

PARSING TEMPORAL AND SPATIAL INFORMATION

CĂTĂLINA MĂRĂNDUC¹, VICTORIA BOBICEV², AND CENEL AUGUSTO PEREZ¹

¹ *Faculty of Computer Science, Al. I. Cuza University, Iași*

² *Technical University of Moldova, Chișinău*

*catalinamaranduc@gmail.com, victoria.bobicev@ia.utm.md,
augusto.perez@info.uaic.ro*

Abstract

In this paper we present a dependency treebank morphologically and syntactically annotated in a specific scheme. We managed to increase the accuracy of the POS-tagger and the syntactic parser used, which led to the increase in the volume of annotated texts. First, we analysed the accuracy with which the syntactic parser recognizes the 14 types of circumstantial complements, especially the temporal and spatial ones. These are the most numerous circumstantial complements, and they are very important for the configuration of a textual world describing reality or proposing a fictitious world, providing information about the type of text. In December 2020 our treebank comprised 42,542 sentences (919,608 words and punctuation). We studied our documents containing fictional and non-fictional narrative. Using a Malt parser optimizer, we extracted dependency chains of time and spatial complements. The number of complements and the degree to which they are precise is related to the type of text, fictional or nonfictional. In order to construct a classifier of texts, one can count the spatial and temporal complements and one can observe if they represent determinations of exact landmarks (with geographical proper names and numbers) - in which case the text is a real narrative, or if they represent imprecise determinations, in which case the narrative is fictional.

Key words — Local complements, narrative fiction, narrative reality, syntactic parser, temporal complement, treebank, type of text.

1. Introduction

The RoDia (Romanian Diachronic) Dependency Treebank was created in 2007 and it increased to 4,600 sentences in 2014 (Perez, 2014). Regarding the basic syntactic format, created in 2007 in accordance with the Dependency Grammar principles (Tesnière, 1959; Mel'Čuk, 1987), we have only made insignificant changes since 2014. The list of complements includes 14 types of circumstances and the coordination is a chain starting from the first coordinate, which also includes the connecting words or punctuation. But the volume of the treebank has increased a lot with the improvement of automatic annotation tools. We have used a hybrid POS-tagger (Simionescu, 2011), which we adapted for Nonstandard Romanian, the list of morphological labels from the MULTEXT-East project (Erjavec, 2012), as well as diverse variants of the MaltParser (Hall *et al.*, 2006), trained on the growing gold corpus that we created. The treebank is corrected manually, but this is getting easier as the number of errors decreases. In December 2020, it comprises 42,542 sentences, with 919,608 words and punctuation.

We have focused on old Romanian texts from the 16th-19th centuries, and we have annotated whole books, because we have noticed that in this way the parser is better trained on more and more specific structures, and the texts are also available for other types of research.

In November 2017, a treebank for Nonstandard Romanian was created on the Universal Dependencies (UD) portal (Mărănduc and Bobicev, 2017). After three years, in November 2020, the treebank is available with 26,221 sentences (572,259 tokens, punctuation included).

Regarding the semantic annotation convention, we failed to create a semantic parser for it and that is why the semantic treebank has only 5,566 sentences with 99,341 tokens (Mărănduc *et al.*, 2018). As a first step towards creating a semantic parser, we have tried to train diverse Malt Parser variants (Smith *et al.*, 2018) available on the UD site, on the basic format of our treebank, with the 14 types of circumstantial complements, attempting to improve their accuracy.

The transformation into the UD convention is done automatically using the Treeops program (Colhon *et al.*, 2017) and the result depends on the correctness of the morphological and syntactic annotation in the basic format.

A method for obtaining a better accuracy is to increase the training corpus with texts rich in the type of complement to which poor accuracy is recorded, because it has too few attestations in the texts. Thus, in order to study the time-space confusions made by the parser, we annotated Neculce's chronicle, which is very rich in spatial complements and quite rich in time complements, being a non-fiction narrative text.

The automatic parser mistakes certain pairs of circumstantial complements:

- time and place;
- associative and instrumental;
- conditional and concessive;
- conditional and consecutive;
- cause and purpose.

