Synthetic scientific report – January-December 2014

Project: The structure and interpretation of Romanian Determiner Phrase in Discourse Representation Theory: the determinants.

Coordinator: Prof. Univ. dr. Alexandra Cornilescu

Contractor: Universitatea din Bucuresti

Stage: 15 decembrie 2014

Main objective of the project: Syntactic and semantic unitary description of a Romanian fragment including the determiners, such that to permit the generation of the semantic description, given the syntactic description of the fragment.

The objective for the year 2014 are:

1. Syntactic description of other Romanian Determiner Phrase, such that to permit an optimal semantic representation at the discourse level. For the targeted determiners, one specifies their features, that is their lexical input. This description is a preceding step for the realization of the electronic lexicon, to be done during this project phase.

2. Typology studies for Romanian, English and Romance languages.

The results of project activity between Typology studies for Romanian, English and Romance languages. January 1st and Decembre 15th:

Research activity:

Objective 1: The syntactic description of other DP in Romanian.

The following types of DPs have been described this year, as can be seen from the attached list of publications:

- DPs with a pronominal center (see Alexandra Cornilescu and Alexandru Nicolae, Classified Pronouns: The View from Romanian, ,BWPL 2, 2014; On the internal structure of pronominal DPs, presented at Annual Conference of English Department ACED 16, June, 3-5, Bucharest, 2014);
- DPs whose had is a proper name (this research has been extended to the diachronic syntax of proper names, see Alexandra Cornilescu and Alexandru Nicolae, Classified Proper Names in Old Romanian: Person and Definiteness, Studies in the Romanian DP, editor Virginia Motapanyana-Hill, 2014)
- DPs headed by nominal article have been discussed in the diachronic evolution this year; a structure that has never been discussed so far in the history of Romanian has been identified: the low definite article. In this structure, the constituent that bears the definite article does not occupy first position, as is obligatory the case in modern Romanian. The low definite article structure offers a strong argument in favour of the hypothesis that the Romanian enclitic definite article has its source in a post-posed Latin demonstrative, not in a pre-posed one, as claimed for other
Romance languages (see Alexandra Cornilescu, The low Definite Article and the evolution of the Romanian DP, presented at Romanian Fest in Geneva, November 2014).

- DPs containing the universal collective quantifier toţi. We have proposed a new analysis of this lexical item, proving that it is more profitable to regard it as a definite adjective with a maximising semantic function (see Alexandra Cornilescu and Anca Dinu, Pronouns and quantifiers: Towards an Analisys of toţi, presented at Annual Conference of English Department ACED 16, June, 3-5, Bucharest, 2014).

- DPs where the specific interpretation is induced by a pronominal adjective. We have proved that in Romance languages there is a considerable semantic difference between pre-nominal and post-nominal adjectives, which belong to different semantic types. One of the semantic effects that a pre-nominal adjective may trigger on the interpretation of the DP is specificity (see Alexandra Cornilescu and Anca Dinu, Adjectivæ and Specificity, RRL LIX, 1, p. 103–128, 2014). This semantic effect obtains for both definite and indefinite DPs.

- DPs containing adjectival and pronominal demonstratives, both for modern Romanian and for old Romanian (see Cornilescu and Nicolae, Diachronic Variation inthe Syntax of Romanian Demonstratives, al 14-lea Colocviu Internaţional al Departamentului de Lingvistică: Variaţia Lingvistică: Probleme Actuale, Bucureşti, 28–29 noiembrie 2014).

- Special attention has been paid to Dative pronominal clitics, which raise special comparative syntactic problems, having a much more extended distribution in Romanian than in other languages. As known, Romanian is among the few Romance varieties which have a special Dative morphology in all the persons, both for personal and reflexive clitics. (see Alexandra Cornilescu, The syntax of datives in unaccusative constructions, in Proceedings of Going Romance 2013: 27th Symposium on Romance Linguistics, Universiteit van Amsterdam, The Netherlands, ed. Petra Sleeman, Benjamins, 2014).

