ISSN 1870-4069

Mapping Dependency Relations onto Semantic Categories

Cătălina Mărănduc[1], Monica Mihaela Rizea[2], Dan Cristea[3]

[1] Al. I. Cuza University, Faculty of Computer Science,
Academic Linguistics Institute I. Iordan – Al. Rosetti Bucharest,
Romania

[2] University of Bucharest,
Solomon Marcus Center for Computational Linguistics,
Romania

[3] Al. I. Cuza University, Faculty of Computer Science,
Institute for Computer Science, Romanian Academy - Iasi Branch,
Romania

{catalina.maranduc,dan.cristea}@info.uaic.ro,
monicamihaelarizea@gmail.com

Abstract. The paper focuses on a dependency treebank that aims to illustrate the Romanian language in more styles of communication and in more geographical and historical variants. The treebank, called UAIC-RoDiaTb, contains 18,630 sentences and it is freely available. The treebank is affiliated to UD (Universal Dependencies), project (with 3,700 sentences illustrating Contemporary Standard Romanian, and with 1,200 sentences illustrating Non-standard Romanian). However, the UD annotation system is simpler than ours, and the affiliation of our treebank is possible only with loss of information. We aim to establish an original system of semantic annotation exploiting all the semantic information contained in our treebank (i.e. in the syntactic categories, in the morphological analysis, in the lexical definition of some words, and in the punctuation). We developed some logical structures similar to the AMR (Abstract Meaning Representation) system, but we intend to maintain the form of the Functional Dependency Grammar (FDG) trees for the semantic layer, in order to preserve the isomorphism with the syntactic one. The chosen solution will be justified and compared with other systems.

Keywords: Dependency treebank, non-standard Romanian, logic-semantic layer of annotation, correspondences syntactic-semantics, similarities with other international systems.

1 Introduction

1.1 Perspectives on Developing UAIC-RoDiaTb

The UAIC Dependency Treebank[1] has become an important corpus for Romanian language, with rich morphologic and syntactic information. This treebank is balanced

[1] Software | UAIC NLP (Natural Language Processing) Group, UAIC-RoDia = ISLRN 156-635-615-024-0.

and attempts to illustrate all the styles of the language; the average is 19.29 words per sentence. We consider that the purpose of Natural Language Processing (NLP) is to model the complexity of the human language, and not only to model the man – computer communication in a simplified way.

Although we have made the transposition table for automatically transposing our conventions into UD ones, part of the UAIC-RoDiaTb was transposed in the UD by the RACAI group (Research Institute of Artificial Intelligence), interested only in Contemporary Standard Romanian.

There are many theoretical problems that differentiate us from UD; for example, the treatment of relational words. The syntactic categories are classified according to the UD conventions in what concerns the morphological classes (i.e., adjectival, adverbial, nominal modifier); additionally, we consider that the syntactic information should be correlated with the semantic one.

The semantic richness of our tags derives from the UAIC annotation conventions that are not modern, but classically syntactic, containing 14 different circumstances carefully checked. The challenge is to build mechanisms for the automatic recognition of these 14 values when they are not determining verbs (i.e., a classification as local, temporal, causal determiner of a noun would be useful). We do not claim the perfection of the syntax-semantics isomorphism; however, we watch it closely in order to discover and to surpass its limitations.

In order to preserve all the information which has been automatically annotated and carefully supervised in all the 18,630 sentences, we propose a project that aims to transform the classical treebank into a semantic layer. The similarity of the syntactic and semantic functions has been widely discussed. Fillmore [8], focusing on the transformational grammar theory, has described the deep structure as a semantic one, with categories such as: Agent, Instrument, Objective, Dative, Locative, and Factitive. In his conception, the syntax is the surface structure. Both structures have a single position for each argument.

