DIFFERENTIAL OBJECT MARKING OF HUMAN DEFINITE DIRECT OBJECTS IN ROMANIAN[1]

EDGAR ONEA[2], DANIEL HOLE[3]

Abstract. In Romanian, direct objects realized as full definite descriptions denoting human entities in a post-verbal position are often marked with the differential object marker *pe*. The literature mentions subtle semantic and pragmatic effects associated with *pe*-marking, but largely ignores two hard syntactic constraints: *pe*-marking is a) blocked by possessor datives external to the object DP, and b) incompatible with overtly definiteness-marked unmodified nominal direct objects. This paper contributes to a theory of differential object marking in Romanian starting out from these 'blocking constraints'. Firstly, we show based on a corpus study that these two constraints isolate the quantitatively most significant domains of variation for the distribution of *pe*-marking. In fact, there is virtually no variation in *pe*-marking of definite human direct objects outside these domains. Secondly, we outline a theoretical analysis at the syntax-semantics interface centered on the syntactic movement of *pe*-marked direct objects. This analysis covers the variation associated with both blocking effects in a unified way.

Keywords: Romanian, Differential Object Marking, Article Drop, Semantic Binding, Possessor Raising, Weak Definites

1. INTRODUCTION

We speak of *differential object marking* (henceforth DOM, term coined by Bossong 1985, 1991) whenever a language overtly case-marks "some direct objects, but not others, depending on semantic and pragmatic features of the object" (Aissen 2003:435). These features include *animacy*, *definiteness* and *topicality* of the direct object (henceforth DO), and are mostly organized along implicational scales (cf. Comrie 1975, Aissen 2003, Croft 2003). In addition, co-argument asymmetries (Primus 2011), temporal, aspectual and modal verbal categories

[1] We gratefully acknowledge the funding provided by the German Science Foundation (DFG) in the context of the Courant Research Centre "Text Structures" (Göttingen) and the project B8 of the SFB 732 (Stuttgart). Moreover, we would like to thank Alexandru Mardale, Gianina Iordachioaia, as well as two anonymous reviewers of this paper for useful comments. All errors are our own.

[2] University of Graz, edgar.onea@gmail.com.

[3] University of Stuttgart, holedan@gmail.com.

BDD-A26971 © 2017 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:01:25 UTC)

(Malchukov and de Hoop 2011) and some additional factors have been argued to play a role in the distribution of DOM in various languages.

The differential object marker in Romanian is the preposition *pe*.[4] The main factor for the distribution of *pe* is the *animacy scale* (cf. e.g. Aissen 2003:437). With some pronominal exceptions and instances of pre-verbal DOs–which we ignore in this paper–only [+HUMAN] DOs can be *pe*-marked, as shown in (1):

(1) a. *L-am* *văzut* *pe* *băiatul* *deştept.* [+HUMAN]
CL.3.SG.ACC-have.1.SG seen PE boy.DEF clever
'I saw the clever boy.'

b. #*L-am* *văzut* *pe* *covorul* *frumos.* [–HUMAN]
CL.3.SG.ACC-have.1.SG seen PE carpet.DEF pretty
intended: 'I saw the pretty carpet.'

For [+HUMAN] DOs, the presence of the marker *pe* depends on further factors: mainly *definiteness* and *specificity* (cf. Farkas 1978, Dobrovie-Sorin 1994, Mardale 2007, Kamp and Bende-Farkas 2006, von Heusinger and Onea 2008, von Heusinger and Chiriacescu 2010 and many others).[5] DOs expressed by proper names and pronouns are always marked with *pe*, definite DPs may or may not be marked depending on further factors, and indefinites may be marked if they are specific.

The distribution of DOM in Romanian, hence, exhibits two domains in which *pe*-marking appears to be optional, viz. with definite and indefinite lexical DPs. This is surprising from a theoretical point of view. To see this, it is useful to consider the distinction between *split* and *fluid* case marking coined by de Hoop and Malchukov (2007). In most DOM languages, DOs are always marked in some realm (e.g. [+DEFINITE +ANIMATE]), while they are never marked in some other realm (e.g. [-DEFINITE, -ANIMATE]). This is called *split* case marking. Violating split case-marking rules leads to ungrammaticality. As opposed to this, there is often an intermediate domain in which object marking appears optional and has an impact on interpretation rather than on grammaticality. This is called *fluid* case marking. Typically, the main categories determining DOM are split categories, while fluid alternations only appear at the fringes of the distribution. In other words: it is not customary to find several semantic contrasts involving several categories associated with the same object marker (cf. also von Heusinger and Kaiser

[4] Case-marked DOs are usually accompanied by clitic doubling. Clitic doubling is widely obligatory with case-marked DOs in modern Romanian (and Romance in general, cf. Kayne 1975, 1991), but the historical distributions of clitic doubling and DOM are different to a certain degree, cf. von Heusinger and Onea (2008), Hill (2013), Mardale (2015), Hill and Mardale (this volume).

[5] Other factors such as individuation and related gender features, topicality and related discourse properties of the DO also play a role, though we will ignore these in this paper. This volume includes an overview on DOM in Romanian to which the reader is referred for further details.

BDD-A26971 © 2017 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:01:25 UTC)

2005 from a different perspective). Crucially, if *pe*-marking is indeed optional for both definites and indefinites and there is some kind of semantic correlate of its presence or absence, *pe* seems to carry precisely this kind of unexpected double duty.[6]

Given that the distribution of DOM with [-DEFINITE] [+HUMAN] DOs has been amply studied in the literature, and given that we know very little about the rules governing the distribution of DOM for [+DEFINITE] [+HUMAN] DOs, we consider the study of the latter an imperative ingredient for a general theory of DOM in Romanian.

In general, it is assumed that *pe*-marking is optional in Romanian for the latter category, though *pe*-marking is considered more common, preferable or some kind of default. The distribution is, however, complicated by two major syntactic constraints on *pe*-marking:

Firstly, definite DPs realized as unmodified nouns and with (nearly any) preposition block the overt definite article (Dobrovie-Sorin and Giurgea 2006, Dobrovie-Sorin 2007, Mardale 2008). Therefore, nominal DOs with the definite article and without modifiers cannot be marked by the preposition *pe*, either. This is shown in (2). This phenomenon, also known as *article drop*, leads to some variation, as speakers may choose to mark the DO with *pe* as in (2b) or to keep the definite article and drop *pe* as in (2c).

