

Triggering and blocking effects in the diachronic development of DOM in Romanian¹

Klaus von Heusinger and Edgar Onea Gáspár
to appear in *Probus* 20 (1) 2008

Abstract

According to recent studies, the diachronic development of differential object marking (DOM) in Romance is expected to start from elements high on the definiteness scale and spread steadily to elements low on the definiteness scale, whereby each development step is facilitated by fine-grained semantic distinctions which act as triggering conditions (cf. von Heusinger and Kaiser 2005). In Romanian the expected line of development can be observed until the 19th century, however this unidirectional evolution is surprisingly reversed around the 19th-20th century: the conditions for the marking of indefinite direct objects with the preposition *pe* are more restrained today than two centuries ago.

In this study we present the diachronic development of DOM in Romanian based on a diachronic corpus consisting of Bible translations from different centuries. We shall contend that the transition of DOM from one referential category to another, such as from proper names to definite NPs and from definite NPs to indefinite NPs is enabled by some distinction in the category DOM is spreading to. For instance in 16th century Romanian DOM spreads to definite NPs such that first NPs having a strong unique interpretation get *pe*-marked while other definite NPs receive DOM only in a later step involving the neutralization of this transitory distinction. However, Romanian data show that such a fine-grained distinction may not only occur in the category DOM is spreading to but also inside of referential categories already *pe*-marked. In this case the transitory distinction may trigger a regress in DOM. This is the case in the 19th century, when specific indefinite NPs are often *pe*-marked and a new semantic distinction between different types of referential anchoring involves a re-interpretation of the semantic import of *pe*-marking such that a certain part of indefinite direct objects systematically lose DOM. We shall argue that the trigger of this new semantic distinction is the independent development of the Romanian clitic doubling system which DOM gets strongly correlated to: the correlation between clitic doubling and DOM combines two slightly different semantic features, which we shall model by the fine structure of specificity. The combination of these features leads to pragmatically motivated language change.

1. Introduction

Romanian like Spanish or Sardinian shows differential object marking (DOM): the direct object in some contexts may or even must be marked by the marker *pe*, mostly accompanied by clitic doubling. The direct object marker *pe* is a homonym with the locative preposition *pe* meaning ‘on’, however the classification of *pe* is not straightforward, since *pe* shares properties of prepositions but does not have a prepositional meaning. There is some consensus in the literature that the most important synchronic conditions of DOM in Romanian are animacy, definiteness, specificity, and topicality (Farkas 1978, Dobrovie-Sorin 1994, Popescu 1997, Bende-Farkas 2002, Mardale 2007, Kamp and Bende Farkas 2006, etc.). Thus *pe*-marking, mostly combined with clitic doubling, only applies to human direct objects if they are definite, specific or topicalised.

According to the general view as shown in (1), animate personal pronouns and proper names are *pe*-marked. In both (1a) and (1b) omitting *pe* leads to ungrammatical sentences. Post-verbal direct objects instantiated by definite NPs are usually *pe*-marked if they denote a human referent and are ungrammatical with *pe* if they denote non-human entities, as shown by the contrast between (2a) and (2b). As opposed to the examples in (1), (2a) is not ungrammatical if the doubled clitic and the object marker are left out, but many speakers would not prefer this alternative. While *pe*-marking is ungrammatical with non-human indefinite direct

¹ We would like to thank Ileana Comorovski, Manuel Leonetti, Victoria Escandell-Vidal, Udo Klein, Elisabeth Stark, Georg Kaiser, Alexandru Mardale and our two anonymous reviewers for critical and constructive comments, which greatly helped to improve the quality of this paper, and especially we would like to thank Georg Kaiser for editing this volume. Needless to say, we are responsible for any and all shortcomings ourselves. The research for this paper has been funded by the German Research Foundation (DFG) as part of the SFB 732/project C2 “Case and Referential Context” at the University of Stuttgart.

objects as well, for human indefinites DOM can distinguish between a specific and a non-specific reading as shown in (3). While (3b) is ambiguous between, on the one hand, a specific reading according to which the speaker is searching for a person who has the property of being a secretary and is probably known to the speaker, and on the other hand, a non-specific reading according to which the speaker is searching for someone having the property of being a secretary, without implying that some secretary actually exists, in (3a) the non-specific reading is ruled out by the presence of *pe*. In fact, the conditions and the semantic import of DOM in present day Romanian are somewhat more complex, and we will take a closer look at them in the following section.

- (1) a. *L-* *am văzut *(pe) el.*
 CL.3.SG.MASC have seen DOM he
 ‘I have seen him.’
- b. *L-* *am văzut *(pe) Mihai.*
 CL.3.SG.MASC have seen DOM Michael
 ‘I have seen Michael.’
- (2) a. *L-* *am văzut pe profesorul tău.*
 CL.3.SG.MASC have seen DOM professor.DEF.MASC your
 ‘I have seen your professor.’
- b. (**L-*) *am văzut (*pe) autobuzul tău.*
 CL.3.SG.MASC have seen DOM bus.DEF.MASC your.
 ‘I have seen your bus.’
- (3) a. *O* *caut pe o secretară.*
 CL.3.SG.FEM search DOM a secretary.FEM
 specific reading: ‘I am searching for a (specific) secretary.’
- b. *Caut o secretară.*
 search a secretary.FEM.
 specific reading: ‘I am searching for a (specific) secretary.’
 non-specific reading: ‘I am searching for some secretary.’

The diachronic evolution of direct object marking in Romanian proceeds along the same lines as in Spanish (Melis 1995, Laca 2002 & 2006, von Heusinger and Kaiser 2007) until the 19th century: the development starts out with highly referential direct objects as personal pronouns and proper names and then spreads, facilitated by some intermediate distinctions, first to definite and as a second step to indefinite noun phrases. However, between the 19th century and the early 20th century this evolution reverses with respect to indefinite NPs, such that today indefinite direct objects are less frequently marked with *pe* than in the 19th century. This can be illustrated by the following contrasting historic translations of the same Bible verse: (in older texts *pe* mostly appears as *pre*)

Matthew 8:16

- (4) 17th *adusără lui îndrăciți mulți...*
 brought he.DAT demonised.PL many...
- 17th *aduseră lui îndrăciți mulți...*
 brought he.DAT demonised.PL many...
- 19th *adus -au Lui pre mulți îndrăciți...*
 brought -have he.DAT DOM many demonised.PL...
- 20th *au adus la El mulți demonizați*
 have brought to he many demonised.PL
 ‘they brought to him many demon-possessed...’

In (4), four Bible translations are compared. While the older translations from the 17th century do not mark the indefinite direct object, the 19th century translation introduces DOM. And even though the referential context can be considered unchanged, in the 20th century translation *pe* is omitted.

In this paper we shall analyze both the diachronic increase of DOM and its decrease based on a small corpus consisting of Bible translations and religious texts from different centuries. Our argument is that the diachronic development of DOM along the definiteness scale occurs through transition points involving a fine-structured semantic differentiation, as argued in von Heusinger and Kaiser (2005). Romanian data suggest, however, that a similar effect can arise due to interaction with an independent phenomenon, such that the transitional semantic differentiation leads to a regress in DOM instead of facilitating its spreading to less referential expressions. This is the case in Romanian between the 19th and the 20th century, when an overlapping effect between clitic-doubling and DOM of indefinite NPs leads to a semantic re-interpretation, resulting in a fine-structured specificity scale linked to the combination possibilities of clitic doubling and DOM. In the second section, we shall present the major synchronic facts about DOM in Romanian, in the third section we shall discuss the diachronic data our paper is dealing with. In the fourth section, we shall develop the notion of fine-structured specificity, distinguishing between different types of specific indefinites whereby specificity will be modeled by means of a scale starting from speaker anchored indefinites and partitives and going down to locally anchored NPs. In the fifth section, we shall be able to analyze the interaction between clitic-doubling and DOM from the 19th to the 20th century and the resulting semantic differentiation using this fine-structured notion of specificity.

