Comparing two methods for adding Enhanced Dependencies to UD treebanks

Gosse Bouma[1,2]
(1) Center for Language and Cognition
University of Groningen
(2) Center for Advanced Study
Norwegian Academy for Science and Letters
g.bouma@rug.nl

ABSTRACT
When adding enhanced dependencies to an existing UD treebank, one can opt for heuristics that predict the enhanced dependencies on the basis of the UD annotation only. If the treebank is the result of conversion from an underlying treebank, an alternative is to produce the enhanced dependencies directly on the basis of this underlying annotation. Here we present a method for doing the latter for the Dutch UD treebanks. We compare our method with the UD-based approach of Schuster et al. (2018). While there are a number of systematic differences in the output of both methods, it appears these are the result of insufficient detail in the annotation guidelines and it is not the case that one approach is superior over the other in principle.

KEYWORDS: Universal Dependencies, Enhanced Universal Dependencies, Dutch, Ellipsis, Coordination, Control.

1 Introduction

While universal dependencies (UD) have proven to be a convenient level of syntactic annotation for many applications, such as language typology (Futrell et al., 2015), stylistics (Wang and Liu, 2017), (cross-lingual) parser comparison (Zeman et al., 2017), relation extraction (Shwartz et al., 2016), and construction of word embeddings (Vulić, 2017), there is also the concern that it is not capturing all relevant syntactic distinctions. The majority of the current treebanks in the UD repository[1] are converted from underlying treebanks that were annotated with a language specific annotation scheme.[2] A frequent observation is that some of the distinctions in the underlying treebank cannot be preserved in the conversion to UD (Lipenkova and Souček, 2014; Pyysalo et al., 2015; Przepiórkowski and Patejuk, 2018). Enhanced universal dependencies (EUD) (see below) can help to diminish the loss of information, especially as it offers the means to annotate syntactic control relations and a more principled solution to annotating elliptical constructions.

EUD should also make it easier to obtain fully specified semantic interpretations from UD annotated text, as explored in Reddy et al. (2017) and Gotham and Haug (2018). Raising and control are syntactic phenomena that influence the semantic interpretation of a sentence. Computing the correct control relations on the basis of UD alone is cumbersome, and may require access to additional lexical resources. Similarly, the semantic interpretation of ellipsis requires identification of the elided material. The UD annotation of ellipsis avoids reference to empty elements, and thus cannot provide all information required for interpreting elliptical constituents. EUD does allow empty nodes, and thus provides a better basis for semantic interpretation.

1.1 Enhanced Universal Dependencies

Enhanced universal dependencies[3] are motivated by the need to obtain a syntactic annotation layer that is more suitable for semantic tasks, such as textual entailment, information extraction, or deriving a full semantic representation of an utterance. In particular, it proposes the following enhancements to basic UD:

- null nodes for elided predicates,
- propagation of conjuncts,
- additional subject relations for control and raising constructions,
- coreference in relative clause constructions.
- modifier labels that contain the preposition or other case-marking information
- conjunction labels that contain the coordinating word.

The EUD guidelines follow the proposals in Schuster and Manning (2016) and the proposed analysis of ellipsis in Schuster et al. (2017).

[1] `http://universaldependencies.org/`

[2] The documentation for 17 of the 102 treebanks in release v2.1 states that dependency relations were manually added, while 59 treebanks contain dependency relations that were created by converting an underlying (manually annotated) treebank.

[3] `http://universaldependencies.org/u/overview/enhanced-syntax.html`

It is clear that the EUD annotation of ellipsis, as well as the addition of control relations, cannot be done on the basis of information contained in the UD annotation only, and thus we present our method for doing this on the basis of the underlying annotation below. Propagation of conjuncts as defined in the guidelines appears to be possible on the basis of the basic UD annotation itself. Nevertheless, we include it in the discussion below, as our treatment is more comprehensive than what is suggested by the guidelines, and also because it interacts with the reconstruction of elided predicates and the addition of control relations. The analysis of relative clauses in EUD adds a dependency from the head of the relative clause to the antecedent noun. The label of this newly introduced dependency is identical to the label of the dependency from the head of the relative clause to the relative pronoun in UD. Thus, this extension can be computed on the basis of UD alone, as long as a relative pronoun is present. However, we do note some challenging cases in the next section. The last two modifications (fine-grained labels for *nmod*, *obl*, *acl*, *advcl*, and *cc*) can be done on the basis of UD annotation alone. We return to these in the section comparing the two conversion methods.

Technically, EUD differs from UD in that it contains null nodes. Furthermore, the annotation graph is no longer guaranteed to be a tree (e.g. the treatment of control means that an element can be the dependent of two predicates), may contain cycles (a consequence of the analysis of relative clauses), and may lack a root node.

