Syntactic Semantic Correspondence in Dependency Grammar

Cătălina Mărănduc
Faculty of Computer Science,
"Al. I. Cuza" University
Iasi, Romania
Academic Institute of Linguistics
"Iorgu Iordan - Al. Rosetti",
Bucharest, Romania
catalinamaranduc@gmail.com

Cătălin Mititelu
Bucharest
Romania
catalinmititelu@yahoo.com

Victoria Bobicev
Technical University
of Moldova
Chişinău
Republic of Moldova
victoria.bobicev@ia.utm.md

Abstract

This paper describes the semantic format of the UAIC Ro-Dia Dependency Treebank, based on the previous classical syntactic annotation. The discussed format exploits all the semantic information annotated in the morphological level. The transformation of syntactic annotation into semantic one is made semi-automatically, using a tool called Treeops, which is a converter of an XML format to another XML format, in accordance with a set of rules. Non-ambiguous syntactic relations are transformed automatically, while ambiguous ones are manually corrected. The paper also contains some explanations of the generic rapport between syntactic and semantic structures. We elaborated a set of types of judgement which govern the selection of semantic roles for the syntactic tags based on the morphological ones, which are ambiguous for the semantic annotation. After the creation of the large enough semantically annotated corpus, a statistical semantic parser will be trained for the further automate annotation of ambiguous syntactic relations.

1 Introduction

Natural language theorists have diversified their studies from three perspectives: The **syntax** is the study of relationships between signs, **semantics** is the study of the relationships between the signs and their denotation, and the **pragmatics** is the study of the relations between the signs and the situation of communication.

But as the linguistic sign is a relationship between a form (a significant) and a signifier (de Saussure, 1916), the syntax that studies the relationships between signs can not ignore their signifier. So it is not possible to make a tangible separation of these linguistic layers; we will observe that morphology also contains semantic and even pragmatic data, because persons 1,2,3 refer to roles in the communication situation, as deictics, interjections and some adverbs. Punctuation also has clearly defined pragmatic and semantic functions (Druguş, 2015).

Transformational syntax also tried to separate a surface level of language, and a deep level. Although computer scientists do had not came to an agreement about this, transformation rules are still written, for example, in some question answering programs that can deduce that "the novel is written by Orwell" means "Orwell wrote the novel".

For Chomsky Chomsky (1965), the deep structure is a simple, logical, general one, and the surface one is an evolution from the first; it generally truncates relationships and eliminates redundancy. Both are syntactic structures. But while he starts from the same data Fillmore Fillmore (1968), states that the deep structure is one of semantic roles. The deep structure would thus contain the relations between the signs and the real world in which their denotations are located. The surface structure that remains to be the syntactic one is in fact more abstract than the deep one. The syntax is obtained by abstracting and generalizing the semantic relations of signs with the real world. Fillmore only considers the verb as the center of the communication, and enumerates six cases that he considers the core relationships of the predicate: Agentive, Dative, Instrumental, Factitive (Result), Objective and Locative.

The number of cases is too reduced in this theory, and all the researchers added other cases to this list. It is not clear why only the core of the sentence should have a semantic structure, as if the optional

Proceedings of the 16th International Workshop on Treebanks and Linguistic Theories (TLT16), pages 167–180,
Prague, Czech Republic, January 23–24, 2018.

Keywords: dependency, format converter, Old Romanian, semantic relations, syntactic relations

dependencies were meaningless. Being less abstract than the syntactic structure, the semantic structure should have more roles than the first one.

2 Related Work

There is no universal consensus about semantic annotation, and about the number of semantic categories. Bonial et al. (2014) made the remark that previously, annotation focused on event relations expressed by verbs, but the meaning of words is not necessarily linked to their morphological value - nouns, adverbs, and interjections can also express an event. They propose to expand the PropBank annotations to nouns, adjectives, and complex predicates. This research is called Predicate Unification.

In the UAIC-FII (Faculty of Computer Science, "Al. I. Cuza" University, Iaşi, Romania) NLP (Natural Language Processing) group, Diana Trandabăţ (2010) has imported about 1,000 sentences from the English FrameNet. She has translated in Romanian the sentences and has imported their semantic annotation. In this way, she has made a first set of semantic annotations in Romanian. Just as the English FrameNet, the NLP annotations only cover the core structure of the sentence, called Semantic Frame, the predicate arguments, called Semantic Roles; the semantic functions of other members of the structure are neglected.

Another group of semantic annotations is related to the VerbNet (Kipper et al., 2006), project based on PWN, or to a combination between the two semantic annotation systems: FrameNet and VerbNet (Shi and Mihalcea, 2005). The problem is the same; they emphasize the importance of the predicate and of the action scenarios (events), but the other words of the language make sense as well, and not all judgments describe events.

The UAIC-RoDia DepTb is annotated in Dependency Grammar, a flexible formalism founded by Tesnière (1959), and actualized as Functional Dependency Grammar by Tapanainen and Jarvinen (1998), Mel'čuk (1988). Actually, a big number of corpora in the world have adopted the same formalism. Looking at the old corpora, that have been united during the few years under at the Universal Dependencies (UD) portal, we find that many of them share the same way of development by going from syntactic to semantic annotation.