2. Conditions of direct object marking in present day Romanian

The observation that the direct object in various languages may or may not be marked has been known since Bossong (1985) under the concept of “differentiated object marking” or DOM. Cross-linguistically, there are at least four parameters that determine the conditions of direct object marking: (i) animacy, (ii) referential categories, (iii) information structure, and (iv) verb semantics. In the typological literature such parameters are mostly conceived as scales or hierarchies (Comrie 1975, Bossong 1985, Haspelmath 1997, Aissen 2003, Croft 2003). DOM is widely supposed to occupy continuous sectors of different length starting at the more prominent part of one or more of the following scales:

Animacy scale:

human > animate > non-animate

Definiteness scale:

pers.pron. > propr.noun > def.NP > indef.spec. NP > indef.non-spec.NP > incorp

Topicality

topical > non topical

Scale of verbal classes according to animacy preferences (von Heusinger and Kaiser 2007)

[+ animate]	>	[± animate]	>	[±/- animate]
Class 1		Class 2		Class 3
<i>beat/kill</i>		<i>see/hear</i>		<i>take/move</i>

While animacy, referential properties and topicality are conceived as hierarchies, the role of verb semantics concerns, among other factors, the selectional properties of the verb with regard to animacy. Accordingly, class 1 verbs usually select animate direct objects, class 2 verbs may have both animate and inanimate direct objects, and class 3 verbs usually select non-animate direct objects (von Heusinger and Kaiser 2007, von Heusinger 2008 *in this vol-*

ume). In addition, different aspects included in the notion of transitivity (Hopper and Thompson 1980), such as affectedness, punctuality, kinesis etc., can be considered. In Romanian all of these factors play a role.

Full personal pronouns and proper names referring to humans are always marked with *pe* in present day Romanian. It should be noted that full personal pronouns usually do not refer to non-human direct objects and that full personal pronouns are only used for emphasis, while weak pronouns suffice for anaphoric reference.

Most of the post-verbal human definite NPs get DOM if there are no semantic (cf. Dobrovie-Sorin 2007) or syntactic restrictions blocking it. The most important and very common restriction is that the structure *pe* + noun + def.art without further modifiers is ungrammatical in Romanian. The reason for this constraint does not have anything to do with DOM itself, since it applies to nearly all Romanian prepositions (the exception is *cu* meaning ‘with’). (Lyons (1999) gives cross linguistic evidence for similar phenomena.) Thus a noun suffixed with the definite article cannot co-occur with a preposition unless it is modified with an adnominal expression. Accordingly both (5a) and (5b) are unacceptable.

- (5) a. **L- am văzut pe băiatul.*
 CL.3.SG have seen DOM boy.DEF.MASC
 ‘I have seen the boy.’
 b. **Am stat lângă băiatul.*
 Have stood next-to boy.DEF.MASC
 ‘I stood next to the boy.’

Some rather familiar functional expressions such as *the mother, the teacher, the priest, the boss*, etc. (as opposed to functional expressions that are rather official, as *the director, the king, the president*, etc.) are exceptions from this constraint: for them the construction prep + unmodified noun + def. article is acceptable (at least in spoken Romanian). Thus (6) is acceptable under the reading that *the boss* is granting contextual uniqueness to its referent:

- (6) *L -am văzut pe șeful.*
 CL.3.SG.MASC have seen DOM boss.DEF.MASC
 ‘I have seen the boss.’

In cases in which the use of the DOM marker is ruled out by this constraint, speakers have to decide whether they will use the definite article and thus omit *pe* or use *pe* and omit the definite article, such that a not fully understood alternation between *pe* + unmodified noun vs. unmodified noun + def.art. can be observed as illustrated in the following contrasts:

- (7) a. *Nu am leșinat, dar m- am emoționat când*
 Not have passed-out but CL.1.SG have got-affected as
*am văzut copilul.*²
 have seen child.DEF
 ‘I did not pass out, but I got affected as I saw the child.’
 b. **Nu am leșinat, dar m - am emoționat când*
 Not have passed-out but CL.1.SG have got-affected as
am văzut pe copilul.
 have seen DOM child.DEF
 ‘I did not pass out, but I got affected as I saw the child.’
- (8) a. *Lapuțin timp l- am văzut pe băiat alergând.*³
 At little time CL.3.SG have seen DOM boy running.
 ‘After a short while I have seen the boy running.’

² <http://www.tangomagazine.ro/content/view/186/41/>

³ <http://www.eva.ro/divertisment/articol347.html>

- b.**Lapușin timp l- am văzut pe băiatul alergând.*
 At little time CL.3.SG have seen DOM boy.DEF running.
 ‘After a short while I have seen the boy running.’

Both (7a) and (8a) stand for different strategies of expressing the same (or at least a very similar) referential category. While some people seem to prefer (8a), it is not totally clear whether there is some semantic difference between them. Note that the presence of the differential object marker renders (7a) ungrammatical as shown in (7b), and the presence of the definite article makes (8a) ungrammatical as well, as shown in (8b). The presence or absence of a clitic pronoun in the b examples would not improve their acceptability. For a proposed explanation of this alternation using OT cf. Popescu (1997).

As pointed out by one of our anonymous reviewers, there are cases in which the presence of *pe* is semantically relevant for definites. In these cases the presence of *pe* rules out generic readings (cf. Dobrovie Sorin 2007), as shown in (9), but this semantic effect may also be related to the presence of the weak pronoun:

- (9) (a) *Ion adoră femeia.*
 Ion worships woman.DEF
 ‘Ion worships women/the woman.’
 (b) *Ion o adoră pe femeie.*
 Ion CL.3.SG worships DOM woman
 ‘Ion worships the/that woman.’

Note that this semantic effect has only the status of a constraint and does not explain the optionality of DOM in the (b) reading or in cases in which generic readings are not available anyway, but in this paper we shall not pursue the discussion of the semantic impact of DOM for definites any further. Hence we can summarize that for post-verbal [+human] definite direct objects *pe*-marking is preferred if no constraints ruling out DOM apply.

For post-verbal indefinite human direct objects DOM is optional if further factors are fulfilled. Such factors include specificity (Farkas 1978, Dobrovie-Sorin 1994, Bende-Farkas 2002, Mardale 2007, Kamp and Bende-Farkas 2006) and discourse prominence (Chiriacescu 2007). As shown in the contrast between (3a) and (3b), DOM disambiguates a specific reading of an indefinite direct object, but unmarked direct objects need not be non-specific.

The distribution of DOM for both definite and indefinite direct objects is significantly different if the direct object is pre-verbal, i.e. either topicalised or focused. We shall not analyze pre-verbal direct objects in this paper any further, for these structures are too rare in our corpus. Moreover, we shall not analyze some other cases of DOM in Romanian either, such as different quantifier expressions, relative pronouns, etc., since we do not have enough diachronic data to make generalizations about these phenomena.

DOM in Romanian is often accompanied by clitic doubling, i.e. the occurrence of a co-indexed weak pronoun. Accusative clitics do not occur with non-marked post-verbal direct objects in Romanian, however for pre-verbal direct objects different rules again apply. If the direct object is post-verbal, the distribution of clitic doubling generally depends on DOM and on the definiteness scale: pronouns and proper names never occur without clitic doubling, there is a strong preference for clitic doubling with *pe*-marked definite direct objects, and some preference for clitic doubling with *pe*-marked indefinite direct objects.