We refer here only to the local and temporal ones, which are very frequent and very important for the configuration of the textual world, be it fictional or non-fictional, and of the text type. If we solve the annotation of one of the 2 categories, we also solved the one with which it is confused. If we manage to solve the correct annotation of time, then all other information that refers to landmarks or directions or sequences is spatial.

Time annotation of circumstantial complements is preserved in the UD format by the existence of three sub-classifications specific for Romanian treebanks: `nmod:tmod`, `advmod:tmod`, `advcl:tcl`.

To increase the accuracy of complement recognition, we have used several methods, including re-correcting those types of complements where we found a large number of errors, based on inconsistencies in the training corpus.

For example, if the erroneous annotation of the constructions: *pe vremea aceea*, *pe cea vreme*, *în vremea ceea* (En: *at that time*, *at the time mentioned*) as being c.c.l. (space

complement) was repeatedly found in the automatic parsing, then we looked for the word *vremea* (En: *time*) in the training corpus to discover further errors not corrected and that generated the current errors. In the first of the five sub-corpora, Neculce’s “Cronicle”, we found 4 errors. These generated 10 errors in the second sub-corpus and an even higher number of errors in the others. The issue would be solved if all of these errors were corrected.

Thus, the lexical elements that should induce the annotation of the dependency relationship as temporal are extracted: *an*, (archaic synonyms: *leat*, *let*, *vleat*), *zi*, *lună*, *dimineața*, *seara*, *noapte*, *ianuarie*, *februarie*, *martie*, *aprilie*, *mai*, *iunie*, *iulie*, *august*, *septembrie*, *octombrie*, *noiembrie*, *decembrie*, *luni*, *marți*, *miercuri*, *joi*, *vineri*, *sâmbătă*, *duminică*, *iarnă*, *vară*, *primăvară*, *toamnă*, *timp*, *anotimp*, *vreme*, *săptămână*, *oră*, *zi*, *veac*, *veci*, *început*, *sfârșit*, *când*, *cândva*, *oricând*, *atunci*, *acum* (*acu*, *acuș*), *mereu*, *totdeauna*, *apoi* (*păi*), *mâine*, *poimâine*, *ieri*, *alaltăieri*, *odată*, *diseară*, *târziu*, *devreme*, *astăzi*, (*azi*), *imediat*, *deocamdată*, *curând*, *pururea* (En: *year*, *day*, *month*, *morning*, *evening*, *night*, *January*, *February*, *March*, *April*, *May*, *June*, *July*, *August*, *September*, *October*, *November*, *December*, *Monday*, *Tuesday*, *Wednesday*, *Thursday*, *Friday*, *Saturday*, *Sunday*, *winter*, *summer*, *spring*, *autumn*, *time*, *season*, *weather*, *week*, *hour*, *day*, *century*, *forever*, *beginning*, *end*, *when*, *sometime*, *anytime*, *then*, *now*, *always*, *all the times*, *then*, *tomorrow*, *the day after tomorrow*, *yesterday*, *the day before yesterday*, *once*, *tonight*, *late*, *early*, *today*, *immediately*, *for now*, *soon*, *forever*).

In order for the parser to memorize them as inducers of the c.c.t. relationship, each of them must be annotated with that relationship at all occurrences with this temporal meaning in the treebank, without any error disturbing the induction process. This is the way for the parser to memorize the words related to the notion of time, following the training with correctly annotated texts.

A solution could also be to link the words in the treebank to the information in the Romanian WordNet, as was done recently with a parsing experiment with which an increase in accuracy of 0.5 percent was obtained (Barbu Mititelu *et al.*, 2016).

2. Parsing Experiments

We parsed our documents with MaltParser (Smith *et al.*, 2018), a data-driven parser. This parser demonstrated the ability to obtain good results for multiple languages and has been widely used in Universal Dependencies projects (Nivre *et al.*, 2016). The first set of experiments was performed with the whole corpus. The whole corpus, except for one document, was used for training, and testing was performed on the document that was excluded from the training set, thus obtaining the data presented in Table 2.

