Graph Rewriting for Enhanced Universal Dependencies

Bruno Guillaume, Guy Perrier

HAL Id: hal-03326492
https://hal.inria.fr/hal-03326492

Submitted on 26 Aug 2021

Graph Rewriting for Enhanced Universal Dependencies

Bruno Guillaume
Université de Lorraine, CNRS,
Inria, LORIA, Nancy, France
`bruno.guillaume@inria.fr`

Guy Perrier
Université de Lorraine, CNRS,
Inria, LORIA, Nancy, France
`guy.perrier@loria.fr`

Abstract

This paper describes a system proposed for the IWPT 2021 Shared Task on Parsing into Enhanced Universal Dependencies (EUD). We propose a Graph Rewriting based system for computing Enhanced Universal Dependencies, given the Basic Universal Dependencies (UD).

1 Introduction

The IWPT 2021 Shared Task on Parsing into Enhanced Universal Dependencies (Bouma et al., 2021) is a second edition of an equivalent shared task in 2020 (Bouma et al., 2020). The goal of the shared task is to produce EUD (Schuster and Manning, 2016), with several new annotation layers expressed on top of UD annotations (Nivre et al., 2020).

In the previous shared task, there were two kinds of approaches: producing EUD annotation from raw text with machine learning methods or producing EUD from UD with a rule-based approach (with or without some learning to optimize rule usage). Like (Heinecke, 2020) or (Dehouck et al., 2020), our proposal corresponds to the second approach: we used an existing tool for producing UD annotations and work only on the conversion from UD to EUD. Unlike other rule-based approaches, we used GREW, a generic Graph Rewriting tool (Bonfante et al., 2018), in order to describe the rules for enhancement.

Another specificity of our work is that we primarily design our rules by following the guidelines. Even if, in a secondary step and in the context of the shared task, we adapt the system to the corpora which diverge from the guidelines (section 2.6), we can easily provide a system closer to the guidelines, adpatable to languages specificities.

Our system achieved 81.58 ELAS score on the task, starting from UDPIPE annotation with an LAS of 85.77. In the paper, we present the system, analyse the results and make some in-depth analysis on French and English of discrepancies between GOLD_{EUD} data and the output of our system starting GOLD_{UD} annotations.

2 Description of the system

2.1 Parsing to UD

For this shared-task, we used the UDPIPE2 (Straka et al., 2016) through the online service[1] to produce the UD annotation of the data. The models used for each language are trained in UD version 2.6 on the following corpora: Arabic-PADT, Bulgarian-BTB, Czech-PDT, English-EWT, Estonian-EDT, Finnish-TDT, French-GSD, Italian-ISDT, Lithuanian-ALKSNIS, Dutch-Alpino, Polish-PDB, Slovak-SNK, Swedish-Talbanken, Tamil-TTB, Ukrainian-IU[2].

2.2 GREW

The transformation UD to EUD is described with the graph rewriting tool GREW[3]. Each rule is defined by a pattern and a set of commands describing how to modify the graph. A dedicated strategy mechanism allows for controlling rules applications (in which order subsets of rules must be applied and how they must be iterated). A global transformation system (rules and strategies) is called a Graph Rewriting System (GRS).

2.3 Representation of EUD annotations

We use here the convention already adopted in the Deep-Sequoia project (Candito et al., 2014), in which dependencies are drawn in black when the

---

[1] `https://lindat.mff.cuni.cz/services/udpipe/`
[2] The only modification done was on the Czech output where 7 obvious errors on the lemmatisation makes the conversion producing non valid data.
[3] `https://grew.fr/`

relation exists both in UD layer and EUD layer, in red when they are present in UD layer only and in blue (and below) when they are present in EUD layer only (see Figure 1). This prevents from having two relations when both layers are identical and make figures easier to read.