While *pe*-marked definite direct objects without clitic doubling seem a very marked/unusual phenomenon in present day Romanian, at the end of the 19th century it was a rather unmarked alternative as shown in the example (10) from a fairy tale written in 1870⁴.

⁴ Eminescu, Mihai(1978). Poezii Proza literara Vol.II. Petru Cretia, Bucuresti. p 268.⁵ We used the 1858 edition. This edition doesn't contain significant changes compared with the original.

- (10) *Hai că mă duc eu să răpesc pe fata Genarului!* (Eminescu)
 Come that me take I to hijack DOM girl.DEF Genar.GEN
 ‘Let me go to hijack the girl of the Genar.’

Pe-marked indefinite direct objects can be observed in present day Romanian both with and without clitic doubling, however DOM occurs significantly more often with clitic doubling than without clitic doubling (cf. Table 16 below in section 5), and for a large number of speakers clitic doubling is obligatory for *pe*-marked post-verbal indefinite direct objects.

3. The diachronic development of DOM in Romanian

The development of DOM in Romanian occurs along similar lines as in Spanish. The historic background of this evolution is similar as well: after the breakdown of the Latin case system Romanian, like Spanish, uses a preposition to mark the direct object.

Studies on Spanish (Melis 1995, Laca 2002, 2006) have shown that in Old Spanish (12th century) only personal pronouns and human proper names are obligatorily marked by the differential object marker *a*. At the same time, definite human direct objects are optionally marked and indefinite direct objects are not marked at all. The historic evolution leads to an increasing amount of DOM with human direct objects rather low on referentiality, such that in Modern Spanish only pseudo-incorporated (non-argumental) indefinite direct objects are systematically not marked (cf. Leonetti 2004). Thus the evolution of DOM only involves a shift down the definiteness scale.

Von Heusinger and Kaiser (2005) argue that the diachronic evolution of DOM in Spanish involves transitional points in which different factors facilitate the development. Accordingly, topicality is a relevant factor opening the way for the marking of definite NPs, but this effect is neutralized in a further step allowing DOM to expand to the whole range of definite NPs. Similarly, specificity is considered a transition point facilitating the marking of indefinite direct objects: first only specific, and in later steps all argumental human indefinite direct objects being *a*-marked. The role of verbal semantics in the development of DOM in Spanish is discussed in von Heusinger and Kaiser (2007) and von Heusinger (2008 *in this volume*). We shall not deal with this problem in this paper.

Based on a corpus consisting of Bible translations and religious texts, we found a very similar line of development for Romanian between the 16th and the 20th century, however the final step observed in Spanish, namely the systematic marking of argumental indefinites, is fulfilled only for a relatively short time in Romanian and is partly reversed at the end of the 19th century. Such effects have been observed in Portuguese as well (cf. Delille 1970, Schäfer-Prieß 2002), where between the 15th to the 17th century DOM is similarly widespread as in Spanish, but in the 17th and 18th centuries DOM systematically disappears in most of the referential contexts, such that in Modern Iberian Portuguese DOM is restricted to certain lexically triggered contexts as for instance *amar* (‘to love’) or the use of nominals extremely high on animacy, such as *Deus* (‘God’). Still, the reasons for this regress are said to be rather a matter of language contact and the tendency to eliminate Spanish influences rather than inherently linguistic forces.

Three interesting aspects are to be treated in the analysis of the Romanian data: the spreading of DOM to definite NPs, the extension of DOM to indefinite NPs and the regress of DOM for indefinites.

3.1. The data-set

In order to gain good evidence about the diachronic development of DOM in Romanian, we have chosen a main corpus consisting of four Bible translations from different centuries. As summarized in von Heusinger and Kaiser (2007), Bible translations provide very good

evidence for diachronic change because they enable the comparison in mostly unchanged referential contexts. The downside of using Bible translations for diachronic analyses is evidently that Bible translations tend to be conservative with respect to language change and Bible texts generally have a relatively restricted and specialized register that notoriously differs from the spoken language. However this disadvantage is not to be overrated, since other older texts in Romanian are also mostly strongly connected to religious topics and show a very restricted register as well.

The oldest available written text in Romanian is known as the letter of Neacșu from 1521. This document does not provide any empirical evidence about direct object marking because it contains not a single clear-cut instance of an animate direct object. The first Bible translation was published about a century later in 1648. Our major corpus analysis refers consequently to the time between 1648 and the end of the 20th century. Since there is one century of documented old Romanian between the very first text and the first complete Bible translation, we provide a brief overview of the situation about direct object marking in the 16th century based on two small religious fragments. Hence our analysis is based on two distinct corpora.

The first corpus includes two religious texts from the 16th century:

Text A: *Pravila lui Coresi* (Brașov 1560-62): a collection of rules for religious punishments.

Text B: *Moltivenicul Diaconului Coresi* (Brașov 1564): a collection of liturgical rules.

The second corpus consists of the first 10 chapters of the New Testament (Matthew 1-10) in 4 Bible translations from different centuries:

Bible A: *Noul Testament de la Bălgrad*. (1648 in Transylvania (Bălgrad))

Bible B: *Biblie.V.T.N.T.* (1688 in Walachia (București))

Bible C: *Biblia de la Blaj* (1795 in Transylvania (Blaj), re-edited 1858)⁵

Bible D: *Biblia ortodoxă* (standard Romanian orthodox Bible from 1994 (București))

The distinction between the corpora is rather a technical matter. Since none of these texts were available to us in electronic format, we analyzed the corpora manually, whereby the analysis method was slightly different for Bible translations, since in this case we could pursue all language changes more exactly and accurately than in other texts. The manual analysis has the obvious disadvantage that we had to restrict our study to a relatively small amount of text, but this goes along with the advantage of a deep and heuristic analysis.

The fact that our corpus consists of Bible translations and that the translators obviously had knowledge of the former translations should be accounted for by two assumptions: on the one hand, traditional Bible translations tend to be conservative with regard to language change, thus with regard to *pe*-marking our translations may reflect the language norm from a few decades earlier than the publishing date would suggest. Thus the Bible translations provide evidence for approximately the following periods of time:

Bible A and B: 17th century⁶

Bible C: second half of the 18th century

Bible D: end of the 19th century, first half of the 20th century

If the translators used the former translations as guidelines, this suggests that any differences in DOM can be considered very important, whereas any predicted difference that does not occur in a later translation may be caused simply by a conservative translation.

An additional restriction of our corpus analysis is that we only looked at [+human] direct objects. One of the reasons for this restriction is that in the Bible translations we used, non-

⁶ We consider differences between Bible A and B to be caused by dialectal differences rather than by diachronic change. It is important to note that we do not assume at any level that the first Bible translations would use archaic language, since in the lack of written documents such a language stage is hardly known to the translator.

human direct objects were notoriously rare. This means that we do not have any data about DOM of [-human] direct objects and none of our generalizations apply to them.