However, the documents in our corpus are quite different and training on one of them and testing on other results in poor accuracy. Thus, we decided to experiment with every document separately. Some of them are relatively small, too small to be used for training, and we selected the three largest documents in our corpus, namely the New Testament *Gospels* and *Acts*, and Neculce’s *Chronicle*.

The MaltParser offers a wide range of parameters for optimization, including nine different parsing algorithms, two different machine learning libraries (each with a number of different learners), and an expressive specification language that can be used

to define arbitrarily rich feature models. In our case we are especially interested in the feature set optimization.

First of all, we used the MaltOptimizer (Ballesteros and Nivre, 2012) to detect the best algorithm and feature set for our documents. The MaltOptimizer processed the documents in three steps. The first step was used to gather information about the various properties of the training set. During the second step, the MaltOptimizer explored a subset of the parsing algorithms implemented in the MaltParser, based on the results of the data analysis to detect the best one for this particular training set. The goal of the third step was optimization of the feature model given by the parsing algorithm chosen. It tested potentially useful features one by one and in combination to ensure that all features in the model actually make a contribution. The result of MaltOptimizer use is presented in Table 1.

Table 1: The best algorithms and the best Labelled Attachment Score (LAS) for three largest documents of our corpus.

Document	Best Algorithm	Best Performance (LAS)
New Testament Gospels	Nivreeager	83,8
New Testament Acts of Apostles	Nivrestandard	78,9
Neculce Chronicle	Nivreeager	84,59

Most of the effort when optimizing MaltParser usually goes into feature selection, that is, in tuning the feature representation that constitutes the input to the classifier. A feature model in MaltParser is defined by a feature specification file in XML. It states that the parsing algorithm uses 32 features including: POSTAG values of the neighbouring tokens around the current token; 4 FORM that presents words around the current token; LEMMA of the current token, 3 DEPREL (dependency relation labels) and 8 complex features that merge two or three features as, for example, morphological label of the word and its dependency relation to the left and to the right.

3. Related Work

The annotation of space and time, as a means of configuring textual worlds or communication situations, is increasingly in the attention of linguists and computer scientists. It is also the basis for the search for time information retrieval, TIR, or geographic information retrieval, GIR. Strötgen (2010) shows how co-occurrences of spatial and temporal information are determinant for the spatio-temporal profiles of documents.

Llorens *et al.* (2009) only deal with the annotation of temporal semantic roles, in accordance with the internationally accepted TimeML scheme, and evaluates a set of time-related MWEs, TIMEX3 in English and Spanish, with an accuracy of 76%, which makes the authors consider that they are likely to be identified in other languages as well. Three years later, Llorens *et al.* (2012) propose an automatic system for identifying time relationships in natural language. The experiments were made on an available English data set annotated with temporal information (TimeBank) in a 10-fold cross-validated evaluation, with an accuracy of 46%.

In the paper (Lefeuvre *et al.*, 2016), a syntactic rather than lexical annotation of time in a treebank in French is described, and the authors make proposals to extend the TimeML scheme. An annotation of temporal dependency structure is performed on a corpus of children’s narratives in (Kolomiyets *et al.*, 2012). The agreement among more annotators is: 0.856 on the event words, 0.822 on the links between events, and of 0.700 on the ordering relation labels.

In Romanian, the English corpus of Time Bank was ported in Romanian (by translation) with all temporal annotations (Forăscu and Tufiş, 2012), having 4715 sentences (65,375 tokens). A conference of the same year, on semantic web data annotation focuses, among other things, on the recognition of TimeML noun events, *i.e.* on a scheme for processing the event and temporal expressions in natural language processing fields (Jeong and Myaeng, 2012).

A chapter in a Springer book is also interested in a database which can manage events that are evolving with time, *i.e.*, the information of spatial objects whose shape and position evolve with time (Xiaoping *et al.*, 2011).

Our corpus consists of texts in Old Romanian and is not annotated with the categories in Time Bank, but it could be because all the information about the modes and tenses are in X-Postag (the morphological annotation specific to our treebank).

