Topic shift and *pe*-marking in Romanian

Sofiana Chiriacescu

University of Stuttgart

Sofiana.chiriacescu@ling.uni-stuttgart.de

1. Introduction

This talk is about some instances of variation with *pe*-marking (as DOM) in the realm of definite unmodified (1) and indefinite descriptions (2).

(1) A: Un băiat merge la doctor. (A boy goes to the doctor.)
   (a) Doctorul   examinează   băiatul.
       Doctor.DEF examines   boy.DEF
       "The doctor examines the boy."
   (b) Doctorul   îl   examinează pe băiat.
       Doctor.DEF CL examines   PE   boy
       "The doctor examines the boy."

(2) (a) Petru l-   a   văzut   pe   un   băiat.
       Peter CL   have   seen   PE   a   boy
       "Peter saw a boy."
   (b) Petru   a   văzut un   băiat.
       Peter   has   seen a   boy
       "Peter saw a boy."

Based on the findings of two web-based story continuation experiments I will show that (i) *pe*-marked referents are referentially more persistent than their unmarked counterparts and that (ii) *pe*-marked referents show a systematic preference to become topics two or three sentences after being introduced in the discourse.

2. *Pe*-marking in Romanian and the referentiality scale

Romanian (see Niculescu1965, Bossong 1985) shows differential object marking (DOM). There is some consensus in the literature that the most important synchronic conditions triggering DOM in Romanian are animacy, definiteness, specificity and topicality (Farkas 1978, Dobrovie-Sorin 1994, Kamp & Bende Farkas (submitted), among others).

pers. pron.	> PN	> def. NP	> spec. indef NP	> non-spec. indef NP	> non-arg NP

Table 1. Referentiality Scale

Full personal pronouns referring to animate entities are always marked with *pe* and doubled by a clitic in synchronic Romanian[1]. Proper names referring to humans or to strongly individuated, personified animals are always *pe*-marked. Modified human definite NPs in direct object position are generally *pe*-marked, while the form without *pe* is rather marginal.

2.1 Definite unmodified NPs

Speakers of Romanian can either drop the marker *pe* and keep the definite enclitic article *–ul* (DEF.masc) as in (3a), or drop the definite article and keep *pe,* as in (3b).

(3) A: Un băiat merge la doctor. (A boy goes to the doctor.)
 (a) Doctorul   examinează   băiatul.
  Doctor.DEF examines   boy.DEF
  "The doctor examines the boy."

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

Both sentences (3a) and (3b) represent different possibilities of expressing very similar referential categories as they are both grammatical and have the same propositional content. The exact distributional contexts for the two forms have not been delimited yet. This variation with definite NPs will be discussed in the second experiment in section 4.

2.2 Indefinite NPs and specificity

The literature (Farkas 1978, Dobrovin-Sorin 1994) assumes that specificity is the main triggering parameter for DOM with indefinite human direct objects in Romanian. Scopal specificity with extensional operators, as in (4) and referential specificity with intensional operators trigger *pe*-marking.

(4)   *Extensional operators*

 (a) Toți bărbații iubesc o femeie.
  All   men   love   a woman
  "All men love a woman." (wide scope/ narrow scope)

 (b) Toți bărbații o   iubesc pe o femeie.
  All men   CL love   PE a woman
  "All men love a woman." (only narrow scope)

---

[1]DOM in Romanian is generally accompanied by clitic doubling, i.e. the occurrence of a co-indexed weak pronoun. A doubling clitic is optional, obligatory or blocked, depending on semantic features of the head noun and further syntactic constraints (see Dobrovie-Sorin 1994, Gramatica Limbii Române 2005).

In plain and transparent contexts, indefinite NPs have been accounted for in terms of epistemic specificity. Sentence (5a) is analyzed as an instance of epistemic specificity, as it is said to reflect the knowledge of the speaker (or of some other salient agent) about the identity of the referent, while (5b) is again ambiguous between a specific and a non-specific reading.