This group contains a Treebank for the Standard Contemporary Romanian, affiliated by the RACAI (Artifical Intelligence Academic Institute or Research) which imported 4,000 sentences from the UAIC Treebank (of the Al. I. Cuza University). However, not all the Natural Romanian Language is a standard one. A small percent of communication acts are in the standard language; spoken language, poetry, regional and old language, Social Media communication are not in the standard language. In all these styles of communication innovation is permitted. We have decided to annotate all kinds of nonstandard language and we have recently become affiliated with the UD as the UD-Romanian Nonstandard Treebank, which was created at the UAIC, but also has contributors from the Republic of Moldova, a country where Romanian is also spoken.

The other treebanks affiliated with the UD have the same problem; it is an enormous advantage that the annotation conventions are strictly the same, but the attention paid to morphology in the classification of the syntactic relations leads to the loss of semantic information previously annotated in the original formats of affiliated treebanks.

In 2003, the PDT authors described the three level structure of their treebank and the Tectogrammatic level (which includes semantic, logical and syntactic information) (Bohmová et al., 2003). They have for a long time been interested in semantics and its links with syntax (Sgall et al., 1986). In a previous paper (Mărănduc et al., 2017), we have shown that our semantic annotation system has affinities with the PDT Tectogrammatic layer. The authors of BulTreebank are also interested in semantics (Simov and Osenova, 2011). The PENN Treebank is also involved in semantics, or in the annotation of entities and events (Song et al., 2015).

3 Semantic Information in Annotated Data

The syntactic annotation in the classic UAIC format (originally created with the intention to serve pedagogical purposes) contains 14 types of circumstantial modifiers: c.c.conc. (concession), c.c.cond.

Judgment	nsubj	dobj	npred	other
Process	ACT	RSLT	-	-
Performance	PERFR	PERF	QLF	-
Actantial	ACT	PAT	-	BEN
Experience	EXPR	EXP	-	BEN
Comunic.	EMT	CTNT	-	RCPT
Definition	DFND	-	DFNS	CNCOP
Chang.idnt	DFND	-	DFNS	CNCOP
Characteriz	CTNT	-	QLF	CNCOP
Existence	QEXIST	-	-	LOC, TIME

Table 1: The semantic core dependencies in relation of the type of judgment.

(condition), c.c.cons. (consecutive), c.c.cumul. (cumulative), c.c.cz. (causal), c.c.exc. (exception), c.c.instr. (instrumental), c.c.l. (local), c.c.m. (modal), c.c.opoz. (opposition), c.c.rel. (relative, referential), c.c.scop. (purpose), c.c.soc. (associative), and c.c.t. (temporal). For these modifiers, we used the semantic tags: CNCS, COND, CSQ, CUMUL, CAUS, EXCP, INSTR, LOC, MOD, OPPOS, REFR, PURP, ASSOC, TEMP. The modal modifier is the the only one ambiguous among those 14 circumstantial modifiers, which can have more values, as: Comparative, Intensifier, Restrictive, Iterative, Privative, Qualifier, Quantitative, for which we used the tags: COMP, INTNS, ITER, PRV, RESTR, QLF, QNT, manually annotated for the moment. The other syntactic relations with semantic meaning are in the UAIC convention: voc. (vocative, addressee), ap. (apposition, resumption), c.ag. (agent complement), incid. (incident), neg. (negative). For these ones, we used the tags: ADDR, RSMP, ACT (the same tag for the active agent), INCID, QNEG (one of the quantifiers in our system). All these tags can be automatically replaced.

However, these values are not necessarily related to verb subordination. There may be nouns from the semantic sphere of these notions or derived from verbs and having such subordinate semantic values. From a syntactic point of view, they are hidden under ambiguous tags, such as noun modifiers. These cases are also manually annotated. Examples:

- *Bani pentru excursie* "Money for the trip" is a nominal modifier with a purpose meaning;
- *Casa de acolo* "The house there" is an adverbial modifier of a noun, with a local meaning;
- *Generaţia de mâine* "The generation of tomorrow" is an adverbial modifier of a noun having a temporal meaning.

Therefore, we can have syntactic tags containing non-ambiguous semantic information and other tags that are not related to any particular semantic information, i.e. they are semantically ambiguous. Our intention is to use a statistical parser to annotate the words with such semantically ambiguous syntactic relations, after we get a sufficiently large training corpus by means of the manual annotations.

The most ambiguous are the core elements of the clause, and for interpreting them, we propose a table of rules and the roles which each type presupposes or admits. The table can be completed with other types if necessary. Our rules are are not similar to frames, because they also take into consideration the sentences that do not describe an event, but an affirmation of the existence of some things, an identification, a description of a state, acts of speech, etc. (see Table 1 3).

Our treebank is annotated and supervised on a multilayer basis. Therefore, we can use the semantic information contained in the fine and correct morphological annotation of the Treebank. The type of the pronoun and pronominal determiners is semantically established: For the possessives, the semantic value is appurtenance (APP), for the demonstratives, the semantic value is deictic (DX), for the interrogatives the semantic value is INTROG, for the negatives, the semantic value is QNEG, and for the emphatic

pronoun and pronominal determiner the semantic value is IDENT. For the indefinite pronoun and a restricted number of them, the value can be, QUNIV; "all", "whichever", are universal quantifiers, and the rest of indefinites have the semantic value UNCTN (uncertain).

