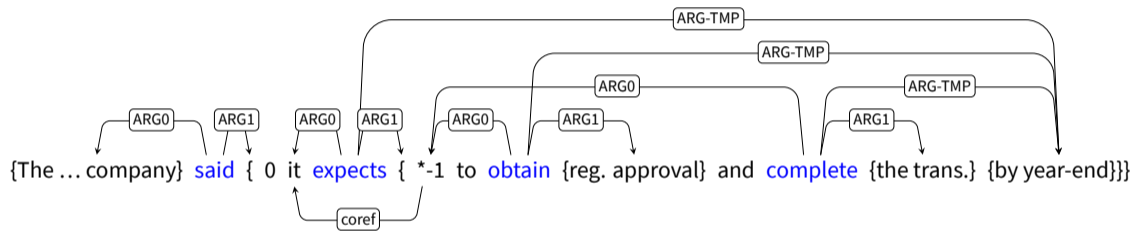


A similar technique is almost impossible to apply to other crops such as cotton, soybean and rice



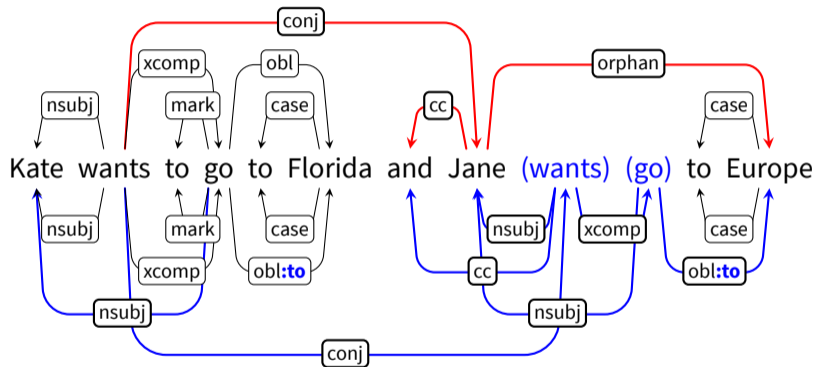
similar technique be almost possible #Benef apply #Gen other crop such_as cotton soybean and rice

Proposition Banks



“The thrift holding company said it expects to obtain regulatory approval and complete the transaction by year-end.”

Enhanced Universal Dependencies



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: 89 Languages and Growing

- I.-E.:  Armenian,  Ancient Greek,  Greek,  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.:  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

Two-Speed Approach

- Automatic part: derived from basic UD
- Optional manual extras

A Mountain of Work



A Mountain of Work

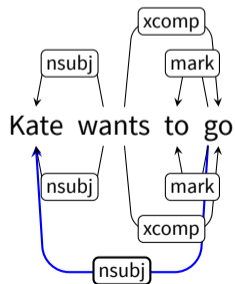
- Work in progress
 - ▶ Only the automatic part now
- WANTED: Feedback

A Mountain of Work

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

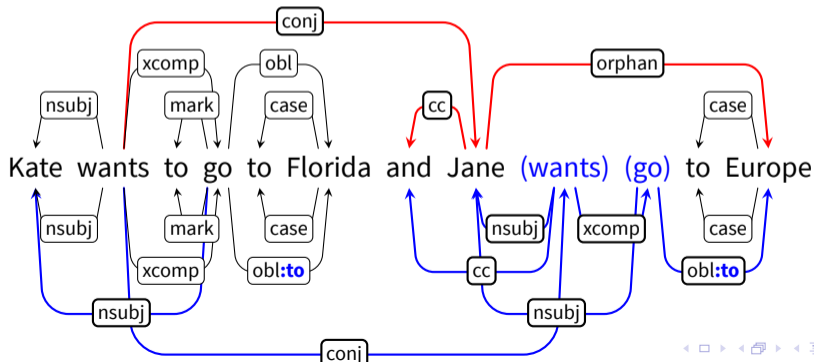
Automatic vs. Manual Annotation

- Automatically derived from UD:
 - ▶ Enhanced Universal Dependencies
 - ★ Grammatical coreference (partially)



Automatic vs. Manual Annotation

- Automatically derived from UD:
 - ▶ Enhanced Universal Dependencies
 - ★ Grammatical coreference (partially)
 - ★ Ellipsis: gapping



Automatic vs. Manual Annotation

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

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

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: 133 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: 133 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.*