(5) *Transparent contexts*

    a.   Petru l-   a   văzut   pe   un   băiat.
           Peter CL   have   seen   PE   a   boy
           "Peter saw a boy." (specific)

    b.   Petru   a   văzut un   băiat.
           Peter   has   seen a   boy
           "Peter saw a boy." (specific/ non-specific)

3. Determinants of discourse prominence: how do referents become accessible?

Most studies dealing with discourse prominence/ accessibility concentrated on the factors that license the use of a pronoun in a certain context, as accessible referents tend to be pronominalized (Givón 1983, Gundel et al. 1993, Ariel 2001, etc.). The discourse history of a certain referent (the relation between antecedent and anaphor) is of utmost importance in such studies. The four more important accessibility enhancing factors distinguished were: Givenness, Continuity, Syntactic prominence and Semantic prominence. Accordingly, a referent which represents old information (as opposed to new information), which was/ will be mentioned more often, in a syntactic prominent position (i.e. subjects vs. objects, the focus of clefts) and in a semantic prominent position (e.g. the Stimulus referent in a transitive event with Stimulus and Experiencer roles) is considered to be highly accessible and correlates with the speaker's referential choice (a pronoun will be presumably used in this case).

Two textual characteristics of the DO referents were investigated[1]:

(i)   *Referential persistence* (Givón 1983, Arnold 1998, Ariel 2001)
      - looks at how often a referent is mentioned in the subsequent discourse;

(ii)   *Topic shift potential* (Givón 1983)
      - calculates whether a given referent has the potential to become a topic (i.e. the grammatical subject) in the subsequent discourse;

---

[1] In this talk I focus on the two textual factors (i) and (ii) below. However, see Chiriacescu & von Heusinger (2010) for an ample discussion of other factors as well.

Prediction 1 (referential persistence):

The referential persistence of the *pe*-marked referents will exceed the referential persistence of their unmarked counterparts.

Prediction 2 (topic shift potential):

In comparison to their unmarked counterparts, case marked direct objects will (i) be mentioned more often in subject position in the subsequent text; and will (ii) appear as subjects earlier in the continuation test (but from the second continuation sentence on).

4. Web-based experiment for prominence with *pe*-marking

4.1 Experiment 1: *pe*-marking with indefinite NPs

4.1.1 Method

4.1.1.1 Participants

23 native speakers of Romanian participated in this experiment. The approximate time needed to complete the experiment was 7 minutes.

4.1.1.2 Materials

Methodology used: sentence continuation task based on three-sentence stories. Six target stories were constructed and no filler stories were added. The first two sentences in each test item (TI for test item) set the context of the story and contained an individual reference to a character (other than the new referent of the direct object introduced in the third sentence). The third TI contained the target direct object realized as an indefinite NP. The first introduced character was the clearly established topic constituent of the story (mentioned at least once in subject position). The subject and object in the target sentences were human and were of different gender. The third stimulus item (TI3 in Table 1 in the Appendix) contained an additional last sentence, in which no relevant referent was mentioned.

I manipulated the realisational form of the direct objects (*pe*-marked direct objects in subexperiment A and unmarked direct objects in subexperiment B), as in (6). The full list of TIs is given in Table 1 in the Appendix.

(6)   (a) Direct Object [with pe]

Anul trecut când am fost la mare am cunoscut un salvamar. Era tot timpul activ. La sfârşitul sejurului meu, a salvat-o pe o fată de la înec.
"Last year when I was at the seaside I met a lifeguard. He was very active all day long. At the end of my stay there, he saved PE-a girl from drowning."

(b) Direct Object [without pe]

Anul trecut când am fost la mare am cunoscut un salvamar. Era tot timpul activ. La sfârşitul sejurului meu, a salvat o fată de la înec.
"Last year when I was at the seaside I met a lifeguard. He was very active all day long. At the end of my stay there, he saved a girl from drowning."

4.1.1.3 Procedure and data analysis

- Participants: were asked to read the given sentences and add five natural-sounding continuation sentences. The given stories were presented in writing and each item was presented individually.

- I analysed 11 continuations for subexperiment A (with *pe*) and 9 for subexperiment B (without *pe*).

- The first five main clauses (including subordinate ones) were coded with respect to the two factors introduced in section 3.

- Example (7) presents one response from subexperiment A for the test item 2 (TI2), and Table 2 illustrates the coding methods used.