It is an open question whether adding EUD to a corpus that already has been annotated with UD can be done automatically, using heuristics for those cases where information is lacking, or whether it requires additional supervision, either in the form of manual correction or information obtained from an underlying treebank that already has annotation in place for elided predicates, shared conjuncts, and raising and control. Here we compare these two possibilities for the Dutch treebanks by comparing the outcome of a rule-based conversion script that takes the underlying treebank annotation as input, with the method of Schuster et al. (2018). The latter paper presents an automatic method for converting from UD to EUD that was tested on English, Finnish, and Swedish. It can be adapted to other languages by providing word embeddings for that language and a list of relative pronoun forms. We ran it with these modifications on the Dutch UD treebanks.

2 Conversion to EUD from an Underlying Treebank

In this section we describe our method for adding enhanced dependency annotation to a corpus for which manually verified language specific syntactic annotation is already in place.

There are currently two Dutch UD corpora. They are based on manually verified treebanks that both follow the guidelines of the Lassy treebank project (van Noord et al., 2013). The Lassy annotation guidelines combine elements from phrase structure treebanks (such as phrasal nodes and use of co-indexed nodes for encoding the syntactic role of fronted WH-constituents, relative pronouns, and ellided phrases) with elements from dependency treebanks (such as dependency labels, discontinuous constituents and crossing branches), similar to the Tiger (Brants et al., 2002) and Negra (Skut et al., 1998) corpora for German.

As of release v2.2 of the Universal Dependencies corpora, (portions of) both Dutch corpora have been converted to UD using the same conversion script.[4] An example of the conversion process is given in Figure 1. For each phrasal node in the underlying annotation, the script first identifies

[4] Available at `https://github.com/gossebouma/lassy2ud`

the head daughter according to UD (indicated by a boxed node in the diagram). In most cases, this is a leaf node labeled with the underlying dependency relation hd. In some cases, such as PPs, however, the UD guidelines and the underlying annotation diverge and (the head of) a non-head daughter is selected as the UD head. Next, labeled edges between non-null leaf nodes are added. For a head node H and a sister non-head node D, an edge $H \xrightarrow{label} D$ is added if H and D are lexical nodes. If H or D is phrasal, the dependency holds between the lexical heads of H and/or D. The label is determined by a mapping from the underlying dependency label of D to the corresponding UD label (where in some cases, the mapping depends on other aspects of the syntactic context). Note that null nodes in the underlying annotation are not annotated in the conversion to UD. The conversion process is described in detail in Bouma and van Noord (2017).

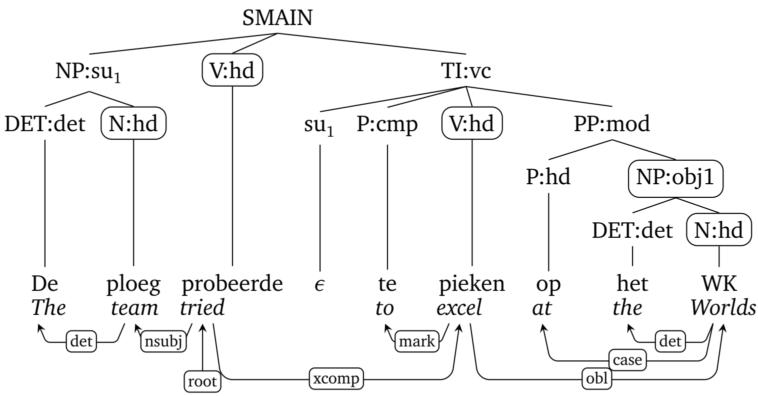

Figure 1: Phrasal annotation and the induced dependency annotation for *de ploeg probeerde te pieken op het WK* (*the team tried to excel at the world championships*).

We add EUD as follows: first, we recover elided predicates in the Alpino dependency tree. Next, we run the standard Alpino to UD conversion script on the expanded tree while ensuring that dependency labels for relations *obl*, *nmod*, *acl*, *advcl*, and *conj* are expanded so as to include the lemma form of the relevant *case, mark* or *cc* dependent or other subtyping (e.g. relatives are labeled *acl:relcl*). Finally, we add dependency relations resulting from the propagation of conjuncts, the modified analysis of relative clauses, and the explicit annotation of control relations. Below, we describe the steps that require access to aspects of the underlying annotation, i.e. recovery of elided predicates, propagation of conjuncts, and adding control relations.

2.1 Recovering elided predicates

The standard UD annotation of gapping and ellipsis (Schuster et al., 2017) involves promoting the highest ranked dependent to head and attaching any remaining dependents to this pseudo head using the *orphan* relation, as illustrated in Figure 2 (top annotation). In EUD, elided predicates are reconstructed by adding additional lexical elements to the input string that are copies of the preceding or following head node to which they correspond. The enhanced annotation can refer to these inserted lexical elements as well, thus making the process that promotes dependents to heads and the *orphan* dependency label superfluous (Figure 2, bottom annotation).