(2) a. *Ion s-a aşezat pe/lângă scaun(*-ul).*
John REFL-has sat on/next to chair(-DEF)
'John sat down on/next to the chair.'
b. *Ion l-a văzut pe băiat (*-ul).*
JohnCL.3.SG-has seen PE boy(-DEF)
'John saw the boy.'
c. *Ion a văzut băiatul.*
Johnhas seen boy.DEF
'John saw the boy.'

Secondly, DP-external possessor datives block *pe*-marking of the possessum DO, as shown in (3). (3a) is a regular transitive sentence in which the DO has a DP internal possessor and is marked by *pe*. (3b) is a similar example, except that the possessor of the DO appears as a dative (clitic pronoun) external to the DP, which is known as *possessive dative*, and theoretically analyzed in this paper as *possessor raising* (henceforth PR, Szabolcsi 1984, Landau 1999). Crucially, as shown in (3c), adding *pe* and the respective accusative clitic to the PR construction

[6] Symptomatically, the way in which, for example, Klein and de Swart (2011) depict the distribution of DOM in Romanian involves fluid alternation only for indefinites, where the semantic effect associated with the alternation is specificity-related. Klein and de Swart (2011) explicitly ignore what they acknowledge to be a fluid alternation in the domain of definite DOs in Romanian, suggesting that it is simply unclear what semantic effects might be associated with the presence or absence of DOM.

BDD-A26971 © 2017 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:01:25 UTC)

is not possible. Again, there seems to be a source for variation, as speakers may either use PR and no *pe* as in (3b), or they may go for an internal possessor and *pe*-marking as in (3a).

(3) a. *Ion îl cunoaşte pe coechipierul lui de fotbal.*
John 3.SG.ACC.CL knows PE teammate.DEF his of football
'John$_i$knows his$_i$football teammate.'

b. *Ion îi cunoaşte coechipierul de fotbal.*
John 3.SG.DAT.CL knows teammate.DEF of football.
'John$_i$knows his$_i$football teammate.'

c. **Ion i-l cunoaşte pe coechipierul de fotbal.*
John 3.SG.DAT.CL -3.SG.ACC.CL knows PE teammate.DEF of football.
intended: 'John$_i$knows his$_i$football teammate.'

These two constraints, hence, isolate *domains of variation* in the distribution of *pe*-marking for [+DEFINITE] [+HUMAN] DOs. We will show, based on a corpus study, that in modern Romanian there does not seem to be any quantitatively significant variation in *pe*-marking that is not within the variation domain associated with these two constraints.

While this may seem intuitive at first sight, it is in fact not obvious that such constraints should give rise to variation at all. When we suggested that speakers may choose between two constructions (one marked with *pe* and one unmarked) in examples like (2) and (3), in fact we did presuppose that *pe*-marking is optional for definite DOs in general. But if empirical evidence suggests (as we argue) that there is not much variation outside these domains, this presupposition seems false. This leads to two puzzles. The one associated with PR is why external possessors are incompatible with *pe*-marking. There is no puzzle as to why DOs with an internal possessor are generally marked with *pe.* Similarly, even though there is a justified question about the conditions governing the choice of internal over external possessors in Romanian, this is not limited in any interesting way to [+HUMAN] DOs and, thereby, to *pe*-marking. As opposed to this, when it comes to article drop, we are not interested in why *pe*-marking is incompatible with the definite article of unmodified nouns. This question is orthogonal to our paper because article drop is not limited to DO marking but concerns many other prepositions as well. What we find puzzling is rather why DOs expressed as unmodified nouns may be used without *pe*, which then indeed leads to a choice between two available constructions, as shown e.g. for (2) above.

One of the main theoretical points of this paper is that these two questions are related and in fact have quite precisely the same answer: the incompatibility of a certain type of VP-internal thematic-role-assigning operators with *pe*-marking. This will lead to a general hypothesis about the rules governing the distribution of DOM for [+DEFINITE] [+HUMAN] DOs in Romanian, which we will be able to phrase in terms of a *split alternation*.

BDD-A26971 © 2017 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:01:25 UTC)

The main part of the article starts with the presentation of a corpus study that shows the major effect of the two blocking constraints on the distribution of *pe*-marking in section 2. In section 3, we then propose a theoretical analysis of the PR constraint, which will be further extended to the article drop constraint in section 4. The concluding section 5 takes a broader perspective and includes a novel generalization about DOM in Romanian.

2. CORPUS STUDY

In this section we present the results of a corpus study that investigates the main empirical question of the paper: what is the distribution of DOM for [+DEFINITE] [+HUMAN] DOs? If asked in the context of a corpus study, this question naturally boils down to a purely quantitative distribution and may have limited relevance to linguistic theory, for well-known reasons. However, given that DOM in Romanian appears to undergo a broad process of language change (cf. von Heusinger and Onea 2008) and to exhibit considerable dialectal and inter-speaker variation (partly because of language change), we judge quantitative corpus data to be the most reliable source of information that is currently available.

In section 2.1 we explain the method by which we extracted the relevant corpus data, followed by a brief presentation of the results in section 2.2, and a general discussion of the findings in section 2.3. Overall, this section shows that the main blocking effects discussed above are responsible for the overwhelming majority of non-*pe*-marked DOs. We suggest that remaining residual instances may actually be neglected for the analysis.

2.1. Method

We conducted a small corpus study using the Romanian TenTen web-corpus crawled by SpiderLing in June 2016, including 3.14 billion tokens and tagged for parts of speech. We developed a particular heuristic method to extract data for the conditions under investigation that required relatively limited manual post-processing. We considered three types of noun phrases as DOs: (i) relational nouns, e.g. *fiul* (‘the son’); (ii) semi-relational nouns, e.g. *copilul* (‘the child’), which have balanced relational and non-relational uses; and (iii) nouns that quite commonly have non-relational usages, e.g. *senatorul* (‘the senator’)[7], which we dub “non-relational” for simplicity. For each of these noun classes, we first searched for strings of the type *pe*+*noun*, for all selected elements of the respective

[7] Of course, most [+HUMAN] nouns have some relational usage, e.g. *the senator of a party*, *the senator of a state*, etc. But these usages were not found to be predominant in our data.