Therefore, we did not give up the annotation of the verbal circumstantial (time) modifiers in our basic syntactic convention; we have tried to see what information we can extract from the syntactically annotated corpus we hold. In the semantic format, we managed to annotate the space and time when they are verbal or nominal determinants, but in this paper we discuss only verbal modifiers.

4. Narrative Corpus Content

Table 2 below presents the documents included in this study, annotated morphologically and syntactically, the number of sentences and tokens, the accuracy of the automatic parser (labelled and unlabelled attachment score, *i.e.* LAS and UAS) and the type of the texts. It is a balanced corpus, *i.e.* the contemporary and the old texts, the regional ones, and the social media communication are all represented. The first word in the title, with capitals, marks these categories. For the old texts, the century is also added. We excluded from this study the legal style, Wikipedia, the lyrical poetry, popular and church (Psalms). For the classification of these types of texts, we need some other criteria (Mărănduc, 2005).

The information on the time and space framing appears in narrative texts. We can study the number of such complements and what would be their form when the narrative is fictional, compared to when the narrative is mystical or a reality one. These complements are also found in dialogues, where they circumscribe the communication situation.

Table 2: The files included in this study

Name of xml	Sent.	Tokens	LAS	UAS	type of text
CHAT	2,579	39,239	82.89	73.57	dialogue narrative reality
CONT_1984_orwell	904	17,608	82.05	73.28	novel narrative fiction
CONT FrameNet	1,092	24,659	79.33	71.10	journal narrative reality
OLD XVI Flower of Gifts	1,083	21,078	82.95	74.68	philos narrative fiction
OLD XVII NewTest. Gospel	5,174	92,440	85.39	79.07	church narrative dialogue
OLD XVII NewTest. Apostles	5,901	122,290	84.50	77.13	church narrative epistolary
OLD XVIII Neculce Chronicle	6,068	157,694	89.16	84.78	chronicle narrative reality
OLD XIX Caragiale Kings +Remember	1,691	48,930	77.09	70.09	novel narrative fiction
POP Ballads Rep Mold	1,112	19,405	80.04	72.37	narrative fiction
POP Ballads Rom	1,486	36,780	79.08	72.64	narrative fiction

Table 3 shows first the total number of occurrences for the principal circumstantial complements in each document, secondly the correctly parsed ones, and thirdly the erroneous occurrences, manually corrected. The temporal and spatial complements are mistaken for each other, the local annotated as temporal being less numerous than the temporal ones annotated as being local. Correctly annotated complements do not exceed half of the total number of occurrences, except in the case of local and modal ones. The table shows that the syntactic parser cannot yet correctly annotate semantic categories of information and semantic relationships.

Table 3: Occurrences of circumstantial complements in the documents, the total number, the correct parsed and the erroneous corrected relations.

c.c.	chat	Neculce	Frame Net	Gospel	Apostle	Flower	Orwell	Carag	Ballads RM	Ballads RO
c.c.l.	983 668 315	6695 4668 2027	558 433 125	2685 1746 939	2921 2120 801	441 350 91	468 332 136	1422 1167 255	1313 873 440	553 343 210
c.c.t.	1216 407 809	3161 1373 1788	763 368 395	1568 747 821	1540 754 786	311 171 140	417 184 233	1105 696 409	490 208 282	234 96 138
c.c.m.	1791 1168 623	4688 2781 1907	816 430 386	2013 906 1157	3628 1489 2139	496 282 214	982 575 407	2165 1572 693	757 427 330	517 316 201
c.c.cz.	328 35 293	1561 693 868	162 34 128	91 47 44	1487 720 767	226 111 115	63 33 30	304 135 169	338 189 149	132 96 36
c.c.	294	1577	99	67	1053	166	40	277	337	144