This conception, underpinning the Semantic Roles of PropBank or FrameNet, is largely used by the computational linguists. However, Fillmore's deep structure has very few semantic categories for the purpose presented above, i.e. only the obligatory verbal dependencies. The modifiers (being mandatory for some verbs) have also semantic roles, the functional words, the markers of tenses, modes, and diathesis, the articles and other morphological categories, and also the punctuation has semantic functions that can be used by other applications such as sentiment analysis, information retrieval, question answering, or temporal structuring of the discourse. This research can be an experiment on Romanian with possible multilingual value.

The syntactic categories are more abstract than the semantic ones; the latter refer to the "deep structure", only in the sense that they are nearest to the communicative purpose. We consider the syntactic and the semantic structures as parallel, since the FDG [14, 20] rejects the concept of deep structure.

For this purpose, we built a new working interface, called Treebank Annotator [11], that allows viewing and comparing two trees of the same sentence in different annotation conventions, and working alternatively with them, see Fig. 1.

Half of the syntactic tags can be automatically replaced with semantic ones. The logico-semantic- argumentative system chosen is able to preserve the annotated and

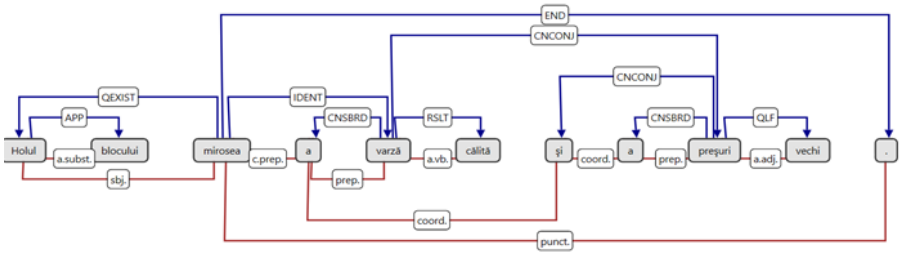

Fig. 1. Comparison between the syntactic and the semantic annotation (The hallway of the block smelled of tanned cabbage and old pretzels).

supervised information of the UAIC-RoDiaTb[2]. The paper also contains a theoretical justification of the system described and an attempt to find future affiliations to international compatible systems. The system is compared with similar works and other solutions will be discussed.

1.2 UAIC-RoDiaTb Syntactic System of Annotation

The UAIC-RoDiaTb is a syntactic treebank based on the FDG grammar. In FDG, the connection is a functional binary relationship between a regent and a dependent. By using this model, the text processing is uniform; any element of the structure, word or punctuation mark, is a node of the tree, in relation to another node.

This kind of treebank is currently being developed for more and more languages. All these corpora respect the FDG rules; however, there are multiple differences between their annotation systems. The coordination is quite difficult to express in this theoretical model, in which the relationship of equality, (horizontal) is not allowed. In our system, the coordination is asymmetrically rendered; the coordinator element is subordinated of the first coordinated and regent for the second.

Another problem with divergent solutions in FDG is the annotation of the relational words. In the Penn treebank, these words are subordinated of the full-meaning words or of the head of clauses that they introduce in the tree. This solution is taken over by many corpora and by the UD system. However, recently, researchers have begun to wonder if this solution does not have more disadvantages than advantages.

(This was one of the proposed themes for the researchers at the Depling 2015 Conference)[3]. In the UAIC-RoDiaTb, the relational words are connectors between the head and the subordinate word, i.e. regents for the word they introduce. The first arguments for this decision are:

- The relational words are marks for the subordination or coordination. The prepositions also impose of the subordinated word the determined or undetermined form, and the case,
- The number of the functional words is small since this lexical category is closed; each of them can be described, they have a number of relations that it can

[2] RoDia is the Romanian word for pomegranate; read on syllables means Romanian Diacronic Treebank.

[3] http://depling.org/depling2015/.

establish, and some formal restrictions imposed of the subordinate. It is possible to formulate a system of rules for a syntactic parser or for a machine learning hybrid or rule-based, given the properties of such relational words.

