Pe-marked definite NPs in Romanian and discourse prominence

Klaus von Heusinger & Sofiana Chiriacescu
Klaus.vonheusinger@ling.uni-stuttgart.de
Sofiana.chiriacescu@ling.uni-stuttgart.de

1. Introduction*

Romanian is a language which exhibits differential object marking (DOM) using the particle pe and (in most cases) a clitic pronoun (Niculescu 1965, Pană-Dindelegen 1997, von Heusinger & Onea 2008, Stark & Sora 2008). Direct object case marking is optional with human indefinite direct objects, as in (1), and almost obligatory with postverbal human definite direct objects, as reflected in the contrast between (2a) and (2b).

(1) (a) Doctorul îl vizitează pe un băiat.
    doctor.DEF CL visits PE a boy
    ‘The doctor visits a boy.’
(b) Doctorul vizitează un băiat.
    doctor.DEF visits a boy
    ‘The doctor visits a boy.’

(2) (a) Doctorul îl examinează pe băiatul bolnav.
    doctor.DEF CL examines PE boy.DEF sick
    ‘The doctor examines the sick boy.’
(b) #Doctorul examinează băiatul bolnav.
    doctor.DEF examines boy.DEF sick
    ‘The doctor examines the sick boy.’

There is an interesting mismatch in the otherwise quite robust distribution of pe-marking with definite noun phrases. Modified human definite direct objects obligatorily receive DOM, as in (2a), but the co-occurrence of pe and the definite article is blocked if the noun phrase is not (further) modified. This blocking effect on pe-marking derives from an independent syntactic rule of Romanian which holds for most prepositions, including pe in its case-marking function, as in (3a). There are, however, two alternative constructions for the ungrammatical (3a), namely (3b), in which the definite article is present and the pe marker is absent, and (3c), in which the marker pe precedes the noun phrase in the absence of the definite article, yielding a definite reading.

(3) (a) *Doctorul îl examinează pe băiatul.
    doctor.DEF CL examines PE boy.DEF
This paper focuses on the contrast between the alternation in (3b) and (3c) and investigates the parameters that determine the choice of one form over the other. We argue that pe-marked definite direct objects are discourse prominent in a sense that is related to the concept of secondary topics introduced by Dalrymple & Nikolaeva (2011). In light of the findings of a web-based story continuation experiment we show that pe-marked definite direct objects qualify for secondary topics since they are (i) referentially more persistent than their unmarked counterparts, and since (ii) they show a systematic preference to become topics two or three sentences after being introduced in the discourse.

The findings with respect to the distribution of pe-marking with definite unmodified noun phrases contribute to the general understanding of DOM and introduce an empirical challenge for the two main approaches on the function of DOM: the Ambiguity Thesis and the Transitivity Thesis. The Ambiguity Thesis (Moravcsik 1978, Croft 1988, Bossong 1985, Aissen 2003) proposes that languages that do not formally distinguish between subject and direct object tend to develop extra markers for direct objects if they are too similar to typical subjects. These approaches focus on the properties of direct objects compared to those of the subjects.

The Transitivity Thesis, or indexing / coding approach (Hopper and Thompson 1980, Naess 2004, 2007) assumes that a direct object is overtly marked if it is a “good” argument in a transitive sentence which represents a “salient event”. The indexing approach comes in different versions. Naess (2004) assumes that affectedness is the relevant notion for making a referent more susceptible for marking, while Dalrymple & Nikolaeva (2011) take information structure as the underlying factor and assume that topicality is the relevant parameter that controls DOM. De Hoop and Narasimhan (2005) and de Hoop & Malchukov (2007) make another distinction that is orthogonal to the two accounts introduced above: They distinguish between split vs. fluid domains of DOM. Split domains are those grammatical contexts in which DOM is obligatory, while fluid domains are those which allow for optionality. They claim that DOM only contributes a function in fluid domains, while it is a kind of agreement marker in split domains. Dalrymple & Nikolaeva (2011) also connect questions of grammaticalization to this contrast between obligatory marking in split domains vs. the variability in fluid domains. Returning now to the relevant data to be investigated, it is generally assumed (Klein & de Swart 2010) that human definite direct object constitute a split domain for DOM. However, as we have seen above, unmodified definite direct objects display optionality. Thus, we have a fluid domain inside a split domain – something we did not expect and which is a challenge for the theories to account for.

This paper addresses two related issues: the alternation between the definite article and pe-marking (with the accompanying clitic), and the fluid behaviour of DOM inside a split domain. The paper is structured as follow: In section 2, we provide a brief overview on the conditions for DOM in Romanian and report from some recent studies on the discourse function of DOM for indefinite direct objects. In section 3, we focus on the conditions for DOM with definite unmodified direct objects, which can be pe-marked and doubled by a clitic, or be headed by the simple definite article alone. We bring evidence that the referential properties of the definite noun phrases cannot justify this alternation. In section 4, we present the findings of the sentence-continuation experiment we conducted to investigate the discourse prominence of the two types of definite noun phrases. Section 5 discusses the
results of the study in light of their contribution for a better understanding of DOM in general. We conclude that this phenomenon is best understood in terms of an ongoing grammaticalization process. Section 6 summarizes the findings made in this paper and points out interesting questions for further research.

2. *Pe*-marking in Romanian

Romanian (see Niculescu 1965, 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, von Heusinger & Onea 2008, Kamp & Bende-Farkas (submitted), among others).

2.1 Strong vs. soft constraints

As exposed in section 1, de Hoop & Malchukov (2007) make another distinction that is orthogonal to the two main accounts for DOM introduced there: They distinguish between split vs. fluid domains of DOM. *Strong or hard constraints (instances of split domains)* obligatorily trigger (or block) DOM. E.g. a human proper name is always *pe-*marked, i.e. the semantics of names obligatorily requires DOM in Romanian. It seems that *pe*-marking does not contribute an additional feature to proper names. *Pe*-marking has grammaticalized to a formal element, like an agreement marker. *Soft constraints (instances of fluid domains)* do not obligatorily trigger (or block) DOM, but (i) show a high significance for DOM, or (ii) are associated with DOM marking. It is claimed that specificity is a soft constraint for *pe*-marking: A specific human direct object is often *pe*-marked (but not always). Differently worded, *pe*-marking signals that the noun is specific.