- Dative clitics have been approached from the perspective of language acquisition; this led to proposing new hypotheses on the syntax and semantics of the clitic doubling construction (see Avram, L., Differential Object Marking in Romanian: The View from Language Acquisition, The 16th Annual Conference of the English Department ACED 16, Universitatea din Bucureşti, 6-8 iunie 2014; Avram, L., M. Coene, A. Sevcenco, Theoretical implications of children’s early production of Romanian Accusative clitics, Lingua, 2014).

- Accusative clitics have been approached from the perspective of language acquisition; this led to proposing new hypotheses on the syntax and semantics of the clitic doubling construction (see Avram, L., Differential Object Marking in Romanian: The View from Language Acquisition, The 16th Annual Conference of the English Department ACED 16, Universitatea din Bucureşti, 6-8 iunie 2014; Avram, L., M. Coene, A. Sevcenco, Theoretical implications of children’s early production of Romanian Accusative clitics, Lingua, 2014).

- Other nominal structures which have been investigated both in Modern Romanian and in Old Romanian, are the partitive and pseudo-partitive constructions. It has been shown that in both old Romanian and modern Romanian there is both an N-de-N (N-of-N) construction common to all Romance languages and a DP-de-DP (DP-of-DP) construction, less frequent, though not absent in the Romance area; it is available in Spanish, for instance (see Mihaela Dogaru, Topicalization and extraposition in visual syntax, in The 8th international conference Language, Culture and Civilization –Converging Perspectives”, The Polytechnic University of Bucharest, 30 –31 May 2014; Mihaela Dogaru, Four silent constructions in Romanian, in 'The Annual Conference of the English Department ACED 16', FLLS, Universitatea Bucureşti, 5-7 iunie 2014; Mihaela Dogaru, On non-canonical uses of the indefinite article in Romanian, in Proceedings of The 12th Conference on British and American Studies, The Department of Theoretical and Applied Linguistics, Transilvania University of Brasov, 4 aprilie 2014).

At this stage of the project we have developed a first version of the electronic lexicon, for the determiners studied so far. The final lexicon is to be produced in 2015, as planned. Thus, we have proposed lexical entries for the following determiners: strong definite determiners (-l, cel, acest), the universal collective quantifiers toţi (all) and amandoi (both), pronominal determiners, proper names, etc. For each of these determiners, we have specified their feature matrix, accompanied by relevant explanations of the features. We present the matrices of the determiners mentioned above:

- The collective determiner toţi.
At the syntactic level, *toţi* is a quantificational adjective, which raises from the lexical NP where it merges in order to satisfy its quantificational properties. From the semantic point of view, *toţi* is best analysed as a semantic maximiser, not as a quantifier. In sum, its features may be represented in the following matrix:

\[
\begin{array}{c}
toţi \\
+{A} \\
\_ [+{\text{def}}]{N/D} \\
uQ
\end{array}
\]

- The enclitic article -(u)l

 -(u)l is a suffix, not a second position clitic. In Modern Romanian, the constituent bearing the definite article (necessarily a +N constituent) occupies first position in the DP (*ticaloasele cuvinte “evil words”). In Old Romanian, the definite constituent could also occupy the second position within the DP (*ticaloase cuvintele). Romanian nominal stems are specified for definiteness (± udef). The definiteness feature is morphologically expressed on the Noun, but is interpretable under D. In this case, D is unvalued interpretable definite (idef_). There is an Agree relation between the determiner and the suffixed noun, in a strictly local configuration (D[idef_:]...NP[u+Def]). The definiteness bearing constituent must be in the specifier immediately under D. Because nominal stems are marked for definiteness in the lexicon, Romanian is positively specified for the definiteness parameter, a nano-parameter proposed by Danon 2010 for Semitic languages.