In our system, the subordinator elements are annotated consequently with the coordinators ones, as subordinated of the regent and regent for the subordinated word or clauses that they introduced.

The syntactic tags of the UAIC-RoDiaTb are classical; for example, there are 14 types of circumstantial modifiers. In fact, the content of this classification of verbal modifiers is more semantic than syntactic, so it contains precious information.

The UD system is based on morphologic information because the morphology part has already been correctly annotated, and the syntactic parsers work better when they make use of this type of information. However, there is redundancy in the UD system that marks twice the morphological information (in the POS-tag and in the dependency tag). Our semantic classification of modifiers is extra information that can be used by future applications.

2 Related Work

2.1 Related Work for Romanian

In the UAIC NLP group, Trandabăț [21] has imported about 1,000 sentences from the English FrameNet. She has translated in Romanian the sentences and has retained their semantic annotation from the English FrameNet. In this way, she has made a first set of semantic annotations on Romanian sentences. Like the English FrameNet, these annotations only cover the core structure of the sentence called Semantic Frame, the (mandatory) predicate arguments, called Semantic Roles, and the semantic function of others members of the structure is not analyzed.

Another research made the classification of verbs like the English VerbNet. The classes of verbs are carefully inventoried for English [13]. The participants at Eurolan 2013 Summer School tried to find some corresponding examples for such classes in eDTLR[4], the electronic transposition of the paper Thesaurus Dictionary [1], to illustrate each class of the Romanian VerbNet [6].

Therefore, in Romanian the most frequent patterns are different and with specific structures by rapport to English. More recent research has shown that there are other parts of speech that can also be logical predicates and can have the same structures of arguments, especially the nouns or adjectives derived from verbs, and the semantic roles must be extended to these other heads (see below).

2.2 Diversity of Semantic Approaches

There is no universal consensus about semantic annotation, and the number of semantic categories is also disputed. Many papers propose a small number of semantic relations

[4] eDTLR was built during 2007-2010 by a group of computer scientists and linguists from three Academic Linguistic Institutes from Iasi, Cluj, and Bucharest, and from two Artificial Intelligence Academic Institutes.

[9-10]. Amaro [2] describes a more extensive corpus work: she analyses a number of 26 relationships from the Portuguese WordNet and studies them in 35,000 contexts, with the aim of creating a semantic annotation based on lexical and syntactic information. In another paper [5], the authors have the purpose of developing semantic annotations for the PAS (Predicate Argument Structure) in the PropBank [17]. They exemplify some semantic tags used by English Vallex, (e.g., ACT (Actor), PAT (Patient), ADDR (Addressee), ORIG (Origin) and EFF (Effect), considering them as being too descriptive for their purpose.

Previously, the annotation effort has focused on event relations expressed solely 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 consider it necessary to expand the PropBank annotations so as to provide coverage for nouns, adjectives, and complex predicates. This research is called Predicate Unification.

The FrameNet annotations for these various logical predicates split them in different frames. For example, fear-noun fall into the 'Fear' frame, fear-verb falls into the 'Experiencer Focus' frame, and afraid-adjective is included in both. As a result, sentences describing the same eventuality would not be recognized as synonymous under the FrameNet annotation.

On the contrary, Ștef[nescu [19], study the semantic similarity experiencing latent semantic analysis models on two large corpora, Wikipedia and TASA (Touchstone Applied Science Associates). The clustering model investigates the similarity between words without proposing any ontology (tags for syntactic relationships). In our NLP group, a similar research is RoPAAS (Romanian Predicate Argument and Adjunct Structure), see [18]. The semantic logic Romanian approaches also consider that the logical predicates can be expressed by adjectives, interjections and adverbs [24].