(7) Example responses and coding methods from the story continuation experiment:

---

*Test Item 2: Anul trecut când am fost la mare am cunoscut **un salvamar**$_1$. **(Pro)**$_1$ era tot timpul activ. La sfârşitul sejurului meu, **(pro)**$_1$ a salvat-**o pe o fată**$_2$ de la înec.*
"Last year when I was at the seaside I met a lifeguard$_1$. He$_1$ was very active all day long. At the end of my stay there, he$_1$ saved PE-a girl$_2$ from drowning."

S1: *M-a mirat ca **(pro)**$_1$ a reusit sa **o**$_2$ salveze, pentru ca **(pro)**$_1$ era un burtos.*
"I was surprised that **he$_1$** could save **her$_2$**, because **he$_1$** was a big-bellied man."

S2: *Dar cum **(pro)**$_1$ a vazut-**o pe fata**$_2$ ca **(pro)**$_2$ striga dupa ajutor, **(pro)**$_1$ nu a stat pe ganduri si **(pro)**$_1$ s-a dus dupa **ea**$_2$.*
"But as soon as **he$_1$** saw **PE-girl$_2$** screaming for help, **he$_1$** did not hesitate and went after **her$_2$**."

S3: *Fata$_2$ era inconstienta cand **(pro)**$_2$ a fost adusa la tarm.*
"**The girl$_2$** was unconscious as **she$_2$** was brought to shore."

S4: *(pro)$_2$ nu avea mai mult de 10 ani.*
"**She$_2$** was not older than 10."

S5: *Parintii$_3$ **ei**$_2$ i-au multumit **salvamarului**$_1$.*
"**Her$_2$** parents thanked **the lifeguard$_1$**."

	Coding methods	Subject		Object		
	Anaphoric forms and grammatical function	refer per		refer per		topic
		item / S	sum	item / S	sum	
S1	(pro$_1$, Cl$_2$) (pro$_1$) (Sub$_1$, DO$_2$) (Sub$_1$)	2	2	1	1	Topic$_1$
S2	(pro$_1$, def NP$_2$) (pro$_2$) [pro$_1$] [pro$_1$, pron$_2$] (Sub$_1$, DO$_2$) (Sub$_2$) [Sub$_1$] [Sub$_1$ DO$_2$]	3	5	3	4	Topic$_1$
S3	[def NP$_2$] (pro$_2$) [Sub$_2$] (Sub$_2$)	0	5	2	6	Topic$_2$
S4	[pro$_2$] [Sub$_2$]	0	5	1	7	(Topic$_2$)
S5	[def mod NP$_3$, pron$_2$, def NP$_1$] [Sub$_3$, attr$_2$, IO$_1$]	1	6	1	8	(Topic$_3$)

Table 2. Coding methods for the continuation sentences in ex. (7)

- The type of referring expressions used to take up the referents (pro, clitic, personal pronoun, definite NP and definite modified NP) of both subject and object referents and their grammatical functions are also listed above.

- Referential persistence is measured by items per sentence (item / S) and the sum of all items up to S5 as an indicator of the stage in the discourse in which more anaphoric expressions are used to refer to the direct object referent than to the subject (in ex. (7) in S3). The persistence of the subject referents was also calculated.

- Topic shift: in which continuation sentence does the direct object referent become a subject? (in ex. (7) in S3).

4.1.2 Results

4.1.2.1 Number of anaphoric references – referential persistence

Figure 1 below displays the average number of anaphoric references of the subjects and objects up to the last continuation sentence (S5) for both subexperiments (subexperiment A [+pe] and subexperiment B [-pe]) for the three tested items.
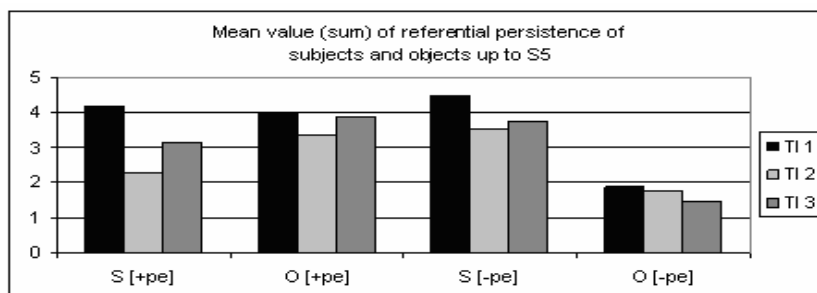

