The Discourse Structuring Potential of Definite Noun Phrases in Romanian

Sofiana Chiriacescu (Sofiana.chiriacescu@ling.uni-stuttgart.de)
Department of Linguistics, Kepler Str. 17
70174 Stuttgart, Germany

Klaus von Heusinger (Klaus.vonHeusinger@ling.uni-stuttgart.de)
Department of Linguistics, Kepler Str. 17
70174 Stuttgart, Germany

Abstract

This paper investigates an alternation found with definite noun phrases in direct object position in Romanian that represents a theoretical puzzle for current theories of Differential Object Marking in this language (Gramatica Limbii Române 2005, Klein & de Swart 2011). When in direct object position and unmodified, definite noun phrases can be accompanied either by the differential object marker *pe*, or by the simple enclitic definite article *-ul*, but not by both at the same time. Based on the findings of a sentence continuation experiment, we show that *pe*-marking is used for noun phrases that show a high discourse structuring potential, which is reflected by their (i) likelihood of subsequent mention (Givón 1983, Arnold 2010) and (ii) topic shift potential (Givón 1983). Furthermore, this paper raises interesting questions about referent tracking in discourse, in particular whether or in what way different types of *definite* noun phrases contribute to the discourse structuring potential of their referents.

Keywords: definite noun phrases; accessibility; referential persistence; topic; production of referring expressions.

Introduction

Romanian is a language that displays Differential Object Marking (DOM, Niculescu 1965, Pană-Dindelegan 1997, Cornilescu 2000). This means that direct objects are sometimes accompanied by the marker *pe*, whereas in other contexts, direct objects remain unmarked. The presence or absence of the *pe*-marker depends on different factors, such as animacy, referentiality and specificity (Farkas 1978, Pană-Dindelegan 1997). Despite the long research tradition on DOM in general and Romanian in particular, the distribution of the marker is still not entirely understood. In this paper we focus on definite unmodified noun phrases as in (1). Here, the noun can be marked with *pe* (e.g. **pe** *băiat* ('**PE** boy') in (1a)), or be marked only with the definite article *-ul* (e.g. *băiatul* (*'the boy'*) in (1b)[1]). Note that *pe*-marking and the definite article cannot co-occur, as reflected by the ungrammaticality of (1c).

In cases in which the noun accompanied by the definite article is modified by an adjective, relative clause, or other modifiers, the presence of the *pe*-marker becomes almost obligatory. Because of this distribution with *pe*-marking, alternations as in (1) were either left unaccounted for (Gramatica Limbii Române, Klein & de Swart 2011), or were explained in terms of genericity (Dobrovie-Sorin 1994) or individualization (Stark 2008).

(1)   (a)   Doctorul   **îl**   examinează   **pe**   băiat.
            Doctor.DEF   **CL**   examines   **PE**   boy
            "The doctor examines the boy."

       (b)   Doctorul   examinează   băiat**ul**.
            Doctor.DEF   examines   boy.**DEF**
            "The doctor examines the boy."

       (c)   * Doctorul   **îl**   examinează   **pe**   băiat**ul**.
            Doctor.DEF   CL   examines   PE   boy.DEF
            "The doctor examines the boy."

In this paper we extend the analysis of indefinite noun phrases in direct object position in Romanian presented in Chiriacescu & von Heusinger (2010) and motivate the distribution in (1) in terms of different discourse structuring potentials of the definite noun phrases. Specifically, in light of the findings of a story-continuation experiment, we show that the presence of the *pe*-marker in (1a) correlates with a higher discourse structuring potential (DSP) of the referents associated with this type of referring expression, compared to those associated with the non-*pe*-marked noun phrase in (1b). The operational definition of DSP employed in this paper is the one developed in Chiriacescu & von Heusinger (2010), where we used two textual characteristics (i.e. referential persistence and topic-shift potential) to determine the discourse status of a referent. DSP is understood as the property of an expression that introduces a discourse referent to provide information about the discourse status of the referent in the subsequent discourse. DSP is measured by means of the two textual characteristics enumerated above, as both metrics pertain to the following discourse

---

[1] Note that in Romanian, *pe*-marked direct objects are usually accompanied by clitic pronouns, which are co-indexed weak pronouns (e.g. the 3.Person, Masc. clitic *îl* in ex. (1a)). Different semantic and syntactic factors are responsible for the presence or absence of clitics with *pe*-marked objects. In this paper, we do not distinguish between clitic doubling and *pe*-marking and address the phenomenon of DOM

as a whole (see Dobrovie-Sorin 1994, or Gramatica Limbii Române 2005 for a discussion on clitic doubling).

