Towards Deep Universal Dependencies

Kira Droganova
Charles University
Faculty of Mathematics and Physics
Praha, Czechia
droganova@ufal.mff.cuni.cz

Daniel Zeman
Charles University
Faculty of Mathematics and Physics
Praha, Czechia
zeman@ufal.mff.cuni.cz

Abstract

Many linguistic theories and annotation frameworks contain a deep-syntactic and/or semantic layer. While many of these frameworks have been applied to more than one language, none of them is anywhere near the number of languages that are covered in Universal Dependencies (UD). In this paper, we present a prototype of Deep Universal Dependencies, a two-speed concept where minimal deep annotation can be derived automatically from surface UD trees, while richer annotation can be added for datasets where appropriate resources are available. We release the Deep UD data in Lindat.

1 Introduction

Universal Dependencies (UD) (Nivre et al., 2016) annotation guidelines have become a de-facto standard for cross-linguistically comparable morphological and syntactic annotation. A significant factor in the popularity of UD is a steadily growing and heavily multilingual collection of corpora: release 2.4 (Nivre et al., 2019) contains 146 treebanks of 83 languages. The UD guidelines have been designed as surface-syntactic, although their emphasis on cross-linguistic parallelism sometimes leads to decisions that are normally associated with deeper, semantics-oriented frameworks (the primacy of content words and the second-class citizenship of function words may serve as an example).

Many theories and annotation frameworks have been proposed that contain a deep-syntactic, tectogrammatical, or semantic dependency layer; to name just a few: Meaning-Text Theory (Žolkovskij and Mel'čuk, 1965), Functional Generative Description (Sgall, 1967), the Proposition Bank (Kingsbury and Palmer, 2002), Sequoia (Candito and Seddah, 2012), or Abstract Meaning Representation (Banarescu et al., 2013). Names vary and so does the extent of 'deep' phenomena that are annotated; the common denominator is that these phenomena are closer to meaning on the meaning-form scale than anything we find in a typical surface-syntactic treebank. By definition, deep representation is more useful for natural language understanding (but it is also more difficult to obtain).

Many of the deep frameworks have been applied to more than one language, sometimes just to demonstrate that it is possible; but none of them is anywhere near the number of languages covered by UD.

UD itself contains a diffident attempt to provide deeper annotations, dubbed the Enhanced Universal Dependencies (Schuster and Manning, 2016). While it is a step in the right direction, it is just the first step: we argue that it should be possible to go deeper. Moreover, Enhanced UD is an optional extension, which is only available in a handful of treebanks (Table 1). Enhanced UD faces the same threat as the other deep frameworks mentioned above: more complex annotation requires more annotation effort, and semantic annotations are often coupled with huge lexical resources such as verb frame dictionaries. Therefore, it is less likely that sufficient manpower will be available to annotate data in a new language. Our principal question is thus the following: is it possible to create a multilingual data collection (and annotation guidelines) that will be as popular and widely used as UD, but deeper?

In our view, the key is to identify a subset of deep annotations that can be derived semi-automatically from surface UD trees, in acceptable quality. These annotations will not be as precise as if they were carefully checked by humans, but they will be available for (almost) all UD languages. More importantly, it will be possible to generate them for new UD languages and the deep extension will thus keep up with the growth of UD. For languages that have better resources available, one could convert them to the deep UD format and provide them instead of the corresponding semi-automatic annotation. Note that there are two dimensions along which a resource can be 'better'. It can provide the same type of annotation as the light, semi-automatic version, just verified by human annotators. But it may also provide additional types of annotations that cannot be obtained automatically. The Deep UD guidelines should thus cover a broad selection of phenomena that are annotated in popular semantic dependency frameworks.

The present paper reports on work in progress. We have prepared the first prototype of the semi-automatic Deep Universal Dependencies, based on UD release 2.4. The resource is available in the LINDAT/CLARIN repository (`http://hdl.handle.net/11234/1-3022`) under the same set of licenses as the underlying UD treebanks. In the following sections we describe what types of annotation this first version contains and how the annotation is derived from the surface trees; we also offer an outlook on possible future development.