Figure (1): Split-fluid tree for DOM in Romanian (Klein & de Swart 2010:10)

2.2 Animacy

In synchronic Romanian, *pe*-marking typically targets those direct objects which denote human entities. This condition is responsible for the acceptability of (4a) and the unacceptability of the sentence (4b).
Pe-marked definite NPs in Romanian and discourse prominence

(4) (a) Am văzut-o pe femeia frumoasă.
   AUX saw CL PE woman.DEF beautiful
   ‘I saw the beautiful woman.’
(b) Am văzut-o * pe cartea frumoasă.
   AUX saw CL PE book.DEF beautiful
   ‘I saw the beautiful book.’

2.3 Pe-marking and the referentiality scale
Besides animacy, DOM in Romanian is sensitive to the type of referring expression of the direct object, according to the Referentiality Scale (Aissen 2003):

Table (1): Pe-marking of postverbal human direct objects in Romanian depending on the Referentiality Scale

<table>
<thead>
<tr>
<th>pers. pron.</th>
<th>&gt; PN</th>
<th>&gt; def. NP</th>
<th>&gt; spec. indef NP</th>
<th>&gt; non-spec. indef NP</th>
<th>&gt; non-arg NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>+/-</td>
<td>+/-</td>
<td>+/-</td>
<td>-</td>
</tr>
</tbody>
</table>

Full personal pronouns referring to animate entities are always marked with pe and doubled by a clitic in synchronic Romanian data. 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.

(5) (a) Doctorul îl examinează pe băiatul bolnav.
   doctor.DEF CL examines PE boy.DEF sick
   ‘The doctor examines the sick boy.’
(b) #Doctorul examinează băiatul bolnav.
   doctor.DEF examines boy.DEF sick
   ‘The doctor examines the sick boy.’

In a previous study (von Heusinger & Chiriacescu 2009), we found the confirmation for this generalization, analysing 650 examples found on Google and in a corpus containing Romanian newspaper articles. As already mentioned in the introductory part of this article, the picture is less homogenous in the domain of definite unmodified definite NPs. Leaving aside syntactic and semantic restrictions which require the pe-marked or the unmarked form, definite unmodified NPs show the optionality encountered in (6) below, which pertains to an independent grammatical rule of Romanian (see section 3).

(6) (a) *Doctorul îl examinează pe băiatul.
   doctor.DEF CL examines PE boy.DEF
(b) Doctorul examinează băiatul.
   doctor.DEF examines boy.DEF
   ‘The doctor examines the boy.’

---

1 Please note that for human, definite modified noun phrases, pe-marking is obligatory (ex. (4)). For human indefinite noun phrases, pe-marking is generally obligatory in cases in which the sentence contains operators. In contexts that lack operators at sentence level, the marker is optional (ex. (7)-(9)).

2 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 facultative, depending on semantic features of the head noun and further syntactic constraints. In this paper we will address the phenomenon of DOM in Romanian as a whole, thus, in the following sections, we will not make an explicit distinction between clitic doubling and pe-marking (but see Dobrovie-Sorin 1994, Gramatica Limbii Române 2005 on clitic doubling).
The literature (Farkas 1978, Dobrovie-Sorin 1994, von Heusinger & Onea 2008) assumes that specificity is the main triggering parameter for DOM with indefinite human direct objects in Romanian. Scopal specificity with extensional operators, as in (7), and referential specificity with intensional operators, as in (8), trigger pe-marking.

(7) Scopal specificity:
(a) Toţi bărbaţii iubesc o femeie. (specific / non-specific)
all men love a woman
‘All men love a woman.’
(b) Toţi bărbaţii o iubesc pe o femeie. (only specific)
all men CL love PE a woman
‘All men love a woman.’

(8) Intensional operators:
(a) Ion caută o secretară. (specific / non-specific)
John looks for a secretary
‘John looks for a secretary.’
(b) Ion o caută pe o secretară. (only specific)
John CL looks for PE a secretary
‘John looks for a secretary.’

In plain and transparent contexts, indefinite NPs have been accounted for in terms of epistemic specificity. Sentence (9a) is analysed 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 (9b) is ambiguous between a specific and a non-specific reading. For an extensive analysis of pe-marked indefinites, see Chiriacescu & von Heusinger (2010).

(9) Epistemic specificity:
(a) Petru l-a văzut pe un băiat. (specific)
Peter CL have seen PE a boy
‘Peter saw a boy.’
(b) Petru a văzut un băiat. (specific/ non-specific)
Peter has seen a boy
‘Peter saw a boy.’

2.4 Discourse prominence and topic shift potential
Dalrymple & Nikolaeva (2011) propose a crosslinguistic analysis of optional case marking in some nominative-accusative languages, elaborating upon former accounts of DOM as topicality. They analyse topicality in the “aboutness” sense, which deals with “the construal of the referent as pragmatically salient so that the assertion is made about this referent”. An important distinction Dalrymple & Nikolaeva (2011) make is between primary topics, which are highly prominent and typically realized as subjects (see Givón 1983, Lambrecht 1994, among others), and between secondary topics, which are less prominent (see Givón 1983). In a preceding article (Chiriacescu & von Heusinger 2010), we pursued a similar line of argumentation, showing that discourse prominence is the factor that plays a major role for pe-marking with indefinite NPs in Romanian. In contrast to other studies, which offered a more or less intuitive definition of the notion of “topic”, we proposed two measurable textual
characteristics to determine the discourse prominence of a particular referent, namely referential persistence (in the sense of Givón 1983, Arnold 1998, Ariel 2001) and topic shift potential (Givón 1983). We showed that pe-marked direct objects are (i) referentially more persistent in the discourse than their unmarked counterparts, and (ii) better candidates to change the current topic (i.e. the grammatical subject) of the upcoming discourse. Based on the empirical findings from a web-based story continuation experiment, we concluded that pe-marking is used as a signal by the speaker to instruct the hearer that more information about the referent will follow.

2.5 Architecture of DOM in Romanian

We can summarize the observations so far as follows: Human direct objects are case marked if they are definite expressions, such as personal pronouns, proper names or definite noun phrases. Specific indefinite direct objects are only case marked if they are discourse prominent, and thus pre-topics, as illustrated in Table (2).