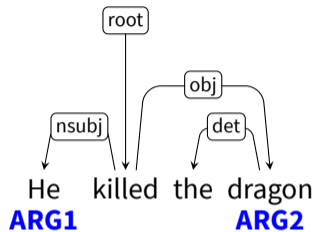
Our “Enhanced Plus”

- 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 `ccomp` infinitives: *He recommended **to replace** the tyres.*
- Adverbial **converbs (gerunds)**
 - ▶ *Terrorists detonated a bomb **killing** five people.*

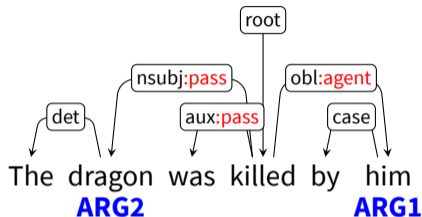
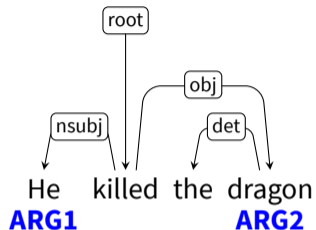
Our “Enhanced Plus”

- 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.*

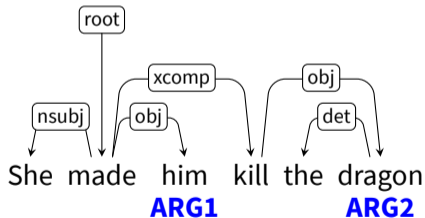
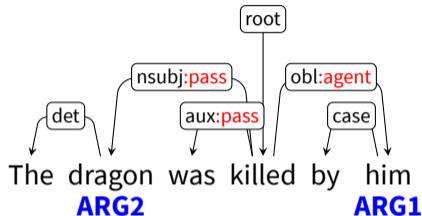
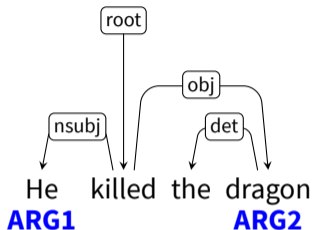
Normalization of Syntactic Alternations



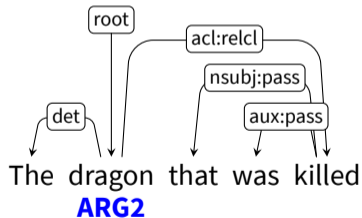
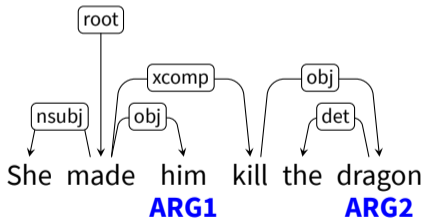
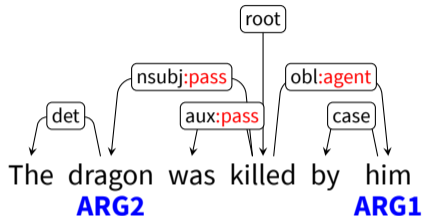
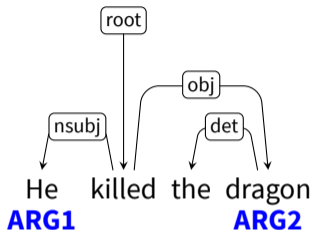
Normalization of Syntactic Alternations



Normalization of Syntactic Alternations



Normalization of Syntactic Alternations



Numbered Arguments

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

Numbered Arguments

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

	Deep UD	PropBank
<i>John broke the window.</i>	<i>John.ARG1 window.ARG2</i>	<i>John.ARG0 window.ARG1</i>
<i>The window was broken by John.</i>	<i>John.ARG1 window.ARG2</i>	<i>John.ARG0 window.ARG1</i>
<i>The window broke.</i>	<i>window.ARG1</i>	<i>window.ARG1</i>

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)

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
- **(Deep UD 2.5 coming soon!)**

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
- (Deep UD 2.5 coming soon!)

- 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

kill

kill ACT₀ PAT₀ ?MEANS₀

- "Kill it," he says.
- "Ford probably would try *trace* to kill the proposal by enlisting support from U.S. takeover-stock speculators and holding out the carrot of a larger bid later, said Stephen Reitman, European auto analyst at London brokers UBS Phillips & Drew.
- John killed Mary with a lead pipe, in the conservatory.