2.4 From UD to EUD

Our goal was to design a GRS, following the EUD guidelines, to convert the UD annotations into EUD annotations, but we observed that the conversion system has to be adapted to each language and to some particular annotation choices.

As the rules of a GRS are organized in strategies, this adaptation is done by activating or not the applications of subsets of rules in the definition of the main strategy of a generic GRS. The rules are freely available[4].

2.4.1 The six types of enhancement

The EUD guidelines identify 6 types of enhancements, and for each type, we have designed a subset of rules to achieve it. We briefly describe the main features of each subset.

Empty nodes for elided predicates   Predicate elision is detected thanks to the presence of `orphan` dependencies. They are mainly found in coordinations and sometimes with parataxis relations. Figure 1 shows the UD annotation of a sentence with ellipsis[5].

First, a null node `N` is created and a dependency `N −obj→ trees` is introduced from `N` to the head of the second conjunct `trees`. The label `obj` is determined by a parallelism with the dependency `sold −obj→ strawberries` [6]. The dependency `trees −cc→ and` is also raised to `N`.

In a second stage, all `orphan` dependencies from the head of the second conjunct are transformed into dependencies from the null node. The labels of the new dependencies are determined from the context. In our example, the dependency `trees −orphan→ winter` is transformed into the dependency `N −obl→ winter`, because `N` is a verb, `winter` a noun with a `case` dependent.

---

[4]https://gitlab.inria.fr/grew/udtoeud

[5]All examples are extracted form the corpora used in the task. Due to lack of space, expressions in square brackets, which are not essential to our purpose, are skipped in the annotation.

[6]Even if the subject is also a candidate, we always favour the object in this case.

Propagation of incoming dependencies to conjuncts   The propagation of incoming dependencies to the conjuncts of a coordination is systematical. The only difficulty concerns modifiers: if a word `H` is modified by a coordination, the label of the dependency from `H` to the head of the first conjunct may need to be changed depending of the POS of the second conjunct. In Figure 2, the dependency `come −obl→ parents` has to be propagated to the head of the other conjunct, the adverb `separately`. Because of this POS, the label has to be changed and the dependency becomes `come −advmod→ separately`.

Many gold corpora of the task do not take into account differences of POS between the conjunct heads and propagate the incoming dependencies without changing their labels.

Propagation of outgoing dependencies from conjuncts   The main problem for this enhancement comes from the ambiguity of the UD annotation schema. It is not possible to distinguish a left dependency on a coordination from a left dependency on the first conjunct of the coordination because both are attached to the head of the first conjunct.

But, it is necessary to remove the ambiguity in order to know if one should propagate a left dependency. This is more or less easy depending on the type of the dependency. In Figure 3, the `nsubj` and `cop` dependencies on the noun `acteur` must be propagated on the head of the second conjunct `protagoniste`. It is easy to design a specific rule for each type of dependency to perform the propagation, what we have done. But, if the dependency is an `advmod` dependency, there is no general criterion for removing the ambiguity. The dependency on `souvent` should be propagated, but not the dependencies on `n'` and `pas`. This depends on the modifier adverb but also on the context. This is a point where our rule-based approach marks its limits compared to learning approaches. Of course, the answer depends strongly on the language and we will see how to take into account the specificity of each language in subsection 2.5.

For right dependencies, there is no ambiguity, because a right dependent on the first conjunct that follow all conjuncts is necessarily a dependent on the coordination.

