

A feature based lexicon for Romanian Determiners

Anca Dinu

University of Bucharest, Faculty of Foreign Languages and Literatures,
Centre for Computational Linguistics

E-mail: anca.dinu@lils.unibuc.ro, anca_d_dinu@yahoo.com

Abstract

In this paper we describe the creation of a detailed feature based lexicon of Romanian Determiners, as a part of developing the project “The structure and interpretation of Romanian Determiner Phrase in Discourse Representation Theory: the determiners”. The main objective of the project is twofold: on the one hand, we aim at producing a unitary description of the Romanian Determiner Phrase (DP), based on the classification of determiners, function of their syntactic and semantic properties; on the other hand, we aim at specifying rules of interpretation (in a model theoretic sense) for Romanian determiners.

The general theoretical framework of the project we have adopted is the minimalist syntax (Chomsky 1999), coupled with Discourse Representation Theory (Kamp and Reyle, 1993).

The motivation for establishing the relevant features for the determiners comes from both syntax and semantics. From the point of view of syntactic theory, specifying the determiner’s relevant features naturally leads to the determination of the parameters of syntactic variation in the DP domain. From the discursive perspective, determinants have a fundamental role, being the most important constituents when it comes to establishing the logical structure of the sentence or of the discourse.

At this stage of the project we have developed a first version of the lexicon, for the determiners studied so far. Thus, we have proposed lexical entries for the following determiners: strong definite determiners *-l* “the”, *cel* “the”, *acest* “this”, *acel* “that”, the universal collective quantifiers *toți* “all” and *amandoi* “both”, pronominal determiners, proper names, etc. For each of these determiners, we have specified their feature matrix, accompanied by relevant explanations/comments for that features. The feature matrix contains morpho-syntactic and semantic features, as they emerged from the studies developed during the project, such as: syntactic category, selectional features, phi-features (person, number, gender), definiteness, quantificational features, cardinality, focus, topic, deixis, proximity, contrastive, location, anaphoric, cataphoric or classifier.

In its broader sense, a feature is simply a formal way of partitioning a syntactic category (such as NP) into sub-categories. A more restricted notion of feature is that of a morpho-syntactic feature, which is a partitioning of a syntactic category into subclasses that are marked (at least partially) morphologically and that are relevant to syntax, most notably by being involved in agreement or government. This covers the phi features such as number, person and

gender, which are all involved in overt agreement, or tense and aspect. Moreover, the term feature also covers semantic features, such as quantificational features, cardinality, etc.

We give here the feature matrices of the determiners mentioned above, as they appear in the constructed lexicon:

- The collective determiner *toți* (all)

$$\begin{array}{c} \text{toți} \\ +A \\ \left[\begin{array}{c} \text{--- [+def]N/D} \\ +\text{number: pl} \\ +\text{gender: masc} \\ uQ \end{array} \right] \end{array}$$

Comment: At the syntactic level, *toți* is a quantificational adjective, which rises from the lexical NP where it merges in order to satisfy its quantificational properties. From the semantic point of view, *toți* is best analysed as a semantic maximiser, not as a quantifier (Cornilescu and Dinu 2013).

- The enclitic article *-(u)l* (the_{masc})

$$\begin{array}{c} \text{-(u)l} \\ +D \\ +N \\ N_ \\ +\text{gender: masc} \\ +\text{number: sing} \\ i\varphi \\ u + \text{def} \end{array}$$

Comment: *-(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-the words”). In Old Romanian, the definite constituent could also occupy the second position within the DP (*ticaloase cuvintele* “evil words-the”). Romanian nominal stems are specified for definiteness (\pm *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 (2008) 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 copilul ăla “poor-the brother-the my, poor-the child-the that”. 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.

- The definite article *cel* (the_{masc})

*cel*₁ (definite article):

$$\left[\begin{array}{c} +D \\ +_QP \\ +gender: masc \\ +number: sing \\ u\varphi \\ u + def \end{array} \right]$$

*cel*₂ (adjectival article):

$$\left[\begin{array}{c} +D \\ +_QP[+focus] \\ +gender: masc \\ +number: sing \\ +EPP \\ u\varphi \\ u + def \end{array} \right]$$

Comment: *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* “the pencil, the red apple”), though it does appear with adjectives when it is an adjectival article or the head of ellipsis (*marul cel rosu, cel rosu* “apple-the the red, the red”). 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*: [[_{Det}[*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 of 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):