2 Related Work

Manual semantic annotation is a highly time-consuming process, therefore a number of authors experimented with (semi-)automatic approaches to semantic annotation. Padó (2007) proposed a method that uses parallel corpora to project annotation to transfer semantic roles from English to resource-poorer languages. The experiment was conducted on an English-German corpus. Van der Plas et al. (2011) experimented with joint syntactic-semantic learning aiming at improving the quality of semantic annotations from automatic cross-lingual transfer. An alternative approach was proposed by Exner et al. (2016). Instead of utilizing parallel corpora, they use loosely parallel corpora where sentences are not required to be exact translations of each other. Semantic annotations are transferred from one language to another using sentences aligned by entities. The experiment was conducted using the English, Swedish, and French editions of Wikipedia. Akbik et al. (2015) described a two-stage approach to cross-lingual semantic role labeling (SRL) that was used to generate Proposition Banks for 7 languages. First, they applied a filtered annotation projection to parallel corpora, which was intended to achieve higher precision for a target corpus, even if containing fewer labels. Then they bootstrapped and retrained the SRL to iteratively improve recall without reducing precision. This approach was also applied to 7 treebanks from UD release 1.4.[1] However, the project seems to be stalled.

Mille et al. (2018) proposed the deep datasets that were used in the Shallow and Deep Tracks of the Multilingual Surface Realisation Shared Task (SR'18, SR'19). The Shallow Track datasets consist of unordered syntactic trees with all the word forms replaced with their lemmas; part-of-speech tags and the morphological information are preserved (available for 10 languages). The Deep Track datasets consist of trees that contain only content words linked by predicate-argument edges in the PropBank fashion (available for English, French and Spanish). The datasets were automatically derived from UD trees v.2.0. Gotham and Haug (2018) proposed an approach to deriving semantic representations from UD structures that is based on techniques developed for Glue semantics for LFG. The important feature of this approach is that it relies on language-specific resources as little as possible.

3 Enhanced Universal Dependencies

The Enhanced UD (Schuster and Manning, 2016)[2] represents a natural point of departure for us. UD v2 guidelines define five types of enhancements that can appear in treebanks released as part of UD. All the enhancements are optional and it is possible for a treebank to annotate one enhancement while ignoring the others. The enhanced representation is a directed graph but not necessarily a tree. It may contain 'null' nodes, multiple incoming edges and even cycles. The following enhancements are defined:

[1] `https://github.com/System-T/UniversalPropositions`

[2] While Schuster and Manning (2016) remains the most suitable reference for Enhanced UD to date, its publication predates the v2 UD guidelines and the proposals it contains are only partially compliant with the guidelines. See `https://universaldependencies.org/u-overview/enhanced-syntax.html` for the current version.

Null nodes for elided predicates. In certain types of ellipsis (*gapping* and *stripping*), multiple copies of a predicate are understood, each with its own set of arguments and adjuncts, but only one copy is present on the surface. Example: *Mary flies to Berlin and Jeremy [flies] to Paris.* The enhanced graph contains an extra node for each copy of the predicate that is missing on the surface. Note that the guidelines do not license null nodes for other instances of ellipsis, such as dropped subject pronouns in pro-drop languages.

Propagation of conjuncts. Coordination groups several constituents that together play one role in the superordinate structure. They are all equal, despite the fact that the first conjunct is formally treated as the head in the basic UD tree. For example, several coordinate nominals may act as subjects of a verb, but only the first nominal is actually connected with the verb via an `nsubj` relation. In the enhanced graph, this relation is propagated to the other conjuncts, i.e., each coordinate nominal is directly connected to the verb (in addition to the `conj` relation that connects it to the first conjunct). Likewise, there may be shared dependents that are attached to the first conjunct in the basic tree, but in fact they modify the entire coordination. Their attachment will be propagated to the other conjuncts, too. (Note that not all dependents of the first conjunct must be shared. Some of them may modify only the first conjunct, especially if the other conjuncts have similar dependents of their own.)

External subjects. Certain types of non-finite, 'open' clausal complements inherit their subject from the subject or the object of the matrix clause. Example: *Susan wants to buy a book.* In the basic tree, *Susan* will be attached as `nsubj` of *wants*, while there will be no subject dependent of *buy*. In contrast, the enhanced graph will have an additional `nsubj` relation between *buy* and *Susan*.

Relative clauses. The noun modified by a relative clause plays a semantic role in the frame of the subordinate predicate. In the basic UD tree, it is represented by a relative pronoun; however, in the enhanced graph it is linked from the subordinate predicate *instead* of the pronoun. (The pronoun is detached from the predicate and attached to the noun it represents, via a special relation `ref`.) This is the reason why enhanced graphs may contain cycles: in *The boy who lived*, there is an `acl:relcl` relation from *boy* to *lived*, and an `nsubj` relation from *lived* to *boy*.