Articles, which come from pronouns, have the same semantic values as these ones: DX, (deictic) for the demonstrative article, APP (appurtenance) for the possessive article, and DEF, UNDEF (defined / undefined) for the determinative articles. The reflexive pronoun can have a restricted number of values, depending on the verb which has this mark and they indicate its possible patterns: impersonal, passive, dynamic, reciprocal or continuant, with the semantic tags: IMPRS, PASS, DYN, RCPR, CTNU. Interjections also have a restricted number of semantic values, in accordance with their word form: affect, alert, imitation, imperative, with the tags: AFF, ALRT, IMIT, IMPER.

As it can be seen, we do not intend to annotate certain entities, such as the ones in information retrieval programs, but semantic categories of great generality and logical connectors or quantifiers. There are similarities with the roles-based models, but we extend this to all the components of the sentence; in addition, judgments are not necessarily seen as events. Our purpose is to make a pattern dictionary of Romanian verbs (PDRoV), taking into consideration, the syntactic relations required for or allowed by each verb. The dictionary will be linked to RoWN (the Word Net for Romanian), and it will take from this dictionary the most particular semantic values for the dependencies.

Verbal dependencies cannot be easily separated into optional and obligatory ones; for some languages, such as Romanian, the presence of the subject in the clause is optional. For some verbs, the presence of local, temporal or quantitative modifiers is mandatory. Examples:

- *to go to Prague* (we cannot say *to go* without showing the target of the movement).
- *The session lasted three hours* (or *a long time*, but not without a temporal determiner).
- *The truck weighs 4 tons*, (or *a lot*, but not without a quantitative determiner).

Of course there are several types of information we have annotated in the semantic format, some are closer to the pragmatic layer and, establishing relations between the participants to the communication act in a certain situation. Interjections, together with deictics, the pronouns of person I-II, with some adverbs and the punctuation link the semantic and the pragmatic levels.

Punctuation has different semantic values when it is at the end of sentence from the cases when it is inside it. In the last position, the dot/full stop marks only the end of the communication, while the exclamation and the question marks indicate both the end and the interrogative or exclamatory forms. The semantic tags for these values are: END, INTROG, EXCL. Inside the sentence, the comma can be the mark of coordination, being a CNCONJ, just as the coordinating conjunctions. Also, the comma can mark the introduction of an explanatory sequence or a topic different from the natural one, some constructions being dislocated. The tags are: ELAB, DISL. A big number of punctuation elements are used to isolate the incident constructions: they are non-appurtenance marks : NOAPP. Other punctuation marks, for example inverted commas, parentheses, dashes, indicate the limits of the text introduced in another text, and we have used for all of them a single semantic tag: QUOT. We have found semantic values in the time and modality of verbs. Some conjunctions or prepositions are specialized for a semantic value: *fiindcă* "because" (CAUS) *pentru* "for" (PURP), etc.

4 Logical-Semantic System

In the UAIC treebank, the relations between clauses are marked with the same labels as those of the words that fulfill the same roles. For a subordinate clause, the tag is annotated as the relation of its predicate with the predicate of the head clause, but it is a relationship of the whole subtree.

Example:

- *Persoanele atente pot învăţa.* "Mindful people can learn." *Persoanele* "People" has the syntactic relation sbj. (and the semantic relation PERFR) subordinated to the root învăţa "learn".

- *Cine are urechi de auzit, poate învăţa.* "Whoever has ears to hear, can learn." The same tags mark the relation of *are* "has", and all the subtree above comma, also subordinated to the root învăţa "learn".

Our trees are not clauses but long sentences, so that their construction can be likened to a logical expression consisting of full-meaning elements and operators to which clauses are connected, and the truth value of the whole sentence can be calculated according to the truth values of the component clauses.

Our system has 6 connectors, the copulative connector (which resembles the logical conjunction), the disjunctive connector (which shows that the clauses are excluded), the opposing connector (which shows that the clauses are opposed without being exclusive), the conclusive connector (which resembles the logical relation of implication), the dependence (subordinate) connector, and the copulative connector. The last usually marks a relationship of equivalence between the subject and the predicative name.

Connectors have the following semantic tags: CNCONJ, CNDISJ, CNADVS, CNCNCL, CNSBRD, CNCOP. The relational words are included among the connected elements in the UAIC syntactic system, being subordinated to the first connected element and simultaneously being the head for the second one. In the UAIC semantic system, we have subordinated them to the second element of the relationship, to emphasize the words with full meaning and especially to conform to most international annotation systems.

Connectors are operators that indicate a relation between two elements. Other operators apply to one element and we call them quantifiers. They form judgments with a general character, which apply to all the set of elements (as universal quantifiers); or they form judgments that apply to at least one element (as an existential quantifier). Other quantifiers modulate the truth value, giving a necessary, possible or impossible character (with negative polarity). Semantic tags used for quantifiers are: QUNIV, QEXIST, QNECES, QPOSIB, QNEG.