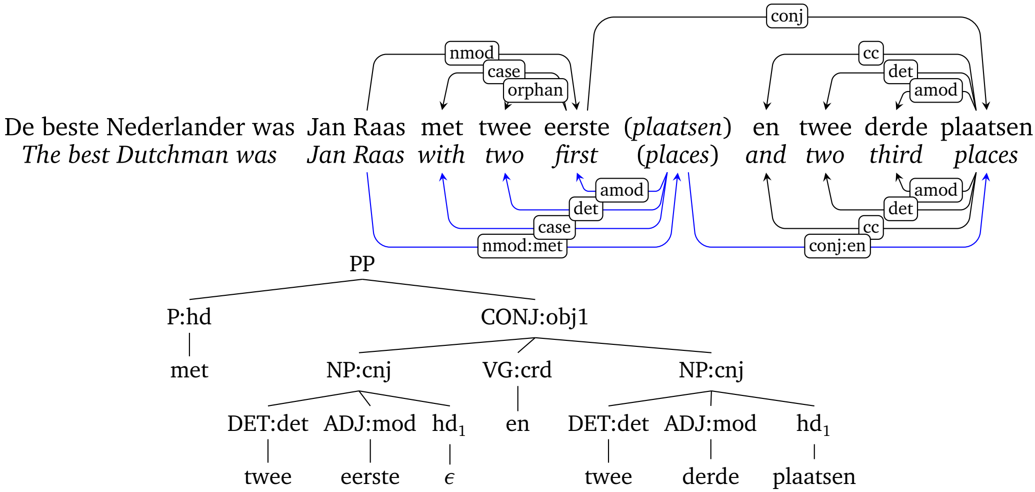

Figure 2: UD (above) and EUD (below) annotation and Alpino dependency tree for the sentence *The best Dutchman was Jan Raas with two first and two third places* containing an NP coordination with an elided nominal predicate in the first conjunct. Co-indexing is shown using subscripts.

As the underlying (Alpino) annotation contains null nodes for elided elements in ellipsis constructions, the recovery of elided predicates can be done as part of the automatic conversion by replacing the relevant null nodes (that is, a co-indexed empty node that is a head according to UD) with a node that is a copy of the head node with which it is co-indexed. With these recovered elements in place, dependency labels can be re-computed and the need to promote non-heads to heads disappears.

2.2 Propagation of Conjuncts

When the subject or object is a conjoined phrase, an additional dependency relation labeled as *subj* or OBJ from (verbal) predicate to the non-head *conj* dependent is added. This addition does not require access to underlying annotation. The guidelines for EUD also state that subjects, objects, and other complements that occur as dependents of a conjoined predicate are attached to both predicates, as illustrated in Figure 3. In the underlying annotation, cases like these (i.e. conjoined predicates which share one or more dependents) are easily recognized as the shared dependent present in one of the conjuncts is co-indexed with an empty node in the other conjunct.

2.3 Raising and control

The guidelines state that for embedded clauses in raising and control contexts, an additional subject relation should be added from the embedded verb to the controlling subject or object in the matrix clause (Figure 4). In the underlying annotation for such cases, the embedded clause contains an empty subject node, co-indexed with a controlling subject, object, or indirect object in the matrix clause. In the conversion to EUD, a dependency relation is added from the head of the embedded clause to (the head of) the controlling NP.

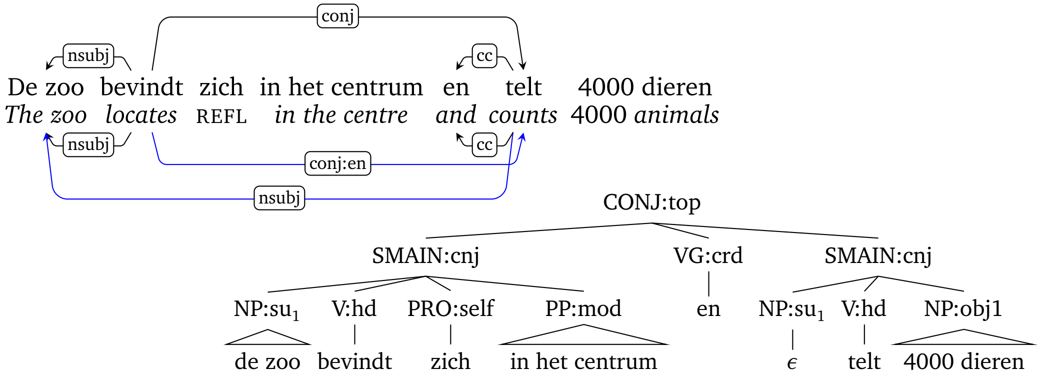

Figure 3: UD (above) and EUD (below) annotation and Alpino dependency tree for a sentence coordination with an elided subject in the second conjunct for the sentence *The zoo is located in the center and has 4,000 animals*.