3.2. 16th century data

The essential data of the first corpus, presented in Table 1, show that the percentage of DOM decreases from the left to the right of the definiteness scale in 16th century Romanian. Thus expressions high on definiteness, as proper names or pronouns, are systematically marked with *pe* whereas other expressions are only marked to a lesser extent. The table has, however, a few peculiarities regarding first and second person pronouns and definite NPs.

	me/you	other pronouns	prop. names	def. NP	indef. NP	Total
+ <i>pe</i>	5	33	3	9	1	51
- <i>pe</i>	5	1	0	36	9	51
total	10	34	3	45	10	102
% + <i>pe</i>	50%	97%	100%	20%	10%	50%

Table 1 Direct object marking with *pe* in 16th century Romanian⁷

The first peculiarity of this overview is that only half of the first and second person singular pronouns are marked with *pe*. At first glance, this seems surprising, since according to Bossong (1985) first and second person pronouns are most likely to be DOM-marked. One possible explanation of this fact may be related to the observation that first and second person singular pronouns retain the special forms for accusative from Latin, thus being morphologically case marked as shown in Table 2. Accordingly, DOM for first and second person personal pronouns may have been redundant in the 16th century, as shown in the contrast between (11a) and (11b) from our text B, where *pe*-marking for first and second person personal pronouns appears to be optional:

	Singular		Plural	
	NOM	ACC	NOM	ACC
I person	eu	pe mine	noi	pe noi
II person	tu	pe tine	voi	pe voi
III person	el/ea	pe el/ea	ei/ele	pe ei/ele

Table 2 Personal pronoun paradigm in present-day Romanian

- (11) a. *In zioa de astăzi **pe** **tine** au număratu întru feții săi.*
 in day of today DOM you.ACC have counted into sons his
 ‘In this day he has counted you to his sons.’
 b. *Iată au curățitu **tine** deîn toate păcatele tale.*
 See have cleaned you.ACC from every sins your.
 ‘See, he has cleaned you from all of your sins.’
- (12) a. *Și întărește **pe** **elu** ănoblicitura jurământului tău.*
 And strengthen DOM he in power swear.GEN your
 ‘And strengthen him with the power of your promise.’
 b. * *Și întărește **elu** ănoblicitura jurământului tău. 071168583140*
 And strengthen he in power swear.GEN your
 intended reading: ‘And strengthen him with the power of your promise’

⁷ Note that generally the +human feature will not be explicitly mentioned throughout the paper, since at every historical step we can reconstruct, this feature seems a necessary condition of DOM marking.

While (11a), (11b) and (12a) are corpus-data, we did not find any examples like (12b) in our texts, and therefore we assume that structures like (12b), in which an unmarked third person personal pronoun is used as a direct object, might have been ungrammatical.

The optionality of *pe* with first and second personal pronouns in 16th century Romanian may be also correlated to the syntactic position (pre- or post-verbal) of the direct object, since we do have both *pe* marked and not *pe* marked first and second person personal pronouns in post-verbal position but only *pe*-marked personal pronouns in a pre-verbal position. However this may also be a matter of chance, since we have a very restricted dataset.

It is not very clear whether the hypothesis that DOM always starts diachronically from the first and second person singular pronouns and spreads towards the less animate and referential pole of the animacy and/or definiteness scale is contradicted by our findings or not: the first and second person singular pronouns do not show *pe*-marking but preserve the old Latin case system, and the use of *pe* as direct object marker may have started at other pronouns and proper names, and while it spread towards the less referential direct objects it also started being used analogously for first and second person pronouns. We assume that the step towards this analogical marking of first and second person pronouns takes place exactly at the time of our first corpus, the 16th century, which would explain the optionality we encountered in examples like (11). In this development the pre-verbal position of the direct object may have been a transitory triggering condition enabling *pe*-marking to spread to first and second person personal pronouns. This would mean that topicality could be regarded as a triggering condition of *pe*-marking in a transitory period. Similar observations are reported about Old Spanish in (Melis 1995, Laca 2002, 2006, von Heusinger and Kaiser 2005). However, due to our restricted data this remains only a hypothesis.

The relatively low percentage of DOM with definite NPs shows that the 16th century can be regarded as a transitional period in which DOM spreads to definite NPs. According to our hypothesis that the development of DOM occurs through transitional points implying a finer differentiation, we expect that some further feature is responsible for the marked definite NPs. We assume that besides topicality certain types of definiteness trigger DOM more than others. As shown in Table 3, NPs with demonstratives and functional expressions are marked as direct objects more frequently, while other definite NPs are marked only marginally.

def.NPs	demonstr. art. + NP	functional expr.	other def. NPs	total
+ <i>pe</i>	3	4	2	9
- <i>pe</i>	3	1	32	36
% + <i>pe</i>	50%	80%	6 %	20%

Table 3 DOM for definite NPs in the 16th century

While 80% of the direct objects expressed by means of functional expressions are marked with *pe*, only 50% of the NPs with demonstratives get DOM. Yet, upon a closer look, from 6 occurrences of NPs with demonstratives, 3 were marked with *pe* and two of the unmarked NPs with demonstratives are fronted and doubled by a resumptive pronoun in the following subjunctive VP, as in example (13) from text B, which may be an independent blocking effect:

- (13) ...*voia* *ți* *-e aciastă fată*, *în legé lu* *Dumnezeu să*
 ...will poss.2.SG is this girl_iFEM in law POSS God to
o_i *iai la tine* *să -ți* *fie căsătorie?*
 CL_i.3.SG take to you.ACC to -poss.CL.2.SG be marriage
 ‘Is it your will to marry this girl in Gods law and to have as your wife?’

Functional expressions like *popa* (‘the priest’) guarantee the contextual uniqueness of their referent. From 4 occurrences of *popa* as direct object, 3 were marked with *pe*. Since other factors may be responsible for the non-marked NPs with demonstratives, we assume

that there is no important difference between the marking of demonstrative NPs and functional expressions. These facts about the marking of definite NPs match the theoretical prediction of von Heusinger and Kaiser (2005). We assume that a common characteristic of functional expressions and demonstratives, namely that they identify a contextually unique referent, is the initial trigger of the marking of definite NPs in Old Romanian, just as topicality was in Old Spanish. Accordingly, functional expressions and demonstratives receive *pe* marking first, as shown in Table 4:

pronouns	proper names	Definite NPs		Indefinite NPs
+pe	+pe	-pe		-pe

pronouns	proper names	functional expressions & demonstratives	other def.NP	Indefinite NPs
+pe	+pe	+pe	-pe	-pe

Table 4 Transitional triggering by certain types of definite NPs

In our corpus we only found one indefinite direct object marked with *pe*. The one example found is not sufficient for any generalization about the eventual very restrained conditions of the marking of indefinite direct objects.

3.3. 17th century-20th century data

In this part of the paper we shall present the main lines of development of DOM for human direct objects from the 17th to the 20th century:

The data collected from the four Bible translations is presented in Table 5:

	pronouns		proper names		def.NP		spec. indef. NP		non-spec. indef. NP	
	-	+	-	+	-	+	-	+	-	+
A 17th	1	52	1	47	22	36	6	0	6	0
B 17th	0	55	1	46	23	40	6	0	7	0
C 19th	0	53	1	46	15	46	0	6	7	3
D 20th	0	17	1	46	15	47	5	2	9	2

Table 5 Quantitative presentation of the diachronic data from the 17th to the 20th century

Comparing these data with the data from the 16th century, it is clear that both tendencies observed in the section above, i.e. the marking of personal pronouns and the marking of definite NPs, have continued. We can consider the marking of pronouns as generally obligatory by the 17th century and the marking of definite NPs as the preferred variant even in the 17th century. However, in the 17th century no indefinite NPs with DOM were found in our corpus. Until the 19th century the marking of definite NPs increases and specific indefinite NPs start being marked significantly, while even some non-specific indefinite NPs get the direct object marker *pe*. The surprising event in the diachronic evolution of *pe*-marking is the significant step back in the marking of specific indefinite direct objects from the 19th to the 20th century. Another striking observation is the radical regress of the number of personal pronouns in the 20th century translation. In the following we shall inspect the data in more detail:

3.3.1. Personal Pronouns

Personal pronouns independent of their number and person are generally marked as early as the 17th century. This rule does not change over the centuries, however a striking regress in the number of personal pronouns in the 20th century can be observed. This decrease in personal pronouns in the 20th century goes back to a main syntactic characteristic of Romanian, namely clitic doubling.