Additional subject relations for control and raising constructions   Raising and control verbs take an infinitive as a `xcomp` dependent and the en-
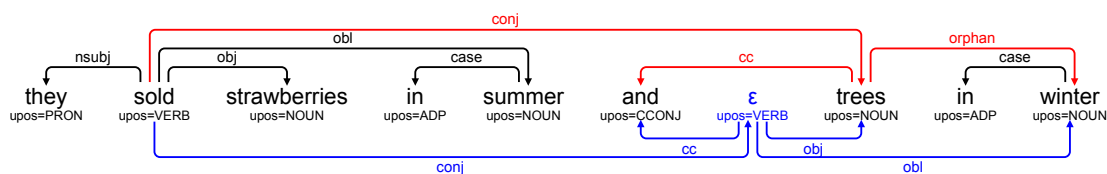
Figure 1: They sold strawberries in summer [,] and [Christmas] trees in [the] winter
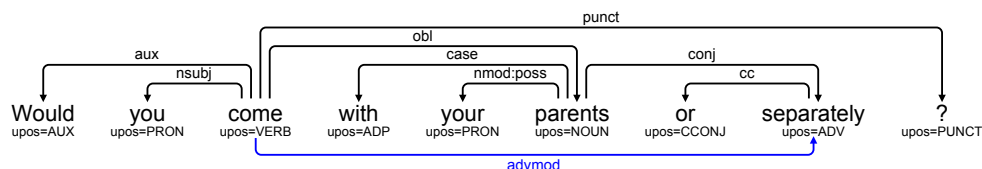

Figure 2: Would you come with your parents or separately ?

hanced subject of the infinitive is either the subject, the direct object or the indirect object of the main verb, if that argument exists. The choice between the three possibilities depends on the lexical information about the raising or control verbs. In our approach, this means that we need lexicons per language. Most often, we do not have such lexicons and choose the subject of the main verb as the subject of the infinitive, as this is the most likely.

Coreference in relative clause constructions
The relative clause enhancement adds `ref` dependencies from the antecedent to the relative pronoun and all dependencies targeting the relative pronoun are moved to the antecedent.

In Figure 4, the sentence contains two relative pronouns, `dont` and `qui`. Let us focus on `qui`. First, rules make an upward path from the relative pronoun by following the dependencies until finding an `acl:relcl` dependency. In the example, this requires crossing a dependency `nsubj` and then `conj`. As soon as the dependency `acl:relcl` is reached, it is possible to add the `ref` dependency, because its source is known, it is the source of the dependency `acl:relcl` and the antecedent of the relative pronoun, and the target of the dependency as well, the relative pronoun, which is kept in memory. In a second stage, all dependencies targeting the relative pronoun are moved to the antecedent. In the example, the dependency `nécessite −nsubj→ qui` is transformed into the dependency `nécessite −nsubj→ entreprise`.

Modifier labels that contain the preposition or other case-marking information   With regard to this enhancement, we strictly follow the guidelines: if a `case` or `mark` dependent on a modifier is a multiword expression, we add the form of the expression to the dependency representing the modification; if it is a single word, we add the lemma of the word.

The guidelines don't cover the case of several `case` or `mark` dependencies having the same source. In Figure 5, the conjunction `because` and the preposition `in` both depend on the same governor `zone` (dependencies in orange). We decide to add the outermost dependent lemma to the modifier dependency `feel −advcl→ zone` because it is related to this dependency, whereas `in` has no relationship to it.

If the two candidate dependents are consecutive, we consider them as a single multiword and add the concatenation of their lemmas to the modifier dependency.

2.4.2   Rule ordering

The six types of enhancements are not totally independent, some of them interact, so the order of application between the six corresponding subsets of rules is not neutral. Figure 1 shows an interaction between the subset implementing ellipsis processing and the subset implementing the propagation of outgoing dependencies of a coordination. If we apply the latter first, it is not possible to propagate the dependency `sold −nsubj→ they` to the second conjunct of the coordination because its verb is elided. We must apply first the subset creating a null node representing this verb.