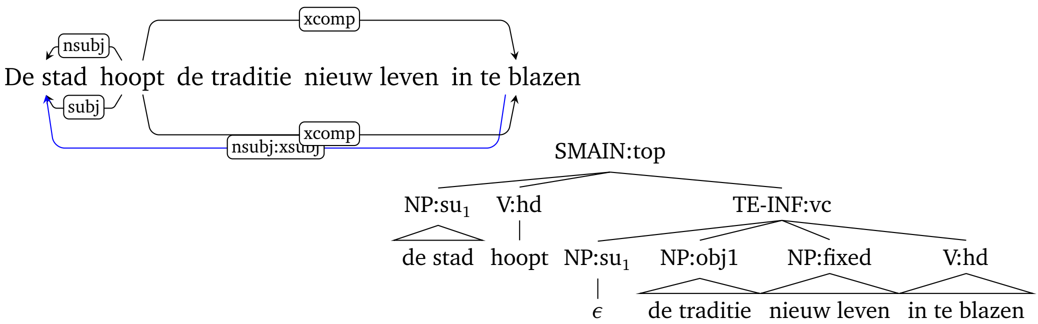

Figure 4: UD (above) and EUD (below) annotation and Alpino dependency tree for the sentence *The city hopes to bring the tradition to life again*. Note that we follow the convention of Schuster et al. (2018) of adding an extension to the *nsubj* label indicating its status.

The conversion takes into account the fact that the controller may itself be co-indexed, as illustrated in Figure 5, where the subject of a passive auxiliary is co-indexed with an elided subject in the second conjunct, which in turn controls a subject in the embedded clause.

2.4 Relative Clauses

In relative clauses, a direct dependency is added from the head of the relative to the nominal antecedent. The dependency is labeled with the dependency relation corresponding to that of its relative pronoun in the standard UD annotation. The dependency to the relative pronoun itself is relabeled as *ref* (Figure 6). Note that the EUD annotation contains a cyclic dependency ($snelweg \xrightarrow{acl:relcl} omcircelt \xrightarrow{nsubj:relsubj} snelweg$).

When the relative pronoun is missing, it seems the correct label for the dependency relation from the head of the relative clause to the antecedent noun requires access to additional information (such as an empty node in the underlying annotation). Unlike English, Dutch does not have

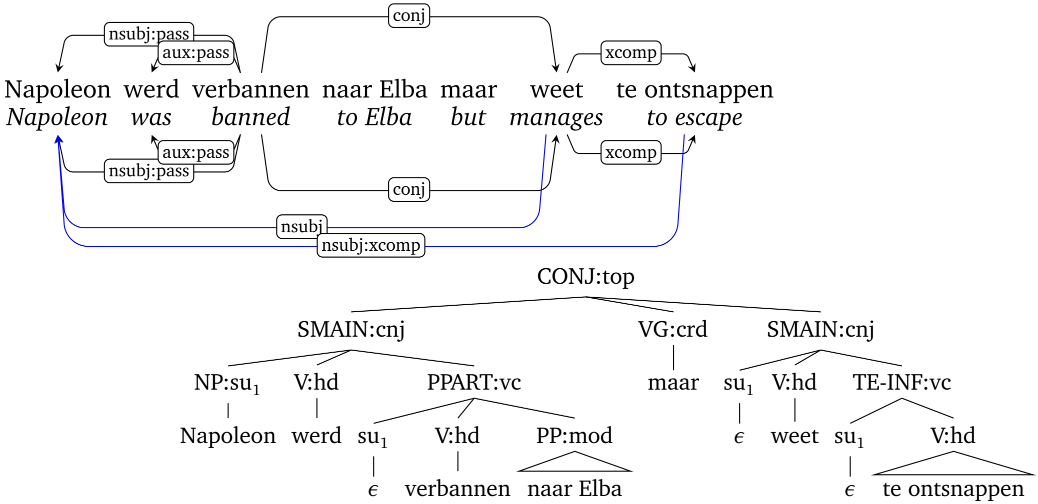

Figure 5: UD and EUD (below) annotation and underlying annotation for the sentence *Napoleon was banned to Elba but manages to escape.*