PARSING TEMPORAL AND SPATIAL INFORMATION

scop	68	684	54	28	481	92	13	157	168	47
	226	893	45	39	572	74	27	120	169	97
c.c. cond.	138	639	74	34	486	136	26	105	76	57
	80	177	43	15	235	60	21	59	37	28
	58	462	31	18	251	76	5	46	39	29
c.c. conc.	39	104	51	79	117	10	39	104	27	22
	1	11	3	4	6	0	6	62	1	1
	38	93	48	75	111	10	33	42	26	21
c.c. cons.	36	1265	59	10	188	55	32	90	42	29
	5	659	21	28	34	26	17	25	14	9
	31	606	38	74	154	29	15	65	28	20
c.c. instr.	112	572	32	26	518	53	33	203	166	123
	43	340	19	12	228	36	17	126	51	49
	69	232	13	13	290	17	16	77	115	74
c.c. soc.	103	1112	23	31	407	65	29	199	78	43
	59	441	11	18	232	23	15	143	23	15
	44	671	12	12	175	42	14	56	55	28

5. Morphological Classification of Temporal and Spatial Circumstantial Complements

Circumstantial complements are generally considered modifiers, which are not part of core dependencies. However, there are certain verbs that by their meaning have such dependencies as necessary and mandatory. An approach related to the realization of dependency relationships through various formal structures can be found in (Bejček *et al.*, 2018), which presents a resource called Forms and Function (ForFun) and the usage possibilities. They present an inventory of the multiple forms that spatial and temporal complements can have and how the preferred selection of some of them can give useful information for others studies. For example, in the classification of texts, a poem that uses narrative or dialogue to highlight the lyric could be classified as lyrical. If we have a narrative that uses the convention of the found notebook, we could classify it as fictional narrative.

5.1. Adverbs

Regarding adverbs, those included in the list in Section 1 can be considered as specific to the temporal relationship, but many others can appear with this relationship although they are non-specific. For example, *cum* (En: *how*), which is specific to the circumstantial modality relation, can have a temporal value, as in Ex 1:

Example 1:

Cum s-a culcat, a și început a sforăi. 'How Refl.Cl.-has lain_down, has even begun to snore' (En: The moment he lay down, he began to snore.)

Other adverbs are ambiguous, because though they express a succession, it can be a spatial or temporal one. For example, the adverb *înainte* (En: *before*) is frequently both spatial and temporal. In the first example, it has a local meaning, and in the second example, it has a temporal one (see Fig. 1 and 2).

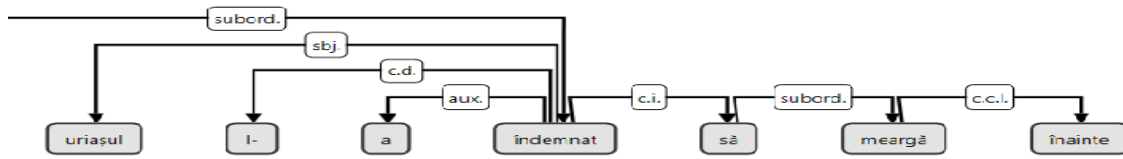


Figure 1. *Înainte* as spatial adverb. *Uriașul l-a îndemnat să meargă înainte.* En: *The giant urged him to go ahead.*

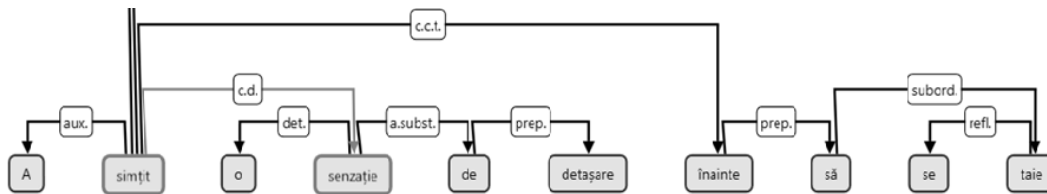


Figure 2. *Înainte* as temporal adverb. *A simțit o senzație de detașare înainte să se taie.* En: *He felt a sensation of detachment before cutting himself.*

If they are in relation to other adverbs that are specialized strictly as adverbs expressing temporal relations, the type of their relation will be disambiguated in accordance with their relation.

5.2. Nouns, Pronouns and Adverbs Preceded by Adpositions

Any of the words in the previous lists can be found with any of the adpositions, and being in circumstantial temporal relation, although some adpositions are more frequently used for a particular relationship, for example *pe*, *spre către* (En: *on*, *to*, *towards*) are used more in local complements than in temporal ones.