BDD-A26971 © 2017 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:01:25 UTC)

class.[8] Based on the results, we established an automatic collocation analysis. This gave us the most frequent verbs and morphological verb forms that are associated with the respective nouns as DOs. This is important, because different verbs are collocated differently with different nouns and different thematic roles, e.g. *help* is strongly collocated with *child* but not with *senator* as a DO. From these hits, we selected some of the most frequent transitive verbs as shown in Table 1. Thereby we manually checked for a low error rate, i.e. a low probability that the noun is a post-verbal subject.

Table 1

The verb-noun combinations in the three relationality categories used for analysis

Category:	**[+RELATIONAL]**	**[±RELATIONAL]**	**[-RELATIONAL]**
Nouns:	*fiică* (‘daughter’), *fiu* (‘son’), *soț* (‘husband’), *soție* (‘wife’), *fratele* (‘brother’), *soră* (‘sister’)	*băiat* (‘boy’), *copil* (‘child’)	*senator* (‘senator), *polițist* (‘policeman’), *portar* (‘keeper’), *politician* (‘politician’)
Transitive verbs:	*ajuta* (‘help’), *cunoscut* (‘know’), *trimis* (‘send’), *luat* (‘take’)	*ajuta* (‘help’), *lasă* (‘leave’), *lovit* (‘hit’)	*întrebat* (‘ask), *surprins* (‘surprised’), *lovit* (‘hit’)

We extracted the relevant data by searching both for the *pe*-marked and for the non-*pe*-marked variants of the respective verb-noun combinations in each of the three categories. Even though our heuristic precautionary measures were useful, the raw data extracted were not error-free. To remedy this, we manually checked a small sample of random hits of non-*pe*-marked tokens to determine the error rate and therefore (linearly) project the number of actual DOs in the sample. At the same time, we also evaluated the number of PR instances in this smaller sample, as well as the number of unmodified nouns, which would stand in competition with *pe+noun* without the definite article due to article drop.

2.2. Results

We show the results of the corpus study in absolute numbers for the three noun types in Table 2. The absolute numbers suggest–quite surprisingly–that *pe*-marking is optional or even dispreferred for [+DEFINITE] [+HUMAN] DOs.

Manual analysis of a random sample of non-*pe*-marked DOs was, as mentioned above, necessary to eliminate erroneous hits in which non-*pe*-marked DPs are not actually DOs. We give one corpus-example each for the noun classes in the [–PE] [+DEF] condition (4a-c), as well as for an error-hit in which the found string is a subject (5).

[8] The choice of the nouns was strongly influenced by an attempt to avoid ambiguities, e.g. the semi-relational noun *fata* (‘the girl’) could not be used because in web-data it can easily be confounded with *fața* (‘face’) or *fată* (‘girl’), the latter having no definite article.

BDD-A26971 © 2017 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:01:25 UTC)

Table 2

Total hits in absolute numbers[9]

	[+PE] [–DEF]	[+PE] [+DEF]	[–PE] [+DEF]
[+ RELATIONAL]	32	739	2013
[± RELATIONAL]	1348	332	5062
[– RELATIONAL]	60	92	29

(4) a. *Senatorul şi- a trimis fiica să ascundă o geantă* […]
senator REFL.DAT.CL has sent daughter.DEF to hide INDEF bag
'The senator sent his daughter to hide a bag […].'

b. *Tata o mai consola zicându -i din când în când:*
"Femeie, lasă băiatul în pace, că o fi ştiind el ce face!"
Woman let boy.DEF in peace because would be knowing he what does
'Father used to comfort her saying from time to time: My wife, leave the boy alone, for he knows what he is doing!'

c. *nu au spus nimic clar din care să reiasă*
cine a lovit poliţistul şi de la ce a pornit scandalul
who has hit policeman.DEF and of at what has started scandal.DEF
'… they said nothing that could have revealed who hit the policeman and what started the scandal'

(5) *Mi- a trimis fiica mea pantofii astia si mai bine ii vindem*
1.SG.DAT.CL has sent daughter.DEF my shoes these and better them sell
'My daughter sent me these shoes and it is better we sell them […].'

The results of the manual analysis are provided in Table 3. The distinction between modified and unmodified nouns was only made for [–PR] hits, yielding a pooled [±MODIFIED] category for the complementary [+PR] cases. The data in Table 2 and Table 3 allow us to linearly project with a reasonable probability the ratio of non-*pe*-marked DOs associated with PR and with article drop, for each of the categories. This is provided in Table 4.

Table 3

Manual analysis of a [-PE] random sample in absolute numbers

	PR	Unmodified Nouns	Modified nouns	Error
Relational nouns	173	2	2	23
Semi-relational nouns	38	59	3	0
Non-relational nouns	1	6	5	17

The relative numbers in Table 4 are calculated with respect to the entire sample, excluding the projected number of erroneous hits (i.e. DPs that are not DOs). For example, the percentage 68.2% in the first cell means that from the entire sam-

[9] [+DEF] abbreviates overt definite articles, [-DEF] abbreviates covert definite articles, (hence, [-DEF] instances are still semantically definite). The combination [-PE] [-DEF] is not grammatical.

BDD-A26971 © 2017 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:01:25 UTC)

ple of (presumably) correct hits for relational nouns, 68.2% are instances which are not *pe*-marked and are at the same time associated with the presence of an external dative possessor. The boldface number 0.7 in the same line represents the proportion of non-*pe*-marked DOs that exhibit no PR and involve a nominal modifier, i.e. data that do not fall under the domain of variation induced by the two blocking effects discussed here.