This paper aims to shed new light on the phenomenon of DOM in Romanian and in general and to bring theoretical and empirical evidence for the fact that different types of definite noun phrases vary with respect to their DSP.

In the next section we provide a brief overview of the major factors discussed in the linguistic and psycholinguistic literature (Givón 1983, Kaiser & Trueswell 2004, Kehler, Kertz, Rohde & Elman 2008) as textual manifestations of accessibility, salience or discourse prominence. Rather than employing a backward-looking perspective on referent resolution, we will account for the distribution of DOM with definite unmodified noun phrases from the perspective of production by adopting a forward-looking perspective. Next, we discuss a sentence continuation pilot study investigating the discourse prominence of direct objects realized as definite noun phrases in Romanian. The last section summarizes the findings made in this paper and points out interesting questions for further research.

Accessibility and the DSP

A body of linguistic and psycholinguistic research has investigated various factors that influence the comprehension and production of different types of referring expressions (Givón 1983, Gundel, Hedberg & Zacharski 1993, Kehler et al. 2008, Arnold 2010). The majority of these studies focused on anaphora resolution, as it is commonly assumed that reduced referring expressions correlate with highly accessible or prominent entities. Furthermore, to determine the accessibility of a referent, researchers have generally employed a backward-looking perspective, determining the factors that license the usage of a particular type of referring expression at a particular stage in the discourse. In other words, given a certain type of referring expression (e.g. a pronoun), the factors that license its use were investigated.

In contrast to personal pronouns, which refer to previously mentioned entities, definite noun phrases display different kinds of *forward-looking* referential properties: First, they can be used for discourse-new entities or for entities introduced by a bridging or inference relation as well (Hawkins 1978, Vieira & Poesio 2000). Second, both types of definite noun phrases (familiar ones and first mentioned ones) change the accessibility or the activation of the associated discourse referents (von Heusinger 2003, 2007).

In this study we extend this *forward-looking* perspective and test the effects of production-driven biases licensed by *pe*-marked nouns and definite nouns in Romanian. We consider only those definite noun phrases that represent hearer-old and discourse-old information, in the sense that they refer back to a referent, which was explicitly mentioned in the preceding discourse. Rather than exploring the properties of the antecedents of definite noun phrases to determine the accessibility of their referents, we investigate two textual characteristics of the referents of definite noun phrases that pertain to the following discourse, namely referential persistence and topic shift potential.

The first metric, referential persistence reflects the likelihood that a particular referent will be picked again in the ensuing discourse (Givón 1983, Kehler et al. 2008). The second metric for DSP, topic-shift potential, is defined in terms of the

likelihood that a referent will be mentioned in grammatical subject position. We focus on the subject position because different linguistic and psycholinguistic studies (e.g. Crawley & Stevenson 1990) have shown that referents mentioned in the syntactic subject position are more salient or accessible in a given discourse than referents mentioned in other syntactic positions (e.g. as direct or indirect direct objects). For the sake of simplicity, the first instance in which a direct object referent becomes the grammatical subject in a matrix clause is treated as an instance of topic shift.

Despite being mentioned in a rather non-preferential grammatical position (i.e. as a direct object), we expect *pe*-marked definite noun phrases to show higher values for both metrics than their non-*pe*-marked counterparts.

In the remainder of this paper, we present the sentence continuation pilot study conducted to test the DSP of referents realized in direct object position.

The Sentence Continuation Experiment

The experiment presented in this section tested the DSP of direct object referents realized as definite noun phrases. We coded two textual characteristics of the referents, namely: (i) referential persistence (i.e. the likelihood that the referent is picked up in the following discourse), and (ii) topic-shift potential (i.e. the tendency of a referent to be mentioned as the grammatical subject in a subsequent matrix clause).

In a previous study on the DSP of indefinite noun phrases in direct object position in Romanian (Chiriacescu & von Heusinger 2010), we concluded that the referents of the *pe*-marked indefinite noun phrases showed higher DSP values than the referents of their non-marked counterparts. Thus, if *pe*-marked definite noun phrases show a higher DSP than direct objects headed by the simple definite article, then we predict that the former will show higher values for referential persistence and topic-shift potential compared to the latter.