Case information. The labels of certain dependency relations are augmented with case information, which may be an adposition, a morphological feature, or both. For example, the German prepositional phrase *auf dem Boden* "on the ground" may be attached as an oblique dependent (`obl`) of a verb in the basic tree. The enhanced label will be `obl:auf:dat`, reflecting that the phrase is in the dative case with the preposition *auf*. This information is potentially useful for semantic role disambiguation, and putting it to the label is supposed to make it more visible; nevertheless, its acquisition from the basic tree is completely deterministic, and there is no attempt to translate the labels to a language-independent description of meaning.

Several extensions of the enhanced representation have been proposed. The *enhanced++* graphs proposed by Schuster and Manning (2016) extend the set of ellipsis-in-coordination types where null nodes are added; they also suppress quantifying expressions in sentences like *a bunch of people are coming*.

Candito et al. (2017) define the *enhanced-alt* graphs, which neutralize syntactic alternations, that is, passives, medio-passives, impersonal constructions and causatives. They also suggest to annotate external arguments of other non-finite verb forms than just open infinitival complements and relative clauses: most notably, for participles, even if they are used attributively. Hence in *ceux embauchés en 2007* "those hired in 2007", *embauchés* heads a non-relative adnominal clause (`acl`) that modifies the nominal *ceux*, but at the same time *ceux* is attached as a passive subject (`nsubj:pass`) of *embauchés*.

4 Pre-existing Enhancing Tools

Enhanced UD contains information that cannot be derived automatically from the basic UD tree; additional human input is needed in order to fully disambiguate all situations. Nevertheless, it is believed that automatic 'enhancers' can get us relatively far. Schuster and Manning (2016) described and evaluated the Stanford Enhancer,[3] which is available as a part of the Stanford CoreNLP suite.

Nyblom et al. (2013) reported on the Turku Enhancer, a hybrid approach (consisting of rule-based heuristics and machine-learning components) to enhancing Stanford Dependencies of Finnish. The enhancements tackled were conjunct propagation, external subjects, and syntactic functions of relativizers; the first two are thus relevant also in Enhanced UD. Their system achieved F_1 score of 93.1; note however that labeled training data is needed for the approach to work.

Nivre et al. (2018) compares the Stanford Enhancer with an adapted version of the Turku Enhancer. They trained it on the Finnish labeled data, but in a delexicalized fashion (only non-lexical features were considered). The Turku Enhancer does not predict null nodes, and for external subjects it only considers subject control (or raising), but not object control. On the other hand, Stanford Enhancer only predicts core arguments as controllers while in some languages non-core dependents can control subjects too. Nevertheless, both enhancers are found usable for other languages, as shown on Swedish and Italian. The paper also evaluates an Italian-specific rule-based enhancer, which does not predict null nodes.

Candito et al. (2017) took a rule-based approach to produce their *enhanced-alt* graphs for French: they developed two sets of rules, using two different graph rewriting systems. However, they only focus on two of the five enhancements (external subjects and conjunct propagation), and they only do it for French. Some of their heuristics are very French-specific and they assume that information needed for disambiguation is available in the source annotation (which is the case of the Sequoia French treebank).

Several other UD treebanks come from sources where some enhanced annotation is available and can be converted to Enhanced UD. Bouma (2018) demonstrates how original annotations from the Alpino treebank can help enhance the Dutch UD treebanks. Patejuk and Przepiórkowski (2018) discuss conversion from an LFG treebank of Polish and note that not only there is more information than in basic UD, some information cannot be captured even by Enhanced UD. Another example is the distinction between private and shared dependents in coordination: for treebanks converted from Prague-style annotation (Arabic, Czech, Lithuanian, Slovak, Tamil), this distinction is readily available.

5 Data Preparation

The first version of Deep UD is based on UD release 2.4 (Nivre et al., 2019) but we intend to generate updates after each future UD release. While we foresee improved semantic annotation for some languages (based on additional lexical resources, for example), the current version is derived just from the annotation available in UD itself (though we use heuristics that may be language- or treebank-specific). UD 2.4 contains 146 treebanks of 83 languages. We exclude 6 treebanks that are distributed, for copyright reasons, as hollow annotations without the underlying text. We further exclude 19 treebanks with incomplete or non-existent lemmatization.[4] Consequently, our resource contains 121 treebanks of 73 languages.