Examples:

- Logical computing with dyadic operator (connector):

  El va trece testele sau va fi eliminat din competiţie. "He will pass the tests or he will be eliminated from the competition." *Sau* "or" is a connector for disjunction (CNDISJ). The expression has the truth value=1 (true) if one of the two clauses is denied. The expression obtained by the affirmation of both clauses or by negation of both, has the truth value=0 (false).

  "And" is a connector for the reunion (conjunction) and the expression formed by "and" shows that both the related clauses have the same truth value.

  He will pass the tests and he will be eliminated=0

  He will not pass the tests and he will not be eliminated=0.

- Logical computing with monadic operator (quantifier):

  Trebuie să trec acest test. "I must pass this test." *Nu este posibil să nu trec acest test.* "It is not possible for me not to pass this test."

  The quantifier necessity (QNECES) is equivalent to the negation (QNEG) of the quantifier possibility (QPOSIB) applied to the negation of the modulated sentence as necessary:

  QNECES (to pass this test) is equivalent to QNEG (QPOSIB(QNEG(to pass this test)))

  i.e. "It is necessary that I pass the test" is equivalent to "It is not possible that I do not pass the test."

5 Treeops - A Tool for Changing the UAIC-Syntactic Format in the UAIC-Semantic Format

All non-ambiguous transformations are done automatically using a tool called Treeops. It is a rule-based XML transformer. Having an XML as input and using a customized set of rules, it produces a new XML structure. This process is similar to the eXtensible Style sheet Language Transformation (XSLT)

process[1]. The set of rules is a function that takes as input an XML structure and produces another XML structure. A non-ambiguous transforming rule can be formulated as an `if-then` statement:

```
if (condition) then action
```

During a transformation process the XML is traversed node by node and the Treeops rule is converted into an if-then statement:

```
if (selector matches node) then action
```

Treeops requires the selector to be an XML Path Language (XPath) expression[2]. The action must be internally defined by taking parameters, for example:

```
changeAttrValue(<new value>)
```

changes the value of the current XML attribute.

For this reason, Treeops is currently working only on the XML format, where it takes the name of the features to be changed. In future, the program could be made to have an XML as input and to display the result in the CONLLU format. Obviously, Treeops is language-independent, while the rules are formulated according to the language of the document, and the result will be in the language that is required by the rules (it may be different from the one in the input).

For example, the rule defined as:

```
//word[@deprel='superl.']/@deprel => changeAttrValue('SUPER')
```

becomes an XSLT template:

```
<xsl:template match="//word[@deprel='superl.']/@deprel">
  <xsl:call-template name="changeAttrValue">
    <xsl:with-param name="new_value" select="'SUPER'"/>
  </xsl:call-template>
</xsl:template>
```

where the `changeAttrValue` template is pre-defined as:

```
<xsl:template name="changeAttrValue">
  <xsl:param name="new_value"/>
  <xsl:attribute name="{name(.)}">
    <xsl:value-of select="$new_value"/>
  </xsl:attribute>
</xsl:template>
```

This is a rule with a single condition for transforming a UAIC syntactic tag into a semantic one. It transforms the syntactic relationship superl. into the semantic tag SUPER (Superlative). This type of rule is used to change the syntactic tag of 13 types of circumstantial complement (except c.c.m., which is semantically ambiguous) and also for the relations: vocative, comparative, subordination, agent complement, negation, and apposition.

There are other rules for transforming non-ambiguous syntactic tags into semantic tags that need to fulfill multiple conditions. Example:

```
//word[@deprel='coord.' and (@lemma='sau' or @lemma='ori'
   or @lemma='ci')]/@deprel => changeAttrValue('CNDISJ')
```

This rule changes the syntactic coordination into a logical-semantic tag for the relation of disjunction, taking into account the conjunctions *sau, ori, ci* "or".

There are also more complex rules for the tree structure transformation. The relational elements are used in the UAIC syntactic structure input as heads, and the semantic structure output has the relational elements subordinated to the meaningfull words. Example:

[1]https://www.w3.org/TR/xslt
[2]https://www.w3.org/TR/1999/REC-xpath-19991116/

```
//word[@deprel='narativ.' and @head=../word/@id]/@head
   => (@head <- $n/../word[@id=$n/@head]/@head)
//word[@id=../word[@deprel='narativ.']/@head]/@head
   => (@head <- $n/../word[@deprel='narativ.']/@id)
//word[@deprel='narativ.' and head=../word/@id]/@deprel
   => (@deprel <- $n/../word[@id=$n/@head]/@deprel)
```

This is a transformation rule which changes the syntactic relation "narrative." In the UAIC convention, narrative connectors are treated as textual ones, they are roots for the sentence, having no relation, and subordinating the principal verb of the sentences, which effects the narrative relationship. The above rule reverses the sense of the relationship, i.e. the head becomes subordinated and vice versa; it creates a relationship for the narrative conjunction, and deletes the relationship of its current head, which becomes the root.