Method and Design

The methodology used was an open-ended sentence continuation task with four test items (Gernsbacher & Shroyer 1989, Chiriacescu & von Heusinger 2010). Participants (n=24) read 4 mini-discourses and were instructed to add 5 logical and natural-sounding sentence continuations to each of them. We manipulated the form of the direct objects in the critical sentence, which resulted in two conditions, i.e. one in which the direct object realized as a definite unmodified noun phrase is *pe*-marked, as in (2), and one in which the same direct object is unmarked, as in (3).

(2) *Sample experimental item for the pe-condition*

La petrecerea de aseara, Andrei$_1$ a cunoscut un politician$_2$ şi un cântăreţ$_3$ de renume. Astăzi **l$_2$-a** intâlnit **pe politician$_2$** în piaţă.
"At yesterday evening's party, Andrew$_1$ met a politician$_2$ and a famous singer$_3$. Today he$_1$ met **PE politician$_2$** at the market".

(3) *Sample experimental item for the non-pe-condition*

La petrecerea de aseara, Andrei$_1$ a cunoscut un politician$_2$ şi un cântăreţ$_3$ de renume. Astăzi a intâlnit **politicianul$_2$** in piaţă.

"At yesterday evening's party, Andrew$_1$ met a politician$_2$ and a famous singer$_3$. Today he$_1$ met **the politician$_2$** at the market".

The fact that we included two sentences in each test item made it difficult for us to control and eventually code every aspect of the mini-discourses, but it provided the greater advantage of creating a natural discourse and combining comprehension and production processes.

Procedure and Data Analysis

(4) Coding methods for an experimental item

*La petrecerea de aseara, **Andrei$_1$** a cunoscut **un politician$_2$** şi* **un cântăreţ de renume$_3$**. *Astăzi **(pro)$_1$** l$_2$-a întâlnit **pe** politician$_2$ în piaţă.*

"At yesterday evening's party, Andrew$_1$ met a politician$_2$ and a famous singer$_3$. Today he$_1$ met PE politician$_2$ at the market."

Table 1: Continuation sentences for an experimental item with coding methods.

Continuation sentences		Ref$_1$ (Sum)	Ref$_2$ (Sum)	Topic
S1	*(pro)$_1$ ştia că acum e şansa lui$_1$.* '**He$_1$** knew that that's **his$_1$** chance.'	2	0	Ref$_1$
S2	*Politicianul$_2$ era un pic grizonat, slăbuţ cu accent baritonal.* '**The politician$_2$** had some greyish hair, was thin with baritone voice.'	0	1	**Ref$_2$**
S3	*Andrei$_1$ s-a dus spre el$_2$, şi (pro)$_1$ i$_2$-a cerut ajutorul să (pro)$_2$ aleagă un pepene bun.* '**Andrei$_1$** went towards **him$_2$** and **he$_1$** asked (**him$_2$**) for help to choose a tasty water melon.'	4	4	(Ref$_1$)
S4	*Politicianul$_2$ s-a întros şi (pro)$_2$ i$_1$-a răspuns cu un aer distrat.* '**The politician$_2$** turned around and (pro)$_2$ responded **him$_1$** in a distracted voice.'	5	6	(Ref$_2$)
S5	*Îl$_2$ chema don Giuseppe şi (pro)$_2$ era inginer zootehnist de meserie.* '**His$_2$** name was don Giuseppe and **he$_2$** was a zootechnician engineer.'	5	8	(Ref$_2$)

We used subscript 1 for the entity that was the subject of the first and the second sentence (e.g. *Andrei* in ex. 2). Subscript 2 was used for the target referent - i.e., the object whose form was manipulated in the critical sentence (e.g. *politician* in ex. 2). Subscript 3 was used for the referent of the singer introduced in the initial mini-discourse. Please note that the critical noun was always mentioned with another conjoined argument in sentence 2.

The first aspect under investigation was ***referential persistence***. We counted the number of times a referent was mentioned in the main and subordinate clauses of the subsequent discourse and the way in which referential persistence relates to grammatical role. For this purpose, we calculated the referential persistence of all referents given in the mini-discourses. The number of times that a referent was mentioned in the continuations was added up to a sum representing the referential persistence of that referent at a particular stage in the discourse (i.e. cumulative value). Comparing the values for referential persistence of all referents, we can gain insights into the exact stage in the discourse at which the critical referent (e.g. politician) has a greater cumulative persistence value than other referents. For example, in the text provided in ex. (4), the referent of the direct object (the politician) is mentioned in S1 through S5 eight times, whereas the referent of the subject is mentioned up to S5 five times.