It is characteristic for Romanian that definiteness is not only a semantic feature, but also a morpho-syntactic feature. As a consequence, Romanian possesses a rule of definiteness agreement between nouns and adjectives, even if, at the current stage in the evolution of Romanian, definiteness is phonologically realized only once, on the first term of the DP. There are however structures where the repetition of the definite marker is obligatory; such structures offer strong evidence that there is definite agreement in Romanian: săracul fratele meu, bietul copiul ăla. As a consequence, we propose that adjectives are also potentially marked as u+def, this feature being acquired by agreement with the noun within a derivation. A second kind of evidence that definiteness is the existence of an expletive definite article on proper names. Expletive article has a specific distribution. It is likely that definiteness has become a syntactic feature, primarily because in Romanian Case is realized on the determiner for masculine nouns (copil, copil+ului) or on both the nominal stem and the determiner in the case of feminine nouns (fata, fete+i). Summing up the description we have presented, the following features compose the matrix of -(u)l:

\[
\begin{array}{c}
-(u)l \\
+{D} \\
+{N} \\
N_-
\end{array}
\]

- **cel** is the second definite article in Romanian, additionally having two other important uses. It may be an adjectival article, as well as a head in elliptical DPs. The important result we have established is that **cel** has a stable selectional property in all its occurrences, specifically, it selects for a QP [__+QP]. This feature manifests itself in different ways, in the three uses mentioned above. The quantificational feature differentiates **cel** from -(u)l, securing the complementary distribution of the two articles. The enclitic article selects nominal constituents (+N, i.e. nouns and adjectives), while the article **cel** selects quantificational phrases (QP, i.e. cardinals, ordinals, degree quantifiers). Cel cannot appear with class N elements when it functions as a definite article (*cel creion, *cel rosu mar*), though it does appear with adjectives when it is an adjectival article.
or the head of ellipsis (marul cel rosu, cel rosu). The diachronic investigation of cel has proved that its use as a definite article is related to the locality constraint on definiteness valuation. The cardinal itself, for instance, cannot bear the enclitic definite article, because it is not ϕ-complete. The presence of a cardinal above a noun blocs a valuation of definiteness in a structure of type Determiner [iDef]+Cardinal+Noun [u+Def]. Since the definiteness feature is too low to reach the determiner, the enclitic definite article is excluded, and the determiner realizes its definiteness feature through the insertion of cel: [[iDef][cel [i+Def]]] + Card+Noun.

Cel as an adjectival article shows a left periphery structure, cel being followed by an obligatorily focused modifier (AP, PP, QP). At the same time, the adjectival article cel construction is a double definite construction, being made up a noun with an enclitic definite article and a focused modifying cel phrase. In our analysis, the noun with the suffix definite article occupies the determiner position as usual, while the focus (AP, PP, QP) is a quantificational projection above the suffixed definite article. Cel is inserted to close off a phrase and show its DP type. The adjectival article cel has a strong EPP feature, which attracts the definite NP to the specifier of cel. What must be retained is the fact that the adjectival article cel must be followed by the contrastive focus phrase, which is a quantificational phrase (Lopez, 2009):

\[
\begin{array}{c}
\text{DP}[	ext{NP} [+\text{def}] [\text{D}[\text{EPP, cel}] [\text{QP} [+\text{focus}]] \text{AP} [\text{DP} \text{NP} [+\text{def}] ]]
\end{array}
\]

marul cel rosu

Ellipsis structures, which have an extended distribution in Romanian, are also focused structures. It has been showed for other languages, as well as for Romanian, that the remnant of ellipsis represents new contrastive information, therefore it is a focused constituent, just as in the case of the adjectival article cel.

Summing up this description, cel has got the following features:

- **Cel 1 (articol hotarat):**
  \[
  \begin{array}{c}
  +D \\
  +_QP \\
  \_u\_φ \\
  [u + \text{def}]
  \end{array}
  \]

- **Cel 2 (articol adjectival):**
  \[
  \begin{array}{c}
  +D \\
  +_QP [+\text{focus}] \\
  +EPP \\
  \_u\_φ \\
  [u + \text{def}]
  \end{array}
  \]