Corpus example(s):

Close [X]

pedt She **ACT** was untrained and, in one botched job **killed a PAT** client **PAT**.

pedt Dallas District Judge Jack Hampton had sparked calls for a judicial inquiry with his remarks to the press last December, two weeks after *-1 sentencing an 18-year-old defendant to 30 years in state prison **for *-2 killing** two homosexual **men PAT** in a city park.

pedt The judge was quoted *-1 as *-4 referring to the victims--PAT as "queers" and *-4 saying **they PAT** wouldn't have been killed *-2 "if they hadn't been cruising the streets *-3 picking up teenage boys."

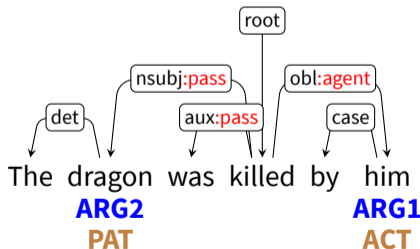
pedt Strong dividend growth, he says *T*-1, is "the black widow of valuation" -- a reference to the female **spiders--ACT** **that ACT** *T*-231 attract **males--PAT** and then **kill them PAT** after mating.

pedt Instead, Mr. Nixon reminded his host, Chinese President Yang Shangkun, that Americans haven't forgiven China's leaders for the military **assault--ACT** of June 3-4 **that ACT** *T*-241 **killed hundreds PAT, PAT** and perhaps **thousands PAT, PAT** of demonstrators.

EngVallex

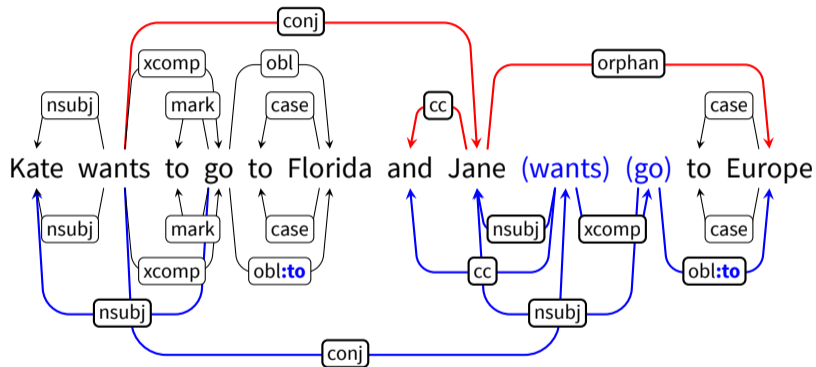
(<http://lindat.mff.cuni.cz/services/EngVallex/EngVallex.html>)

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

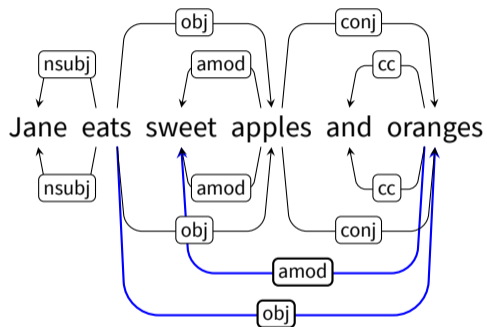


Thanks!
Tack!

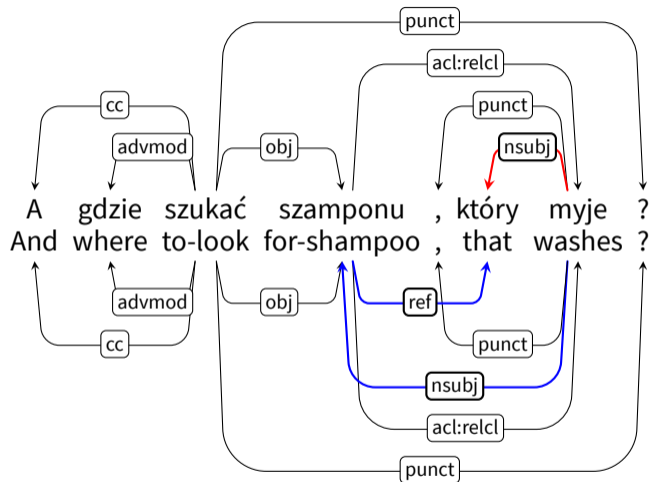
Enhanced UD



Enhanced UD



Enhanced UD



“And where to look for shampoo that works?”