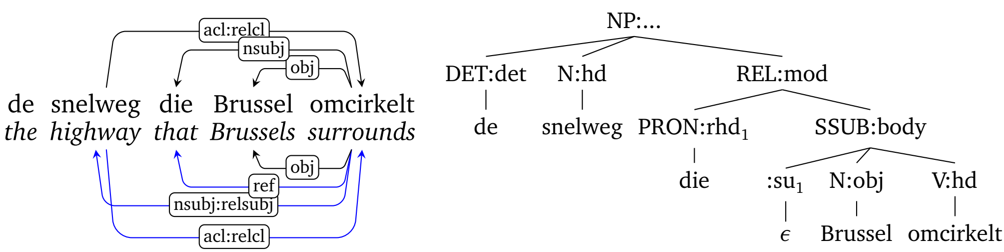

Figure 6: UD and EUD annotation and underlying annotation (right) for the NP *the highway that surrounds Brussels.*

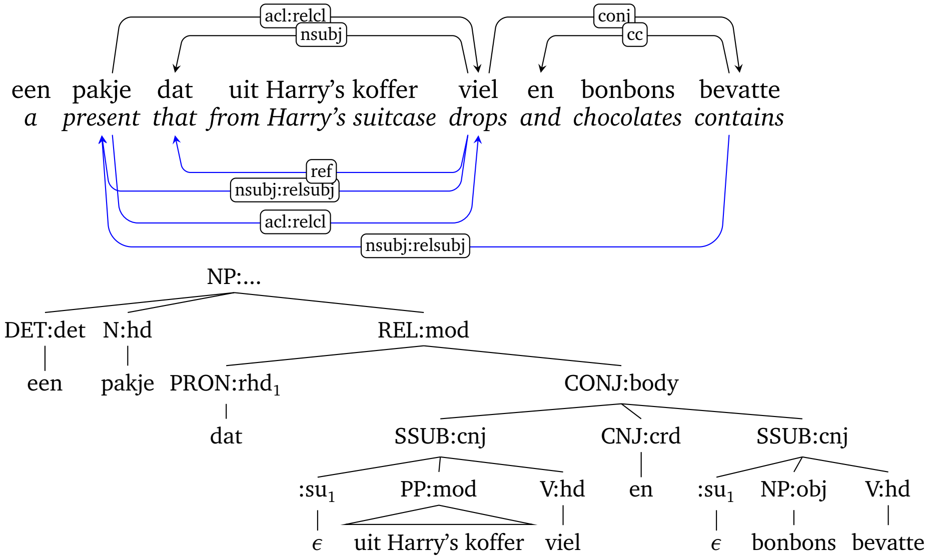

Figure 7: UD and EUD (top) annotation and underlying annotation (bottom) for the NP *a present that drops from Harry's suitcase and contains chocolates.*

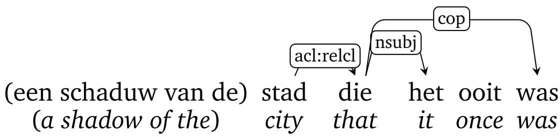

Figure 8: UD annotation for the NP *a shadow of the city that it once was*, where the relative pronoun heads the relative clause.

bare relatives and thus, at first sight, this problem does not arise. It turns out, however, that coordinated relative clauses, where the relative pronoun heads a conjunction and is "*extracted across-the-board*", do give rise to a similar problem. In these cases, illustrated in Figure 7, the grammatical role of the antecedent noun in the second conjunct is not explicitly marked in UD, as the relative pronoun is a dependent of the head of the first conjunct only. All instances of this phenomenon in the Dutch treebanks exhibit strict parallelism (i.e. the relative pronoun fulfills the same grammatical role in both conjuncts) but whether this is a requirement has been a matter of some debate (Williams, 1978; Hartmann et al., 2016).

Another complication arises in cases where the relative pronoun is a predicate in a copula construction. UD considers the predicate to be the head in copula constructions, and thus the relative pronoun is the syntactic head of the relative clause in cases such as Figure 8. The relative pronoun thus is an *acl:relcl* dependent of the antecedent noun. The EUD guidelines for relative clauses potentially introduce a reflexive dependency in such cases (from antecedent noun to antecedent noun, replacing the UD dependency pointing to the relative pronoun). As this is not allowed (and its interpretation would be unclear), the UD annotation is preserved in EUD in such cases.[5]

[5]The current annotation guidelines acknowledge this problem.

	UD deps	EUD deps	EUD=UD	elided
AE	75,100	78,203	63,491	84
SE	75,100	78,200	63,398	38

Modified deps	deps	%
Agreement	13,463	85.7
Disagreement	2,251	14.3
Total	15,714	100.0

relation	AE	SE	relation	AE	SE
acl:relcl	501	533	nsubj	3,645	3,761
advcl:om+acl:om	89	19	nsubj:pass	721	742
advcl:te	7	106	nsubj:pass:relsubj	53	64
advmod:*	2,492	2,538	nsubj:relsubj	276	295
amod:*	4,513	4,376	nsubj:xsubj	194	403
aux:pass	762	753	obl:*	4,680	4,655
det	7,944	7,815	ref	484	395
nmod	6,272	6,164	root	5,788	5,733

Table 1: Statistics for the effect of going from UD to EUD. The top-right table shows (dis-) agreement between AE and SE only for those cases where the EUD of either AE or SE is not equal to UD.