Figure 1. Referential persistence of subjects and objects up to S5 for indefinite NPs

Prediction 1 concerning the referential persistence of the direct objects was confirmed: the *pe*-marked direct object is more persistent in the subsequent discourse than the unmarked DO.
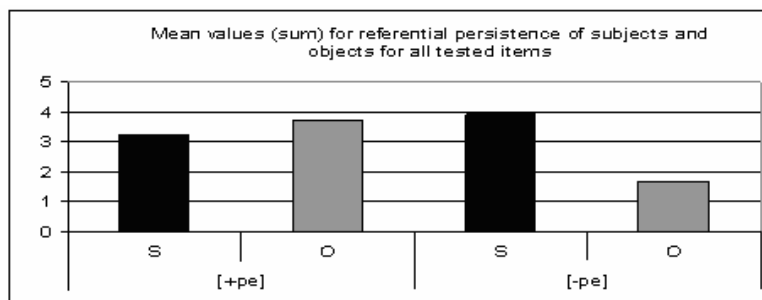

Figure 2. Mean values for referential persistence of subject and object in both subexperiments
for indefinite NPs

Subexperiment A: the referential persistence of the *pe*-marked DO referent exceeds even the referential persistence of the subject referent (Figure 2). In sum, participants preferred a continuation story which evolved around the referent of the subject, thus taking it up more often, unless the direct object referent was *pe*-marked.

4.1.2.2 *Topic shift potential*

Prediction 2 was confirmed. 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.
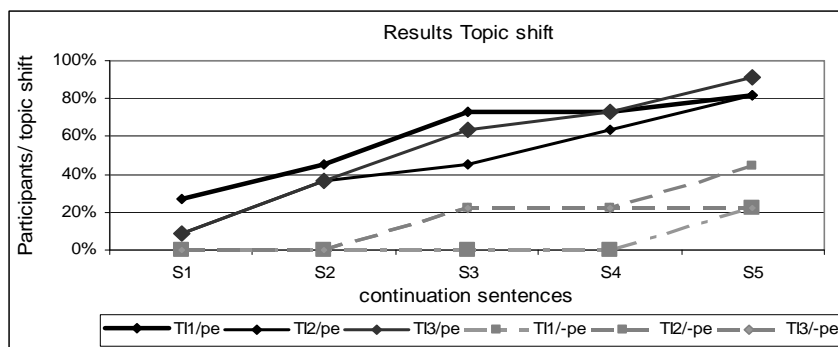

Figure 3. Results: Topic shift for indefinite NPs

As predicted, the *pe*-marked referents became the new topics earlier in the subsequent discourse than their unmarked counterparts. Few direct objects (whether marked or unmarked) became subjects in S1 and S2 (→ due to the priviledged status of the subject referent).

4.1.3   Interim discussion

Prediction 1 (referential persistence) and Prediction 2 (topic shift potential) are confirmed. *Pe*-marked referents are preferred to be continued with in comparison to the unmarked direct objects.

They also make good candidates for topics (i.e. the new subjects) after the second continuation sentence on- contrasted to the unmarked DO referents.

4.2 Experiment 2- definite NPs
4.2.1 Method
4.2.1.1 Participants
24 native speakers of Romanian participated in the experiment. The approximate time needed to complete an experiment was 9 minutes.

4.2.1.2 Design, Procedure and Materials
The same methodology and coding methods as in Experiment 1 were used to test the discourse behaviour of definite unmodified NPs in direct object position.10 participants completed Subexperiment A (with *pe*) and 10 other participants completed subexperiment B (without *pe*).