Table 4

Projected relative numbers based on random sample annotation

	PR	article drop	residual[-PE]	[+PE]	Total
Relational nouns	68.2%	.7%	**.7%**	30.2%	100%
Semi-relational nouns	28.5%	44.2%	**.2%**	24.9%	100%
Non-relational nouns	.6%	3.65%	**3%**	92.6%	100%

In a nutshell, (i) PR with relational (uses of) DOs and unmodified/articleless nouns accounts for the bulk of non-*pe*-marked definites; (ii) non-relational (uses of) DOs strongly correlate with *pe*-marking.

2.3. General discussion

Our corpus study leads to two core observations that will be further investigated in the rest of the paper:

The first observation is that the residual [-PE] data, i.e. those instances of [+DEFINITE][+HUMAN] DOs that are not *pe*-marked and that are not accounted for by to article drop or by PR, are marginal (around 2% of the data on average). Given this small number of residuals, one can consider the residual [-PE] items in the corpus as indicative of language change underway that will eventually lead to obligatory marking. Indeed, many of the [-PE] examples are archaic. Only very few examples, like the one in (6), are fully acceptable and not archaic. Hence, all significant variation in *pe*-marking is located in the domain delimited by the two syntactic blocking effects. As we see it, any theoretical analysis should focus first and foremost on these domains. Put differently, we suggest split *pe*-marking for [+DEFINITE][+HUMAN] (modulo some remnants of language change): whenever there is no syntactic reason to avoid *pe*-marking, *pe*-marking applies either as a grammatical rule or as a default that is in the process of becoming a grammatical rule.

(6) [*Căsătoria a fost una dintre cele mai strălucite afaceri din viața sa*:]
a luat fiica unui alt magnat al finanțelor europene […]
has taken daughter.DEF a.GEN other tycoon of finances European
'[The marriage was one of his best businesses:] he took the daughter of another European finance tycoon […].'

BDD-A26971 © 2017 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:01:25 UTC)

The second observation is that for nouns that tend to have a relational reading, PR is the main predictor of non-*pe*-marking. In fact, typical relational nouns essentially only alternate between the position of the possessor, internal arguments leading to *pe*-marking and external possessor blocking *pe*-marking. This makes PR the primary target of our investigation, thus we turn to it in the next section. As opposed to this, we find a lot of variation in the case of DOs expressed by unmodified nouns, which suggests that the correct theoretical approach might be to search for some correlating semantic feature.

3. POSSESSOR RAISING IN ROMANIAN

In this section, we first provide background on external possessor datives in Romanian. We then briefly motivate our analysis as PR, in fact as true PR, a process that involves syntactic movement of the possessor. Finally, we suggest a way to account for the incompatibility of Romanian PR and DOM in the theory, and we discuss some of the virtues of this analysis.

To recap the key data, in Romanian, possessors of a DO can either be realized inside the DP, i.e. as internal possessors, as in (7), or they can appear in a dative position, i.e. as external possessors, as witnessed in (8) and (9). While in the case of internal possessors *pe*-marking is at least preferred, when the possessor appears in the dative position, *pe*-marking becomes clearly and categorically ungrammatical. Given that in other contexts, a dative and an accusative clitic may co-occur, it is clear that the culprit for the ungrammaticality of (9) is the presence of *pe*. This calls for a theoretical explanation.

(7) *Ion* $^{?}$(*îl*) *cunoaşte* $^{?}$(*pe*) *coechipierul lui de fotbal.*
John 3.SG.ACC.CL knows PE teammate.DEF his of football
'John$_i$knows his$_i$football teammate.'

(8) *Ion îi cunoaşte coechipierul de fotbal.*
John 3.SG.DAT.CL knows teammate.DEF of football.
'John$_i$knows his$_i$football teammate.'

(9) * *Ion i-l cunoaşte pe coechipierul de fotbal.*
John 3.SG.DAT.CL -3.SG.ACC.CL knows PE teammate.DEF of football.
intended: 'John$_i$knows his$_i$football teammate.'

As a starting point for our analysis, we follow Cinque and Krapova (2008), Dumitrescu (1990) and Avram and Coene (2000, 2008) in assuming that Romanian has *true* PR. True PR involves movement of the possessor out of an object DP into a dative position. There are various arguments in the literature showing that true PR in Romanian involves movement. We do not repeat them here, but cf. Cinque and Krapova (2008). Instead, a clear distinction between *true* PR and *false* PR appears more important for our purposes.

BDD-A26971 © 2017 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:01:25 UTC)

False PR has mainly been discussed for French and other Romance languages (Kayne 1977, Guéron 1985, Vergnaud and Zubizarreta 1992), but is also attested in a much wider set of languages, including Germanic and Slavonic (König and Haspelmath 1998, Hole 2006, 2012, 2014) and also in many non-Indoeuropean languages around the world as documented, for instance, in contributions to Aikhenvald et al. (eds.) (2001). Two core properties of 'false' PR constructions are listed in (10), with French examples given in (11).

(10) **Properties of *false* possessor raising**:
a. theta-marking of the dative DP with implications for the main event
b. plural dative DPs co-occur with singular possessum DPs

(11) a. *Le médecin leur a radiographié l'estomac/ *les estomacs.*
the physician DAT.3.PL has x-rayed DEF-stomach/ DEF.PL stomachs
'The doctor x-rayed their stomachs.'
b. *Le médecin a radiographié *leur estomac/ leur estomacs.*
the physician has x-rayed their stomach/ their stomachs
'The doctor x-rayed their stomachs.'

Property (10a) amounts to saying that the dative referents in (11a) are participants of the event of x-raying their stomachs. This is not the case in (11b), which could be used to describe a situation where models of stomachs owned by the possessor referents were x-rayed without the possessors' participation in the event. The difference also surfaces in a scenario where the possessors of the stomachs are no longer alive. In such a scenario (11b) can truthfully and felicitously be uttered. (11a), by contrast, is infelicitous and maybe even false under such circumstances. Property (10b) allows for singular marking on the possessum DP in (11a) even though the dative DP is plural; in fact, according to Vergnaud and Zubizarreta (1992), plural marking on the possessum is ungrammatical in French. By contrast, DP-internal possessors as in (11b) must agree with their head nouns in number. There is a longstanding controversy in the literature as to whether the dative in PR constructions is initially merged inside the possessum DP and moves out secondarily (cf. recently Lee-Schoenefeld 2006 for German), or whether it is merged directly at a higher position on the main projection line (Hole 2012, 2015). We will follow Hole (2012, 2015) here, suggesting that false PR does not involve movement of the possessor from the possessed DP.