3 Comparison

We ran both the rule-based method outlined above which takes the underlying language specific (Alpino) annotation as starting point (referred to as AE in this section) and the heuristic method of Schuster et al. (2018) (SE) on the training section of the UD Dutch LassySmall corpus.[6]

Table 1 (top-left) shows that the EUD has approx. 4% more relations than UD for both AE and SE, but also that the relations present in EUD are identical to UD in only approx. 84% of the cases. The number of newly introduced elided nodes differs significantly between the two methods. The top-right table gives statistics for the level of agreement between AE and SE, for those cases where EUD $\neq$ UD in either AE or SE.

The bottom table in Table 1 gives an overview of dependency relations whose frequency differs most strongly between AE and SE. The differences can be attributed to the following phenomena:

- **Propagation of conjuncts**: if the underlying annotation indicates that modifiers and function words should be spread across conjunts, AE applies the rule for propagation of conjuncts, whereas SE does not. This explains the higher numbers for *amod,det, nmod* and *obl* in AE. It accounts for phrases such as *verlaten straten en ruines (deserted streets and ruins), talrijke persoonlijke documenten en brieven (numerous personal documents and letters), de herkomst en geschiedenis van woorden (the origin and history of words)* and *invoer of uitvoer van buitenlandse deviesen (import and export of foreign currencies)*.
- **Lack of propagation of conjuncts**: If a predicate in a copula construction is conjoined (as in *Mandeville was arts en filosoof (Mandeville was writer and philosopher)*), no co-indexed

[6]There was one sentence, wiki-8628.p.21.s.3, for which SE did not produce a result. It contains a coordinated relative clause, with a severely elliptical second conjunct (where in the underlying treebank, the relative pronoun, passive auxiliary, and passive participle are elided). It is excluded in the comparison below.

empty subject is present in the underlying annotation (as in that annotation, the copula is the head and not the predicate). As a consequence, no propagation of conjuncts takes place in AE, whereas these cases are covered by SE. This explains the higher number of *nsubj* in SE. Also, SE spreads conjoined relative clauses where AE does not. This explains why SE contains more *acl:relcl* and *nsubj:relsubj* dependents.

- **Adding control relations**: SE adds a control relation to all *xcomp* dependents. This includes non-verbal cases such as *het stripboek was bedoeld als reclame (the comic was intended as advertisement), de partij kwam in opspraak (the party became controversial), door een campagne te voeren met als thema ....(by running a campaign with as theme ....)*. As these have no co-indexed empty subject in the underlying annotation, AE does not add a control relation in these cases and thus it has significantly fewer *nsubj:xcomp* dependents. On the other hand, AE correctly identifies the indirect object as controller in cases like *de priester gaf Johannes te drinken (the priest gave Johannes to drink)*, where *Johannes* is a *iobj* and *te drinken* a non-finite verbal *xcomp*.

- **Disagreements about lexical extensions:** The dependencies *acl* and *advcl* are extended with the lemma of their *case* or *mark* dependent. However, *advcl* and *acl* clauses can contain both, as in *een zondag om nooit te vergeten (a sunday to never forget)*. AE adds the *case* lemma *om* in these cases, where SE adds the marker *te*. This explains the almost complementary distribution of *acl:om* + *advcl:om* and *acl:te* + *advcl:te*. A similar issue arises with coordinations containing two *cc* elements, such as *zowel op straat als in woningen (both on the street and in houses*), where AE adds the lemma of the first *cc* to the *conj* dependency and SE the lemma of the second. Finally, some prepositional phrases consist of a preposition as well as a postpositional particle (*op twee na (except two)*). Again, the two methods give different results for the lexical extension.

- **Graphs without root:** A consequence of the analysis of relatives is that some sentences do not have a root element. I.e. in *Kuifje is een reporter die in Brussel woont (Tintin is a reporter who lives in Brussels)*, *reporter* is the root in UD, but it is a dependent of *lives* in EUD, thus loosing its root status according to SE but not according to AE. The documentation for UD simply states that the *root* dependency points to the root of the sentence, while EUD does not raise the issue whether a EUD graph can be without root or even have multiple roots, thus leaving open the question what exactly defines a root in a potentially cyclic graph (i.e. is it the element that is not a dependent of anything else or the element(s) from which all other elements can be reached?).