4.2.2 Results
4.2.2.1 Referential persistence

- The referential persistence of all referents (subjects, objects and other referents) in the two subexperiments is roughly the same (i.e. 8.2 in subexperiment A and 7.5 in subexperiment B).
- Prediction 1 (referential persistence) was again confirmed: The referents of the *pe*-marked direct objects are mentioned more often in the subsequent discourse, than the unmarked ones (see Figure 4.).
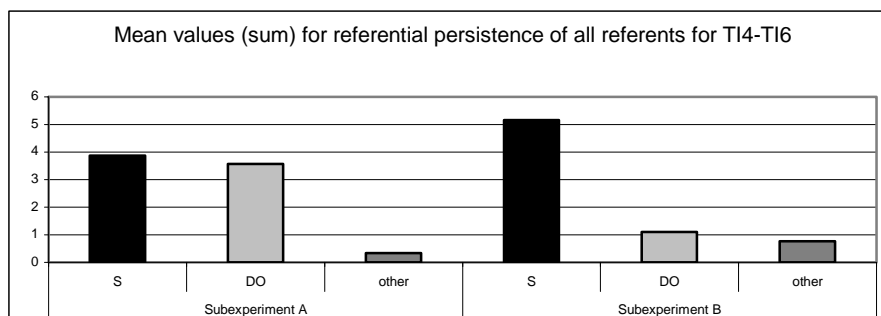

Figure 4. Referential persistence of the referents of the definite NPs

- The continuation sentences for TI5 and TI6 show that the *pe*-marked referents outrank the subject referents in referential persistence.
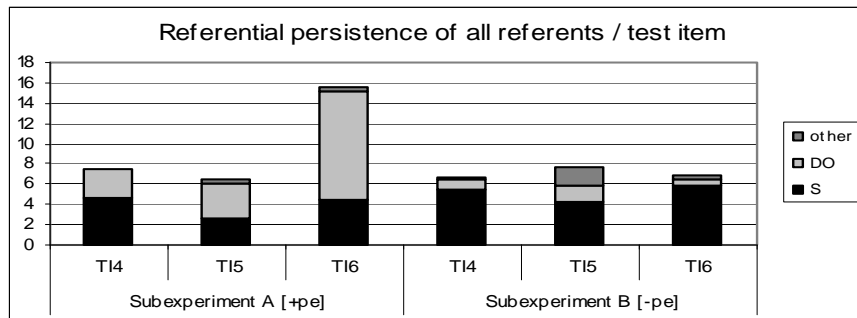
Figure 5. Referential persistence of the referents of the definite NPs for all tested items

4.2.2.2 Topic shift potential

- *Pe*-marked referents display an overall higher preference to be mentioned as subjects in the subsequent discourse, in comparison to the unmarked referents (Prediction 2 confirmed).

- Marked direct objects start being picked up as a subject from the first continuation sentence on. On the contrary, unmarked direct objects are very rarely mentioned in subject position in the first two continuation sentences, as figure (6) shows:
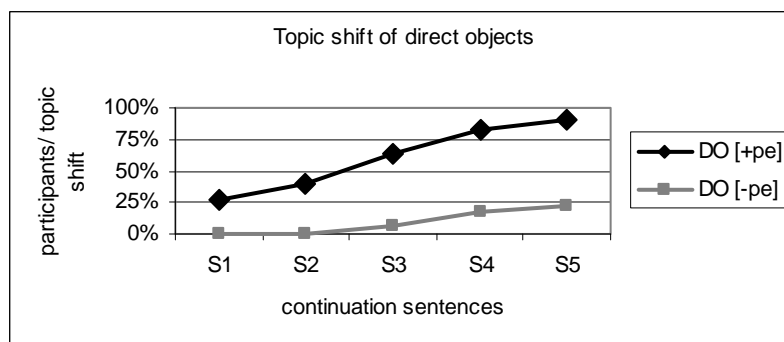

Figure 6. Topic shift for all tested items in both subexperiments

- We saw in section 4.2.3.1. above that the *pe*-marked direct objects in TI4 showed a lower referential continuity than TI5 and TI6. We furthermore notice that the same DO referents are the least probable to be mentioned as subjects early in the continuation sentences (i.e. S1 and S2). In contrast to that, the referentially persistent DOs in TI5 and T6 become subjects earlier in the continuation sentences (i.e. in S1 and S2).

- In sum: Predictions 1 and 2 confirmed.

4.3 General findings of Experiment 1 and Experiment 2

- Referential persistence and topic shift underlined the privileged status of the *pe*-marked referents (whether expressed by and indefinite NP or by a definite unmodified NP) and thus confirmed Predictions 1 and 2.