We take enhanced UD graphs (Section 3) as the point of departure for deep UD. However, only a small fraction of the UD treebanks have some enhanced annotation, and if they do, then they often omit one or more of the five types of enhancements defined in the guidelines. There are 24 treebanks of 16 languages that have enhanced graphs (Table 1). We will refer to these enhanced graphs as *trusted enhanced annotations*. Some of them were converted from non-UD manual annotations, some were probably generated with the help of automatic enhancers, but at least they were overseen by the teams responsible for the given language.

We use the Stanford Enhancer[5] to generate enhanced graphs for corpora that lack them. For the six treebanks in Table 1 that contain trusted annotation of all five enhancement types, we take the trusted annotation. For the other 18 treebanks in the table, ideally we should merge the trusted annotation with the output of the enhancer so that all enhancement types are present. However, merging may not be trivial

[3]The Stanford UD Enhancer was adapted from an older tool that was designed to work with the Stanford Dependencies, a predecessor of UD.

[4]Note that we do not exclude some other treebanks where lemmas exist but have been assigned by a stochastic model instead of human annotators.

[5]The README file of the released data provides details on what version we used and how we ran it.

Language	Treebank	Gapping	Coord	XSubj	RelCl	CaseDeprel
Arabic	PADT		yes			
Bulgarian	BTB		yes	yes	yes	yes
Czech	CAC		yes			
Czech	FicTree		yes			
Czech	PDT		yes			
Dutch	Alpino	yes	yes	yes	yes	yes
Dutch	LassySmall	yes	yes	yes	yes	yes
English	EWT	yes	yes	yes	yes	yes
English	PUD	yes	yes	yes	yes	yes
Estonian	EWT	yes				
Finnish	PUD	yes	yes			
Finnish	TDT	yes	yes	yes		
Italian	ISDT		yes	yes	yes	yes
Latvian	LVTB	yes	yes	yes		yes
Lithuanian	ALKSNIS		yes			
Polish	LFG		yes	yes		yes
Polish	PDB		yes			
Polish	PUD		yes			
Russian	SynTagRus	yes				
Slovak	SNK		yes			
Swedish	PUD	yes	yes	yes	yes	yes
Swedish	Talbanken	yes	yes	yes	yes	yes
Tamil	TTB		yes			
Ukrainian	IU	yes	yes	yes	yes	

Table 1: Overview of enhanced annotations in UD 2.4 treebanks. Gapping: there are empty nodes representing elided predicates. Coord: dependencies (both incoming and outgoing) are propagated to all conjuncts. XSubj: higher argument is linked as the subject of a controlled verb. RelCl: nominal modified by a relative clause is linked as argument or adjunct in that clause. CaseDeprel: case markers are added to the dependency labels of adverbial and oblique dependents.

in sentences where multiple enhancement types interact, and we leave it for future work. In the current version, the enhanced graphs in these 18 treebanks are replaced by the output of the Stanford Enhancer.

Note that using the Stanford Enhancer does not guarantee that the resulting annotation identifies all five types of enhancements—even if the phenomenon exists in the language and the treebank is large enough to provide examples. Identification of relative clauses relies on a language-specific list of relative pronouns and on the optional dependency label `acl:relcl`, but some treebanks use `acl` instead. Gapping, besides being relatively rare, is not annotated properly in the basic representation of some UD languages. Consequently, only 58 enhanced treebanks have some null nodes (gapping) and only 54 treebanks have edges specific to relative-clause enhancements. Most treebanks have the other three types; a remarkable exception is Japanese where the three treebanks have only one enhancement type, namely the case-augmented dependency relations. 37 treebanks feature all five types. We plan to expand the relative clause annotation to other treebanks in the future; listing relative pronouns (a closed class) is quite feasible, and we can utilize the morphological feature `PronType=Rel` where available.

6 Delving Deeper

There are numerous phenomena that various semantic frameworks strive to capture. Without precluding any of them from future versions of Deep UD, we believe that the core of sentence understanding is its predicate-argument structure. We start with verbal predicates and identify their arguments, if present in the same sentence. We number the arguments roughly reflecting their decreasing salience and making