*Demonstrative pronouns and adjectives.*

From a semantic perspective, demonstrative pronouns and adjectives are strong determiners, just like the definite articles. The specific difference between demonstratives and the definite articles is the obligatory presence of a deictic feature [+Deixis], which anchors the referent in the utterance context or discourse. Additionally, a demonstrative which is used ostensively also places the referent with respect to the speaker in the local context of the speech act. Such uses of demonstratives are assumed to exhibit a [+location]. Another characteristic is the presence of locative PPs, which reinforce the location ostensive feature (cartea asta de aici “this book here”). “Placement in space” may also be textual, manifesting itself as an anaphoric or cataphoric feature: [+Def], [+Deixis, ±Proximate], [+Contrastive (Locatie)], [+Anaforic], [+Cataphoric]. All demonstratives show a strong emphasis feature (focus or topic), checked at the left periphery.
From a syntactic perspective, typical of Romanian is the existence of **specialized** pre-nominal and post-nominal demonstratives. It may be shown that the two demonstratives (not augmented/augmented) *acest/acesta* or *acel/acela* differ more than in their position with respect to the noun. Unaugmented demonstratives *acest/acel* are syntactic heads (*X₀*), not phrases (*Xₚs*), which explains why they cannot be used as pronouns, but only as determiners (*acest creion/ *acest*). *Acest/acel* (this/ that) cover the whole distribution of the two definite articles -(*u*)l and *cel*, since they are compatible with both nominal and quantificational constituents. There is no restriction on the complements selected by *acest/acel*, except for the already mentioned fact that *acest/acel* occur only with lexical NPs, never with empty NPs. We have thus uncovered the features: [+*X₀*], [+__NP(lexical)]. Preposed demonstratives mostly belong to the written language and are rarely used with the deictic locative value. Preposed demonstratives (*acest/acel*) which have a locative interpretation must be contrastively stressed.

Augmented demonstratives *acesta/acela* have very different syntactic properties. In the first place, morphologic augmentation has been interpreted as a process of incorporating a locative element, such as the Latin adverb illac. From the point of view of the phrasal status, augmented demonstratives are phrases (*Xₚs*) and, therefore, move as phrases, creating a number of locality problems. We give just one example: the movement of the augmented demonstratives towards a focus position, where they check their deictic or location features, is blocked by prenominal adjective. Compare *acest fost ministrul acesta fosta* / *ministrul acesta fost*. It is also interesting to find an answer to the question why augmented demonstratives must always be preceded by a noun suffixed by the definite article. As apparent, unlike unaugmented demonstratives, augmented demonstratives select a nominal definite phrase [+__NP[+def]]. Examining Old Romanian, a plausible hypothesis is that a postposed demonstrative was assimilated to a postposed adjective, which is always indefinite (*trandafir frumosul*). It is the presence of the definite article on the noun which makes the DP definite, satisfying a requirement typical of the Romanian phrase: the definite constituent obligatory occupies the first position of the DP. Summing up, *acest/acel* și *acest/acesta* have the following feature matrices:

```
acest
+D
[+X₀]
[+Def]
φ
[+__NP[−def]]
[+Deixis, +Proximitate]
[+Contrastiv (Locatie)]
( [+Anaforic])
( [+Cataforic])
```

```
acel
+D
[+X₀]
[+Def]
φ
[+__NP[−def]]
[+Deixis, −Proximitate]
[+Contrastiv (Locatie)]
( [+Anaforic])
( [+Cataforic])
```
• Personal pronouns, as, for instance, noi, are also determiners at the end at the derivation, but there is evidence, coming from the distribution of adjectives (el frumosul), that these formatives merge as NPs, triggering definiteness agreement with the adjective. Their characteristic features are person and definiteness. Here is our proposal for personal pronouns:

![Diagram of noi]

- Clitic pronouns differ from strong pronominal forms in as much as they are heads (X₀ constituents) and in as much as Romanian allows the clitic doubling construction (l-am vazut pe Ion). We have interpreted this structure on the model of the “big DP” (Uriagereka, 1995, among many), where the clitic selects for a definite DP which doubles it (D₀^DP). Consequently, a clitic pronoun like l- has the following features:

![Diagram of l-]