What we understand by *a semantically ambiguous syntactic relation* does not mean that the sentence may have more interpretations, but the same general or morphologically defined syntactic relation (eg, nominal modifier) can be transposed into a large number of less general semantic relations. The table in the appendix contains correspondences between UD syntactic tags, UAIC syntactic tangs, and the semantic tags of the formats described here. 214 lines of the table do not mean that the system has 214 semantic tags, but that there are 214 combinations of the 45 UAIC syntactic tags, the 53 tags of Romanian specific subclassifications in the syntactic UD system, and the 96 semantic tags. Empty boxes in the table of judgment types are marked with DASH because there is no specific syntactic or semantic relation to that position in that type of judgment. For example, there is no direct object if the type of judgment requires a predicative name. Examples from the table: On row 199, column 3, we have the *sbj.* tag, which annotates the subject in the UAIC syntactic convention. On the row 199, column 2, four values correspond at it in the UD syntactic convention, those for a word subject, a clause subject, each of them active or passive: *nsubj, csubj, nsubj:pass, csubj:pass*. The tags of the 199 row, if repeated, can correspond to the 199 - 211 rows on the 4th column, i.e. there are 12 possible semantic tags, which demonstrates that the subject is a semantically ambiguous syntactic relationship. The QUOTES mark in this table the repetition of the previous row.

The Treeops program was used both to get an automatically semi-transformed variant of the semantic format (5,566 sentences are completely manually transformed in the semantic format), and, by writing another set of rules, to transform the treebank from the XML-UAIC syntactic format into XML-UD syntactic format. Another program performed the transformation from the XML-UD into CONLLU UDV2, the format required to introduce the first 1,200 phrases in the UD, under the name UD-Romanian Nonstandard. Currently, the UAIC treebank has 18,000 sentences, (except the 4,000 earlier + 1,200 now), 12,800 to be added on the upcoming releases.

6 Applications

These annotations are now applied to 5,200 sentences, most of them the four New Testament Gospels of 1648, the first published in Romanian, with Cyrillic letters, which were obtained by an Optical Character Recogniser (OCR) built at the Institute of Mathematics and Computer Science of Chisinau (Colesnicov et al., 2016). Various research could be applied to the corpus, such as those on incident texts nested in one another. Example:

Iară Iisus zise ucenicilor Săi: Adevăr zic voao: anevoe va întra bogatul întru Împărăţiia Ceriurelor. "And Jesus said unto his disciples, Verily I say unto you: the rich shall scarcely enter the Kingdom of the Heavens."

In this example, the first part, up to the first colon, has the evangelist Matthew as emitter (EMT) and the reader as receiver (RCPT). Jesus and the disciples are designated here by the third person. The second text introduced by the verbum dicendi has as emitter Jesus and as receivers, the disciples. Here Jesus is designated by the first person, and the disciples by the second person. The third text, introduced in the second one by another verbum dicendi and another colon, is the content (CTNT) of Jesus' teaching, a general judgment that does not refer either to himself or to the disciples, but to a generic character, the

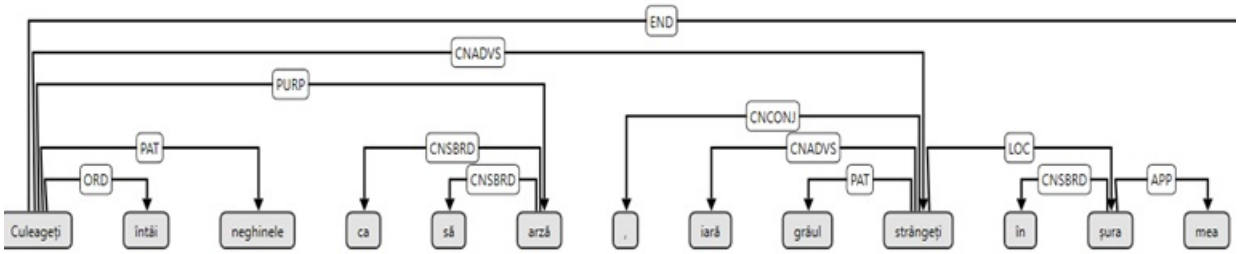

Figure 1: The parabolic original text.

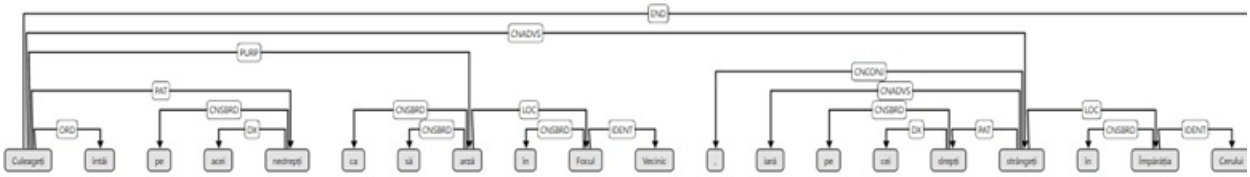

Figure 2: The text obtained by replacing the words with their parabolic key.

rich man, designated by the third person. The structure of roles in the three texts is different, as the verbal persons selection shows.

Another study can analyze the parabolic levels. This means that fully meaningful words are replaced by completely different ones and the connectors are preserved; we have two isomorphic parallel stories, a surface one, and another one containing the meaning. Example:

Omul samănă sămânţă bună în holda sa. Veni duşmanul şi sămănă între grâu neghini. La vreamea secerişului, stăpânul va porunci secerătorilor: Culeageţi întâi neghinele ca să arză, iară grâul strângeţi în şura mea. "Man sow good seed in his field. Come the enemy, and sow the tares among the wheat. At the time of the harvest, the owner will command to the reapers: First reap the tares to burn, and the wheat gather in my barn."