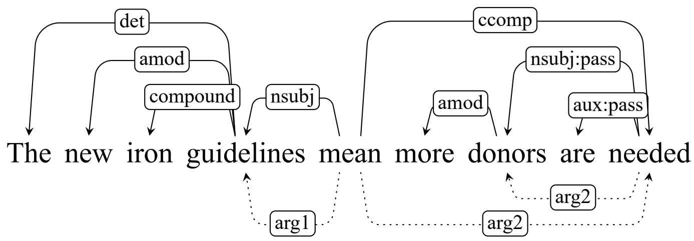

Figure 1: An example of a deep graph for the English sentence *The new iron guidelines mean more donors are needed.*

sure that for the same predicate (sense), the argument with a particular semantic role will always get the same label/number, regardless the syntactic environment. That means that we have to neutralize valency-changing operations such as passivization; here we are very close to the *enhanced-alt* representation proposed by Candito et al. (2017). For example, in *George killed the dragon* as well as in *The dragon was killed by George*, *George* will be `arg1` and *the dragon* will be `arg2`. We do not label the actual semantic roles (i.e., agent / actor / killer for *George* and patient / killed for *the dragon*) directly in the text. Instead, the predicate instance can be linked to a frame dictionary (if available) where the corresponding frame will provide interpretation of the numbered arguments. Linking of frame instances to dictionary frames will not be trivial and the concrete approach will depend on the language and on the nature of the target lexical resource. Valency frame dictionaries often contain information on morphological and syntactic properties of the arguments. A verbal lemma will typically correspond to several (sometimes dozens of) different frames. Sometimes the forms of the arguments (their morphological case, preposition etc.) will narrow down the search; but full disambiguation may not be possible without a statistical model or a human annotator. Once we have the correct frame, identification of individual arguments is (again) just matching their properties against those specified by the frame.

We follow the CoNLL-U Plus file format[6] with two new columns: DEEP:PRED and DEEP:ARGS. These columns contain annotation we add on top of Enhanced UD; without them, the file is still a valid CoNLL-U file. The value in DEEP:PRED identifies the predicate. It can be a reference to a particular sense (frame) in a dictionary but we currently use just the lemma of the verb, possibly augmented with other lemmas if it is a compound verb (e.g. Germanic phrasal verbs such as *come up*). The value in DEEP:ARGS points to the head nodes of subtrees that represent the arguments. For example, `arg1:33|arg2:12,27` means that the most salient argument (possibly the agent) is headed by node 33, while the second most salient argument (possibly the patient) is coordination and the conjuncts are headed by nodes 12 and 27, respectively. See Figure 1 for an example of a deep graph.

Thanks to Enhanced UD, the annotation resolves some instances of grammatical coreference (Zikánová et al., 2015), i.e., situations where one node serves as an argument of multiple verbs, and it can be inferred from the grammatical rules of the language. On the other hand, the current version does not attempt to address textual coreference, e.g., a personal pronoun that is coreferential with a noun. Arguably, textual coreference cannot be resolved without a human annotator or a trained model.

Some arguments are not accessible through Enhanced UD; similar to Candito et al. (2017), we are experimenting with heuristics that yield additional enhanced dependencies for non-finite verbs:

Infinitives that are not `xcomp`. They can be ordinary clausal complements (`ccomp`) and then we cannot identify their subject, as in Dutch: *Zijlaard adviseerde te gokken op de sprint* (lit. *Zijlaard advised to bet on the sprint*) "Zijlaard advised betting on the sprint". But they can be also adverbial clauses (`advcl`), or adnominal clauses (`acl`), if the main clause's predicate is a light verb with a noun, as in Dutch: *had moeite om zich te concentreren* (lit. *had trouble so himself to concentrate*) "struggled to concentrate". The infinitive *concentreren* "to concentrate" in this case works similarly to an `xcomp`, that is, it should inherit the subject from the matrix clause.

[6] `https://universaldependencies.org/ext-format.html`

Participles. An attributively used participle modifies a noun. If it were a relative clause, the enhanced graph would identify the noun as the "subject" argument of the participle; but it is an `amod` rather than a clause, and no external subject relation is present. A Dutch example: *de afgelopen week* (lit. *the expired week*) "last week". We add a heuristic that participles attached as `amod` shall take the modified noun as their argument; note that we need to distinguish active and passive participles in order to find out whether the noun is argument 1 or 2. Currently we only look for the morphological feature `Voice=Pass` but it is not always available, and some verb forms can be used both in active and passive clauses. Consider English: *the shares reflected on your statement*; *reflected* is used as a passive participle but `Voice=Pass` is not present, it is just a "past participle" without any voice feature. We may need to estimate whether a verb is transitive, and if it is, the participle will be considered passive, otherwise it will be considered active. Nevertheless, no such heuristic was applied to the current version of the data.