177
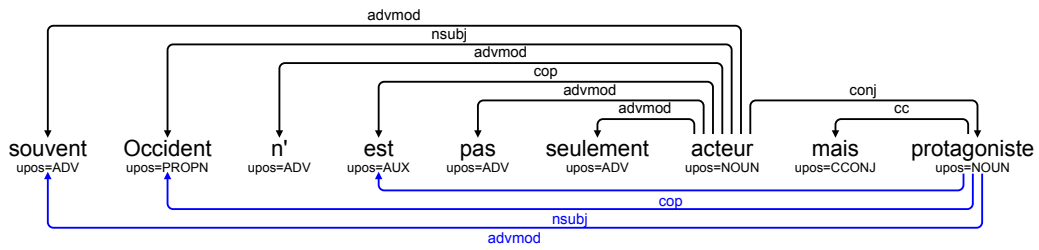
Figure 3: souvent [, l'] Occident n' est pas seulement [l'] acteur [,] mais [aussi le] protagoniste [des violations des droits de l' homme] (often[, the] West is not only [the] actor[,] but [also the] protagonist [of human rights violations])

Figure 4: Une entreprise dont on reparlera et qui nécessite [un budget important] (A business that will be talked about again and that requires [a large budget])

Figure 5: I feel so bad because I was [so] in the zone [that I didn't even get her name]

178

The subset of rules adding case or mark information to modifier dependencies must be applied at the end because it does not create any possibility to apply another subset later.

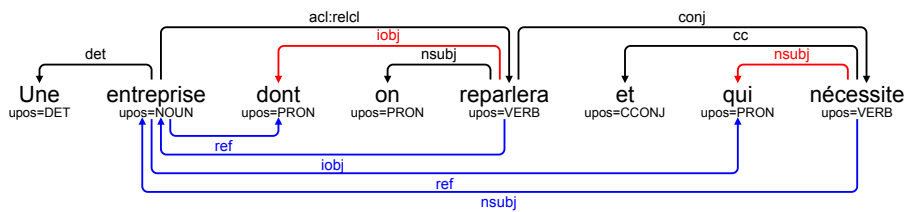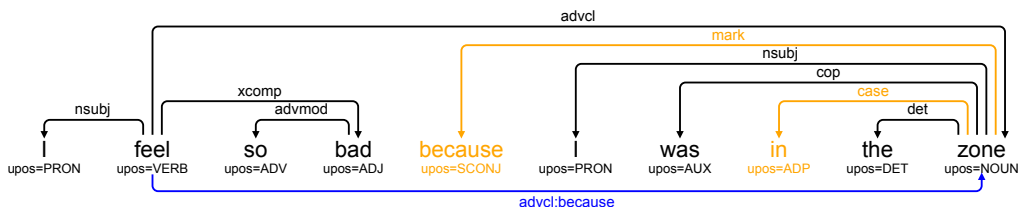The order between the application of the five other subsets is relatively free. Figure 4 shows an example of interaction between the subset relating to relative clauses and the subset implementing the propagation of incoming dependencies of a coordination. We applied the former subset first and to apply the second subset, we only need to propagate the dependency `entreprise −acl:relcl→ reparlera` to the second conjunct of the coordination. But there is no problem to reverse the order between the two subsets.

Even if the order chosen between the 6 subsets aims at minimizing the number of rule applications, we cannot avoid applying some subsets a second time. Figure 6 shows an example of repetition in the subset application. To add the dependency `build −nsubj→ I`, we need first to apply the subset related to control verbs. We obtain the dependency `sell −nsubj→ I`. Then we apply the subset related to the propagation of outgoing dependencies of a coordination. We obtain the dependency `use −nsubj→ I`. Finally, we apply the subset relative to control verbs a second time and we obtain the last dependency `build −nsubj→ I`.

All rules presented in this subsection constitute the generic GRS used to convert the UD annotation into the EUD for the 17 languages.

2.5 Adaptation to the specificities of languages

Rule packages are added to the generic GRS to express specificities of language groups. In order to be applied, they are inserted in the strategy at carefully chosen positions in the generic strategy.

In this way, strategies can be designed adapted to particular languages, by activating or not these new packages.

Now, let us examine which types of rules can be added to express specificities of certain languages.