Romanian has false PR constructions as well, as is shown in (12).

(12) a. *I-a rupt degetul mic.*
3.SG.DAT.CL-has broken finger.DEF little
'(S)He broke his little finger.'
lit: '(S)He broke him the little finger.'
b. *A rupt degetul lui mic.*
has broken finger.DEF his small
'(S)He broke his little finger.'

As expected, in (12a), the thematic involvement of the possessor of the finger is required, whereas this is optional in (12b), which makes the example somewhat

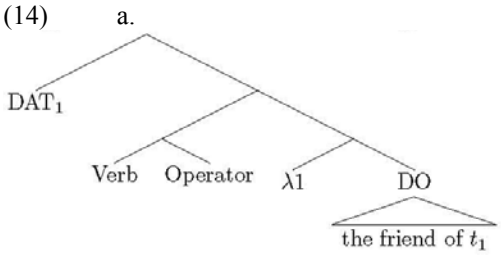
BDD-A26971 © 2017 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:01:25 UTC)

bizarre.[10] Crucially, however, neither of the properties in (10) applies to the example (8) in Romanian above, which is a good indicator that (8) is not an instance of false PR, but rather true PR.

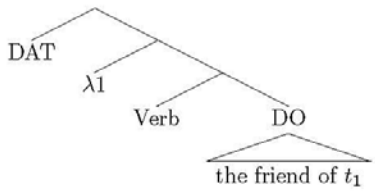
With this background, we now turn to the interaction between PR and DOM in Romanian. We make two additional assumptions: firstly, we (hopefully uncontroversially) assume that *pe*-marked DOs move to a higher position (at LF) which must be higher than the VP.[11] Secondly, we assume that in the case of true PR, the first landing site of the possessor argument is inside the VP. One well-known candidate landing site for the possessor is a low applicative head in the sense of Pylkkänen (2002), which is above the DO first merge position; see also Cornilescu (2015). This leads to an illicit semantic-binding configuration (in the sense of Büring 2005), because the trace t_j in (13) is not c-commanded by the possessor argument.

(13) * [... [pe [$_{DP}$ the N of t_i]]$_i$ [... [Possessor$_j$ [... t_i]] ...]

(14) a.



b.



We will propose a slightly more specific theory here, however. In particular, we assume that there is a semantic operator applied to the verbal head (but equivalently it could also be positioned right below the verbal head) which makes sure that a thematic argument slot is created, which is not only filled by the dative but also binds the possessor argument of the DO as part of its core function.[12] This is

[10] The number neutrality does not apply to (12a) because the possessor is singular. In other examples, a number marking contrast does occur, though we admit that the contrast is not as clear as in French. Since false PR does not interact with DOM (because we cannot generally get false PR with animate DOs), we do not pursue the investigation of the exact properties of false PR in Romanian any further.

[11] Potential reasons for this movement include referential features of the kind discussed in Diesing (1990), classical case-assignment movement, or various conceivable semantic type shifts induced by the *pe*-marker that are only compatible with a higher assignment of thematic role outside the VP. The motivation of movement is orthogonal to our analysis and a natural topic for further research.

[12] We do not discuss the details of this analysis in the paper, partly because for a theory of DOM, any of the two structured in (14) will suffice; we intend to cover the missing details in future research on PR in Romanian. Also note that PR is blocked under this implementation because whenever the DO moves (i.e. if it is *pe*-marked), the operator applies vacuously, leading to ungrammaticality in the spirit of the prohibition against vacuous quantification proposed by Kratzer (1995:131).

BDD-A26971 © 2017 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:01:25 UTC)

structurally illustrated in (14a), which is a more complex version of what would–from a compositional perspective–be minimally needed. The leaner structure is shown in (14b).

There are two main reasons to adopt this (type of) analysis. The first one is that such a structure generalizes to cases of false PR as well, which does not involve movement of the possessor but requires its thematic involvement. With the structure in (14a) (modulo the missing trace and a slightly different operator), both versions of PR can be captured very similarly. This predicts that in various languages we might find interactions between DOM and even false PR. To our knowledge, this has not yet been studied empirically in any systematic way. However, very first preliminary data from Spanish suggest that such interaction may exist in other languages indeed. Rodríguez-Mondoñedo (2007:256ff) argues that in small clauses involving the verb *tener*, DOM with the marker *a* blocks inalienable possession readings with an (implicit) DP-external possessor, as illustrated in (15).

(15) a. *Tiene (*a) un hermano*
has A INDEF brother
'(S)He has a brother.'
b. *Tiene (a) un hermano en la cárcel*
has A INDEF brother in the jail
'(S)He has a brother in jail.'

The second reason for the kind of analysis proposed here is that it conceptually relates to a more general theory of DOM. Primus (2011) suggests that the functional reason for DOM is that in the most prototypical cases, there are asymmetries between subject and DO that often boil down to co-argument dependencies (a typical one being possession of the object by the subject). DOM obtains exactly when these asymmetries are not obvious.

One natural way to think of the operator in (14a) is in analogy to a reflexivizer (cf. also Hole 2014): it creates a structurally encoded co-argument dependency between the dative argument and the DO. Crucially, in most of our corpus data, the possessive dative was co-indexed with the subject, as, for example, in (16). Hence PR creates a co-argument dependency (encoded as modification of verbal semantics by an operator) between the subject and the DO. This is precisely the kind of constellation discussed by Primus (2011). Hence, our analysis fits a more general theory and makes the blocking effect less surprising.

(16) *Nervos peste măsură, Gheorghiță şi- a luat fratele la bătaie.*
Angry above measure Gheorghită REFL.DAT.CL has taken brother.def to beating
'Being very angry, Gheorghiță$_i$started beating up his$_i$brother.'

BDD-A26971 © 2017 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:01:25 UTC)

4. ALTERNATION FOR UNMODIFIED NOUNS

In this section, we provide a semantic analysis of the variation observed for unmodified DOs with and without the definite article (article drop). To recap the crucial data, [+DEFINITE] [+HUMAN] direct objects without further modifiers can be *pe*-marked when the definite article is covert as shown in (17a) but appear without *pe*-marking whenever the definite article is overt, as shown in (17b).