Converbs (gerunds). English: *X did Y..., killing several people*. The syntactic annotation does not tell us that X is the argument 1 of *killing*. We work with the hypothesis that a gerund or converb attached as `advcl` inherits the subject of the matrix clause. This is a rule at least in some languages but we have yet to evaluate to what extent the rule may be universal.

Language-specific heuristics. A number of heuristics will be needed that are language- or even treebank-specific. For example, passivization of English ditransitive clauses promotes the indirect object rather than the direct object *(what I was asked)*.[7] Therefore, if there is a direct object in a passive clause, the subject should be considered argument 3 and not 2.

7 Conclusion and Outlook

We presented a prototype of Deep Universal Dependencies, a deep-syntactic annotation layer that can be derived semi-automatically from surface UD graphs. Our plan is to accommodate rich semantic annotations in languages where necessary resources are available, and automatically generate the core part for other languages after each UD release. Our contribution at the current stage is threefold: 1. While UD releases still contain Enhanced UD only for a few treebanks, we make sure that enhanced graphs are available everywhere; 2. to find more arguments, we do additional enhancements (infinitives, gerunds, participles) internally but we do not show them in the enhanced graphs so that the graphs stay within the current guidelines; 3. we normalize diathesis and show the numbered arguments (canonical subject and object in the terms of Candito et al. (2017)).

The list of possible future directions is much longer than we can accommodate in a short paper; for instance, we want to take advantage of oblique argument marking in treebanks where it is available, improve recognition of passives and other diathesis alternations, or implement other enhancements from Schuster and Manning (2016)'s *enhanced++*. Nevertheless, the most important next step is to evaluate the quality of the generated annotation (both the output of the Stanford Enhancer and the additional heuristics we applied to the enhanced graphs). Since there is no gold-standard labeled data suitable for such evaluation, we will have to manually inspect random samples of the output, or compare the predicate-argument patterns with existing valency dictionaries (in languages where they exist).

Acknowledgements

This work is partially supported by the GA UK grant 794417, the SVV project number 260 453, and the grant no. LM2015071 of the Ministry of Education, Youth and Sports of the Czech Republic.

We also wish to thank the three anonymous reviewers for their valuable comments.

References

Alan Akbik, Marina Danilevsky, Yunyao Li, Shivakumar Vaithyanathan, Huaiyu Zhu, et al. 2015. Generating high quality proposition banks for multilingual semantic role labeling. In Proceedings of the 53rd Annual Meeting of

[7] Of course, one could then question whether *I* is an indirect object in the active clause if it can be promoted by passivization; here we follow the actual approach of the English UD treebanks.

the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), volume 1, pages 397–407.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract Meaning Representation for Sembanking. In Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse, pages 178–186.

Gosse Bouma. 2018. Comparing two methods for adding enhanced dependencies to UD treebanks. In Proceedings of the 17th International Workshop on Treebanks and Linguistic Theories (TLT 2018), pages 17–30, Oslo, Norway. Linköping Electronic Conference Proceedings.

Marie Candito and Djamé Seddah. 2012. Le corpus Sequoia: annotation syntaxique et exploitation pour l'adaptation d'analyseur par pont lexical. In TALN 2012-19e conférence sur le Traitement Automatique des Langues Naturelles.

Marie Candito, Bruno Guillaume, Guy Perrier, and Djamé Seddah. 2017. Enhanced UD dependencies with neutralized diathesis alternation. In Proceedings of the Fourth International Conference on Dependency Linguistics (Depling 2017), pages 42–53, Pisa, Italy.

Peter Exner, Marcus Klang, and Pierre Nugues. 2016. Multilingual supervision of semantic annotation. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 1007–1017.

Matthew Gotham and Dag Trygve Truslew Haug. 2018. Glue semantics for Universal Dependencies. In Miriam Butt and Tracy Holloway King, editors, Proceedings of the LFG'18 Conference, pages 208–226, Wien, Austria. CSLI Publications.

Paul Kingsbury and Martha Palmer. 2002. From TreeBank to PropBank. In LREC, pages 1989–1993.

Simon Mille, Anja Belz, Bernd Bohnet, and Leo Wanner. 2018. Underspecified universal dependency structures as inputs for multilingual surface realisation. In Proceedings of the 11th International Conference on Natural Language Generation, pages 199–209.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. 2016. Universal Dependencies v1: A Multilingual Treebank Collection. In Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016), pages 1659–1666, Paris, France. European Language Resources Association.