Table (2): Pe-marking of postverbal human direct objects in Romanian depending on the Referentiality Scale and the discourse prominence

<table>
<thead>
<tr>
<th>human</th>
<th>pers. pron.</th>
<th>&gt; PN</th>
<th>&gt; def. NP</th>
<th>&gt; spec. indef NP</th>
<th>&gt; non-spec. indef NP</th>
<th>&gt; non-arg NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-topic</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>?</td>
<td>-</td>
</tr>
<tr>
<td>No pre-topic</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

In the next section, we will discuss the distribution of pe-marking with definite unmodified noun phrases in more detail and show that the referential meanings of the definite noun phrases cannot account for the alternation between the pe-marked form and the form preceded by the definite article alone.

3. Definite unmodified NPs

The blocking effect which constitutes the focus of this presentation not only applies to the DOM-marker, but to almost all nouns which are preceded by most prepositions in Romanian. This rule is responsible for the ungrammaticality of the enclitic definite article on the unmodified noun phrase in the presence of other prepositions. While most prepositions always block the attachment of the enclitic definite article on unmodified nouns and do not allow for an alternative construction without, note the form la doctor ‘to doctor’ in (10a), in the case of pe as a case marker, the above mentioned constraint also holds but gives rise to an alternation (see discussion in 3.1). Speakers of Romanian can either drop the marker pe and keep the definite enclitic article -ul (DEF.masc), as in (10b), or drop the definite article, as in (10c) and keep pe.

(10) (a) Un băiat merge la doctor.
    a boy goes to doctor
    ‘A boy goes to the doctor.’

(b) Doctorul examinează băiatul.
    doctor.DEF examines boy.DEF
    ‘The doctor examines the boy.’

3 Gramatica Limbii Române (2005) lists several prepositions that block the apparition of the definite article on the N.
(c) Doctorul îl examinează pe băiat.

doctor.DEF CL examines PE boy

‘The doctor examines the boy.’

Both sentences (10b) and (10c) represent different possibilities of expressing very similar referential categories, as they are both grammatical and have the same propositional content. However, the exact distributional contexts in which each direct object form is used have not been delimited yet.

Before we further investigate the alternation between the pe-form and the article-form of differentially marked definite direct objects, we will discuss some additional blocking effects that trigger the one form or the other. In what follows, we offer two examples for such blocking effects: (i) the lexical semantics of the NP, and (ii) a particular construction (the possessive dative). There are probably many other processes that block the one or the other form, which we will not discuss here (see for metonymical shifts as a major blocker of the pe-marked construction Chiriacescu 2007, von Heusinger & Onea 2008).

3.1 Hard constraints: lexical and syntactic restrictions

Archaic usages of certain terms in direct object position found in written texts at the beginning of the 20th century can be simultaneously suffixed by the definite article and pe-marked, even in the absence of further modifiers (Chiriacescu 2007, Chiriacescu & von Heusinger 2009). However, a direct object like in (11), in which the functional expression șeful (‘the boss’) is suffixed by the definite article and simultaneously pe-marked, is not a recommended one in synchronic Romanian.

(11) L-am văzut pe șeful.

CL have see PE boss.DEF

‘I have seen the boss.’

Another marginal exception is found in the case of expressions for kinship relations (e.g. the father / the aunt). The referents of these NPs are usually uniquely identifiable entities in the context of the utterance, thus signalizing a high degree of individualization:

(12) (a) Îl văd pe tată.

CL see PE father.DEF

‘I see the father.’

(b) Merg la mama.

go to mother.DEF

‘I go to mother.’

Again, this exception is not only found in combination with differentially marked direct objects, but also in combination with other prepositions, as it becomes obvious in (12b).

At the sentence level, pe-marking is ruled out whenever the definite article is modified by a possessive preverbal (13a) or postverbal dative (13b), even in cases where the NP is further modified by an adjective:

(13) (a) Maria își înțelege (*pe) buna prietenă.

Maria DAT understands PE good.DEF friend

‘Maria understands her good friend.’
(b) Întelegându-şi (* pe) frumoasa soţie a făcut [...].
understanding-DAT PE beautiful.DEF wife has made
‘Understanding his beautiful wife, he made [...].’

The noun involved in such a possessive relation is strongly individuated and combines with the definite article. These are the constructions representing the unmarked modality to convey possession. Nevertheless, besides the examples in (13), there coexist other constructions to express possession in which the noun is *pe*-marked and appears with a possessive pronoun in the genitive:

(14) Maria o înţelege pe prietena ei [dar nu pe a mea]
Maria CL understands PE friend.DEF her [but not mine]
‘Maria understands her friend [but not mine].’

Only sentence (14) emphasizes the fact that the direct object prietena (‘the friend’) is Maria’s friend and not mine/ yours/ etc. So, the DOM marker adds a discursive contrast to the object it precedes in these contexts.

3.2 Soft constraints: referential properties

The alternation between a *pe*-marked direct object and one in which the definite article is suffixed on it may depend on the referential properties of the definite description. Following Hawkins’ (1978) classification, we can distinguish along four basic uses of definites: (i) anaphoric, (ii) immediate situational uses, (iii) larger situational uses and (iv) associative uses of definites. With Hawkins’ classification in place, we investigated whether these different uses of definite descriptions play a role in DOM marking in Romanian, or not.

3.2.1 Anaphoric definite NPs

If a definite is called anaphoric, its meaning has to be dependent on the interpretation of a previously occurring (and typically indefinite) noun phrase. We looked at examples in which a referent was introduced with an indefinite nominal in the discourse and then picked up by a definite one. In a first survey, we found out that the *pe*-marked and the unmarked DO form overlap and we did not find any significant difference between the alternate forms. This is illustrated in the examples in (15) through (18).

(15) Ion a cunoscut un politician şi un scriitor.
John has met a politician and a writer
‘John met a politician and a writer.’
(a) A doua zi a văzut politicianul la televizor.
the next day has seen politician.DEF on TV
(b) A doua zi l- a văzut pe politician la televizor.
the next day CL has seen PE politician on TV
‘He saw the politician on TV next day.’

(16) Mihai a invitat un prieten şi un cunoscut la masă.
Mihai has invited a friend and an acquaintance at dinner.
‘Mihai has invited a friend and an acquaintance for dinner.’
(a) Ne-a arătat prietenul într-o poză înainte de a veni.
us has shown friend.DEF in a picture before to come
(b) Ni l- a arătat pe prieten într-o poză înainte de a veni.
us CL has shown PE friend in a picture before to come
‘He showed us a photo of the friend before he came.’
(17) Am citit o carte despre un tenor sicilian și o soprană rusă.
have read a book about a tenor Sicilian and a soprano Russian
‘I read a book about a Sicilian tenor and a Russian soprano.’