Key:

omul=Fiul omenesc "man=Human Son"

holda=lumea "the field=the world"

sămânţă bună=cei drepţi "good seed=the righteous"

neghinele=cei nedrepţi "the tares=the unrighteous"

duşmanul=diavolul "enemy=devil"

vremea secerişului=sfârşenia veacului "the time of the harvest=the end of the world"

secerătorii=îngerii "reapers=angels"

săarză=să arză în focul veşnic "to burn=to burn in the Eternal Fire"

şura mea=Împărăţia Ceriului "my barn=the Kingdom of the Heaven"

Replacement:

Fiul omenesc samănă pe cei drepţi în lume. Veni diavolul şi sămănă între cei drepţi pe acei nedrepţi. La sfârşenia veacului, Fiul omenesc va porunci îngerilor: Culegeţi întâi pe cei nedrepţi ca să arză în Focul veşnic, iară pe cei drepţi strângeţi în Împărăţia Cerului. "The human Son sows the righteous in the world. The devil comes and sows those unrighteous among the righteous. At the end of the world, the Human Son will command to the angels: First reap those unrighteous to burn in the Eternal Fire, and the righteous gather in the Kingdom of Heaven."

In Figures 1 and 2 it can be seen that although the semantic contents of the nodes change, the basic structure remains the same (see Figure 1 and 2).

7 Conclusion and Future Work

In this paper we discussed the transformation process of UAIC RoDia-Dep-Treebank syntactic annotation into the logical-semantic annotation. This transformation is done automatically for non-ambiguous syntactic relations, and manually for ambiguous relations. We also described the applications created for the annotation format tranformation. We show the examples of the linguistic and pragmatic research using corpora with semantic annotation.

In future, we plan to annotate morphologically and syntactically the second part of the New Testament, the Acts of Apostles, and to transform all the syntactic treebank into the new format. We plan to train a statistical parser on this corpus, in order to transform the ambiguous syntactic relations.

References

Alena Bohmová, Jan Hajič, Eva Hajičova, and Barbora Hladka. 2003. *The Prague Dependency Treebank: A Three-Level Annotation Scenario. Text, Speech and Language Technology*. Springer Publisher, Prague.

Claire Bonial, Julia Bonn, Kathryn Conger, Jena D. Hwang, and Martha Palmer. 2014. Propbank: Semantics of new predicate types. In *Proceedings of LREC*. pages 3013âĂŞ–3019.

Noam Chomsky. 1965. *Aspects of the Theory of Syntax*. MIT Press, Cambridge, UK.

Alexandru Colesnicov, Ludmila Malahov, and Tudor Bumbu. 2016. Digitization of romanian printed texts of the 17th century. In *Proceedings of the 12th International Conference Linguistic Resources and Tools for Processing the Romanian Language*. Alexandru Ioan Cuza University Press, pages 1–11.

Ferdinand de Saussure. 1916. *Cours de Linguistique generale*. Payot, Paris.

Ioachim Druguş. 2015. Metalingua: a metalanguage for the semantic annotation of natural languages. In *Proceedings of the 11th International Conference Linguistic Resources and Tools for Processing the Romanian Language*. Alexandru Ioan Cuza University Press, pages 79–94.

Charles J. Fillmore. 1968. The case for case. In *Universals in Linguistic Theory*. Holt, Rinehart, and Winston, pages 1–88.

Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2006. Extensive classifications of english verbs. In *Proceedings of the 12th EURALEX International Congress*.

Cătălina Mărănduc, Monica Mihaela Rizea, and Dan Cristea. 2017. Mapping dependency relations onto semantic categories. In *Proceedings of the International Conference on Computational Linguistics and Intelligent Text Processing (CICLing)*.

Igor Mel'čuk. 1988. *Dependency Syntax: Theory and Practice*. The SUNY Press, Albany, NY.

Petr Sgall, Eva Hajičová, and Jarmila Panevová. 1986. *The meaning of the sentence in its semantic and pragmatic aspects*. Academic Press and Reide, Prague, Dordrecht.

Lei Shi and Rada Mihalcea. 2005. Putting pieces together: Combining framenet, verbnet and wordnet for robust semantic parsing. In *Proceedings of the Sixth International Conference on Intelligent Text Processing and Computational Linguistics*. Mexico.

Kiril Simov and Petya Osenova. 2011. Towards minimal recursion semantics over bulgarian dependency parsing. In *Proceedings of the RANLP 2011 Conference*. Hissar, Bulgaria.

Zhiyi Song, Ann Bies, Stephanie Strassel, Tom Riese, Justin Mott, Joe Ellis, Jonathan Wright, Seth Kulick, Neville Ryant, and Xiaoyi Ma. 2015. From light to rich ere: Annotation of entities, relations, and events. In *Proceeding of 14th Annual Conference of the North American Chapter of the Association for Computational Linguistics*.

Pasi Tapanainen and Timo Jarvinen. 1998. Towards an implementable dependency grammar. In *CoLing-ACLâĂŹ98 workshop Processing of Dependency-based Grammars, Montreal*.

Lucien Tesnière. 1959. *Eléments de syntaxe structurale*. Klincksieck, Paris.