(17) a. *Ion l-a văzut pe băiat(*-ul).*
John CL.3.SG-has seen PE boy(-DEF)
'John saw the boy.'
b. *Ion a văzut (*pe) băiatul.*
John has seen PE boy.DEF
'John saw the boy.'

We suggest that it is not article drop itself that is responsible for this variation. It is rather a coincidental fact that unmodified definite DOs lend themselves to weak definite interpretations. Weak definites, analyzed here on a par with incorporation, are claimed to be incompatible with *pe*-marking because of the necessity to move *pe*-marked DOs out of VP.

When considering the variation in *pe*-marking for unmodified DOs, it is immediately visible that generic usages of unmodified nominal DOs, like the one in (18), constitute the bulk of the [-PE] data in our corpus of sentences with human object nominals.[13]

(18) *Concursurile de frumusețe ajută copilul să-și facă prieteni.*
Competitions.DEF of beauty help child.DEF to-REFL.3.SG make friends
'Beauty contests help children to make friends.'

We analyze the DOs in such examples as *weak definites* in the sense of Carlson and Susmann (2005), exemplified in (19). As one important test, co-reference in ellipsis (expected for regular referential definites) does not apply to examples like (18). In (20), one cannot conclude that the elliptic object of the second clause relates to the same child(ren) as the object of the first one.

(19) a. *John went to the doctor.*
b. *John is reading the newspaper.*

[13] Similarly, von Heusinger and Onea (2008:73) discuss the contrast between (i-a) and (i-b), suggesting that (i-a) has a prominent generic reading that (i-b) lacks.

i) a. *Ion adoră femeia.*
John worships woman.DEF
'John worships women/the woman.'
b. *Ion o adoră pe femeie.*
John CL.3.SG.ACC worships PE woman
'John worships the/that woman'

BDD-A26971 © 2017 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:01:25 UTC)

(20) *Concursurile de frumusețe ajută copilul să -şi facă prieteni.*
competitions.DEF of beauty help child.DEF to-REFL.3.SG make friends
şi concursurile de dans la fel.
and competitions.DEF of dancing as well
'Beauty contests help children to make friends, as do dancing competitions.'

Weak definites, as a category, are limited to unmodified nouns, as discussed for example in Aguilar-Guevara and Zwarts (2010:181). Hence, we may be able to naturally explain the variation 'induced' by article drop without posing explanatory burden on article drop itself. Importantly, however, weak definites may occasionally also appear with modifiers when the modifier establishes a subtype of the modified noun, as illustrated in (21). Another tradition of terminology would speak of "relational" or "classificatory" adjectives.

(21) a. *Lola is in the new hospital.* (no weak reading)
b. *Lola is in the medical hospital.* (weak reading ok)
c. *You should see the doctor who works in the medical center.* (no weak reading)
d. *You should see the eye doctor.* (weak reading ok)

This predicts that we should find relatively fewer examples of non-*pe*-marked DOs within the very same verb-noun combinations investigated above when the object is modified. Crucially, when the adjective creates a contextually relevant (pragmatic) subtype of the noun, significantly more non-*pe*-marked DOs should remain.

We tested this at least at a cursory level by manually analyzing all hits of the form *ajută (pe) copilul* [MODIFIER] ('helps the [modifier] child').[14] As shown in Table 5, we found 67 hits of modified [-PE] DOs of this type in the corpus, but a majority of the data (50%) involves PR. From the remaining [-PE] hits, the modifiers are, in general, either subtype-creating adjectives such as (e.g. *small child*, *unborn child*) or subtype-creating adjectives in the relevant pragmatic context, sensu Dayal (2004), (e.g. *sick child, obese child*). We only found five instances which were not clearly subtype-creating adjectives (e.g. 'the aggressed child', 'the child reaching the age of adolescence'), and even these are debatable. Importantly, a majority of subtype-creating modification structures was not marked with *pe* (6:26), whereas most cases of modification not defining subtypes led to *pe*-marking (33:5). (The contrast is significant: χ-squared = 29.9423, df = 1, p-value = 4.451e-08.) This allows us to ask the crucial question: Can the movement assumption developed in section 3 account for the observation that weak definites are not *pe*-marked in Romanian?

[14] The choice of the lexical material is justified by the strong collocation between *help* and *child* as *pe*-marked DO in Romanian. In our data we had 1016 co-occurrences with T-score=31.783.

BDD-A26971 © 2017 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:01:25 UTC)

Table 5

Manual analysis of all hits for *helps (pe)* [modifier] *child*

	Subtype Modification		Pragmatic Subtype Modification		No Subtype Modification		Error	Total
	+PR	–PR	+PR	–PR	+PR	–PR		
[+PE]	0	2	0	4	0	33	0	39
[–PE]	6	17	19	9	6	5	5	67

In a recent analysis, Aguilar-Guevara and Zwarts (2010) suggest that weak definites (and some generics alike) denote kinds in the sense of Krifka (2003), McNally and Boleda (2004). This leads to a very obvious problem: the objects in examples such as (18) obviously relate to individuated countable concepts (potentially under some generic quantification) but not to kinds per se. In order to solve this problem, Aguilar-Guevara and Zwarts (2010) follow the pseudo-incorporation analysis of Dayal (2004) and further developed in Espinal and McNally (2011) for bare nouns as property denoting expressions, suggesting that when expressions denoting kinds receive a thematic role, an operator that takes the kind and the thematic role of a particular event as arguments can establish the thematic participation of instantiations of that kind in the relevant event (see also Dobrovie-Sorin et al. 2006). Without going into further technical details, it is obvious that this operation is directly linked to the lexical interpretation of the verb (which is directly modified by the relevant operator), as is usual with (pseudo-)incorporation. Once we assume that *pe*-marked DOs move outside the VP, the operation is no longer available;[15] at least if we exclude kind-traces from our semantic repertoire (which is motivated by the distribution of bare nouns cross-linguistically).