Null Subject Languages   Arabic, Bulgarian, Czech, Estonian, Finnish, Italian, Polish, Russian, Slovak, Tamil and Ukrainian are null subject languages. Their grammar permits verbs to lack an explicit subject. This can be a problem for the propagation of subjects of coordinated verbs.

Consider the Polish sentence *"Moje gospodarstwo daje mi zabezpieczenie, mam gdzie wrócić*

(My farm gives me security, I have a place to come back to)". The general rules of subject propagation will propagate the subject *gospodarstwo* from the verb *daje* to the coordinated verb *mam*, which is incorrect because *mam* is at first person and does not require any explicit subject. In order to avoid the propagation, a specific rule marks all first and second person finite verbs, so that they cannot receive a subject dependency. For the third person, there is an ambiguity. In the Polish sentence *"Chłopiec wstaje, otwiera drzwi"*, there are two correct translations in English: *"The boy gets up, opens the door"* and *"The boy gets up, he opens the door"*, because one can propagate the subject *boy* or not. We chose to propagate the subject, which means that in this case, there is no difference between null subject languages and others.

Case addition to the dependency labels for modifiers   For case-based languages, the labels of the dependencies targeting modifiers are augmented with their cases. The rule package implementing this enhancement is trivial.

Left dependents of a coordination   We designed rules to propagate the left dependents of a coordination by dependency types. As we said before, a left dependent on the first conjunct of a coordination is ambiguous: it can depend on the whole coordination or only on the first conjunct. In order to determine, for a given language if a given type of dependency must be propagated or not, we tested the two alternatives on the dev corpus of the shared task and keep the alternative yielding the highest score.

This method has important limits because it depends on the annotation of the gold corpus. Moreover it is very coarse; for a given dependency type, not all dependencies have the same behaviour: some must be propagated, others not. It would be necessary to refine the conversion rules but for that, we need linguistic knowledge about the concerned language.

Raising and control verbs   The default rule we use is to consider that the subject of the raised or controlled verb is the subject of the main verb but this is not always true. A language-specific lexicon should indicate for each of these verbs which argument of the main verb is the subject of the raised or controlled verb. From the training and development corpora available for the task, we have created lexicons for a five languages: Dutch, English, French,
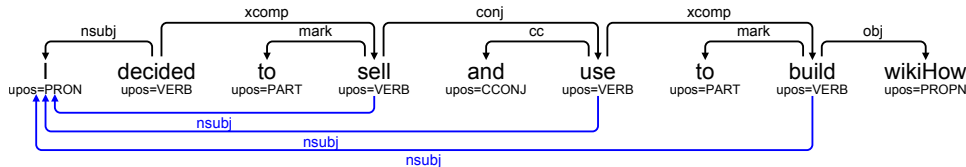
Figure 6: [So] I decided to sell [eHow] and use [the proceeds] to build wikiHow.

Italian and Polish.

2.6 Adaptation to annotation choices

Some annotators of the gold corpora do not strictly follow the guidelines. We have adapted our GRS to their choices on some very specific points.

Dependency label extension specific to one language In Dutch, enhancements for relative clauses distinguish antecedents of relative pronouns that play the deep role of subject and direct object in the relative clause with the extensions `relsubj` and `relobj`. We have designed specific rules to add this extension.

Enhancements partially taken into account We have taken into account the fact that enhancements are only partially achieved for Arabic, Bulgarian, Estonian, French, Russian and Tamil.

Coordinating conjunction raising For the Arabic, Dutch, English, Italian and Swedish treebanks, the names of coordinating conjunctions are added to the corresponding `conj` dependencies, in the same way as for prepositions and subordinating conjunction. We have taken this into account even though it is not indicated in the guidelines.

Propagation of `root` dependencies According to the guidelines, `root` incoming dependencies of a coordination should be propagated like all incoming dependencies, but some treebanks do not and we take this into account.