2.3 Comparison between UAIC Semantic Ontology and the Tectogrammatic Layer of the Prague Dependency Treebank (PDT)

The PDT[5] is a long-term project, started in 1996. In 2003, in [4], the Czech researchers described them as a three-level annotated corpus of 1.8 mil. tokens. The first level is the morphological annotation; the second is the superficial syntactic annotation, affiliated to UD, and the third one is called the tectogrammatical level, or the level of linguistic meaning (based on the framework of Functional Generative Description).

The UAIC-RoDiaTb also has a first morphological level, but it is included in the syntactic level. The superficial syntactic layer of PDT has less information than the old classical syntactic layer of the UAIC-RoDiaTb, and the verbal modifiers are not classified. The tectogrammatic layer of PDT is obtained after some transformations. There are a big number of relations called "functions" abbreviated "func":

/ACT/PAT/ADDR/EFF/ORIG/ACMP/ADVS/AIM/APP/APPS/ATT/BEN/CAUS/CNCS/COND/CONJ/COMPL/CPR/CRIT/CSQ/CTERF/DENOT/DES/DIFF/DIR1/DIR2/DIR3/DISJ/ETHD/EXT/FRWH/GRAD/ID/INTF/INTT/HER/LOC/MANN/MAT/MEANS/MOD/NORM/PAR/PREC/REAS/REG/RESL/RESTR/RHEM/RSTR/SUBS/TFHL/THL/THO/TOWH/TPAR/TSIN/TTILL/TWHEN/VOC/VOCAT/NA/SENT/

[5]http://ufal.ms.mff.cuni.cz/pdt/pdt.html

The information is organized in more attributes, each of its values, while in our system, all semantic information is encoded in the deprel attribute: modifiers are classified in a similar way as in the classical syntactic convention of UAIC- RoDepTb and in the ontology proposed below: CNCS = Concession, CAUS = Causative, CSQ = Consequence, ADDR = Addressee, COND = Conditional, etc.; moreover, an automatic transposition of the UAIC semantic system of annotation in the PDT system will be possible.

In a report published on PDT site [16], the recent modifications and the direction of the development of this large resource have been described. The PDT texts are in the journalistic style, illustrating the contemporary language.

The complexity of the information annotated in the third layer of the PDT allows for many different lines of research.

A preoccupation for functional words, and the intention of annotating the meaning of these words is common for PDT.02 and the ontology presented below; i.e., the meaning of grammatical categories that these words form can be semantically annotated as past, passive, reflexive, reciprocal, continuous, etc.

Different sections of the report describe the attributes *grammatemes: typgroup, factmod, diatgram.* The *typgroup* aims at refining the category of singular or plural, the *factmod* is conceived to annotate the meaning of the mode of the verb, which expresses a real action, or a possible, uncertain, claimed action.

In the ontology proposed below, there are the following categories: Past, Future, Continuous for annotating the *grammatemes*, and the categories Generic, Uncertain, Imperative, and Optative for the *factmod*.

The attribute *diatgram* formalizes the meaning of the verb diathesis (voices): Passive, Reciprocal in our system. For the PDT Refl 1 and 2, UAIC have the tags: Continuative, Dynamic, and Impersonal.

The values of *sentmod* attribute are: *enunc, excl, dezid, imper, inter,* [22] which annotate the type of the sentence, from the pragmatic perspective of the sentence emitter. The system that we present has also the categories Interrogative, Imperative, Exclamation.

The development of this important resource, PDT, especially after the introduction of the tectogrammatic layer, shows how vast the prospects that open the semantic annotation for the future reuse of the corpus are.

2.4 Comparison between UAIC Semantic Ontology and the Abstract Meaning Representation (AMR) Project

In this paper, „ontology" refers to the set of semantic-logic concepts chosen as tags for the semantic annotation, and to the attribution of these tags to syntactic relations, to morphologic categories, or to particular words.