Hence, within the domain of article-drop, we see variation in Romanian that has nothing to do with article-drop as such. Instead, our analysis boils down to a fairly systematic semantic reason for variation: *pe*-marking excludes reference to kinds (naturally associated with unmodified nouns) because *pe*-marked DOs must move past the thematic-role-assigning operator needed to transform kinds into their instantiations.

5. CONCLUSION AND OUTLOOK

In this paper, we have discussed the distribution of DOM with [+DEFINITE] [+HUMAN] DOs in Romanian. We started out from the observation that the apparent optionality of *pe*-marking is surprising on theoretical grounds. Based on a cor-

[15] Cornilescu and Dobrovie-Sorin (2007) suggest that non-*pe*-marked DOs in Romanian may have the semantic type of properties (<*e*, *t*>) whereas [+PE] [+HUMAN] DOs have type *e*. To a certain degree, our analysis captures the same intuition, however our assumptions that [+PE] [+HUMAN] definites move and that [-PE] [+HUMAN] definites denote kinds are more cautious and parsimonious at the same time. They explain why a definite article is used (kind denotation involves uniqueness) and they keep more analytical possibilities open for other DOs, e.g. indefinites or pronouns. For similar ideas for Spanish cf. Bleam (2006).

BDD-A26971 © 2017 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:01:25 UTC)

pus study, we suggested that the variation in Romanian is–at least in quantitative terms–marginal (i.e. *pe*-marking is either obligatory or about to become obligatory), except for two domains of variation delimited by two syntactic blocking effects: PR and article drop. We sketched an analysis for each of the blocking effects and suggested that they can be explained at the syntax-semantics interface by the assumption that *pe*-marking of DOs invariably triggers movement of the object to a position above the VP.

We wish to close this paper with a speculative remark that outlines possible further research. According to our analysis, both types of non-*pe*-marked DOs induced by these blocking effects, involve a special semantic composition of the verb and the DO in terms of a VP-internal operator that mitigates the connection between verb and DO; this operator modifies the semantic role creation properties of the verb. Going one step further, we may then say that *pe*-marking is lacking in Romanian when the verb does not take its DO as a plain, standard argument. If this observation is correct, we can generalize as in (22).

(22) **Tentative Generalizations for DOM in Romanian**
a) All [+DEFINITE] [+HUMAN] DOs that are combined with the transitive verb in a standard way are marked with *pe*.
b) No [+DEFINITE] [+HUMAN] DO that is combined with the transitive verb by means of some local thematic role assignment operator may be marked with *pe.*

This would, indeed, be the kind of split alternation system that more general theories would expect, with fluid alternation limited to indefinites and specificity effects. This raises the more general question of how our proposal connects to the older cross-linguistic idea championed by Hopper and Thompson (1980) that DOM is a marker of prototypical transitivity: prototypical transitivity may well be connected with a certain (standard) mechanism at the syntax-semantics interface that combines verbs with their object arguments.

REFERENCES

Aguilar-Guevara, A., J. Zwarts, 2010, "Weak definites and reference to kinds", *Proceedings of SALT,* 20, 1–15.

Aikhenvald, A., R.M.W. Dixon, M. Onishi (eds), 2001, *Non-canonical marking of subjects and objects*, Amsterdam, John Benjamins.

Aissen, J., 2003, "Differential Object Marking: Iconicity vs. Economy", *Natural Language and Linguistic Theory*, 21, 435–483.

Avram, L., M. Coene, 2000, "Dative/genitive clitics as last resort", *Balkanistica*, 15, 1–33.

Coene, M., L. Avram, 2008, "Can children tell us anything we did not know about parameter clustering? ", in: T. Biberauer (ed.), *The Limits of Syntactic Variation*, Amsterdam, John Benjamins, 459–482.

Bleam, T., 2006, "The role of semantic type in differential object marking in Bare Plurals, Indefinites, and Weak-Strong Distinction", *Belgian Journal of Linguistics*, 19, 3–27.

BDD-A26971 © 2017 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:01:25 UTC)