The French and the Polish treebanks, the latter partially, not only add subjects for raising and control verbs, as mentioned in the guidelines, but add deep subjects for modifier infinitive and participial clauses. Since this goes beyond the guidelines, we do not consider these enhancements.

Table 1 summarizes the three kinds of adaptation to the different languages.

3 Results

Data we submit to the task, called GREW(UDPIPE), is the output of the application of a language spe-
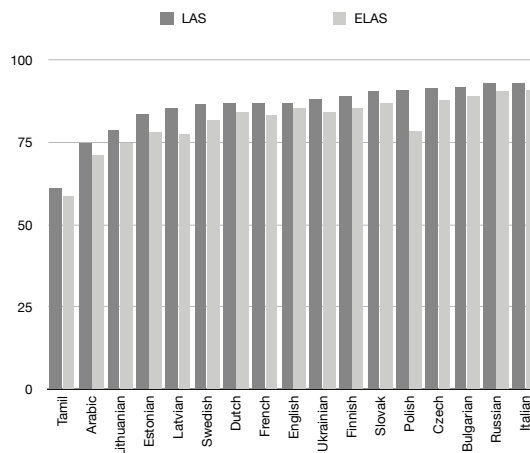
Figure 7: LAS of UDPIPE and ELAS of GREW(UDPIPE)

cific GRS to the output of UDPIPE. Figure 7 shows for each language, the LAS of UDPIPE and the ELAS of GREW(UDPIPE). We can observe that the final result highly depends on the UD annotation quality produced by UDPIPE.

In addition to the results provided by the orgnanisers, we make complementary analysis, focusing on the UD to EUD transformation. In Table 2, we report the ELAS score obtained by our system applied on the gold UD annotation as input. We also recall the same measure reported last year by two others systems (Dehouck et al., 2020; Heinecke, 2020). The score of our system is in most of the case between the two other proposal, closed to Dehouck's system except for Estonian and Swedish.

4 Analysis of discrepancies with the gold annotation In French and English

In order to better understand the behaviour of our GRS, we made some manual inspection of the difference between the annotation GREW(GOLD$_{UD}$) and GOLD$_{EUD}$. We focused on the two languages for which we had the grammatical skills to analyze the discrepancies: French and English.

	ar	bg	cs	nl	en	et	fi	fr	it	lv	lt	pl	ru	sk	sv	ta	uk
null subject	X	X	X			X	X		X			X	X	X		X	X
left dep propag							X			X							
left aux dep propag												X					
left case dep propag										X		X					
left cop dep propag										X		X					
left mark dep propag										X		X					
case raising			X			X	X			X	X	X	X	X		X	X
subj control raising				X	X			X	X			X					
specific extens				X													
partial enhancement	X	X				X		X					X			X	
coord conj raising				X	X				X						X		
root propagation											X	X	X	X		X	X

Table 1: The three kinds of adaptation of the system. Top: adaptation to languages; middle: adding lexical information; bottom: adaptation to specific annotations observed in dev data.

Language Shared task	Dehouck 2020	Heinecke 2020	Our 2021
Arabic	98.8	95.2	98.5
Bulgarian	98.6	97.8	97.6
Czech	97.9	94.7	97.6
Dutch	98.9	94.4	97.6
English	99.5	98.0	99.0
Estonian	99.2	92.6	93.9
Finnish	97.3	94.4	96.9
French	98.9	96.4	99.0
Italian	99.5	98.4	98.8
Latvian	95.7	91.0	92.1
Lithuanian	98.8	94.6	98.2
Polish	94.9	91.1	95.2
Russian	98.6	95.4	98.2
Slovak	98.8	95.4	98.1
Swedish	98.8	96.1	94.7
Tamil	99.3	97.0	98.3
Ukrainian	95.8	94.6	95.9
Average	**98.2**	**95.1**	**97.0**

Table 2: Evalutation of the rule-based systems on Gold UD data: Dehouck (Dehouck et al., 2020), Heinecke (Heinecke, 2020) and our system

4.1 Discrepancies in French

For the French corpus, we observed 589 discrepancies[7] of the computed annotation with the gold annotation, and we manually analyzed the first 100. Table 3 details this analysis.