AMR is a semantic representation language proposed by Bănărescu [3] that uses graph notations for computer processing and a modified form of the PENN annotation [11] for human reading and writing. AMR graphs are rooted, labeled, directed, acyclic (DAGs); they are able to represent various linguistic phenomena, such as semantic roles, co-reference, questions, modals and negation, named entities, copula, reification, and so on [23]. The nodes of an AMR graph are labeled with concepts while the edges are labeled with relations.

Similarly, to the UAIC ontology, AMR provides full sentence deep semantic representations, not only the mandatory relations. AMR annotates sentences independent of context (i.e. it takes the sentence, and not the text as the unit of annotation). Even if some discourse relations, such as contrast ('but') and concession ('even though'), and co-reference are already represented in the AMR annotations at intra-sentential level, the future directions for AMR imply new possibilities for representing the inter-sentential co-reference and discourse relations.

Another similarity is the use of a relatively large number of semantic relations. AMR uses approximately 100 relations, (the UAIC system has 96) that include frame arguments adopted from the PropBank annotations in OntoNotes [12], and other semantic relations. The paper [3] give a comprehensive list of these relations:

- **General semantic relations**: :accompanier, :age, :beneficiary, :cause, :concession, :condition, :consist-of, :degree, destination, direction, domain, duration, etc.,
- **Relations for quantities**: :quant, :unit, :scale,
- **Relations for date-entities**: :day, :month, :year, etc.

The difference between AMR and the UAIC system is that the first one does not annotate word tokens in a sentence, but concepts. This means that AMR generalizes over morpho-syntactic idiosyncrasies such as word category, word order, or morphological variation. Consequently, content words are annotated as concepts (i.e. they drop such information as plurality, articles, or tense and aspect) and they can correspond either to predicate-like elements (e.g.: 'teach', 'teacher', 'attractive', 'acquainted'), or to special (English) keywords (e.g.: 'person', 'name', special entity-types 'distance-quantity' and logical conjunctions 'and', or', etc.). The function words are either annotated by means of the semantic relations they represent, or omitted if they do not contribute to the meaning of a sentence.

AMR differs from other ontologies since it combines multiple layers of linguistic annotation in a single structure with the aim of obtaining a high degree of generalization, both at one-language level and cross-linguistically. Keeping a single structure rather than multiple layers is also correlated with losing a lot of language-specific information since AMR is generally considered to 'abstract away' from the morphological and syntactic variations that are present in a language, and this accounts for many of the cross-lingual differences.

Although this approach leads to one of the AMR strengths, which is the ability to encode in a single representation multiple sentences sharing the same meaning even if not identically worded, it also results in some 'side-effect' limitations. These refer, for example, to the fact that AMR does not distinguish between real events and hypothetical, future, or imagined ones since it does not encode tense and aspect features because they do not generalize well cross-linguistically [3, 23].

However, the AMR quality of collapsing more ways of saying things made it interesting for MT (Machine Translations) experiments, in order to see if this representation can serve, e.g., as a useful, minimally divergent transfer layer in machine translation [26].

Table 1. The syntactic and semantic roles depending on the type of judgment.

Type of judgment	SSubject	Direct object	Predicative noun	Ot **Other**
Process	Agent	Result	none	none
Performance	Performer	none	Qualifier	Circonstances
Actantial	Agent	Patient	none	none
Experience	Experiencer	Experience	none	Circonstances
Existence	Existent	none	none	Circonstances
Communicative	Emitter	Content	none	Recipient
Definition	Definiens	none	Definiendum	copula
Identity changing	Definiens	none	Definiendum	copula
Possession	Possessor	Posseded	none	none
Characterization	Theme, content	none	Qualifier	copula

3 Mapping the Classical Syntactic Relationships onto the Semantic Categories Proposed

3.4 Monosemantic Syntactic Tags

In the UAIC-RoDiaTb there are 44 syntactic tags, 20 of which having a unique translation into semantic tags: superl. (superlative), comp. (comparative), ap. (apposition), incid. (incident), neg. (negation), voc. (vocative), c.ag. (agent complement), and 13 circumstantial modifiers, except the modal one, which can have more values.