9

- The higher likelihood of a referent to be mentioned in the subsequent discourse is not a reflex of the high activation of the referent, but rather represents a mechanism employed by the speaker to link the hearer's attention to an entity which will be further elaborated upon.
- Psycholinguistic research (Elman 1993, Saffran et al. 1996, Bicknell et al. 2008, Levy 2008) has convincingly shown that statistical regularities are observed at different levels of linguistic output. It seems that hearers identify frequency patterns in order to predict what is likely to occur in the following context. The referential persistence of the *pe*-marked NPs analyzed here shows that language users make use of such regularities at the discourse level as well.

5. Conclusion

- The cases of optional *pe*-marking with indefinite NPs and definite unmodified NPs were accounted for in terms of discourse prominence. More concretely, it was shown that *pe*-marking is used to signal the topic shift potential of the direct object referents.
- The function of *pe*-marking in 'neutral contexts' was shown to be the same for indefinite as well as for definite unmodified NPs.
- If the findings concerning *pe*-marking in Romanian are right, then we should find similar distinctions in other languages as well.
- Two main questions arise at this point. The first question is concerned with the relation between the property of referential persistence and other referential properties expressed in a sentence, such as definiteness and specificity. If referential persistence signals the referential intention of the speaker to use the introduced referent in the subsequent discourse, this intention will also determine the semantic-pragmatic properties of the expression in the sentence in which it is used.
- The second question pertains to the diachronic development of DOM in Romanian and the role discourse prominence played in the grammaticalization process of *pe*-marking.

Selected references

Ariel, Mira. 2001. Accessibility theory: An overview. In Ted J.M. Sanders, Joost Schilperoord & Wilbert Spooren (eds.), *Text representation: Linguistic and psycholinguistic aspects,* 29–87. Amsterdam: John Benjamins.

Arnold, Jennifer. 1998. *Reference Form and Discourse Patterns*. PhD dissertation, Stanford University.

Bicknell, K., Elman, J.,Hare, M., McRae & Kutas. 2008. Online expectations for verbal arguments conditional on event knowledge. In Love, Bernd, McRae, K., & Sloutsky, V. M. (eds.), *Proceedings of the 30th annual conference of the cognitive science society*, 2220–2225. Austin, TX: Cognitive Science Society.

Bossong, Georg. 1985. *Empirische Universalienforschung. Differentielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen: Narr.

Chiriacescu, Sofiana and Klaus von Heusinger. *Discourse prominence and pe-marking in Romanian.* In: the International Review of Pragmatics 2(2). 298- 332.

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

Elman, J. 1993. Learning and development in neural networks: The importance of starting small. *Cognition* 48: 71-99.

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

Givón, Talmy. 1983. Topic continuity in discourse: An introduction. In T. Givón (ed.), *Topic Continuity in Discourse. A Quantitative Cross-Language Study*, 1-41. Amsterdam, Philadelphia: Benjamins.

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

Grosz, Barbara, Joshi Aravind & Scott Weinstein. 1995. Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics* 21: 203-226.

Gundel, Jeanette, Nancy Hedland & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69: 274-307.

Kamp, Hans & Agnes Bende Farkas. (submitted). Epistemic specificity from a communication- theoretic perspective. Submitted to *Journal of Semantics*.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106: 1126-1177.

Niculescu, Andrei. 1965. Obiectul direct prepozitional in limbile romanice. In *Individualitatea limbii romane intre limbile romanice.* Editura Stiintifica: Bucuresti.

Saffran, Jenny, Richard Aslin & Newport. 1996. Statistical learning by 8-month-old infants. *Science* 274: 1926–1928.

Appendix A

Full list of stimulus items used:

Experiment 1 (Indefinite descriptions)

	DO [with *pe*]	DO [without *pe*]
TI1	*Aseară a fost extraordinar de cald. Pentru că nu mai rezista în casă, Graur₁ s-a hotărât să se ducă în oraş. Pe drum l- a văzut pe un copil₂ intrând într-un magazin.* "It was extraordinarily warm outside yesterday evening. Because it was unbearable for him to stay home anymore, Graur₁ decided to go downtown. On his way there he saw **PE a child₂** entering a store."	*Aseară a fost extraordinar de cald. Pentru că nu mai rezista în casă, Graur₁ s-a hotărât să se ducă în oraş. Pe drum a văzut un copil₂ intrând într-un magazin.* "It was extraordinarily warm outside yesterday evening. Because it was unbearable for him to stay home anymore, Graur decided to go downtown. On his way there he saw **a child₂** entering a store."
TI 2	*Anul trecut când am fost la mare am cunoscut un salvamar₁. Era tot timpul activ. La sfârşitul sejurului meu, a salvat-o pe o fată₂ de la înec.*	*Anul trecut când am fost la mare am cunoscut un salvamar₁. Era tot timpul activ. La sfârşitul sejurului meu, a salvat o fată₂ de la înec.*