Most of the marked direct objects are doubled by co-indexed clitics in present-day Romanian (cf. Dobrovie-Sorin 1994), even though it is not completely clear whether the lack of clitic doubling results in ungrammatical or merely not preferred sentences. In fact, the lack of clitic doubling of marked proper names is less disturbing with less transitive verbs in the sense of (Hopper and Thompson 1980), which recently has been linked to the selectional properties of verbs regarding animacy (see scale of verbal classes in section 2, see also von Heusinger 2008 *in this volume*), as shown by the contrast below:

- (14) a. *L- am văzut pe Ion.*
 CL.3.SG have seen DOM John
 ‘I have seen John.’
 b. ^{??}*Am văzut pe Ion.*
 have seen DOM John
 ‘I have seen John.’

- (15) a. *L -am bătut pe Ion.*
 CL.3.SG have beaten DOM John.
 ‘I have beaten up John.’
 b. ^{*}*Am bătut pe Ion.*
 have beaten DOM John.
 ‘I have beaten up John.’

(14a) is strongly preferred by most Romanian speakers, but the lack of clitic doubling as in (14b) leads only to an archaic syntax, as opposed to (15b) which can be considered completely ungrammatical. In fact even speakers who would completely reject (14b) acknowledge that (15b) is significantly worse. This shows that clitic doubling has, just like DOM, quite subtle triggering conditions.

In our 16th century texts the co-occurrence of personal pronouns and clitics was a sporadic phenomenon mostly triggered by fronting of the strong pronoun. A strong preference can be observed for the choice of either the strong pronoun or the clitic but not to both, as illustrated in (16) from text B. While (16a) and (16b) are authentic examples from text B, we did not find examples like (16c) in our 16th century corpus:

- (16) a. *unde pre noi până în ciasulu de acmu ai ținutu*
 where DOM we until in hour from now have hold
 ‘where you have taken care of us until now.’
 b. *și nu ne lăsa și astăzi și pururile*
 and not CL-3.PL leave and today and always
 ‘and do not leave us now and forever’
 c. [?]*și nu ne lăsa pre noi și astăzi și pururile*
 and not CL-3.PL leave DOM we and today and always
 ‘and do not leave us now and forever’

In contrast, in the Bible translations A and B, i.e. in the 17th century, there are many cases in which the *pe*-marked personal pronoun co-occurs with weak pronouns. Still, there is a slight preference for strong pronouns: clitics generally occur less frequently than strong pronouns. Weak pronouns become predominant in the 19th century Bible, but mostly they are accompanied by *pe*-marked strong pronouns. In the 20th century Bible, however, strong pronouns are being systematically omitted in favour of clitics.

This diachronic observation is illustrated in Table 6, where a very clear shift from the left (–[clitic] +[pe] +[strong pronoun]) to the right (+ [clitic] –[pe] –[strong pronoun]) can be ob-

served.⁸ While in the 19th century Bible most of the accusative clitics were accompanied by strong pronouns with obligatory *pe*-marking, in the 20th century Bible weak pronouns suffice for anaphoric purposes and strong pronouns are very rare.⁹ Strong pronouns, if they are present at all, are necessarily subject to DOM. They are, however, mostly perceived as emphasis, which is also shown by the fact that focus particles like *numai* ('only') or *și* ('too') require the presence of strong pronouns.

	-[clitic] +[strong pron.]	+ [clitic] +[strong pron.]	+ [clitic] -[strong pron.]
A 17 th	32	31	22
B 17 th	34	21	16
C 19 th	10	43	15
D 20 th	1	16	51
Total	77	111	104

Table 6 Evolution of clitic pronouns from the 17th to the 20th century

The diachronic change with regard to clitic doubling is illustrated in (17):

Matthew 8:15

(17) A17 th	<i>și</i>	<i>lăsă</i>	<i>pre</i>	<i>ea</i>	<i>frigul</i>	-[cl] +[pe] +[pron.]		
	and	left	DOM	she	fever			
B17 th	<i>și</i>	<i>lăsă</i>	<i>pre</i>	<i>ea</i>	<i>frigurile</i>	-[cl] +[pe] +[pron.]		
	and	left	DOM	she	fever.PL			
C19 th	<i>și</i>	<i>o</i>	<i>au</i>	<i>lăsat</i>	<i>pre</i>	<i>ea</i>	<i>frigurile</i>	+ [cl] +[pe] +[pron.]
	and	CL.3.SG	have	left	DOM	she	fever.PL	
D20 th	<i>și</i>	<i>au</i>	<i>lăsat</i>	<i>-o</i>	<i>frigurile</i>	+ [cl] -[pe] -[pron.]		
	and	have	left	CL.3.SG	fever.PL			
	'and the fever left her...'							

3.3.2 Proper names

Proper names are generally marked in Romanian as early as the 16th century. From the 17th century until now, no important change has occurred in this domain. The only exception found in our corpus is Matthew 4:23, where the direct object is expressed by a city name. Geographic names are a very systematic exception from DOM even in present day Romanian. Geographic names do not allow DOM even in typical cases in which they metonymically stand for the inhabitants, as for instance "blessing/curing/enslaving Babylon".

3.3.3 Definite NPs

As a contrast to the 16th century data definite NPs are marked in the 17th century at a very high percentage. Hence, we can consider the fine-graded differentiation from the 16th century as a transitional phenomenon enabling DOM to spread to definite direct objects in general. From here, the tendency to mark definite NPs as direct objects shows a striking steadiness over the centuries, which is illustrated in Table 7:

⁸ Note that in Table 6 only three of the four theoretically possible combinations are listed. This is due to the fact that for post-verbal direct objects clitic doubling never occurs if the direct object is not *pe*-marked.

⁹ As it has been pointed out in section 3.1, the 20th century Bible does not accurately reflect 20th century Romanian but rather the diachronic state of the first half of the 20th century or even the last half of the 19th century, and similarly the 19th century Bible stands rather for 18th century Romanian. Hence, as one of our anonymous reviewers points out, in the 19th century clitics were not systematically doubled by strong pronouns and sufficed for anaphoric purposes. At this point it is also important to note that weak pronouns could be used for anaphoric purposes alone even in earlier stages of the language as shown in Table 6.

definite descriptive NPs	16 th	A 17 th	B 17 th	C 19 th	D 20 th
– <i>pe</i>	36	22	23	15	14
+ <i>pe</i>	9	36	40	46	47
Percentage of marking	20%	62%	63%	75%	77%

Table 7 Evolution of DOM for descriptive def. NPs from the 17th to the 20th century

While in the 16th century only a small and well-defined part of the definite NPs (functional expressions and NPs with demonstratives) was marked with *pe*, there are (at least regarding our corpus data) no more significant differences between the semantics of marked and unmarked NPs in the 17th century. The transitional distinction has been neutralized and the whole category of definite NPs is systematically marked as direct objects. This allows for the continuation of Table 4 as shown in Table 8:

Starting point:

pronouns	proper names	Definite NPs	Indefinite NPs
+ <i>pe</i>	+ <i>pe</i>	– <i>pe</i>	– <i>pe</i>

Transitional distinction :

pronouns	proper names	functional expressions & demonstratives	other def.NP	Indefinite NPs
+ <i>pe</i>	+ <i>pe</i>	+ <i>pe</i>	– <i>pe</i>	– <i>pe</i>