[_{DP}[NP [+def] [_D[EPP]cel] [_{QP}[+focus]_{AP} [_{DP} NP[+def]
 marul cel rosu
 “apple-the cel red”

- Demonstrative pronouns and adjectives *acest*, *acel/acesta*, *acela* (this, that)

acest
 +D
 [+X0]
 [+Def]
 +gender: masc
 +number: sing
 $u\phi$
 [+__NP[−def]]
 [+Deixis, +Proximate]
 [+Contrastiv (Locative)]
 ([+Anaphoric])
 ([+Cataforic])

acel
 +D
 [+X0]
 [+Def]
 +gender: masc
 +number: sing
 $u\phi$
 [+__NP[−def]]
 [+Deixis, −Proximate]
 [+Contrastiv (Locative)]
 ([+Anaphoric])
 ([+Cataforic])

acesta

+D
[+XP]
[αDef]
+gender: masc
+number: sg
<i>uφ</i>
[+__NP[+def]]
[+Deixis, +Proximate]
[+Contrastiv Locative]
([+Anaphoric])
([+Cataphoric])

acela
+D
[+XP]
[αDef]
+gender: masc
+number: sg
<i>uφ</i>
[+__NP[+def]]
[+Deixis, -Proximate]
[+Contrastiv Locative]
([+Anaphoric])
([+Cataphoric])

Comment: 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, ±Proximity], [+Contrastive (Location)], [+Anaphoric], [+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 (X^0), not phrases (XPs), which explains why they cannot be used as pronouns, but only as determiners (*acest creion/ *acest* “this pencil/this”). *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 ministru* / **ministrul acesta fost* “this former minister/ minister-the this former”.

- Personal pronouns, such as *noi* (we):

[noi]
[+N]
[+D]
[+u + def]
[+pers: 1st person]
[+number: pl]
[<i>uφ</i>]
[+clasificator]

Comment: Personal pronouns as, for instance, *noi* “we”, are also determiners at the end of the derivation, but there is evidence, coming from the distribution of adjectives (*el frumosul* “he beautiful-the”), that these formatives merge as NPs, triggering definiteness agreement with the adjective. Their characteristic features are person and definiteness.

- The clitic pronoun *l-* (*him_{CL}*)

[<i>l-</i>]
[+D]
[+__DP]
[+u + def]
[+pers: 3rd person]
[+number: sing]
[<i>uφ</i>]

Comments: 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* “*him_{CL}*-have seen *pe Ion*”). We have interpreted this structure on the model of the “big DP” (Uriagereka, 1994, among many), where the clitic selects for a definite DP which doubles it (D0^DP).

- Proper names

Proper Name

$$\left[\begin{array}{c} +N \\ +pers \\ +u + def \\ +classifier \end{array} \right]$$

Comments: Romanian proper names show the same properties as in other languages only in part. Specifically, they must have Classifiers in their extended projection and they check a [person] feature, obligatorily rising to the DP domain. Additionally, in Romanian, proper names are syntactically definite, definiteness being required for marking Case.

Conclusions

We have summarized some of the outputs of the project “The structure and interpretation of Romanian Determiner Phrase in Discourse Representation Theory: the determiners”, stressing the creation of a lexicon for Romanian Determiners, which contains the determiners studied so far and their feature matrix. We are not aware of any other work that thoroughly specifies the morpho-syntactic and semantic features of Romanian Determiners. This lexicon may be used in a wide range of areas, from theoretical linguistics to natural language processing, especially in the case of Romanian, a language with underdeveloped electronic resources.

Acknowledgements:

This work was supported by a grant of the Romanian National Authority for Scientific Research, CNCS - UEFISCDI, project number PN-II-ID-PCE-2011-3-0959.

Selected References:

1. Danon, G. Definiteness spreading in the Hebrew construct state, *Lingua* 118(7), 872-906, 2008.
2. Chomsky, N. *Derivation by phase*. Cambridge, MA: MIT, 1999.
3. A. Cornilescu and A. Dinu. Adjectives and Specificity, *RRL*, LVIII, 4, 455–480, 2013.
3. H. Kamp and U. Reyle. *From Discourse to Logic. An Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, Kluwer Academic Publishers, Dordrecht Netherlands, 1993.
4. Lopez, L. *A Derivational Syntax for Information Structure*, Oxford Studies in Theoretical Linguistics 23, Oxford University Press, 2009.
5. Uriagereka, J. Aspects of Clitic Placement in Western Romance, *Linguistic Inquiry*, 1994.