Certain word groups are usually time complements: *până ieri*, *până la anu*, *de mâine*, *pe urmă*, *în acel an*, *după aceea*, *după ce*, etc. (En: *until yesterday*, *until next year*, *as of tomorrow*, *afterwards*, *that year*, *afterwards*, *after* etc.)

Also, an adposition followed by a noun with a cardinal or ordinal numeral is a circumstantial complement of time.

5.3. Clauses with Temporal and Spatial Information

Clauses that fulfil the spatial or temporal modifier relation, as well as the words accompanied by prepositions, can select typical or common introductory elements with other types of relations. They show an action or state that is in a relationship of spatial or temporal succession to the action or state in the regent clause.

A large number of temporal clauses are introduced by relative pronouns with adpositions: *după ce*, *după care*, *până ce* (En: *after*, *after which*, *until*).

6. Types of Local and Temporal Complements Specific to Text Types

Table 4 shows the relationship among the number of sentences in a document, the type of text, and the number of temporal and local complements. In the case of temporal

PARSING TEMPORAL AND SPATIAL INFORMATION

complements, the precision of the determination is achieved by the existence of a number, whether it is a cardinal or ordinal numeral. In the case of local complements, the precision of the determination is achieved by the existence of a proper noun in the word chain, especially if it is a geographical name.

If we build a program that classifies into text types morpho-syntactically annotated texts entered into a database, we should enable it to look for certain parameters, such as the ratio of the number of sentences to the number of time modifiers. If the number of time complements were less than half the number of sentences, then we would have a fictional narrative text or a lyrical text. The value of 0.5 that we provisionally give this parameter can be found if we evaluate the same number of texts whose text type is already established.

Table 4: Correspondence among the type of the text, the number of temporal and local complements, with/without numerals and proper names

Name of xml	Sent.	c.c.t.	c.c.t. with NUM	c.c.l.	c.c.l. with Np	type of text
CHAT	2,579	1216	103	983	94	dialogue
CONT 1984 Orwel	904	417	9	468	10	novel narrative fiction
CONT Frame Net	1,092	763	38	558	81	journal narrative reality
OLD XVI Flower of Gifts	1,083	312	4	441	34	philosophy narrative fiction
OLD XVII New Test. Gospel	5,174	1,568	44	2,685	269	church narrative dialogue
OLD XVII New Test. Apostles	5,901	1,540	59	2,921	527	church narrative epistolary
OLD XVIII Neculce Chronicle	6,068	3,158	144	6,696	2129	chronicle narrative reality
OLD XIX Caragiale Princes	1,691	939	14	1,190	133	novel narrative fiction
POP Ballads Rep Mold	1,112	317	4	717	14	narrative fiction
POP Ballads Rom	1,486	482	16	1,325	52	narrative fiction

In the same way, the type of an unknown text can be assessed by comparing the number of sentences to the number of local complements. It remains to be seen whether these local and temporal complements are precise or imprecise.

In the case of dialogue (like in the case of the CHAT texts), the participants relate the information to the communication situation, using the adverbial complements *here* and *now* as well as their synonyms.

Example 2:

Acu doarme aici lângă calculator. (En: Now she is sleeping here near the computer.

As the table shows, our social media texts have a big number of temporal complements, but only 103 precise ones, with numerals: *pe 14* En: *On the 14th* (date of the day), *pe la 11*; En: *Around 11* (the hour).

Local modifiers are a bit less numerous, and contain only a small number of geographical names: *la Cairo, la Romexpo, spre Calea Victoriei, de la Iași, în București*. En: *to Cairo, at Romexpo, towards Calea Victoriei, from Iasi, in Bucharest*.

In Orwell's novel "1984", being a fictional narrative, the complements of time and place will be in a small number and without much precision. Only 9 temporal modifiers contain numerals and 10 local modifiers contain proper names. The names are almost all character names. There are also invented geographical names, such as *Airbase One*. By comparison, in FrameNet, where there are articles in journals that report various real facts, the number of time and place modifiers and the number of numerals (38) and proper names (81) is higher. Geographical proper names are present and we can see that they are often territorial subdivisions or precisely delimited confined spaces: *at Roseberry Road, in Ewood Park, to Godstowe, at Greysteel*.