(a) Recent am văzut tenorul în piată.
recently have seen tenor.DEF in market
(b) Recent l-am văzut pe tenor în piată.
recently CL have seen PE tenor in market
‘I recently saw the tenor at the market.’

(18) Ieri am făcut cunoștință cu unul dintre consilierii locali.
yesterday have made acquaintance with one of the town counsellors.
‘Yesterday I met one of the town counsellors.’

(a) Seara am auzit consilierul vorbind despre criza economică.
evening.DEF have heard counsellor.DEF talking about crisis.DEF economic
(b) Spre seară l-am auzit pe consilier vorbind despre criza economică.
around evening CL have heard PE counsellor talking about crisis.DEF economic
‘Yesterday evening I heard the counsellor talking about the economic crisis.’

Since both forms are felicitous in (15)-(18), the examples provide evidence that the pe-marked and the non-pe-marked form are not in complementary distribution.

3.2.2 Immediate situation use

In connection to the anaphoric uses, the additional demonstrative use will be briefly discussed:

(19) Context: Mary and Peter are watching a show about the mayor of Brașov.
Peter says to Mary:

(a) Un prieten de-al meu cunoaște primarul.
a friend of mine knows mayor.DEF
(b) Un prieten de-al meu îl cunoaște pe primar.
a friend of mine CL knows PE mayor.
‘A friend of mine knows the mayor.’

In (19), the mayor is indirectly present in the utterance context since he appears on TV and it seems that referring to him demonstratively is possible only with pe. For a pe-marked construction to be felicitously used in contexts like (19), the referent should be previously introduced in the discourse or be accompanied by a pointing gesture. In this case, the pe-marked construction has a visibility requirement built into it.

Immediate situation uses in the sense of Hawkins (1978) involve reference to an individual or entity which is present in the utterance context and is unique in that situation in meeting the descriptive content of the definite description. In comparison to demonstrative uses as (19) above, which can be felicitously used only when the referent is visible to both participants in the conversation, in the immediate situation use, the referent of the entity or individual in question does not have to be visible to both participants. However, it is vital that the hearer should be able to see the intended object, as in (20) below:
(20) Context: A policeman is standing guard outside a prison, which is surrounded by a twenty-foot wall. Suddenly, he hears the voice of a colleague policeman from the other side:
   (a) Prinde -l pe fugar.
       catch CL PE runaway
   (b) Prinde fugarul.
       catch runaway.DEF
       ‘Catch the runaway.’

Both sentences (a) and (b) are felicitous in context (20), however, they give rise to different interpretations on the side of the hearer, i.e. if he can see the escaping prisoner or not. In (20a), the hearer is not only informed of the existence of the escaped prisoner, but he is also instructed to locate the referent in the immediate situation of utterance. If the policeman within the prison utteres sentence (20a), then the policeman outside the prison must see the runaway. Otherwise, he would ask a wh-question to find out which prisoner escaped, where he went, etc. On the other hand, if the first policeman utters sentence (20b), the hearer does not have to see the runaway to understand what happened and how he should react. He would have to look where (and who) the prisoner actually is.

3.2.3 Larger and global situation use

Turning to the Romanian examples in light of Hawkins’ classification, both larger and global situation uses render different readings of the definite NPs.

(21) Context: At home, looking out of the window
   A: What’s wrong?
   B: (a) L -am văzut pe postaș.
       CL have seen PE mailman
   (b) Am văzut poștașul.
       have seen mailman.DEF
       ‘I have seen the mailman.’

The mailman in (21a) and (21b) is a typical pe-referent: it refers to a particular functional role that is no more required to be introduced in domestic contexts, exactly like unique entities as the moon. If speaker B uses the unmarked direct object form postasul (‘the mailman’), then it is the function which the referent fulfils that is important, and not its identity (e.g. in a context in which speaker B is waiting for a letter from the dean). Another example, which makes this distinction clearer, is presented in (22) below:

(22) (a) La inaugurarea aeroportului Brașov, cetățenii vor invita primarul.
       at inauguration.DEF airport.DEF Brașov citizen.DEF will invite mayor.DEF
(b) La inaugurarea aeroportului Brașov, cetățenii îl vor invita pe primar.
       at inauguration.DEF airport.DEF Brașov citizen.DEF CL will invite PE mayor
       ‘At the inauguration of the Brasov airport, the citizens will invite the mayor.’

Again, both (22a) and (22b) are felicitous in this context, but with different connotations. (22a) refers to whoever person might occupy the mayor position at the time the airport will be opened (a particular function). In (22b), pe is not tight to the function its referent designates, but to the individual that occupies this position, say Mr. Jones.
3.2.4 Associative use or bridging

The last major class of uses of definite descriptions is that of associative use, also known as *bridging*. This use gives rise to subtle differences between the *pe*-marked and the unmarked construction:

(23) Part-whole bridging:

Biserica din centrul orașului m-a impresionat mult.
‘The church from the centre of the town impressed me much.’

(a) L-am auzit *pe preot* vorbind așa de frumos despre [...].
   have heard PE priest talking so to nicely about
(b) Am auzit *preotul* vorbind așa de frumos despre [...].
   have heard priest.DEF talking so to nicely about
   ‘I have heard the priest talking so nicely about [...].’

In the situation above, the *pe*-marked construction is the preferred continuation alternative, since the (23b) answer would allude to another priest and not necessarily to the one encountered in the church mentioned.

3.2.5 Transparent vs. opaque uses

Another dimension found in relation to definites is the distinction between transparent and opaque contexts (cf. Keenan & Ebert 1973), as in the sentence: *We will interview the winner*, where *the winner* can be understood as: a.) the actual winner or b.) the one who will win. In such a context, we would expect that speakers will use the unmarked construction for the b.) reading and the *pe*-marked construction if they actually know who the winner is. An additional dimension is the referential vs. attributive reading (Donnellan 1966):

(24) (a) Trebuie să prindem ucigașul lui Paul.
   must to catch murderer.DEF of Paul
   ‘We have to catch the murderer of Paul.’
(b) Trebuie să-l prindem pe ucigașul lui Paul.
   must to -CL catch PE murderer.DEF of Paul
   ‘We have to catch the murderer of Paul.’