Topic-shift was the second aspect tested. We did not take into consideration whether this shift was maintained in the following discourse or not. In other words, we did not explore whether the referent of the critical item was mentioned after this point in subject position or not. Furthermore, we considered topic-shifts that occurred in matrix clauses alone, as different studies have shown that referents mentioned as subjects in subordinate clauses do not make good topics (Emonds 1970, among others).

Results

24 participants provided continuations for the initial mini-discourses. The results from the two metrics for DSP, referential persistence and topic-shift potential, indicate the preferential discourse status of the referents headed by *pe*, compared to those headed by the simple definite article. In the following sections, we discuss the findings of the two textual characteristics in detail.

Referential Persistence

Figure 1 displays the mean values for referential persistence of all referents of the test items 1-4 (TI1-TI4). We found a highly significant difference between *pe*-marked and unmarked direct objects ($p < 0.0001$) with respect to referential persistence. Specifically, for the *pe*-condition, we notice a strong likelihood for the direct object referents to be highly recurrent in the following discourse. In contrast, in the non-*pe*-marked condition direct object referents are mentioned less often in the subsequent discourse. Thus, these results confirm our predictions regarding the higher DSP of *pe*-marked referents, as these referents display a strong tendency to be picked up in the subsequent discourse and a high referential persistence in the ensuing discourse.
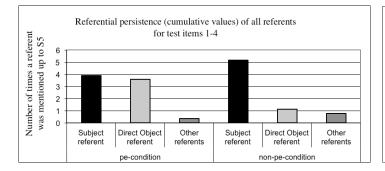
Figure 1: Referential persistence of all referents introduced in the initial mini-discourses in all test items up to S5.
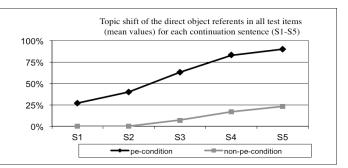

Figure 2: The topic shift potential of referents mentioned in both conditions, measured as the cumulations of the *first* occurrences of the direct object in subject position in S1-S5.

Furthermore, Figure 1 shows that the difference in referential persistence between subject and direct object referents is smaller in the *pe*-condition: The black subject bar and the light gray direct object bar have similar heights in the *pe*-condition, but in the non-*pe*-condition, the subject bar is much higher that the direct object bar. In other words, it seems that the referent of the *pe*-marked referent becomes a better competitor for the subject referent in terms of referential persistence.

This observation brings up an interesting question that pertains to the interaction of the referential persistence values of a discourse's referents. More concretely, is the referential persistence of the direct object referents suppressing (in the *pe*-condition), or enhancing (in the non-*pe*-condition) the referential persistence of other referents (e.g. of the subject referents)? Or are the two values rather independent of each other? (for related discussions, see Gernsbacher & Shroyer 1989).

In sum, we find that the *pe*-marker reverses the general low likelihood or expectancy found with referents realized in direct object position, as these referents tend to be picked up frequently in the ensuing discourse.

Topic Shift
The second textual characteristic investigated was the topic shift potential of direct object referents. Recall that the first time the referent of a direct object is mentioned in grammatical subject position in a matrix clause is counted as an instance of topic shift (Givón 1983). The counts for the topic shift potential are cumulative.

The findings condensed in Figure 2 reveal several patterns. We found a significant difference between *pe*-marked direct objects and non-*pe*-marked objects with respect to topic shift ($p<0.021$). Specifically, the referent of the *pe*-marked direct object displays a stronger preference to become a subject in the continuation sentences (S1-S5) than the referent of the non-*pe*-marked direct object referent. While almost all participants mentioned the referent of the *pe*-marked direct object sooner or later as a subject in the continuation text, the unmarked direct object became a subject in less than 25% of cases.

Figure 2 shows that the referent of the unmarked direct object was never picked up in subject position in the first two continuation sentences (S1 and S2) provided by the participants. In contrast, when we look at the *pe*-marked conditions, we see that the referent of the *pe*-marked direct object was occasionally mentioned in subject position in the first two continuation sentences, even though the rate was not very high (around 35%).

The findings concerning the topic shift potential of direct objects realized as definite unmodified noun phrases confirmed our prediction that *pe*-marked referents are expected to be more prone to shift the topic of the following discourse (i.e. to become the grammatical subject) compared to the unmarked referents.

Overall, the results of the sentence-continuation pilot study presented in this section parallel those reported in Chiriacescu & von Heusinger (2010) and Chiriacescu (2011) about the discourse behaviour of indefinite noun phrases in direct object position, as *pe*-marking contributes to the higher DSP of the noun phrase that it precedes.