Neutralization effect:

pronouns	proper names	Definite NPs	Indefinite NPs
+ <i>pe</i>	+ <i>pe</i>	+ <i>pe</i>	– <i>pe</i>

Table 8 Transitional distinction and neutralisation effect for definite NPs

As can be observed in Table 7, from the 17th to the 20th century the evolution of DOM for definites slows down significantly. Nevertheless, there seems to be a certain preference shift over time regarding the already mentioned syntactic incompatibility of definite articles with *pe* for unmodified nouns. This restriction blocks an important number of expected DOM occurrences, and the speakers usually have to choose between solutions including *pe* and excluding the definite article and solutions excluding *pe* and including the definite article as shown in the contrast between (7a) and (8a) in section. *Pe*-marking as in (8a) becomes increasingly preferred in newer Bible translations, while the use of noun + definite article without *pe* as in (7a) gets less preferred, as illustrated in (18). The exact nature of this preference shift is not very clear. The spreading of the strategy *pe* + unmodified noun may either be related to the similar construction used with other prepositions or with some semantic shift of *pe*.

Matthew 9:34

- (18) A17th *Cu craiul dracilor scoate dracii.*
 With king devils.GEN extracts devils.DEF
- B17th *Cu stăpânul dracilor scoate dracii.*
 With master devils.GEN extracts devils.DEF
- C19th *Cu domnul dracilor scoate dracii.*
 With lord devils.GEN extracts devils.DEF
- D20th *Cu domnul demonilor scoate pe demoni.*
 With lord demons.GEN extracts DOM demons.DEF
 ‘It is by the lord of demons that he drives out demons’

3.3.4 Indefinite NPs

Indefinites start being systematically marked as direct objects only after the 17th century and reach a high frequency of DOM in less than 100 years. After this point for some reason a large part of the indefinite NPs are not marked anymore. The overall evolution can be summarized as in Table 9:

indefinite NPs	A 17 th	B 17 th	C 19 th	D 20 th
– <i>pe</i>	12	13	7	14
+ <i>pe</i>	0	0	9	4
Percentage of +<i>pe</i>	0%	0%	58%	26%

Table 9 Evolution of *pe*-marking for indefinite descriptive NPs from the 17th to the 20th century

It is hard to tell the exact reasons why indefinites start being marked somewhere in the 18th century, since we do not have relevant data in our corpus. In other words, our data from the 19th century Bible capture the situation *after* the stipulated neutralization of some transitory distinction which may have facilitated the spreading of DOM to indefinites. In fact it is not only the transitional distinction we cannot make any statement about, but the conditions of *pe*-marking for indefinite direct objects in the 19th century Bible are also somewhat problematic, even though at first glance the data shown in Table 5 suggest that the triggering condition might have been specificity. Examples like (4), in which presumably non-specific direct objects get DOM, show that specificity is at least not a straightforward condition of DOM for indefinite direct objects in the 19th century. It seems that *pe*-marking was generally optional with argumental (i.e. non-incorporated) direct objects in the 19th century but more likely to occur with specific direct objects.

The regress of indefinite direct object marking from the 19th to the 20th century clearly shown by our data (58% → 26%) is an interesting phenomenon of diachronic micro-variation that may be useful for the understanding of the general mechanism of the diachronic evolution of DOM. The phenomenon has already been illustrated in (4), repeated here for convenience:

Matthew 8:16

- (4) 17th *adusără lui* *îndrăciți* *mulți...*
 brought he.dat demonised.pl many...
 17th *aduseră lui* *îndrăciți* *mulți...*
 brought he.dat demonised.pl many...
 19th *adus -au Lui* *pe* *mulți* *îndrăciți...*
 brought -have he.dat DOM many demonised.pl...
 20th *au adus la El* *mulți demonizați*
 have brought to he many demonised.pl
 ‘they brought to him many demon-possessed...’

Since we assume that the Bible translation D strongly relies on Bible translation C, the fact that *pe* has been left out in the same context in the younger translation is not a matter of contingency. Moreover, based on speakers’ intuition and the relevant literature (e.g. Farkas 1978, Dobrovie-Sorin 1994, Bende-Farkas 2002) one may argue that (4) would not yield the intended reading in present-day Romanian if *pe* was maintained, because the direct object is a non-specific indefinite NP.

The situation is however somewhat more complex, since there are some specific indefinite NPs marked in the 19th century with *pe* which are not being marked with *pe* anymore in the 20th century. This is the case in example (19), in which an unspecific reading of the direct object is not available, since the indefinite NP refers to a very specific individual, named Matthew:

Matthew 9:9

- (19) C19th *trecând Iisus de acolo, au văzut pre un om*
going Jesus from there, has seen DOM a man
șăzând la vamă, anume Mateiu.
sitting at toll, namely Matthew
D20th *plecând Iisus de acolo, a văzut un om*
going Jesus from there, has seen a man
care ședea la vamă, cu numele Matei.
who sat at toll, with name.DEF Matthew
'As Jesus went on from there, he saw a man named Matthew sitting at the tax collector's booth.'

Moreover, there are other examples in which something that could be informally dubbed a “referential competition” may very well explain the lost *pe* in the 20th century translation. Such a case is clearly illustrated in the contrast between (20) in which DOM survives in the 20th century and (21) in which DOM is eliminated in the 20th century:

Matthew 4:18

- (20) C19th *...au văzut pre doi frați...*
...has seen DOM two brothers
D20th *...a văzut pe doi frați...*
...has seen DOM two brothers
'...he has seen two brothers...'

Matthew 4:21

- (21) C19th *...au văzut pre alți doi frați...*
...has seen DOM other two brothers...
D20th *...a văzut alți doi frați,*
...has seen other two brothers...
'...he has seen two other brothers...'

While in both examples (20) and (21) the exact names of the brothers are given in the sentence immediately following, there is a difference between the two examples with regard to familiarity. This goes back to the operator “other”, which is anaphoric and expresses a certain link to the context itself and makes *pe*-marking of the indefinite superfluous.

The fact that the regress in *pe*-marking cannot be directly related to the notion of specificity does not contradict the mainstream view that *pe* disambiguates between specific and non-specific readings of indefinites in present-day Romanian. In fact we shall argue that it is not specificity itself but the fine structure of specificity that is relevant for the regress in DOM for indefinites from the 19th century to present-day Romanian.

As shown by our examples above, it is not trivial to give an explanation to the regress in DOM of Romanian indefinites and such an explanation will be the main task of the remainder of this paper. While the diachronic data from our Bible translations have sufficed to show this diachronic regress, our corpus does not deliver enough data to explain or model the phenomenon. Therefore, in the next section we present more synchronic data that exhibit an interesting variation which we trace back to the diachronic development of DOM and the evolution of clitic doubling.

3.4. More synchronic data

Our diachronic corpus provides important data for the main lines of the evolution of DOM but it is far from being representative for present-day Romanian. On the one hand, the amount of data processed and the thematic constraints on biblical texts are not adequate to

make strong generalizations about subtle semantic differences, and on the other hand, biblical texts tend to be linguistically conservative thus reflecting the state of a language that is already conceived as archaic: our 1994 Bible translation reflects probably the state of the language from the first half of the 20th century, or even of the second half of the 19th century. Moreover, since judgments about *pe*-marked indefinites in present-day Romanian are very subtle and may differ to some extent for different speakers, synchronic corpus data are important because they deliver examples already embedded in specific referential contexts.