- **Position of elided predicates:** A last systematic source of conflict between AE and SE, not immediately reflected by the statistics in Table 1, is disagreement about the position of an elided item. Apart from the fact that AE inserts more elided elements than SE, it is also the case that AE and SE often disagree about the exact position of the elided material. Remember that in the CONLL-U format[7], elided nodes are entered as additional lines in the annotation, indexed with a subscripted (N.1 *etc.*) node. The annotation guidelines do not specify explicitly where to place this element in the string. A human annotator would most likely place the elided predicate at the position which results in a grammatical sentence. The AE script, however, recovers elided predicates by spotting index nodes without lexical content that lack attributes indicating their string position. The internal XML-format of the

[7] `universaldependencies.org/format.html`

Alpino annotation trees employs a normal form where, by default, empty index elements are leftmost daughters of the phrasal node by which they are dominated. The AE script therefore inserts recovered elided elements in front of the words making up the phrase in the corresponding CONLL-U format. SE usually opts for a more natural word order. Note that as a consequence, there not only is a conflict regarding the elided predicate itself, but also all dependents of this element will have conflicting EUD annotations. It is not clear to us however whether the AE strategy violates the guidelines or not.

Most of the conflicts between the two methods can be resolved by making the guidelines more explicit (i.e. by listing which dependents should be included in spreading of conjuncts, by deciding whether every *xcomp* should have a subject, by providing a definition of *root*, and by specifying where elided elements should be placed). We do not think that there are decisions that would be hard to implement by either the rule-based or the heuristic approach. The rule-based approach can obviously be modified so as to include spreading of conjoined predicative heads and relatives, to filter *root* dependency labels of nodes that have another dependency label as well, and to add a controller to non-verbal XCOMP's. The most challenging problem might be the adequate placement of elided predicates, but then, it should be noted that it is unclear to what extent this is relevant for downstream (semantic) applications. The heuristic strategy might seem to be hard pressed to predict propagation of conjuncts that are modifiers, such as *amod*, *nmod*, and *obl*. A brief corpus investigation suggests however, that, for instance, spreading of *amod* dependents occurs almost always in those cases where the *amod* is part of the first conjunct, and the second conjunct is not introduced by a determiner and almost never contains an *amod* itself. Spreading of *obl* occurs almost always in conjoined finite clauses where the phrase headed by *obl* occurs as first (fronted) constituent and the second conjunct is verb-initial (i.e. does not contain a fronted element).

4 Conclusion

In this paper, we have outlined a rule-based approach for converting a treebank with rich but language-specific dependency annotation to EUD and compared this with a method that converts from UD to EUD. Our hypothesis was that the richer annotation in the underlying treebank would help to produce more accurate annotation than the heuristic method that uses information from UD only. Comparing the output of both conversion scripts on the training section of UD Dutch LassySmall revealed that there are systematic differences between the two. These appear to be mostly due to insufficient detail in the annotation guidelines for EUD. Resolving these would probably reduce the number of conflicts in the output of the two conversion methods to a very small number, suggesting that both the AE and the SE method are able to produce accurate EUD annotation.

Some issues regarding the guidelines may require considering once more which problem EUD is supposed to solve exactly. Candito et al. (2017) cover a somewhat different set of syntactic phenomena that might also be incorporated in the enhanced layer. They propose to specify control also for certain adjectives and non-finite verbal modifiers of nouns and furthermore argue for a canonical representation that undoes the effect of diathesis alternations. From a semantic point of view, these are valuable enhancements, and it seems to fit in the general philosophy of UD that says that comparable sentences should receive comparable annotation across languages. Schuster and Manning (2016), on the other hand, discuss certain enhancements of basic UD such as a reanalysis of partitives, light noun constructions, and phrasal prepositions, that are specific to English. More generally, one can imagine that each language specifies what counts as a partitive, light verb, phrasal preposition, etc. (not unlike specifying what counts as an auxiliary in the current UD annotation scheme), and that the annotation or conversion can proceed automatically and uniformly across languages on the basis of such lists. The utility of such enhancements is demonstrated for instance in Bar-Haim et al. (2007), which contains a series of normalization and inference rules over syntactic (dependency) trees for recognizing textual entailment. At the same time, it seems that EUD ideally should also function as a uniform syntactic representation that can be used to derive full semantic interpretations of sentences without requiring language-specific lexica or rules. It remains to be seen whether the complex graphs that result from adding all conceivable enhancements are indeed suitable for such systems as well.

Acknowledgments

This research was carried out while the author was visiting scientist at the Center for Advanced Study at the Norwegian Academy of Science and Letters. It has benefitted from presentations at the Center as well as from feedback by the anonymous reviewers for this conference.

References

Bar-Haim, R., Dagan, I., Greental, I., Szpektor, I., and Friedman, M. (2007). Semantic inference at the lexical-syntactic level for textual entailment recognition. In *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, pages 131–136. Association for Computational Linguistics.

Bouma, G. and van Noord, G. (2017). Increasing return on annotation investment: the automatic construction of a Universal Dependency treebank for Dutch. In Nivre, J. and de Marneffe, M.-C., editors, *NoDaLiDa workshop on Universal Dependencies*, Gothenburg.