Joakim Nivre, Paola Marongiu, Filip Ginter, Jenna Kanerva, Simonetta Montemagni, Sebastian Schuster, and Maria Simi. 2018. Enhancing universal dependency treebanks: A case study. In Proceedings of the Second Workshop on Universal Dependencies (UDW 2018), pages 102–107, Bruxelles, Belgium. Association for Computational Linguistics.

Joakim Nivre, Mitchell Abrams, Željko Agić, Lars Ahrenberg, Gabrielė Aleksandravičiūtė, Lene Antonsen, Katya Aplonova, Maria Jesus Aranzabe, Gashaw Arutie, Masayuki Asahara, Luma Ateyah, Mohammed Attia, Aitziber Atutxa, Liesbeth Augustinus, Elena Badmaeva, Miguel Ballesteros, Esha Banerjee, Sebastian Bank, Verginica Barbu Mititelu, Victoria Basmov, John Bauer, Sandra Bellato, Kepa Bengoetxea, Yevgeni Berzak, Irshad Ahmad Bhat, Riyaz Ahmad Bhat, Erica Biagetti, Eckhard Bick, Agnė Bielinskienė, Rogier Blokland, Victoria Bobicev, Loïc Boizou, Emanuel Borges Völker, Carl Börstell, Cristina Bosco, Gosse Bouma, Sam Bowman, Adriane Boyd, Kristina Brokaitė, Aljoscha Burchardt, Marie Candito, Bernard Caron, Gauthier Caron, Gülşen Cebiroğlu Eryiğit, Flavio Massimiliano Cecchini, Giuseppe G. A. Celano, Slavomír Čéplö, Savas Cetin, Fabricio Chalub, Jinho Choi, Yongseok Cho, Jayeol Chun, Silvie Cinková, Aurélie Collomb, Çağrı Çöltekin, Miriam Connor, Marine Courtin, Elizabeth Davidson, Marie-Catherine de Marneffe, Valeria de Paiva, Arantza Diaz de Ilarraza, Carly Dickerson, Bamba Dione, Peter Dirix, Kaja Dobrovoljc, Timothy Dozat, Kira Droganova, Puneet Dwivedi, Hanne Eckhoff, Marhaba Eli, Ali Elkahky, Binyam Ephrem, Tomaž Erjavec, Aline Etienne, Richárd Farkas, Hector Fernandez Alcalde, Jennifer Foster, Cláudia Freitas, Kazunori Fujita, Katarína Gajdošová, Daniel Galbraith, Marcos Garcia, Moa Gärdenfors, Sebastian Garza, Kim Gerdes, Filip Ginter, Iakes Goenaga, Koldo Gojenola, Memduh Gökırmak, Yoav Goldberg, Xavier Gómez Guinovart, Berta González Saavedra, Matias Grioni, Normunds Grūzītis, Bruno Guillaume, Céline Guillot-Barbance, Nizar Habash, Jan Hajič, Jan Hajič jr., Linh Hà Mỹ, Na-Rae Han, Kim Harris, Dag Haug, Johannes Heinecke, Felix Hennig, Barbora Hladká, Jaroslava Hlaváčová, Florinel Hociung, Petter Hohle, Jena Hwang, Takumi Ikeda, Radu Ion, Elena Irimia, Ọlájídé Ishola, Tomáš Jelínek, Anders Johannsen, Fredrik Jørgensen, Hüner Kaşıkara, Andre Kaasen, Sylvain Kahane, Hiroshi Kanayama, Jenna Kanerva, Boris Katz, Tolga Kayadelen, Jessica Kenney, Václava Kettnerová, Jesse Kirchner, Arne Köhn, Kamil Kopacewicz, Natalia Kotsyba, Jolanta Kovalevskaitė, Simon Krek, Sookyoung Kwak, Veronika Laippala, Lorenzo Lambertino, Lucia Lam, Tatiana Lando, Septina Dian Larasati, Alexei Lavrentiev, John Lee, Phương Lê Hồng, Alessandro Lenci, Saran Lertpradit, Herman Leung, Cheuk Ying Li, Josie Li, Keying Li, KyungTae Lim, Yuan Li, Nikola Ljubešić, Olga Loginova, Olga Lyashevskaya, Teresa Lynn, Vivien Macketanz, Aibek Makazhanov, Michael Mandl, Christopher Manning, Ruli Manurung, Cătălina Mărănduc, David Mareček, Katrin Marheinecke, Héctor Martínez Alonso, André Martins, Jan Mašek, Yuji Matsumoto, Ryan McDonald, Gustavo Mendonça, Niko Miekka, Margarita Misirpashayeva, Anna Missilä, Cătălin Mititelu, Yusuke Miyao, Simonetta Montemagni, Amir More, Laura Moreno Romero, Keiko Sophie Mori, Tomohiko Morioka, Shinsuke Mori, Shigeki Moro, Bjartur Mortensen, Bohdan Moskalevskyi, Kadri Muischnek, Yugo Murawaki, Kaili Müürisep, Pinkey Nainwani, Juan Ignacio Navarro Horñiacek, Anna Nedoluzhko, Gunta Nešpore-Bērzkalne, Lương Nguyễn Thị, Huyền Nguyễn Thị Minh, Yoshihiro Nikaido, Vitaly Nikolaev, Rattima Nitisaroj, Hanna Nurmi, Stina Ojala, Adédayọ̀ Olúòkun, Mai Omura, Petya Osenova, Robert Östling, Lilja Øvrelid, Niko Partanen, Elena Pascual, Marco Passarotti, Agnieszka Patejuk, Guilherme Paulino-Passos, Angelika Peljak-Łapińska, Siyao Peng, Cenel-Augusto Perez, Guy Perrier, Daria Petrova, Slav Petrov, Jussi Piitulainen, Tommi A Pirinen, Emily Pitler, Barbara Plank, Thierry Poibeau, Martin Popel, Lauma Pretkalniņa, Sophie Prévost, Prokopis Prokopidis, Adam Przepiórkowski, Tiina Puolakainen, Sampo Pyysalo, Andriela Rääbis, Alexandre Rademaker, Loganathan Ramasamy, Taraka Rama, Carlos Ramisch, Vinit Ravishankar, Livy Real, Siva Reddy, Georg Rehm, Michael Rießler, Erika Rimkutė, Larissa Rinaldi, Laura Rituma, Luisa Rocha, Mykhailo Romanenko, Rudolf Rosa, Davide Rovati, Valentin Roșca, Olga Rudina, Jack Rueter, Shoval Sadde, Benoît Sagot, Shadi Saleh, Alessio Salomoni, Tanja Samardžić, Stephanie Samson, Manuela Sanguinetti, Sarah McGuinness, Abigail Walsh, Dage Särg, Baiba Saulīte, Yanin Sawanakunanon, Nathan Schneider, Sebastian Schuster, Djamé Seddah, Wolfgang Seeker, Mojgan Seraji, Mo Shen, Atsuko Shimada, Hiroyuki Shirasu, Muh Shohibussirri, Dmitry Sichinava, Natalia Silveira, Maria Simi, Radu Simionescu, et al. 2019. Universal dependencies 2.4. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Jenna Nyblom, Samuel Kohonen, Katri Haverinen, Tapio Salakoski, and Filip Ginter. 2013. Predicting conjunct propagation and other extended Stanford Dependencies. In Proceedings of the Second International Conference on Dependency Linguistics (DepLing 2013), pages 252–261, Praha, Czechia. Matfyzpress.