Diana Trandabăţ. 2010. *Natural Language Processing Using Semantic Frames - PHD Thesis*. Faculty of Computer Science, Al. I. Cuza University, Iaşi.

A Table of semantic tags, their explanation, and their correspondence whit UAIC and UD syntactic tags

Nr.crt.	UD syntactic	UAIC syntactic	UAIC semantic	Explanation
1	amod, det, nummod	a.adj.	COMP	Comparative
2	”	”	QEXIST	Quantifier:existential
3	”	”	QUNIV	Quantifier:universal
4	”	”	DX	Deictic
5	”	”	IDENT	Identifier
6	”	”	INTROG	Interrogative
7	”	”	QNEG	Quantifier:negative
8	”	”	APP	Appurtenance
9	”	”	QLF	Qualifier
10	”	”	QNT	Quantity
11	”	”	UNCTN	Uncertain
12	advmod	a.adv.	LOC	Local
13	”	”	MOD	Modal
14	”	”	PRV	Privative
15	”	”	RESTR	Restrictive
16	”	”	ITER	Iterative
17	”	”	TEMP	Temporal
18	appos	ap.	RSMP	Resumption
19	nmod	a.pron.	QUNIV	Quantifier:universal
20	”	”	QEXIST	Quantifier:existential
21	”	”	DX	Deictic
22	”	”	IDENT	Identifier
23	”	”	INTROG	Interrogative
24	”	”	QNEG	Quantifier:negative
25	”	”	APP	Appurtenance
26	”	”	UNCTN	Uncertain
27	nmod	’a.subst.	ASSOC	Associative
28	”	”	CAUS	Causative
29	”	”	CNCS	Concessive
30	”	”	COND	Conditional
31	”	”	CSQ	Consequence
32	”	”	CUMUL	Cumulative
33	”	”	DFNS	Definiens
34	”	”	EXCP	Exception
35	”	”	INSTR	Instrumental
36	”	”	LOC	Local
37	”	”	MOD	Modal
38	”	”	OPPOS	Opposite
39	”	”	PARS	Pars
40	”	”	APP	Appurtenance
41	”	”	POLIT	Politness
42	”	”	PRV	Privative
43	”	”	PURP	Purpose
44	”	”	REFR	Reference
45	”	”	TEMP	Temporal

Nr.crt.	UD syntactic	UAIC syntactic	UAIC semantic	Explanation
46	aux	aux.	ABIL	Ability
47	”	”	FTR	Future
48	”	”	OPTV	Optative
49	aux:pass	”	PASS	Passive
50	aux	”	PAST	Past
51	”	”	POTN	Potentiality
52	acl	a.vb.	ASSOC	Associative
53	”	”	CAUS	Causative
54	”	”	CNCS	Concession
55	acl	a.vb.	COND	Condition
56	”	”	CSQ	Consequence
57	”	”	CUMUL	Cumulative
58	”	”	DFNS	Definiens
59	”	”	EXCP	Exception
60	”	”	INSTR	Instrumental
61	”	”	LOC	Local
62	”	”	MOD	Modal
63	”	”	OPPOS	Opposite
64	”	”	PARTV	Partitive
65	”	”	RESTR	Restrictive
66	”	”	PAT	Patient
67	”	”	POLIT	Politness
68	”	”	APP	Appurtenance
69	”	”	PRV	Privative
70	”	”	PURP	Purpose
71	”	”	QLF	Qualifier
72	”	”	REFR	Reference
73	”	”	TEMP	Temporal
74	”	”	ASSOC	Associative
75	”	”	CAUS	Causative
76	nmod:agent	c.ag.	ACT	Actant, Agent
77	obl, advmod, advcl	c.c.conc.	CNCS	Consequence
78	”	c.c.cond.	COND	Condition
79	”	c.c.cons.	CSQ	Consequence
80	”	c.c.cumul.	CUMUL	Cumulative
81	”	c.c.cz.	CAUS	Causative
82	”	c.c.exc.	EXCP	Exception
83	”	c.c.instr.	INSTR	Instrumental
84	”	c.c.l.	LOC	Condition
85	”	c.c.m.	MOD	Modal
86	”	”	INTNS	Intensifier
87	”	”	ITER	Iterative
88	”	”	PRV	Privative
89	”	”	RESTR	Restrictive
90	”	”	MOD	Modal
91	”	”	QLF	Qualifier
92	”	”	QNT	Quantity
93	”	c.c.opoz.	OPPOS	Opposite
94	”	c.c.rel.	REFR	Reference