By introducing more synchronic data into our analysis we can both check on the predictions our diachronic corpus allows about DOM for indefinites in present-day Romanian and introduce new data into our analysis:

According to our expectations the observation that 26% of the indefinites are *pe*-marked in the 1994 Bible edition predicts that somewhat less or the same percentage of the indefinite NPs are being marked in present-day Romanian, since the trend may have continued in the last few decades, while we do not see any reason for the trend to have changed.

Accordingly, we can compare the prediction of our corpus analysis with a simple test of present-day Romanian based on a *Google* search. Of course, *Google* is not necessarily a representative corpus, however it has some important advantages: it is easily accessible, it reflects mostly relatively new texts and it can be processed electronically, which makes working with data much easier than with our manually processed analogue data.

Our *Google* query is simple yet effective in dealing with a very specific question. We took five typical transitive verbs generally selecting human direct objects: *a ajuta* ('to help'), *a întreba* ('to ask'), *a suna* ('to phone'), *a chema* ('to call') and *a bate* ('to beat up') and checked for the frequency of *pe*-marked and non-*pe*-marked indefinite direct objects they occur with. To exclude any false results we only searched for past tense instances with auxiliaries agreeing with 1.sg/pl and 3.pl subjects and masculine singular direct objects. This way we could exclude the possibility that the unmarked direct object was in fact the subject of the sentence with an unusual (OVS) word order. We shall illustrate the search method with the example of the verb *to ask*:

- (22) a. i. *Am întrebat pe un...*
 Have.1.SG/PL ask.PERF DOM INDEF.ART.MASC
 'I/We have asked a...'
 ii. *Am întrebat un...*
 Have.1.SG/PL ask.PERF INDEF.ART.MASC
 'I/We have asked a...'
 b. i. *Au întrebat pe un...*
 Have.3.PL ask.PERF DOM INDEF.ART.MASC
 'They have asked a...'
 ii. *Au întrebat un...*
 Have.3.PL ask.PERF INDEF.ART.MASC
 'They have asked a...'

In (22a.i) and (22b.i) the direct object is marked with *pe*, thus a confusion with the subject is excluded, whereas in (22a.ii) (22b.ii) the direct object is unmarked and could thus be mistaken for the subject. However, since in Romanian verbs agree with the subject in number and person, in our ii examples the subject cannot be confused with the direct object, since the direct object is in third person singular, while the subject is in first person singular/plural and third person plural respectively. Accordingly, we only searched for instances like in (22), where the subject agrees with the verb in another way than the direct object would (if it were a subject) and ignored instances with third person subjects. Note also that we only searched for masculine direct objects avoiding the problem with the clitic placement, since the feminine object-clitic *-o* is after the verb, while the masculine object-clitic *-l-* is in front of the aux-

iliary. We also eliminated phraseologies and some older texts since our target was present-day Romanian.

Our findings show (as illustrated in Table 10) that our prediction holds quite well for modern Romanian: slightly less human indefinite direct objects are being marked than in the 1994 Bible translation (and definitely less than in our 19th century translation):

VP	<i>pe</i> -marked	not marked	Total	Percentage of marking
to call 1	6	128	132	5%
to call 3pl	3	106	109	3%
to phone 1	29	72	101	29%
to phone 3pl	1	3	4	25%
to ask 1	131	274	405	32%
to ask 3pl	5	12	17	29%
to beat up 1	4	19	23	17%
to beat up 3pl	11	66	77	14%
to help 1	13	81	94	14%
to help 3pl	5	61	66	8%
Total	208	822	1028	20%

Table 10 *pe*-marking in present-day Romanian for human indefinite NPs

The additional data show that different verbs have a very different affinity to the marking of their indefinite direct objects. This shows, as expected according to von Heusinger and Kaiser (2007) for Spanish, that verbal semantics have some impact on the distribution of DOM in Romanian. Another interesting observation, that has to our knowledge not been investigated so far, is that first person VPs generally have a higher rate of direct object marking than third person VPs. Both observations are compatible with the assumption that the fine structure of specificity is relevant for the distribution of marked indefinite direct objects, since different verbs have a different affinity to specific objects and generally first person subjects have a higher chance of having specific objects, since first person subjects are at the same time local and speaker anchors for indefinites. We shall come back to this distinction in section 4.

3.5. The evolution of clitic doubling

The final set of data required for the analysis of the regress in marked indefinite direct objects is an overview of the evolution of clitic doubling. We have already seen above that clitic doubling leads to a significant regress in strong pronominal direct objects but we did not yet analyze the diachronic data about the distribution of clitic doubling with different expression types.

The table below presents the historic evolution of clitic doubling with strong pronouns, definite and indefinite NPs over the analyzed centuries based on our data. For present-day Romanian we only give impressionistic estimations:

	[- clitic] [+ str. pron.]	[+clitic] [+str. pron.]	[+clitic] [- str. pron.]	[- clitic] def. NP	[+clitic] def. NP	[- clitic] indef. NP	[+clitic] indef. NP
17 th	42%	33%	24%	99%	1%	100%	0%
18–19 th	14%	63%	22%	99%	1%	100%	0%
19–20 th	1%	23%	75%	preferred	optional	preferred	marginal
present	ungrammatical	few	predominant	few	predominant	optional	optional

Table 11 Clitic doubling and DOM in diachronic stages along the definiteness scale

This overview shows clearly that clitic doubling evolves along the same lines as DOM, however with a significant delay. Accordingly, while until the 18th or the 19th century clitic doubling was only sharing the pronominal domain with DOM, for some reason clitic doubling

spreads quickly after the 19th century into the domain of descriptive nouns. Our data from the 20th century Bible show in fact archaic data with respect to clitic doubling and while they make the right prediction about DOM they fail completely on clitic doubling in the domain of full nouns. The reason for this discrepancy is that in religious texts for a matter of style clitic doubling, except for pronominal direct objects, is strongly not preferred. An impressionistic lecture of different texts from the 19th century and the first half of the 20th century suggests that at this time clitic doubling with definite NPs was possible and even preferred by certain authors, while rarely used by others. But it is only in the second half of the 20th century that clitic doubling with definite direct objects becomes predominant – yet not completely obligatory. Moreover, the evolution of clitic doubling with indefinite NPs starts obviously at the very end of the 19th century and slowly increases, but only with *pe*-marked indefinite direct objects. Thus in the 20th century indefinite direct objects can occur in three distinct structures:

Type 1 +[pe] +[cl-doubling]	<i>L- am întrebat pe un băiat.</i> CL.3.SG have asked DOM a boy 'I have asked a boy'
Type 2 +[pe] –[cl-doubling]	<i>Am întrebat pe un băiat.</i> Have asked DOM a boy 'I have asked a boy'
Type 3 –[pe] –[cl-doubling]	<i>Am întrebat un băiat.</i> Have asked a boy 'I have asked a boy'

Table 12 Combination of functional elements with indefinite DOs

Our brief *Google* query has revealed the following distribution of the three syntactic structures involving indefinite direct objects:

verb/subject	Type 1 [+pe] [+ clitic doubling]	Type 2 [+pe] [– clitic doubling]	Type 3 [–pe] [– clitic doubling]
to call 1	5%	0%	95%
to call 3pl	1%	2%	97%
to phone 1	27%	2%	71%
to phone 3pl	13%	12%	75%
to ask 1	27%	5%	68%
to ask 3pl	17%	12%	71%
to beat up 1	17%	0%	83%
to beat up 3pl	11%	3%	86%
to help 1	10%	4%	86%
to help 3pl	2%	6%	94%
Total	13%	5%	83%

Table 13 Distribution of the three types of indefinite direct objects.