"Flower of Gifts" is a 16th century series of fictional narratives with moral teachings. It has very few local and temporal precise determinations, only 4 temporal modifiers with numerals, and 34 local modifiers with proper nouns. The proper names are of historical leaders or philosophers cited: *Aesop, of Alexander*.

One ontological problem is the classification as a reality narrative of the biblical text. It is historically attested, presented as real by its authors but challenged by some readers. The number of temporal and especially spatial determinations is large, as in reality narratives. The determination by numerals (*Gospel 44, Apostles 59*) and the geographical proper names are frequent and detailed, all the territorial subdivisions covered by Jesus's journey and then the journeys of the Apostles are named. Examples: *In Bethany, in Jerusalem, in Erihon, in Samaria, in Damascus, in Israel, in Mesopotamia, in Cappadocia, in Egypt, in Phrygia* etc. As these statistics show and in accordance with the intention of the emitter's communication, we therefore consider the religious text a real narrative.

"The Chronicle" by Neculce is a reality narrative of the 18th century. This chronicler is a memoirist; he recounts historical events of his life. Being a military leader, he knows precise data about the route taken by various armies and describes them in detail, using chains of local determiners and proper geographical names. The number of local determiners and proper names is higher than in any other text in our database, 2,129 local modifiers with proper names. This specific aspect of the text is related to the memorial character. Being a chronicle, naturally neither of the specified time complements, 144, which contain numerals (days, years), are missing.

As for the popular ballads, these are fictional narratives, the number of time complements is small, and the local modifiers are imprecise, being descriptions of nature that participate in the plot and characterization. Proper names are character names. In the case of other types of text, lyrical, scientific, legal, they will have a small number of spatial and temporal complements, just as in fictional texts, but other classification criteria must also be applied.

7. Conclusions

Time and space verbal modifiers are very important for configuring the textual world. These are the most numerous circumstantial complements in texts, along with the modal ones, which represent a semantically heterogeneous category. In order to prevent

parsers from mistaking time and space verbal modifiers, they need to memorize a large number of examples and the entire training corpus must be correctly annotated. By linking words to a semantic dictionary we can get a better accuracy parser.

In order to construct a classifier of texts, one can count the spatial and temporal complements and one can observe if they represent determinations of exact landmarks (with proper names and numbers) - in which case the text is a real narrative -, or if they represent imprecise determinations, in which case the narrative is fictional. The number and the precision of time and space verbal determiners can demonstrate the real, nonfictional character of a narration.

The classification must be continued adding other criteria, as: the logical relations of condition and consequence in legal texts, the figurative meaning in lyrical texts, the designation of the emitter and receiver in direct speech, etc.

References

- Ballesteros, M. and Nivre, J. (2012) MaltOptimizer: An Optimization Tool for MaltParser. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 58–62.
- Barbu Mititelu, V., Ion, R., Simionescu, R., Scutelnicu, A. and Irimia, E. (2016). Improving parsing using morpho-syntactic and semantic information. *The Romanian Human-Computer Interaction Journal*, 9(4):285-304.
- Bejček, E., Hajičová, E. and Mikulová, M. (2018). The Relation of Form and Function in Linguistic Theory and in a Multi-layer Treebank. In *Proceedings of the 16th International Workshop on Treebanks and Linguistic Theories*, 56–63.
- Colhon, M., Mărănduc, C. and Mititelu, C. (2017). A Multiform Balanced Dependency Treebank for Romanian. In *Proceedings of Knowledge Resources for the Socio-Economic Sciences and Humanities, (KnowRSH) workshop at the Recent Advances in Natural Language Processing (RANLP)*, 9-18.
- Erjavec, T. (2012). MULTEXT-East: Morphosyntactic Resources for Central and Eastern European Languages. *Language Resources and Evaluation*, Vol. 46/1. Netherlands. Springer Publisher, 131–142.
- Forăscu, C. and Tufiş, D. (2012) Romanian TimeBank: An Annotated Parallel Corpus for Temporal Information. In *Proceedings of Language Resources and Evaluation*, LREC, 3762-3766.
- Hall, J., Nivre, J. and Nilsson, J. (2006): MaltParser: A Data-Driven Parser-Generator for Dependency Parsing. In *Proceedings of the Fifth Intelligence Conference on Language Resources and Evaluation*, LREC, Genoa, Italy, 2216-2219.
- Jeong, Y. and Myaeng, S-H., (2012). Using Syntactic Dependencies and WordNet Classes for Noun Event Recognition. In *Proceedings of the premier international forum, for the Semantic Web Linked Data Community*, 41-50.
- Kolomiyets, O., Bethard S. and Moens, M.-F. (2012). Extracting Narrative Timelines as Temporal Dependency Structures. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2012)*, 88-97.