Imagine a first context in which the police are at the crime scene wondering who might have murdered Paul. On this attributive reading, both, (24a) and (24b) could be felicitously used. Now imagine a second context in which the police are at the crime scene. After taking into account all evidence and proofs found there, they conclude that the serial killer John Smith must have murdered Paul. In this referential context, only the *pe*-marked construction in (24b) can be felicitously used, while the unmarked construction in (24a) is misleading.

3.2.6 Functional readings and scope

Another difference between the *pe*-marked and the unmarked form is the different scopal behaviour under distributive operators, such as *all*:

(25) (a) Toți chiriașii salută proprietarul bogat.
    all renters salute owner.DEF rich
    ‘All renters salute the rich owner.’
(b) Toți chiriașii îl salută pe proprietar / pe proprietarul bogat
    all renters CL salute PE owner / PE owner.DEF rich
    ‘All renters salute the owner / the rich owner.’
In (25a), the phrase *proprietarul* (‘the owner’) could be interpreted as “Each renter salutes his/her owner”, even if the NP is further modified by the adjective *bogat* (‘rich’). However, the *pe*-marked DO in (25b) clearly underlines the fact that the mentioned owner is the same for each renter. In this case, we consider that this might be a secondary effect of an underlying feature. We assume that this feature relies upon the discourse prominence of the definite noun phrase.

### 3.2.7 Kind referring uses

Other uses of definites are kind referring terms. When referring to a kind, the direct object referent has to be marked by *pe* as in (26b):

(26) (a) In Mongolia am fotografiat nomadul.
    in Mongolia have photographed nomad.DEF
(b) In Mongolia l-am fotografiat pe nomad.
    in Mongolia CL have photographed PE nomad
    ‘I have photographed the nomad in Mongolia.’

We consider that this example strongly supports the hypothesis that a kind referring NP refers to an entity, and that this entity must be well-established in the shared knowledge of speaker and hearer to be felicitously used.

### 3.2.8 Intermediate summary

In the last subsections, we have tested different contrasts with respect to the referential properties of definite noun phrases, which are summarized in table 3.

Table (3): Referential properties of definite noun phrases and use of definite article vs. *pe*-marking for human direct objects in Romanian.

<table>
<thead>
<tr>
<th>Marker/Use</th>
<th>anaphoric</th>
<th>immediate situational</th>
<th>larger situational</th>
<th>associative (bridging)</th>
<th>referential contexts</th>
<th>functional readings</th>
<th>generic readings</th>
</tr>
</thead>
<tbody>
<tr>
<td>definite article</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>attributive</td>
<td>functional</td>
<td>-</td>
</tr>
<tr>
<td><em>pe</em>-marking</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>attributive/referential</td>
<td>absolute</td>
<td>+</td>
</tr>
</tbody>
</table>

The overview presented above represents contexts in which the two types of definite noun phrases sometimes trigger different readings. However, for the time being, we do not test these different readings in much depth and conclude that the two forms cannot be derived from the referential properties of the definite noun phrase. The alternation between the *pe*-marked and the unmarked form seems to be due to the discourse function of the definite noun phrase. Definite noun phrases do have a “cataphoric” force and thus express a certain (additional) level of activation of the referent they are associated with (see von Heusinger 2007). In the following, we test this option in a sentence-continuation task.

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

To investigate whether the presence of the *pe*-marker boosts the prominence or salience of the referents associated with the direct objects realized as definite unmodified noun phrases, we used the metrics for discourse prominence developed for the experiment with indefinite noun

---

4 In this context, the definiteness of the object NP must be formally marked, if the NP is intended to refer to the kind *nomad*. The reason probably is that in postverbal position, bare NPs tend to be interpreted as indefinite, so that there must be a formal indicator if a definite interpretation is intended. We will not go into this case further at this time.
phrases (cf. Chiriacescu & von Heusinger 2010). More precisely, we analysed whether *pe*-marked definite noun phrases are (i) referentially persistent in the subsequent discourse (i.e. whether the referent headed by *pe* is likely to be continued), and (ii) more susceptible to shift the topic (i.e. in the sense of Givón 1983, Ariel 2001, among others) of the current discourse. Let us now consider how we predict participants’ responses to pattern with respect to the two different metrics we tested. First, in light of the findings from the *pe*-marking experiment with indefinite noun phrases (Chiriacescu & von Heusinger 2010) and other experimental investigations, which showed that accessible / salient referents are more likely to be subsequently mentioned (Givón 1983, Gernsbacher & Shroyer 1989, Arnold 1998, among others), we predict that referents headed by *pe* will be referentially more persistent in the ensuing discourse, compared to referents marked with the simple definite article.

Second, given the observation that important or salient referents tend to be mentioned in topic position (which in English generally corresponds to the grammatical subject position, e.g. Ariel 2001, Arnold 1998), we predict that in comparison to their unmarked counterparts, *pe*-marked direct objects will (i) be mentioned more often in topic position in the subsequent text, and will (ii) become the new topic in the following discourse.

4.1 Method

Participants
Twenty-four native speakers of Romanian participated in this experiment. They received no incentive for taking part in the survey. It took about fifteen minutes to complete an experiment.

Materials
The methodology used in this experiment was an open-ended sentence-continuation task. Participants were presented target items consisting of mini-discourses, as in Table (4). Their task was to read the given story fragments and add five logical and natural-sounding sentence continuations for each of them. Although the inclusion of two to three sentences in each test item made it difficult to control every aspect of these discourses, it provided the advantage of creating a more natural discourse (e.g. Gernsbacher & Shroyer 1989). The first two sentences of each test item set the context of the story and contained individual references to two characters. The first character was the clearly established topic of the mini-discourse, as it was mentioned in subject position at least once and was the entity the story was about. In the last sentence of the mini-discourses, the referent was introduced as a definite noun phrase in direct object position.

---

5 Independently of these two textual characteristics, we considered the type of referring expression used to pick up the referent of the direct objects. We will not discuss the findings of this metric in this paper, but note that the likelihood of subsequent mention does not point in the same direction as the likelihood of being realized with a pronoun (see also the discussions in Kehler et al. 2008, Kaiser 2010, Chiriacescu & von Heusinger 2010). For the purposes of this paper, just note that pronominalization does not reflect discourse prominence as defined in this paper.
Table (4): The only difference between the two versions of the test items 1-3 (TI1, TI2 and TI3) in this table is the presence vs. absence of the pe-marker on the direct objects in the left column and the presence of the definite article -ul on the direct objects in the right column.