Nr.crt.	UD syntactic	UAIC syntactic	UAIC semantic	Explanation
95	”	c.c.scop.	PURP	Purpose
96	”	c.c.soc.	ASSOC	Associative
97	nmod:tmod, adv-mod:tmod, advcl:tcl	c.c.t.	TEMP	Temporal
98	obj, expl, ccomp	c.d.	QPOSIB	Quantifier:possibility
99	”	”	BEN	Beneficiary
100	”	”	CTNT	Content
101	”	”	EXP	Experience
102	”	”	GREET	Greeting
103	”	”	INSTR	Instrumental
104	”	”	APP	Appurtenance
105	”	”	PURP	Purpose
106	”	”	OBJ	Object
107	”	”	RSLT	Result
108	”	”	PAT	Patient
109	iobj, expl, xcomp	c.i.	PERF	Performance
110	”	”	RSLT	Result
111	”	”	BEN	Beneficiary
112	”	”	EXPR	Experiencer
113	”	”	RCPT	Receiver, Recipient
114	”	”	APP	Appurtenance
115	advmod	comp.	COMP	Comparative
116	cc, conj	coord.	CNCNCL	Connect:conclusion
117	”	”	CNDISJ	Connect:disjunction
118	orphan	-	EQVH	Ellipse, Equivalent with the head
119	”	”	EQVHP	Equivalent with the head, but positive
120	”	”	EQVHZ	Equivalent with the read, but negative
121	cc, conj	”	CNADVS	Connect:adversative
122	”	”	CNCONJ	Connect:reunion
123	cop	-	CNCOP	Connect:copulative
124	nmod:pmod	c.prep.	ASSOC	Associative
125	”	”	BLAM	Blam
126	”	”	BEN	Beneficiary
127	”	”	CAUS	Causative
125	”	”	BLAM	Blam
128	”	”	CNCS	Concession
129	”	”	COND	Condition
130	”	”	CSQ	Consequence
131	”	”	CTNT	Content
132	”	”	CUMUL	Cumulative
133	”	”	EQVL	Equivalent
134	”	”	EXCP	Exception
135	”	”	EXP	Experience

Nr.crt.	UD syntactic	UAIC syntactic	UAIC semantic	Explanation
137	”	”	OPPOS	Opposite
138	”	”	PURP	Purpose
139	”	”	RCPT	Recipient, Receiver
140	”	”	REFR	Reference
141	det	det.	UNDEF	Undefined
142	”	”	DEF	Defined
143	”	”	DX	Deictic
144	”	”	APP	Appurtenance
145	xcomp	el.pred.	QNT	Quantity
146	”	”	UNCTN	Uncertain
147	”	”	DFND	Definiendum
148	”	”	EQVL	Equivalent
149	”	”	EXPR	Experiencer
150	”	”	IDENT	Identifier
151	”	”	PERF	Performance
152	”	”	APP	Appurtenance
153	”	”	RESTR	Restrictive
154	”	”	RSLT	Result
155	”	”	QLF	Qualifier
156	expl	-	EXPL:APP	Expletive:appurtenance
157	”	”	EXPL:BEN	Expletive:beneficiary
158	”	”	EXPL:EXP	Expletive:experience
159	”	”	EXPL:EXPR	Expletive:experiencer
160	”	”	EXPL:OBJ	Expletive:object
161	”	”	EXPL:DFND	Expletive:definiendum
162	”	”	EXPL:PAT	Expletive:patient
163	”	”	EXPL:RCPT	Expletive:receiver
164	parataxis	incid.	INCID	Incident
165	discourse	interj.	AFF	Affect
166	”	”	ALRT	Alert
167	”	”	IMIT	Imitation
168	”	”	IMPER	Imperative
169	cc	narativ.	CNCNCL	Connect:conclusion
170	mark	”	CNSBRD	Connect:subordination
171	cc	”	CNDISJ	Connect:disjunction
172	”	”	CNADVS	Conect:adversative
173	”	”	CNCONJ	Connect:reunion
174	-	n.pred.	RSMP	Apposition
175	”	”	EMT	Emitter
176	”	”	DFNS	Definiens
177	”	”	EXP	Experience
178	”	”	IDENT	Identifier
179	”	”	APP	Appurtenance
180	”	”	PRV	Privative
181	”	”	QLF	Qualifier
182	mark	part.	GNR	Generic
183	”	”	GREET	Greeting
184	”	”	POTN	Potentiality
185	”	”	IMPER	Imperative
186	punct	punct. (non-final)	DISL	Dislocation

Nr.crt.	UD syntactic	UAIC syntactic	UAIC semantic	Explanation
187	”	”	QUOT	Quotation
188	”	”	NOAPP	Non-appurtenance
189	”	”	CNCONJ	Connect:reunion
190	”	”	ELAB	Elaboration
191	”	punct.(final)	END	End
192	”	”	EXCL	Exclamation
193	”	”	INTROG	Interrogative
194	expl:pv, expl:poss	refl.	CTNU	Continuant
195	”	”	DYN	Dynamic
196	”	”	RCPR	Reciprocal
197	expl:pass	”	PASS	Passive
198	expl:impers	”	IMPRS	Impersonal
199	nsubj, csubj, nsubj:pass, csubj:pass	sbj.	ACT	Actant, Agent
200	”	”	PERFR	Performer
201	”	”	PERF	Performance
202	”	”	DFND	Definiendum
203	”	”	EMT	Emitter
204	”	”	QEXIST	Quantifier:existence
205	”	”	QUNIV	Quantifier:universal
206	”	”	QUPOSIB	Quantifier>possibility
207	”	”	QUNECES	Quantifier>necessity
208	”	”	EXPR	Experiencer
209	”	”	EXP	Experience
210	”	”	PAT	Patient
211	”	”	RCPT	Receiver
212	mark	subord.	CNSBRD	Connect:subord
213	advmod	superl.	SUPER	Superlative
214	vocative	voc.	ADDR	Addressee