	"Last year when I was at the seaside I met a lifeguard[1]. He was very active all day long. At the end of my stay there, he saved **PE a girl[2]** from drowning."	"Last year when I was at the seaside I met a lifeguard[1]. He was very active all day long. At the end of my stay there, he saved **a girl[2]** from drowning."
TI 3	*Oaspeţii stăteau în sufragerie neştiind cu cine să înceapă o conversaţie. Ana[1] l-a văzut pe un bărbat[2] stând singur lângă canapea. Între timp, chelnerii serveau mâncarea.* "The guests were anxiously sitting in the living room, not knowing whom to start a conversation with. Ana[1] saw **PE a man[2]** sitting alone on the couch. Meanwhile, the waiters were serving food."	*Oaspeţii stăteau în sufragerie neştiind cu cine să înceapă o conversaţie. Ana[1] a văzut **un bărbat[2]** stând singur lângă canapea. Între timp, chelnerii serveau mâncarea.* "The guests were anxiously sitting in the living room, not knowing whom to start a conversation with. Ana[1] saw **a man[2]** sitting alone on the couch. Meanwhile, the waiters were serving food."

Table 1. Test items for the first experiments (indefinite NPs)

Experiment 2 (DOM with definite NP)

	DO [with *pe*]	DO [without *pe*]
TI4	*Daniela[1] lucrează atât la spital cât şi într-o clinică privată. O ambulanţă a intrat pe poarta spitalului la ora 23, iar Daniela a fost chemată de acasă pentru a-l opera pe pacient[2].* "Daniela[1] works at the hospital and in a private clinic. An ambulance entered the hospital's gate at 11 p.m. and Daniela was asked to come and operate on **PE patient[2].**"	*Daniela[1] lucrează atât la spital cât şi într-o clinică privată. O ambulanţă a intrat pe poarta spitalului la ora 23, iar Daniela a fost chemată de acasă pentru a opera **pacientul[2].*** "Daniela[1] works at the hospital and in a private clinic. An ambulance entered the hospital's gate at 11 p.m. and Daniela was asked to come and operate on **the patient[2].**"
TI5	*După accident, Cristina[1] a stat de vorbă cu un martor[2]. In scurt timp, la fata locului a sosit un echipaj de politie[3]. Acestia **l-au luat pe martor[2]** de-o parte si au oprit traficul in zona pentru scurt timp.* "After the accident, Cristina[1] talked to a witness[2]. A police team[3] arrived at the site shortly afterwards. They took **PE witness[2]** aside and stopped the traffic in the area for a short while."	*După accident, Cristina a stat de vorbă cu un martor. In scurt timp, la fata locului a sosit un echipaj de politie. Acestia au luat **martorul** de-o parte si au oprit traficul in zona pentru scurt timp.* "After the accident, Cristina[1] talked to a witness[2]. A police team[3] arrived at the site shortly afterwards. They took **the witness[2]** aside and stopped the traffic in the area for a short while."
TI6	*La petrecerea de aseara, Andrei[1] a cunoscut un politician[2] si un un cantaret de renume[3]. Astazi **l-a intalnit pe politician[2]** in piata.* "At yesterday evening's party, Andrew[1] met a politician[2] and a famous singer[3]. Today he met PE politician[2] at the market."	*La petrecerea de aseara, Andrei[1] a cunoscut un politician[2] si un un cantaret de renume[3]. Astazi a intalnit **politicianul[2]** in piata.* "At yesterday evening's party, Andrew[1] met a politician[2] and a famous singer[3]. Today he met the politician[2] at the market."

Table 2. Test items for the second subexperiment (without *pe*)