- Lefevre-Halftermeyer, A., Antoine, J.-Y., Couillaud, A., Schang, E., Abouda, L., Savary, A., Maurel, D., Eshkol-Taravella, I. and Battistelli, D. (2016). Covering various Needs in Temporal Annotation: a Proposal of Extension of ISO TimeML that Preserves Upward Compatibility. In *Proceedings of Language Resources and Evaluation Conference*, 2016, 3802-3806.
- Llorens H., Navarro-Colorado, B. and Saquete Boró, E. (2009). From Semantic Roles to Temporal Information Representation. In *Advances in Artificial Intelligence and Soft Computing 14th Mexican International Conference on Artificial Intelligence (MICAI-2009)*, 36–43.
- Llorens H., Saquete Boró, E. and Navarro-Colorado, B. (2012). Automatic system for identifying and categorizing temporal relations in natural language. *International Journal of Intelligent Systems* 27(7), 680-703.
- Mărănduc, C., (2005). *Norms of correct text formation - from a pragmatic perspective*, National Foundation for Science and Art Publishing House, Bucharest.
- Mărănduc, C. and Bobicev, V. (2017). Non-Standard Treebank Romania – Republic of Moldova in the Universal Dependencies. In *Proceedings of Conference on Mathematical Foundations of Informatics (MFOI)*, 111-116.
- Mărănduc, C., Mititelu, C. and Bobicev, V. (2018). Syntactic Semantic Correspondence in Dependency Grammar. In Jan Hajic (Ed.), *Proceedings of the 16th International Workshop on Treebanks and Linguistic Theories*, 167-180.
- Mel'Čuk, I. A., (1987). *Dependency Syntax: Theory and Practice*, Buffalo, Suny Press.
- Nivre, J., de Marneffe, M.-C., Ginter, F., Goldberg, Y., Hajič, J., Manning, C. D., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N., Tsarfaty, R. and Zeman, D. (2016). Universal Dependencies v1: A Multilingual Treebank Collection. In *Proceedings of Language Resources and Evaluation, LREC'16*, 1659-1666.
- Perez, C.-A. (2014). Linguistic Resources for Natural Language Processing. PhD dissertation, Iași, Al. I. Cuza University.
- Simionescu, R. (2011). Hybrid POS Tagger. In *Proceedings of the Workshop on Language Resources and Tools in Industrial Applications*, Euroalan 2011 summer school, Cluj-Napoca, Romania, 21-28.
- Smith, A., Bohnet, B., Lhoneux, M., Nivre, J., Shao, Y. and Stymne, S. (2018). 82 treebanks, 34 models: Universal dependency parsing with multi-treebank models. In *Proceedings of the CoNLL2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, 113–123.
- Strötgen, J. (2010). Extraction and exploration of spatiotemporal information in documents. In *Proceedings of the 6th Workshop on Geographic Information Retrieval*.
- Tesnière, L.: (1959) *Elements of structural syntax*, Paris, Klincksieck.
- Xiaoping, Ye., Peng, Z. and Guo, H. (2011). Spatio-Temporal Data Model and Spatio-Temporal Databases. *Temporal Information Processing Technology and Its Application*, Springer, Berlin, Heidelberg, 91–112.