In order to explain mislabeling in the propagation of incoming dependencies of a coordination, let us return to the example in Figure 2. In propagating incoming dependencies of a coordination that is a modifier, we cannot automatically propagate the label related to the first conjunct to the other conjuncts, because we have to take into account the POS of the heads of these conjuncts. This was not taken into account for the two errors mentioned in the table.

In the table, we have also distinguished errors related to subject or object attributives from errors related to raising and control verbs, because of their particular property: the attributives may have other POS than verb. For exemple, consider the sentence *ils laissent les troupes de la KFOR en paix (they leave the KFOR troops in peace)*. The noun *peace* has an enhanced subject, which is *troops*. The gold annotation ignores this type of subject.

44 discrepancies come from the fact that the gold annotation implements enhancements that are not considered by the guidelines. Columns ¬**EUD** gives a detailed analysis of these enhancements with the number by type. Let us give an example to explain the last type of these non-standard enhancements. Consider the phrase *l'occasion également pour J.-P. Bruneau de présenter ses voeux (the occasion also for J.-P. Bruneau to present his wishes)*. The gold annotation indicates that *J.-P. Bruneau* is the enhanced subject of *present*, which is not

---

[7]Number of differences computed by the `diff` Unix tool: differences in consecutive lines are merged as one difference.

	EUD	UD	GRS	¬EUD
Total	**47**	**1**	**8**	**44**
Non-propagated incoming dependencies of a coordination	29			
Non-propagated left outgoing dependencies of a coordination	6			
Wrongly propagated left outgoing dependencies of a coordination			1	
Non-propagated right outgoing dependencies of a coordination	2			
Label errors in the propagation of incoming dependencies of a coordination	2			
Forgotten subject of controlled verb	3		1	
Verbs considered as control verbs by error			2	
Forgotten subjects of subject attributes	3			
Forgotten subjects of object attributes	2			
Grammaticaly ill-formed sentence			1	
Incorrect handling of light verb constructions			3	
Subjects of epithet participles				28
Subjects of modifier infinitive or participial clauses				14
Subjects of infinitives in constructions NOUN + PREP + NOUN + de + INF				2

Table 3: Manual inspection of the 100 first discrepancies in French between GOLD$_{EUD}$ and GREW(GOLD$_{UD}$). Columns are: errors in EUD gold annotation, errors in UD gold annotation, errors produced by our GRS, EUD gold annotations non described in guidelines.

considered by the guidelines.

4.2 Discrepancies in English

For the English test corpus, we observed 815 discrepancies of the computed annotation with the gold annotation, and we also manually inspected the first 100. Table 4 gives a detailed analysis of these errors with their number by type.

5 Conclusion

We have observed that many conversion problem arise with the CASEDEPREL layer. This layer is of course highly dependent of the language (because lexical information is used in relation definitions). This prevent the new relation to be universal and we believe that this is counterproductive in the objective of a universal description among a large set of languages.

In this paper, we have proposed a rule-based system for computing EUD annotation from UD. Our raw results are far behind the best systems of the task. This can be explained by the fact that we are dependent of the basic UD annotation provided by another tool. Moreover, the manual inspection we have made shows that, at least on English and French, the GOLD test data used in the task are not error-free and contains several annotations that are not described in the guidelines. We can suspect that this is in favour of the learning based approaches which are designed to adapt to the annotated data, completely ignoring the guidelines.