<table>
<thead>
<tr>
<th>TI1: pe-marked condition</th>
<th>TI1: non-pe-marked condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daniela1 lucrează atât la spital cât și într-o clinică privată. O ambulanță a intrat pe poarta spitalului la ora 23, iar Daniela1 a fost chemată de acasă pentru a-l opera pe pacient2.</td>
<td>Daniela1 lucrează atât la spital cât și într-o clinică privată. O ambulanță a intrat pe poarta spitalului la ora 23, iar Daniela1 a fost chemată de acasă pentru a opera pacientul2.</td>
</tr>
<tr>
<td>‘Daniela1 works at the hospital and in a private clinic. An ambulance entered the hospital’s gate at 11 p.m. and Daniela1 was asked to come and operate on PE patient2.’</td>
<td>‘Daniela1 works at the hospital and in a private clinic. An ambulance entered the hospital’s gate at 11 p.m. and Daniela1 was asked to come and operate on the patient2.’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TI2: pe-marked condition</th>
<th>TI2: non-pe-marked condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>După accident, Cristina1 a stat de vorbă cu un martor2. La scurt timp, la fața locului a sosit un echipaj de poliție3. Acestia3 l-au luat pe martor2 de-o parte și au oprit traficul în zonă pentru scurt timp.</td>
<td>După accident, Cristina a stat de vorbă cu un martor. La scurt timp, la fața locului a sosit un echipaj de poliție. Acestia au luat martorul de-o parte și au oprit traficul în zonă pentru scurt timp.</td>
</tr>
<tr>
<td>‘After the accident, Cristina1 talked to a witness2. A police team3 arrived at the site shortly afterwards. They3 took PE witness2 aside and stopped the traffic in the area for a short while.’</td>
<td>‘After the accident, Cristina1 talked to a witness2. A police team3 arrived at the site shortly afterwards. They3 took the witness2 aside and stopped the traffic in the area for a short while.’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TI3: pe-marked condition</th>
<th>TI3: non-pe-marked condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>La petrecerea de aseară, Andrei1 a cunoscut un politician2 și un cântăreț3 de renume3. Astăzi1-a întâlnit pe politician2 în piață.</td>
<td>La petrecerea de aseară, Andrei1 a cunoscut un politician și un cântăreț de renume. Astăzi a întâlnit politicianul în piață.</td>
</tr>
<tr>
<td>‘At yesterday evening’s party, Andrew1 met a politician2 and a famous singer3. Today he1 met PE politician2 at the market.’</td>
<td>‘At yesterday evening’s party, Andrew1 met a politician2 and a famous singer3. Today he1 met the politician2 at the market.’</td>
</tr>
</tbody>
</table>

We constructed six target stories and four fillers. We manipulated the realisation form of the direct objects, which resulted in two conditions: one in which pe heads the direct object and one in which the direct object remains unmarked by pe.

Procedure and data analysis
The first five main clauses (including subordinate ones, if there were any) of each continuation story provided by the participants were analysed. We coded two aspects of the definite direct objects: (i) their referential persistence and (ii) their topic shift potential. Example (27) represents an example response for test item 3 (TI3) for the pe-condition, and table 5 illustrated the coding methods used.
Sample experimental item and coding methods for the pe-condition for test item 3

Test Item 3: *La petrecerea de aseară, Andrei1 a cunoscut un politician2 și un cântăreț de renume3. Astăzi (pro)1 l-a întâlnit pe politician2 în piață.*

‘At yesterday evening’s party, Andrew1 met a politician2 and a famous singer3. Today he1 met PE politician2 at the market.’

S1: (pro)1 știe că acum e șansa lui1.

‘He1 knew that that’s his1 chance.’

S2: Politicianul2 era un pic grizonat, slăbunch cu accent baritonal.

‘The politician2 had some greyish hair, was thin with baritone voice.’

S3: Andrei1 s-a dus spre el2 și (pro)1 i-a cerut ajutorul să (pro)1 aleagă un pepene bun.

‘Andrei1 went towards him2 and he1 asked (him2) for help to choose a tasty watermelon.’

S4: Politicianul2 s-a intrus și (pro)2 i-ă răspuns cu un aer distrat.

‘The politician2 turned around and (pro)2 responded him1 in a distracted voice.’

S5: Il2 chema don Giuseppe și (pro)2 era inginer zootehnist de meserie.

‘His2 name was don Giuseppe and he2 was a zootechnician engineer.’

Table (5): Coding methods for the continuation sentences in (27) provided for TI3

<table>
<thead>
<tr>
<th>Coding methods</th>
<th>First referent (Andrei)</th>
<th>Target referent (the politician)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anaphoric forms</td>
<td>refer per item / S</td>
<td>sum</td>
</tr>
<tr>
<td>S1</td>
<td>[pro1] (pron1)</td>
<td>2</td>
</tr>
<tr>
<td>S2</td>
<td>[def NP2]</td>
<td>0</td>
</tr>
<tr>
<td>S3</td>
<td>[PN1, pron2] [pro1, CL2] (pro1)</td>
<td>3</td>
</tr>
<tr>
<td>S4</td>
<td>[def NP2] [pro2, CL1]</td>
<td>1</td>
</tr>
<tr>
<td>S5</td>
<td>[CL2] [pro2]</td>
<td>0</td>
</tr>
</tbody>
</table>

We used subscript 1 for the first referent, Andrei, subscript 2 for the referent of the target item, un politician (‘a politician’), and subscript 3 for the referent of the singer. The type of referring expression (e.g. pro, clitic, personal pronoun, definite NP and definite modified NP) of the sentences’ referents is listed in Table (5). Round brackets mark subordinate clauses and square brackets main clauses. Referential persistence is measured by referents mentioned per sentence (referent / S) and the sum of all items up to S5 (i.e. a cumulative measure). Comparing the sums indicates at what stage in the discourse we have more anaphoric expressions referring to one referent compared to another. In our example (27), the referent of the target item (i.e. the referent of the politician) exceeds in persistence the referent of the first referent (i.e. Andrei) in the last continuation sentence (S5). Furthermore, we verified in what sentence the target referent (i.e. the politician) becomes the subject of a main clause, and thus the topic constituent. In example (27), this happens in sentence continuation 2 (S2).