3.5 Categories Dependent on the Morphological Tag or on the Word Form

There are categories which can be strictly separated considering the meaning of their morphological classification, i.e. we can formulate rules for the correspondence of syntactic and morphologic annotation with semantic tags. Using this information, the syntactic deprels can be automatically changed in semantic ones by rules with two or more conditions. Examples:

- The types of articles, annotated det., can have the following semantic values: cel, cea, cei, cele, etc. (En: *the* + adjective)= Deictic; un, o, niște, etc. (En: *a*) = Undefined; al, a, ai, ale, etc., (En: *of the*) = Possessive,
- For aux. (auxiliary), the occurrences can have the following semantic values (in agreement with the meaning of verbal forms obtained with these auxiliaries): Optative, Future, Past, Passive. For the auxiliary putea (En: *can*) the semantic values are: Potentiality, Ability, or Competence,
- Frequently, the subordination marks, prepositions, conjunctions, or Rw, Dw, Pw are marks of the semantic values: pentru (for) has the value Purpose; fiindcă,

deoarece, căci (because) has the value Causative; deși, măcar că, (although) has the value Concession; dacă (if) has the value Condition, etc,

- The syntactic-morphological tag a.pron. can have the following semantic values: Possessive, Deictic, Negative, Interrogative, Emphatic, Undefined or Quantifier: universal for: toți, fiecare, oricare, (all, any, every),
- The tag punct. can have the following semantic values: Exogene, Unnecessary, Dislocation, Connect-reunion, End, Exclamation, Interrogation. In Druguș [7] the punctuation elements are considered as linguistic connectors, having logical and semantic values; we have already developed this theory by interpreting semantically each punctuation element, which, in FDG, must be annotated as part of the syntactic or semantic tree.

3.6 Syntactic Tags Semantically Polyvalent

The polyvalent relations are: a.adj, a.adv., a.subst., a.vb., c.d., c.i., c.prep., a.subst., c.c.m., sbj., n.pred., el.pred. For the sbj. and n.pred./el.pred., c.d./c.i. (subject, predicative noun/element, direct /indirect object); there is a reduced number of possible values, depending on the type of judgment (see Section **4.1.**).

However, the a.vb. (verbal attribute), a.subst. (attribute expressed by noun), and c.prep. (prepositional object) are syntactic tags without semantic value, established according to formal morphological criteria. They can have almost every semantic value.

4 Types of Judgments and Connectors

4.4 Judgments

The judgments contained in a sentence are not focused only on an event. We have proposed below a classification of sentences including not only events, but also definitions, descriptions, speech acts, or existential affirmations. For each type of judgment, the roles are different and we can use this information for the detection of roles by a semantic parser hybrid or rule based. The roles for each type of judgment are shown in Table 1.

4.5 Connectors

We can consider a many words and punctuation elements as logical connectors. Emanuel Vasiliu has devoted most of his books to the relationship between logical artificial languages and natural language studies, either in the clause or in the sentence [24-25], especially regarding the translation of logical connectors and categories into natural language words. The translations of the logical connectors into the natural language are: ~ = (negation) ”not”; U = (reunion) și “and”; ∩= (disjunction) = sau ”or”; ⊃ = (implication) = deci ”so”.

In terms of logics, they form horizontal expressions, and in terms of the dependency UAIC convention, there are oblique descending lines, the connector being posted between the connected words, i.e. subordinate of the first and regent for the second.