Sebastian Padó. 2007. Cross-lingual annotation projection models for role-semantic information. Saarland University.

Agnieszka Patejuk and Adam Przepiórkowski. 2018. From Lexical Functional Grammar to Enhanced Universal Dependencies. Instytut Podstaw Informatyki Polskiej Akademii Nauk, Warszawa, Poland.

Sebastian Schuster and Christopher D. Manning. 2016. Enhanced English Universal Dependencies: An Improved Representation for Natural Language Understanding Tasks. In Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016), Paris, France. European Language Resources Association.

Petr Sgall. 1967. Functional sentence perspective in a generative description. Prague Studies in Mathematical Linguistics, 2:203–225.

Lonneke Van der Plas, Paola Merlo, and James Henderson. 2011. Scaling up automatic cross-lingual semantic role annotation. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2, pages 299–304. Association for Computational Linguistics.

Šárka Zikánová, Eva Hajičová, Barbora Hladká, Pavlína Jínová, Jiří Mírovský, Anna Nedoluzhko, Lucie Poláková, Kateřina Rysová, Magdaléna Rysová, and Jan Václ. 2015. Discourse and Coherence. From the Sentence Structure to Relations in Text. Studies in Computational and Theoretical Linguistics. ÚFAL, Praha, Czechia.

Aleksandr K. Žolkovskij and Igor A. Mel’čuk. 1965. O vozmožnom metode i instrumentax semantičeskogo sinteza (on a possible method and instruments for semantic synthesis). Naučno-texničeskaja informacija, 5:23–28.