• Proper names.
Romanian proper names show the same properties as in other languages in part. Specifically, they must have Classifiers in their extended projection and they check a [person] feature, obligatorily raising to the DP domain. Additionally, in Romanian, proper names are syntactically definite, definiteness being required for marking case. A tentative feature matrix for Romanian proper names is shown below:

| nume propriu |
| +N           |
| +pers        |
| +u + def     |
| +classifier  |

**Objective 2.** Typology studies for Romanian, English and Romance languages.

Members of the team have computationally investigated problems such as language similarity, showing the typological properties of Romanian in comparison with Romance languages and Turkish. Romanian lexical similarity has entailed proposing original, automatic techniques of investigating etymons and cognate pairs, a result which has led to sketching a map of resemblance between Romanian, Romance languages and Turkish, at the lexical level. This investigation has been extended to a study of intelligibility and similarity between Romanian and other languages, from the perspective of text understanding [8] [7]. This is a highly topical problem, having a multitude of applications in automatic translation, acquisition of a second language, language acquisition, language evolution, etc. A proposal have been made for the automatic extraction of cognate pairs between Romanian and any other language with electronic resources; in this manner, cognate pairs have been determined exhaustively for Romanian, Romance languages and Turkish. We have also investigated the ability to decide whether two words are cognate or not, by employing an alignment algorithm [9] and by using an adequate metric [6]. We could thus identify the traces left by words which enter the language at some time point, as well as the evolution of these words [5].

**Administrative and organizational activities.**

Members of the team have held regular meetings to discuss problems related to the assignment of research topics, the presentation of the results and the establishing of the conferences or journals where the results were to be presented and submitted. It may be considered that all the activities related to monitoring the project have been successfully carried out.

**Dissemination of the results.**

This year there have been two conferences of the project whose proceedings are to be published. In June 2014 a workshop of the project has been organized within the International Conference of The English Department – ACED 16, 2014, held in Bucharest. All the project members have presented papers: Alexandra Cornilescu, Anca Dinu, Larisa Avram, Mihaela Dogaru si Liviu Dinu. The second session of the project was represented by the workshop Romanian Fest in Geneve, jointly organized by the Centre for the Study of Language Development, represented by our project and by Univerity of Geneva, at the invitation of the Linguistics Department of this University. Participants in this event have presented two-hour long conferences addressed to all the linguistics students and to all the teaching staff of the Department. The proceedings of this
workshop have been accepted for publication in the linguistic journal of the Geneva University, Grammaire Generative de Geneve (GGG).

We have also had editorial activities. We have prepared a volume 16 I of BWPL, intituled “Specificity and (in)definiteness”, which contains some of the papers presented at the workshop on Specificity in 2012. The BWPL volume has been reviewed and is in press.

A large number of research articles have been published in journals of high impact such as RRL, Lingua, Fundamenta Informaticae, BWPL, Neural Computing and Applications, and in the proceedings of conferences (ACED16, Conferinta Internationala Anuala a facultatii de Litera, Conferinta Internationala a FLLS, Going Romance, SCL 2014, The Workshop Romanian Fest in Geneva 2014, The International Conference on the Acquisition of Romance Languages „The Romance Turn 6”, The 2th Conference on British and American Studies, The13th International Congress for the Study of Child Language, The 52nd Annual Meeting of the Association for Computational Linguistics, The9th International Conference on Language Resources and Evaluation, The14th Conference of the European Chapter of the Association for Computational Linguistics 2014, TheConference on Empirical Methods in Natural Language Processing, The 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations, Workshop on Language Knowledge and Development in Native and Non-Native Speakers, The 8th international conference Language, Culture and Civilization –Converging Perspectives, The 12th Conference on British and American Studies, etc.) and a series of results already represents the nucleus of paper still in manuscript, which are to be continued and finalized in the next and last stage of the project.

Conclusions:

- The objectives of this stage have been fully realized.
- It is necessary to continue research, without modifying the current research plan.

Articles and book chapters published by the project members in 2014:


22. Avram, L., *Differential Object Marking in Romanian: The View from Language Acquisition*, The 16th Annual Conference of the English Department (ACED 16),