Despite its weakness, we believe that our system have several benefits:

- It has highlighted some places where the guidelines require precisions, like the presence of several `case` or `mark` on the same head;

- It can be used for improving the existing EUD data in the project by identiying annotation error in the current EUD annotations; using a different approach, we can guess that the errors reported will be complementary to the ones that can spotted with other methods;

- Thanks to the modular aspect of the GRS with rules packages adpated to language specificities, it is usable as a starting point for adding EUD annotation layer on languages where there is no such data and where learning based methods cannot be used.

References

Guillaume Bonfante, Bruno Guillaume, and Guy Perrier. 2018. *Application of Graph Rewriting to Natural Language Processing*, volume 1 of *Logic, Linguistics and Computer Science Set*. ISTE Wiley.

Gosse Bouma, Djamé Seddah, and Daniel Zeman. 2020. Overview of the IWPT 2020 shared task on parsing into enhanced Universal Dependencies. In *Proceedings of the 16th International Conference on Parsing Technologies and the IWPT 2020 Shared Task on Parsing into Enhanced Universal Dependencies*, pages 151–161, Online. Association for Computational Linguistics.

Gosse Bouma, Djamé Seddah, and Daniel Zeman. 2021. From raw text to enhanced universal dependencies:

	EUD	UD	GRS	Guidelines
Total	**54**	**17**	**10**	**19**
Non-propagated incoming dependencies of a coordination	9			
Non-propagated outgoing dependencies of a coordination	18		4	
Propagation of incoming dependencies of a coordination	7		1	
Verbs considered as control verbs by error	1			
Forgotten subjects of subject attributes	2			
Forgotten subjects of object attributes	1			
Wrongly processed relative clause constructions	2			
Errors in raising of single labels for case or mark dependencies	11		2	
Errors in raising of fixed labels for case or mark dependencies	3			
`Pos` extension forgotten in `nmod:pos` labels			1	
Wrongly processed ellipsis			2	
Relatives pronouns annotated as interrogative		17		
Two `case`/`mark` on the same source				15
Non alphabetic `ADP` (like "@")				4

Table 4: Manual inspection of the 100 first discrepancies in English between gold annotation and GRS applied on GOLD$_{UD}$. Columns are: errors in EUD gold annotation, errors in UD gold annotation, errors produced by our GRS, cases unspecified in the guidelines.

The parsing shared task at iwpt 2021. In *Proceedings of the 17th International Conference on Parsing Technologies (IWPT 2021)*, pages 146–157, Bangkok, Thailand (online). Association for Computational Linguistics.

Marie Candito, Guy Perrier, Bruno Guillaume, Corentin Ribeyre, Karën Fort, Djamé Seddah, and Éric de la Clergerie. 2014. Deep syntax annotation of the sequoia French treebank. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 2298–2305, Reykjavik, Iceland. European Language Resources Association (ELRA).

Mathieu Dehouck, Mark Anderson, and Carlos Gómez-Rodríguez. 2020. Efficient EUD parsing. In *Proceedings of the 16th International Conference on Parsing Technologies and the IWPT 2020 Shared Task on Parsing into Enhanced Universal Dependencies*, pages 192–205, Online. Association for Computational Linguistics.

Johannes Heinecke. 2020. Hybrid enhanced Universal Dependencies parsing. In *Proceedings of the 16th International Conference on Parsing Technologies and the IWPT 2020 Shared Task on Parsing into Enhanced Universal Dependencies*, pages 174–180, Online. Association for Computational Linguistics.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajič, Christopher D. Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers, and Daniel Zeman. 2020. Universal Dependencies v2: An evergrowing multilingual treebank collection. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 4034–4043, Marseille, France. European Language Resources Association.

Sebastian Schuster and Christopher D. Manning. 2016. Enhanced English Universal Dependencies: An improved representation for natural language understanding tasks. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2371–2378, Portorož, Slovenia. European Language Resources Association (ELRA).

Milan Straka, Jan Hajič, and Jana Straková. 2016. UDPipe: Trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, POS tagging and parsing. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4290–4297, Portorož, Slovenia. European Language Resources Association (ELRA).