The first aspect under investigation was the referential persistence of the story’s referents. We counted (i) how many times each referent was mentioned in the main and subordinate clauses of the continuation sentences (see columns 3 and 5 in Table (5)), and (ii) how referential persistence relates to grammatical role. For this purpose, we calculated the referential persistence of all referents given in the test items. The times a referent was mentioned in a continuation sentence were added up to a sum representing the referential persistence of that referent at a particular stage in the discourse (see columns 4 and 6 in Table...
Additionally, we calculated the mean values for the referential persistence in all test items of the (i) pe-marked referents, (ii) non-pe-marked referents, (iii) subject referents, and of (iv) other referents (Figure 4).

Topic shift was the second aspect tested. For the sake of simplicity, the first instance in which a direct object referent became the grammatical subject in a main clause was treated as an instance of topic shift (in example (27), this happens in S2). We did not take into consideration whether this shift in topic was maintained after this sentence or not.

4.2 Results

24 participants provided continuations for the initial story fragments. The results from the two metrics, referential persistence and topic shift potential reflect the discourse status of the stories’ referents. In the following sections we discuss the findings of the two textual characteristics in detail.

4.2.1 Number of anaphoric references – referential persistence

The first textual characteristic investigated was referential persistence. Figure (2) displays the mean values for referential persistence of all referents of the test items 1-3 (TI1-TI3). For the pe-condition, we notice a strong likelihood of the referent to me mentioned in the following discourse. On the contrary, the direct object referents in the non-pe-marked condition are picked up in the subsequent discourse less often. The predictions concerning this metric are confirmed, as the pe-marked referents were picked up more often in the subsequent discourse than the referents of the unmarked direct objects.

In sum, participants preferred a continuation story that evolved around the referent of the subject, thus taking it up more often, unless the direct object referent was pe-marked. In such a case, the referent of the pe-marked referent becomes a better competitor for the subject referent in terms of referential persistence.

4.2.2 Topic shift potential

The second textual characteristic investigated was the topic shift potential of direct object referents. Recall that each mention of a direct object in grammatical subject position was counted as an instance of topic shift. The counts for the topic shift potential are cumulative.

The findings condensed in Figure (3) reveal several patterns. First, 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. Second, 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 (3). More than 80% of the participants mentioned the *pe*-marked referents in grammatical subject position, whereas the referents of the non-*pe*-marked direct objects were mentioned as subjects in roughly 25% of the cases.

Third, Figure (3) 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. On the contrary, the referent of the *pe*-marked direct object was picked up in the first two continuation sentences, even though the rate was not high.

The findings concerning the topic shift potential of direct objects confirmed the initial predictions, as the referents of the *pe*-marked direct objects displayed a higher expectancy to be mentioned again as topics in a main clause (i.e. in subject position) in comparison to the unmarked ones.

4.2.3 Discussion

The findings with respect to the discourse structuring potential of direct objects realized as definite noun phrases parallel those reported in Chiriacescu & von Heusinger (2010, 2011) about the discourse behaviour of indefinite noun phrases in direct object position. Up to the last continuation sentences (S5), the *pe*-marked referents (i) exceeded their unmarked counterparts in referential persistence (76% vs. 24%), and (ii) became the topic of the discourse more often than the non-*pe*-marked referents (in 80% vs. 15% of the cases).

5. DOM in Romanian

The investigation of the discourse function of *pe* showed that *pe* signals discourse prominence of a direct object nominal, if it is realized as an unmodified human definite NP or human indefinite NP. The particular contextual circumstances for DOM in Romanian allow us to formulate some general conclusions on DOM with respect to (i) its function in general, (ii) the particular parameters that interact for DOM marking, (iii) the contribution of the DOM marker, and (iv) the grammaticalization path of this marker.

5.1 Theories in competition

Two main approaches to Differential Object Marking (DOM) in general are currently under discussion: the Ambiguity Thesis and the Transitivity Thesis. The Ambiguity Thesis (Comrie 1975, Moravcsik 1978, Croft 1988, Bossong 1985, Aissen 2003) proposes that languages that do not formally distinguish between subjects and direct objects tend to develop extra markers to indicate such direct objects that are too similar to typical subjects. The proponents of this approach focus on the properties of the direct object and on the way these properties contrast with those of the subject. This view is also known as the discriminatory / disambiguating / distinguishing account.
The Transitivity Thesis, or indexing / coding approach (Hopper and Thompson 1980, Naess 2004, 2007), in contrast, assumes that a direct object is overtly marked if it represents a "good" argument in a transitive sentence that expresses a "salient event". The characteristics that make a direct object a better candidate for DOM marking are different in every version of the indexing approach. Hopper & Thompson’s notion of “salience”, for example, is instantiated by different other notions that make reference to additional restrictions. Naess (2004) assumes that “affectedness” is the relevant notion, while Dalrymple & Nikolaeva (2011) take information structure as the underlying structure and assume that “topicality” is the relevant parameter that controls DOM.

The two approaches also differ in the assumptions about the semantic and pragmatic properties of a typical subject argument and a typical direct object argument. The Ambiguity Accounts assume that a typical subject is highly individualized, human, definite and topical, while a typical direct object is not very highly individualized, inanimate and indefinite. Thus, if a typical object shows properties of a typical subject, it is likely to be marked in order to avoid ambiguity. The indexing account assumes that a typical direct object is discourse prominent, human and definite, and a DOM marker signals, according to this approach, that the direct object is such a typical argument.

Our findings suggest that the discourse function of *pe* in Romanian is best accounted for in a theory of indexing (Hopper and Thompson 1980, Naess 2004, 2007): *Pe* signals a high discourse prominence of the direct object and it signals that the argument is a typical direct object.

5.2 Parameters and Architecture

DOM in Romanian follows a complex pattern of syntactic, semantic, pragmatic, and discourse parameters. DOM in Romanian is obligatorily determined by (i) DP-type, (ii) animacy, and (iii) definiteness. Furthermore, *pe*-marking is optional for indefinites and signals specificity and/or discourse prominence. De Hoop and Narasimhan (2005) and de Hoop & Malchukov (2007) make another distinction that is orthogonal to the two accounts introduced above: They distinguish between split vs. fluid domains of DOM. Split domains are those grammatical contexts in which DOM is obligatory, while fluid domains allow for optionality. They claim that DOM can only contribute a function in fluid domains, while it is similar to an agreement marker in split domains. Dalrymple & Nikolaeva (2011) connect questions of grammaticalization to this contrast between obligatory marking in split domains vs. the variability in fluid domains.