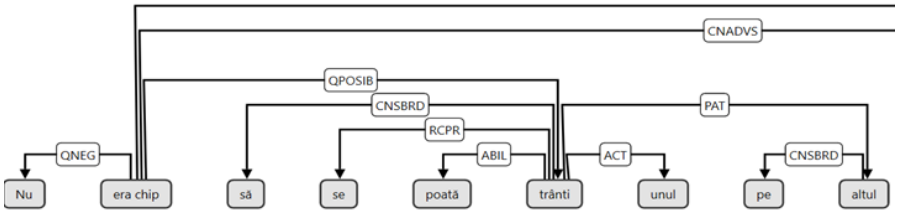

Fig. 2. The operator possibility: “There was no way they could prevail against one another”.

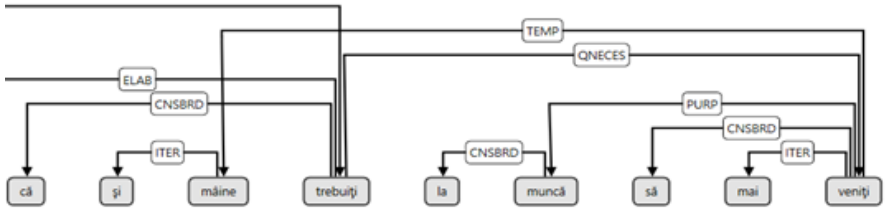

Fig. 3. The operator necessity: “that tomorrow you must come to work longer”.

The binary connectors will be annotate as: Connect:reunion, Connect:opposition, Connect:disjunction; Connect: subordination.

The connectors: ~ = “not”; ∀= “all” are not dyadic, but monadic. We decide to call it Quantifiers, because they do not connect two elements. The Quantifier:negation is subordinated of the word which is negated.

The Quantifier:universal is subordinate of a noun. The modal connectors, also monadic, consequently annotated as Quantifier:necessity (the symbol = □) and Quantifier:possibility (the symbol = ◊) are expressed in natural language by a sentence head that must have a subordinate subjective clause. Examples: Trebuie (să) = □, este probabil (că) = ◊. “We need (to ...), It is likely (that...).”. The modal quantifiers will be considered as regents of the clause which they modalise, see Fig. 2 and Fig. 3.

5 Conclusions

This paper proposes a type of semantic annotation with more categories since we aim to keep all the information that has been annotated in the classical syntactic layer; as we have argued in this paper, this information is important since it can be exploited by other applications. Another purpose was to find an international annotation with similar categories in view of a future affiliation.

The similarities with the tectogrammatic layer of PDT and with the AMR logical categories are obvious. However, there are also differences since the resultant graph of the AMR semantic annotation is not a dependency tree, and the nodes are not words, but concepts. In order to show the isomorphism between the syntactic and the semantic structures, we chose to build a corpus of semantic dependency trees, which is similar to the tectogrammatic layer of the PDT.

The set of annotations has been successfully experimented. The UAIC treebank has a parallel corpus in semantic format, having now 5,500 sentences. Moreover, in a

collection of 400 sentences, the syntactic relations with a high degree of ambiguity have been manually annotated by three experts, and their agreement has been of 85% (with the same solutions adopted and without different annotation for similar situations).