Bossong, G., 1985, *Empirische Universalienforschung: Differentielle Objektmarkierung in den neuiranischen Sprachen*, Tübingen, Narr.
Bossong, G., 1991, "Differential object marking in Romance and beyond", in: D. Wanner, D. Kibbee (eds) *New Analyses in Romance Linguistics (XVIII Linguistic Symposium on Romance Languages 1988)*, Amsterdam, Benjamins, 143–170.
Büring, D., 2005, *Binding Theory*, Cambridge, Cambridge University Press.
Carlson, G., R. Sussmann, 2005, "Seemingly indefinite definites", in: S. Kepsar, M. Reis (eds), *Linguistic Evidence*, Berlin, de Gruyter, 26–30.
Chiriacescu, S., K. von Heusinger, 2010, "Discourse prominence and pe-marking in Romanian", *The International Review of Pragmatics*, 2, 298–322.
Cinque, G., I. Krapova, 2008, "Clitic reduplication constructions in Bulgarian", in: D. Kallulli, L. Tasmovski (eds), *Clitic Doubling in the Balkan Languages*, Amsterdam, John Benjamins, 257–287.
Comrie, B., 1975, "Definite and Animate DOs: A Natural Class", *Linguística Silesiana*, 3, 13–21.
Cornilescu, A., V. Hill, 2015, *Formal approaches to DPs in old Romanian*, Leiden, Brill.
Cornilescu, A., C. Dobrovie-Sorin, 2008, "Clitic Doubling. Complex Heads and Interarboreal Operation", in: D.Kalluli, L. Tasmowski (eds), *Clitic Doubling in the Balkan Languages,* Amsterdam/Philadelphia, John Benjamins, 289–319.
Croft, W., 2003, *Typology and Universals*, Cambridge, Cambridge University Press.
Dayal, V., 2004, "Scope Marking as Indirect Wh Dependency", *Natural Language Semantics*, 2, 137–170.
de Hoop, H., A. Malchukov , 2007, *On fluid differential case marking: A bidirectional OT approach*, *Lingua*, 117, 9, 1636–1656.
Diesing, M., 1990, *The syntactic roots of semantic partition*, PhD Dissertation, Amherst.
Dobrovie-Sorin, C., 1994, *The Syntax of Romanian*, Berlin, de Gruyter.
Dobrovie-Sorin, C., 2007, "Article-Drop in Romanian and Extended Heads", in: G. Alboiu, A. Avram *et al.* (eds.), *Pitar-Moş. A biulding with a view. Papers in honour of Alexandra Cornilescu*, Bucureşti, Editura Universităţii din Bucureşti, 99–106.
Dobrovie-Sorin, C., I. Giurgea, 2006, "The suffixation of Definite Articles in Balkan Languages", *Revue romanine de linguistique,* 51, 1, 113–135.
Dobrovie-Sorin, C. *et al.*, 2006, "Noms nus, nombre et types d'incorporation", in: C. Dobrovie Sorin (ed.), *Noms nus et généricité*, Paris, Presses Universitaires de Vincennes.
Dumitrescu, D., 1990, "El dativo posesivo en español y en rumano", *Revista Española de Linguística,* 20, 403–429.
Espinal, M., L. McNally 2011, "Bare nominals and incorporating verbs in Spanish and Catalan", *Journal of Linguistics*, 47, 87–128.
Farkas, D.,1978, "Direct and inDO 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), Chicago, 88–97.
Guéron, J., 1985, "Inalienable possession, PRO-inclusion and lexical chains", in: J. Guéron, H.-G. Obenauer *et al.* (eds), *Grammatical Representation*, Dordrecht, Foris, 43–86.
von Heusinger, K., G. Kaiser, 2005, "The evolution of differential object marking in Spanish", in: K. von Heusinger, G. Kaiser, E.Stark (eds), *Proceeding of the Workshop Specificity and the Evolution / Emergence of Nominal Determination Systems in Romance*. Konstanz, Fachbereich Sprachwissenschaft der Uni. Konstanz, 33–69.
von Heusinger, K., E. Onea, 2008, "Triggering and blocking effects in the diachronic development of DOM in Romanian", *Probus*, 20, 67–110.
Hill, V., 2013, "The DO Marker in Romanian: A Historical Perspective", *Australian Journal of Linguistics*, 33, 140–151.
Hole, D., 2006, "Extra argumentality–affectees, landmarks, and voice", *Linguistics*, 44, 383–424.
Hole, D., 2012, "German free datives and Knight Move Binding", in: A. Alexiadou, T. Kiss, *et al*. (eds), *Local Modelling of Non-Local Dependencies in Syntax*, Tübingen, Niemeyer, 213–246.

BDD-A26971 © 2017 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:01:25 UTC)

Hole, D., 2014, *Dativ, Bindung und Diathese*, Berlin/Boston, de Gruyter.
Hopper, P., S. Thompson, 1980, "Transitivity in grammar and discourse", *Language*, 56, 4, 251–299.
Kamp, H, A. Bende-Farkas, 2006, "Epistemic specificity from a communication-theoretic perspective", Ms., Stuttgart University.
Kayne, R., 1975, *French Syntax: The Transformational Cycle*, Cambridge MA, MIT Press.
Kayne, R., 1977, *Syntaxe du français: Le cycle transformationnel*, Paris, Editions du Seuil.
Kayne, R., 1991, "Romance Clitics, Verb Movement, and PRO", *Linguistic Inquiry*, 22, 647–686.
Klein, U., P. de Swart, 2010, "Case and referential properties", *Lingua*, 122, 1, 3–19.
König E., M. Haspelmath, 1998, "Les constructions à possesseur externe dans les langues d'Europe", in: J. Feuillet (ed.), *Actance et valence dans les langues de l'Europe*, Berlin, Mouton de Gruyter, 525–606.
Kratzer A., 1995, "Stage-level and indivual-level predicates", in: G. Carlson, F. G. Pelletier (eds), *The Generic Book*, Chicago, Chicago University Press, 125–175.
Krifka, M., 2004, "Bare NPs: Kind-referring, Indefinites, Both, or Neither?", in: R. Young, Y. Zhou (eds), *Proceedings of SALT* 13, Cornell, CLC Publication, 180–203.
Landau, I., 1999, "Possessor Raising and the Structure of VP", *Lingua,* 107, 1–37.
Lee-Schoenefeld, V., 2006, "German possessor datives – raised and affected", *Journal of Comparative Germanic Syntax,* 9, 101–142.
Malchukov, A., H. de Hoop, 2011, "Tense, aspect, and mood based differential case marking", *Lingua*, 121, 1, 35–47.
Mardale, A, 2007, *Les prépositions fonctionnelles du roumain: étude comparative. Thèse de doctorat*, Université Paris 7 and Université de Bucarest.
Mardale, A., 2008, "Prépositions et article défini en roumain", in: A. de Saint-Pierre, M. Thibeault (eds), *Actes des 21èmes Journées de Linguistique*, Québec, CIRAL, 78–93.
Mardale, A., 2015, "Differential Object Marking in the First Original Romanian Texts", in: V. Hill (ed.), *Formal Approaches to DPs in Old Romanian*, Leiden, Brill, 200–245.
McNally, L., G.Boleda, 2004, "Relational adjectives as properties of kinds", in: O. Bonami, P. Cabredo Hofherr (eds), *Empirical Issues in Syntax and Semantics*, 5, 179–196.
Primus, B., 2011, "Animacy, generalized semantic roles, and differential object marking", in: M. Lamers, P. de Swart (eds), *Case, Word Order, and Prominence. Interacting Cues in Language Production and Comprehension*, Dordrecht, Springer, 65–90.
Pylkkänen, L., 2002, *Introducing arguments*, PhD Dissertation, Massachusetts Institute of Technology, Cambridge, MA. University of Connecticut.
Rodríguez-Mondoñedo, M., 2007, *The syntax of objects: Agree and differential object marking,* Ph.D. Dissertation.
Szabolcsi, A., 1984, "The possessor that ran away from home", *The Linguistic Review*, 2, 89–102.
Vergnaud, J.-R., M. Zubizarreta, 1992, "The definite determiner and the Inalienable Constructions in French and English", *Linguistic Inquiry*, 23, 595–652.

BDD-A26971 © 2017 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:01:25 UTC)