It has been assumed (Klein & de Swart 2010) that human definite direct objects constitute a split domain for DOM in Romanian, i.e. that the *pe*-marker occurs obligatorily with definite noun phrases. However, we pointed out throughout this paper that the distribution of *pe*-marking with definite modified direct objects differs from its distribution with definite unmodified direct objects. While the former get *pe*-marking, the latter are optionally *pe*-marked. Thus, in the case of definite unmodified direct objects, Romanian DOM shows a fluid domain inside a split domain. In light of the existing theories, this observation is unexpected and challenging to account for.
Figure (4): Split-fluid tree for DOM in Romanian (Klein & de Swart 2010:10)

\[
\text{split I: DP-type} \\
\text{[-pro]} \quad \text{[+pro]} \\
\text{split II: animacy} \\
\text{[-anim]} \quad \text{[+anim]} \\
\text{split III: definiteness} \\
\text{[-name]} \quad \text{[+name]} \\
\text{fluid: specificity} \\
\text{[± spec]} \quad \text{[+spec]}
\]

Figure (5): Split-fluid tree for DOM in Romanian – modified version

\[
\text{split I: DP-type} \\
\text{[-pro]} \quad \text{[+pro]} \\
\text{split II: animacy} \\
\text{[-anim]} \quad \text{[+anim]} \\
\text{split III: definiteness} \\
\text{[-def]} \quad \text{[+def]} \\
\text{fluid: topicality} \\
\text{[-topical]} \quad \text{[+topical]} \\
\text{+ syntactic constraint} \\
\text{fluid: topicality} \\
\text{[-topical]} \quad \text{[+topical]} \\
\emptyset \quad \text{pe} \quad \emptyset \quad \text{pe}
\]

Figure (5) shows that Romanian DOM-marking does not spread along a continuous path. In other words, even though pe-marking is found with indefinite noun phrases, not all definite noun phrases allow for the presence of the marker. One possible explanation for this “irregular” or idiosyncratic behaviour of DOM with definite noun phrases could be derived from different syntactic restrictions found with prepositions in Romanian. However, this explanation cannot account for the contribution of pe in Romanian (and of DOM in general).

5.3 Contribution of pe in Romanian

In light of the findings of the experiment presented in section 4, we argued that pe signals the discourse structuring potential of the referents it precedes. Such referents were shown to be more recurrent in the following discourse and to be more prone to shift the topic of the discourse.
In light of the findings presented in section 4, we need a discussion about the status of pre-topics and the criteria that distinguish them from non-topics. There are in general two views with respect to the function of DOM-markers in the languages of the world: A fixed contribution vs. a contrastive contribution. In the fixed contribution accounts it is assumed that a DOM marker has always the same function, even though in certain context this function cannot be detected any more. In the contrastive-function approaches, it is assumed that the marker just signals some contrast between the forms. The relevant contrast is then decided depending on the type of referring expression. For example, for definite noun phrases, the contrast is often topicality, while for indefinites, specificity plays an important role. In general, the functions of DOM-markers are visible with respect to one of the referential categories (i.e. the fluid domain(s) represented in a particular language), and we usually lack enough data with respect to the diachronic development of DOM. In other words, the factors that might have initially triggered DOM are often unknown. Thus, it is difficult to distinguish between the two functions of DOM.

5.4 Grammaticalization path

Figure (6) summarizes von Heusinger & Kaiser’s (2005:45) findings with respect to the grammaticalization path of the DOM-marker in Spanish. In their view, α-marking in Spanish has a contrasting function.

To summarize the findings for the grammaticalization path presented in Figure (6), we can say that DOM in Spanish has extended from marking animate pronouns and proper names to marking animate and specific NPs. It seems that at one intermediate step there was a clear distinction between topicalized definite and non-topicalized definites. Once the evolution has affected the whole definite cell, topicality was neutralized as a DOM-trigging feature. The next step of the evolution affects the indefinite cell. Here, the additional feature ±specific allows for a smoother grammaticalization of DOM. This observation has led von Heusinger & Kaiser (2005) to conclude that the evolution of DOM is facilitated by intervening or “transitional” categories such as topicality and specificity. These categories are active only for the category to which DOM is developing: topicality for definite NPs, specificity for indefinite NPs. We can only speculate why we find such pairs: Topicality expresses a prominent contrast that (most often) affects definite NPs, while specificity expresses a contrast that (most) often affects indefinite NPs.
The question for pe in Romanian is whether we can assume a function as topic marker, which is the pragmatic function and pressure of the marker. The function however is neutralized once the marker is grammaticalized and has only the function of agreement (pronouns, proper names, definite modified NPs). But the exceptions seem to show that the underlying function is still there and can be used by speakers to make certain differences.

6. Conclusion

Differential Object Marking (or pe-marking) in Romanian is optional with definite and indefinite noun phrases. Indefinite noun phrases that are preceded by the pe-marker occupy a higher place on the referentiality scale, as they show a strong preference for referential readings (e.g. specific and wide scope readings). In a recent study (Chiriacescu & von Heusinger 2010), we accounted for the seemingly optionality with indefinite noun phrases in terms of discourse prominence. In the realm of definite NPs, Romanian has an independent syntactic rule that blocks the use of pe together with the definite article if the noun phrase is not further modified. This rule gives rise to two alternative uses: pe-marking and the omission of the definite article or the use of the definite article and the omission of pe. To investigate the factors that trigger this alternation, we conducted a story-continuation experiment, the findings of which revealed two patterns: (i) pe-marked definite direct objects are referentially more persistent in the following discourse, and (ii) pe-marked direct objects show a preference to become topics two or three sentences after being introduced in the discourse. These results parallel the assumptions made in Dalrymple & Nikoleava (2011), namely that DOM can be explained in terms of information structure. We consider that pe-marked direct objects in Romanian signal discourse prominence, in terms of high referential persistence and topic shift potential, a function that is visible only with such types of referring expressions that still allow for variation, namely indefinite and definite noun phrases. This additional aspect of DOM, namely its discourse structuring potential, might represent the missing link to understand the reasons for the beginning of the grammaticalization of different markers of DOM and their initial function.

7. References