References

1. Academy of Romania: Thesaurus-Dictionary of Romanian Language. Compound by 2 series: I: 1913–1949. Dictionary of Romanian Language. Socec, Universul Publishing, Bucharest, and II: 1965–2010. Dictionary of Romanian Language. Romanian Academy Publishing, Bucharest (1913–2010)
2. Amaro, R.: Extracting semantic relations from portuguese corpora using lexical-syntactic patterns. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), pp. 3001–3005 (2014)
3. Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., Schneider, N.: Abstract meaning representation for sembanking. In: Proceedings of the 7th linguistic annotation workshop and interoperability with discourse, pp. 178–186 (2013)
4. Böhmová, A., Hajič, J., Hajičová, E., Hladká, B.: The prague dependency treebank: A three-level annotation scenario. Text, Speech and Language Technology, Springer, vol. 20, pp. 103–127 (2003) doi: 10.1007/978-94-010-0201-1_7
5. Bonial, C., Bonn, J., Conger, K., Hwang, J. D., Palmer, M.: PropBank: Semantics of new predicate types. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), pp. 3013–3019 (2014)
6. Cristea, D., Mihăilă, C., Forăscu, C., Trandabăț, D., Husarciuc, M., Haja, G., Postolache, O.: Mapping Princeton wordnet synsets onto Romanian wordnet synsets. Romanian Journal on Information Science and Technology, vol. 7, no. 1–2, pp. 125–145 (2004)
7. Fillmore, Ch. J.: The case for case. Universals in linguistic theory, (Part Two). Bach, E. and Harms, R. T. (eds), Holt, Rinehart and Winston Publishing, pp. 1–25 (1968)
8. Fukuda, S., Nanba, H., Takezawa, T.: Extraction and visualization of technical trend information from research papers and patents. D-Lib magazine, vol. 18, no. 7–8 (2012) doi: 10.1045/july2012-fukuda
9. Gupta, S., Manning, C.: Analyzing the dynamics of research by extracting key aspects of scientific papers. In: Proceedings of 5th International Joint Conference on Natural Language Processing, pp. 1–9 (2011)
10. Hovy, E., Marcus, M., Palmer, M., Ramshaw, L., Weischedel, R.: OntoNotes: The 90% solution. In: Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers, pp. 57–60 (2006)
11. Levin, B.: English verb class and alternations. A Preliminary Investigation. University of Chicago Press (1993)
12. Mel'Cuk, I. A.: Dependency syntax: Theory and practice, Buffalo, Suny Press (1988)
13. Matthiessen, C. M. I. M., Bateman, J. A.: Text Generation and systemic-functional linguistics: Experiences from english and japanese. Printer Publishers (2002)
14. Mikulová, M., Bejček, E., Mírovský, J, Nedoluzhko, A., Panevová, J., Poláková, L., Stranák, P., Ševčíková, M., Žabokrtský, Z.: From PDT 2.0 to PDT 3.0 (modifications and complements), ÚFAL technical report (2013) https://ufal.mff.cuni.cz/techrep/tr54.pdf
15. Palmer, M., Gildea, D., Kingsbury, P.: The proposition bank: An annotated corpus of semantic roles. Association for Computational Linguistics, vol. 31, no. 1, pp. 71–105 (2005)
16. Perez, C. A., Mărănduc, C., Simionescu, R.: Ro-PAAS a resource linked to our UAIC-Ro-Dep-treebank. Advances in Artificial Intelligence and Soft Computing, In: Proceedings of 14th Mexican International Conference on Artificial Intelligence. Springer, vol. 9413, pp 29–46 (2015) doi.org/10.1007/978-3-319-27060-9_3

17. Ștefănescu, D., Banjade, R., Rus, V.: Latent semantic analysis models on Wikipedia and TASA. In: Proceedings of Ninth International Conference on Language Resources and Evaluation, pp. 1417–1422 (2014)
18. Tapanainen, P., Jarvinen, T.: A non-projective dependency parser. In: Proceedings of the 5th Conference on Applied Natural Language Processing, pp. 64–71 (1997)
19. Trandabaț, D.: Natural language processing using semantic frames. PHD Thesis, Faculty of Computer Science, Al. I. Cuza University, Iași (2010)
20. Urešova, Z.: Building the PDT-Vallex valency lexicon. In: Proceedings of the Corpus Linguistics Conference (2009)
21. Vanderwende, L., Menezes, A., Quirk, Ch.: An AMR parser for english, french, german, spanish and japanese and a new AMR-annotated corpus. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, pp. 26–30 (2015) doi: 10.3115/v1/N15-3006
22. Xue, N., Bojar, O., Hajic, J., Palmer, M., Uresova, Z., Zhang, X.: Not an interlingua, but close. Comparison of English AMRs to chinese and czech. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), pp. 1765–1772 (2014)