Conclusion

The findings of the pilot study described in this paper extend an expectancy-driven approach to language processing (e.g. Kehler et al. 2008, Arnold 2010). We have shown that referents realized in positions that are otherwise low in accessibility (i.e. referents realized as definite noun phrases in direct object position) will show high DSPs, whenever this potential is encoded by a special marker, such as *pe* in Romanian. The consequences of the *pe*-marking on definite unmodified noun phrases, as exemplified in (1), are explained by showing that the relevant discourse contribution of *pe* is to signal to the addressee that further information about the referent marked in this way will follow and that the same referent is more likely to be picked up in grammatical subject position (i.e. as a topic constituent) in the following discourse.

In light of the present findings, we expect different markers of definite noun phrases to cross-linguistically vary with respect to the discourse structuring potential of the referents they are associated with (confirmation of prior findings on other types of referring expressions, e.g. Chiriacescu & von Heusinger (2010) and Chiriacescu (2011) on indefinite noun phrases in Romanian, German and English).

References

Arnold, J. (2010). How speakers refer: the role of accessibility. *Language and Linguistic Compass*, 4, 187-203.

Chiriacescu, S. (2011). The discourse structuring potential of indefinite noun phrases. Special markers in Romanian, German and English. Ph.D. dissertation, University of Stuttgart.

Chiriacescu, S., & von Heusinger, K. (2010). Discourse prominence and *pe*-marking in Romanian. International Review of Pragmatics 2(2), 298-332.

Cornilescu, A. (2000). Notes on the Interpretation of the Prepositional Accusative in Romanian. *Bucharest Working Papers in Linguistics* (1), 91−106.

Crawley, R., Stevenson, R., & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. Journal of Psycholinguistic Research (19), 245-264.

Dobrovie-Sorin, C. (1994). *The Syntax of Romanian*. Berlin: Mouton de Gruyter.

Emonds, J. (1970). Root and structure-preserving transformations. Doctoral dissertation. MIT, Cambridge, Mass.

Farkas, D. (1978). Direct and indirect object reduplication in Rumanian. In: D. Farkas, W.M. Jacobsen & K.W. Todrys (eds.). *Papers from the Seventeenth Regional Meetithe Chicago Linguistics Society* (*CLS 14*). Chicago: Chicago Linguistic Society, 88-97.

Gernsbacher, M., & Shroyer, S. (1989). The cataphoric use of The indefinite 'this' in spoken narratives. *Memory and Cognition* 17: 536 - 540.

Givón, T. (1983). Topic continuity in discourse: An introduction. In T. Givón (Ed.), *Topic Continuity in Discourse. A Quantitative Cross-Language Study*. Amsterdam, Philadelphia: John Benjamins.

Gramatica Limbii Române. (2005). Vol I. Cuvantul. Bucureşti: Editura Academiei.

Gundel, J., Hedland, N., & Zacharski, R. (1993). Cognitive Status and the form of referring expressions in discourse. *Language, 69,* 274-307.

Hawkins, J. (1978). Definiteness and Indefiniteness- A Study in the Reference and Grammaticality Prediction. Croom Helm, London.

von Heusinger, K. (2003). The Double Dynamics of Definite Descriptions. In: J. Peregrin (ed.). *Meaning in the Dynamic Turn*. Amsterdam: Elsevier, 150-168.

von Heusinger, K. (2007). Accessibility and Definite Noun Phrases. In: M. Schwarz-Friesel & M. Consten & M.Knees(eds.). *Anaphors in Text: Cognitive, Formal and Applied Approaches to Anaphoric Reference*. Amsterdam: Benjamins, 123-144.

Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word order language. *Cognition, 94*, 113–147.

Kehler, A., Kertz, L., Rohde, H., & Elman, J. (2008). Coherence and Coreference Revised. *Journal of Semantics, 25*, 1-44.

Klein, U., & de Swart, P. (2011). Case and Referential Properties. *Lingua* 122 (1), 3-19.

Niculescu, A. (1965). Obiectul direct prepozitional in limbile romanice. In: Individualitatea limbii romane intre limbile romanice. Bucuresti: Editura Stiintifica.

Pană-Dindelegan, G. (1997). Din noun despre statutul Prepozitiei. Cu referire specială la prepoziţia PE. *LRXLVI. Nr. 1-3*, 27-55.

Stark, E. (2008). What is marked by differential object marking in Romance. Paper presented at the workshop on Case Variation, University of Stuttgart, 19 June 2008.

Vieira, R., & Poesio, M. (2000). An empirically-based system for processing definite descriptions. *Computational Linguistics, 26(4)*: 539–593.