Brants, S., Dipper, S., Hansen, S., Lezius, W., and Smith, G. (2002). The TIGER treebank. In *Proceedings of the workshop on Treebanks and Linguistic Theories*, volume 168.

Candito, M., Guillaume, B., Perrier, G., and Seddah, D. (2017). Enhanced UD dependencies with neutralized diathesis alternation. In *Depling 2017-Fourth International Conference on Dependency Linguistics*.

Futrell, R., Mahowald, K., and Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, 112(33):10336–10341.

Gotham, M. and Haug, D. (2018). Glue semantics for universal dependencies. In *Proceedings of the 23rd international Lexical-Functional Grammar Conference*, Vienna.

Hartmann, J., Konietzko, A., and Salzmann, M. (2016). On the limits of non-parallelism in ATB movement: Experimental evidence for strict syntactic identity. *Quantitative Approaches to Grammar and Grammatical Change: Perspectives from Germanic*, 290:51.

Lipenkova, J. and Souček, M. (2014). Converting Russian dependency treebank to Stanford typed dependencies representation. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, volume 2: Short Papers*, pages 143–147.

Przepiórkowski, A. and Patejuk, A. (2018). From LFG to enhanced universal dependencies. In *Proceedings of the 23rd International LFG conference*, Vienna.

Pyysalo, S., Kanerva, J., Missilä, A., Laippala, V., and Ginter, F. (2015). Universal dependencies for Finnish. In *Proceedings of the 20th Nordic Conference of Computational Linguistics (Nodalida 2015)*, pages 163–172.

Reddy, S., Täckström, O., Petrov, S., Steedman, M., and Lapata, M. (2017). Universal semantic parsing. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 89–101.

Schuster, S., Lamm, M., and Manning, C. D. (2017). Gapping constructions in universal dependencies v2. In *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies (UDW 2017)*, pages 123–132.

Schuster, S. and Manning, C. D. (2016). Enhanced English universal dependencies: An improved representation for natural language understanding tasks. In *Proceedings of LREC*.

Schuster, S., Nivre, J., and Manning, C. D. (2018). Sentences with gapping: Parsing and reconstructing elided predicates. In *Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2018)*.

Shwartz, V., Goldberg, Y., and Dagan, I. (2016). Improving hypernymy detection with an integrated path-based and distributional method. In *Proceedings of the 54th Annual Meeting of the ACL 2016*, pages 2389–2399, Berlin.

Skut, W., Brants, T., Krenn, B., and Uszkoreit, H. (1998). A linguistically interpreted corpus of German newspaper text. *arXiv preprint cmp-lg/9807008*.

van Noord, G., Bouma, G., van Eynde, F., de Kok, D., van der Linde, J., Schuurman, I., Sang, E. T. K., and Vandeghinste, V. (2013). Large scale syntactic annotation of written Dutch: Lassy. In Spyns, P. and Odijk, J., editors, *Essential Speech and Language Technology for Dutch: the STEVIN Programme*, pages 147–164. Springer.

Vulić, I. (2017). Cross-lingual syntactically informed distributed word representations. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, volume 2, pages 408–414.

Wang, Y. and Liu, H. (2017). The effects of genre on dependency distance and dependency direction. *Language Sciences*, 59:135 – 147.

Williams, E. (1978). Across-the-board rule application. *Linguistic Inquiry*, 9(1):31–43.

Zeman, D., Popel, M., Straka, M., Hajic, J., Nivre, J., Ginter, F., Luotolahti, J., Pyysalo, S., Petrov, S., Potthast, M., Tyers, F., Badmaeva, E., Gokirmak, M., Nedoluzhko, A., Cinkova, S., Hajic jr., J., Hlavacova, J., Kettnerová, V., Uresova, Z., Kanerva, J., Ojala, S., Missilä, A., Manning, C. D., Schuster, S., Reddy, S., Taji, D., Habash, N., Leung, H., de Marneffe, M.-C., Sanguinetti, M., Simi, M., Kanayama, H., dePaiva, V., Droganova, K., Martínez Alonso, H., Çöltekin, c., Sulubacak, U., Uszkoreit, H., Macketanz, V., Burchardt, A., Harris, K., Marheinecke, K., Rehm, G., Kayadelen, T., Attia, M., Elkahky, A., Yu, Z., Pitler, E., Lertpradit, S., Mandl, M., Kirchner, J., Alcalde, H. F., Strnadová, J., Banerjee, E., Manurung, R., Stella, A., Shimada, A., Kwak, S., Mendonca, G., Lando, T., Nitisaroj, R., and Li, J. (2017). Conll 2017 shared task: Multilingual parsing from raw text to universal dependencies. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 1–19, Vancouver, Canada. Association for Computational Linguistics.