The data show that Type 1 indefinites are much more common than Type 2 indefinites, while both are much less common than Type 3 indefinites. Moreover, Type 1 indefinites are very sensitive to the distinction between first and third person subjects, which on the other hand implies that Type 2 indefinites are similarly sensitive to this distinction. Accordingly, Type 1 indefinite direct objects are more likely to occur with first person subjects while Type 2 direct objects are more likely to occur with third person subjects. We shall explore the exact implication of these correlations in the following section.

Before we proceed to the analysis of this correlation some remarks are necessary here: Many speakers of present-day Romanian, quite independently of their geographic prove-

nience, reject *pe*-marking of indefinite direct objects in singular, while they do admit *pe*-marking in certain quantificational contexts. Moreover, most of the speakers who do admit and actually use *pe*-marking with (some) indefinite direct objects, categorically reject Type 2 structures. The general judgment seems to be that Type 2 structures are simply archaic and not acceptable anymore. Accordingly, these speakers – who can be considered the majority – are not likely to perceive semantic differences between Type 1 and Type 2 structures, since they would use Type 1 in any context in which Type 2 occurs in our corpus. Moreover, we hardly found any speakers who would clearly distinguish between Type 1 and Type 2.

The distributional peculiarities observed in Table 13 may be, hence, either due to dialectal or to extremely subtle semantic differences to be investigated in the future, and for a synchronic analysis of DOM in Romanian they may not be very important. However, for a diachronic study they are crucial, since they not only show that there is a nearly finished language change process with regard to DOM in contemporary Romanian, but also that this diachronic change depends on semantic factors: according to our data Type 2 structures are more likely to occur with specific verbs and they are generally more frequent with third person subjects. In fact, speakers who categorically reject Type 2 structures still admit that there is a significant difference in acceptability between verbs such as ‘to ask’ and verbs like ‘to see’, such that Type 2 structures are much worse if the verb is the latter. Accordingly, we shall use the data from present-day Romanian to construct a hypothesis about the semantic differentiation that may have caused the regress in DOM observed from the 19th to the 20th century.

4. The fine structure of specificity

In the literature on DOM in Romanian, specificity is considered as one of the major factors that synchronically triggers DOM for animate indefinite direct objects. Thus in most papers the following observation is made and considered to be of major importance: a non-marked indefinite direct object is ambiguous between a specific and non-specific reading as shown in (23a), while a marked direct object only has a specific reading as shown in (23b):

- (23) a. *Caut un elev.*
 Search-1.SG a pupil
 specific: ‘I am looking for a specific pupil.’
 non-specific: ‘I am looking for some pupil.’
 b. *Îl caut pe un elev.*
 CL.3.SG search-1.SG DOM a pupil.
 specific: ‘I am looking for a specific pupil.’

This observation captures the intuition of most speakers, but it is an important question what exactly is meant by *specificity*.

The concept of specificity was initially introduced to transfer the *de re* – *de dicto* distinction between definite NPs under verbs of propositional attitudes, as in (24), to indefinite NPs, as in (25):

- (24) a. *Joan wants to marry the winner* – but he doesn’t want to see her. [de re]
 b. *Joan wants to marry the winner* – so she needs wait for the race to end. [de dicto]
 (25) a. *Peter intends to kiss a French girl* – even though she doesn’t love him. [specific]
 b. *Peter intends to kiss a French girl* – though he hasn’t met one yet. [non-specific]

The Romanian equivalents of these examples suggest, as noted above, that *pe* can mark a specific reading as in (26a), while the lack of *pe* leaves the sentence ambiguous:

- (26) a. *Petru vrea să o sărute pe o franțuzoaică*
 Peter wants to CL.3.SG kiss DOM a French-woman
 Specific: ‘Peter wants to kiss a certain French woman.’
- b. *Petru vrea să sărute o franțuzoaică.*
 Peter wants to kiss a French-woman
 Specific: ‘Peter wants to kiss a certain French woman.’
 Non-Specific: ‘Peter wants to kiss some French woman.’

The notion of specificity covers a number of distinct phenomena: (i) scopal specificity, (ii) epistemic specificity, (iii) partitive specificity, and (iv) relative specificity (see Farkas (1995) for (i)-(iii) and von Heusinger (2002) for (iv), which in Farkas and von Heusinger (2003) is called “anchored specificity”). It is unclear whether there is one single concept of specificity with these four subclasses or whether these are different, though related, concepts.

With regard to the **scope interaction** between an indefinite and some other operator, such as a verb of propositional attitude, negation or a universal quantifier, one can distinguish between indefinites taking on a wide scope which are also called scopally specific, and indefinites taking a narrow scope which are regarded as scopally non-specific, as illustrated in (27), where both a specific and a non-specific reading are possible:

- (27) *Bill visits a museum every day.*
 specific reading: There is a certain museum which Bill visits everyday.
 non-specific reading: Every day, Bill visits some museum or another.

While a *pe*-marked Romanian correspondent of (27) as in (28a) can only have a specific reading, the translation without *pe* is ambiguous between the specific and non-specific readings as shown in (28b):

- (28) a. *Bill o vizitează pe o fată în fiecare zi.*¹⁰
 Bill CL.3.SG visits DOM a girl in every day
 Specific reading: ‘There is a certain girl who Bill visits every day.’
- b. *Bill vizitează o fată în fiecare zi.*
 Bill visits a girl in every day
 Specific reading: ‘There is a certain girl who Bill visits every day.’
 Non-specific reading: ‘Every day, Bill visits some girl or another.’

Dobrovie-Sorin (1994) argues that specific NPs are referentially closed expressions that cannot be affected by external quantifiers. Thus, instead of the widespread hypothesis, that specificity (at least concerning scopal specificity) involves a wide-scope reading, she proposes that specific indefinites are merely not-narrow-scoped. Her argument is that if specific indefinites had a wide scope, then they would be subject to quantifier raising (QR), which however is blocked by clitic doubling (and also by *pe*). In effect, her analysis yields logically equivalent semantic representations concerning the phenomena treated in this paper.

Examples that show the same (intuitive) contrast but do not contain operators as in (29) fall under the category of **epistemic specificity**. In (29a) and (29b) possible continuations on the specific vs. non-specific readings are given. This contrast is also often described as referential vs. non-referential.

- (29) A student in Syntax 1 cheated on the exam. (Fodor and Sag 1982)
- | | |
|--|----------------|
| a. His name is John. | [specific] |
| b. We are all trying to figure out who it was. | [non-specific] |

¹⁰ Since non-human direct objects are usually not DOM-marked, we changed the example to *Bill visits a girl every day*.

A modified Romanian version of this example with an indefinite NP as direct object shows that *pe* can disambiguate an epistemic specific reading as well:

- (30) a. *L -au arestat pe un băiat.*
 CL.3.SG have arrested DOM a boy
 specific reading: ‘They arrested a boy.’ (Bill).
 b. *Au arestat un băiat.*
 have arrested a boy
 specific reading: ‘They arrested a boy.’ (Bill)
 non-specific reading: ‘They arrested a boy.’ (We don’t know yet who it was).

The notion of **partitive specificity** goes back to the distinction of Milsark (1974) according to which indefinite NPs can either receive a weak (existential) interpretation as in (31a) or a strong (presuppositional) interpretation as in (31b), which is generally called “partitive”:

- (31) a. There are some ghosts in this house.
 b. Some ghosts live in the pantry, others live in the kitchen.

Enç (1991) develops this idea of specificity as partitivity and argues, based on examples like (32), that the accusative marker in Turkish marks this type of specificity:

- (32) *Oda-m-a birkaç çocukgir-di*
 room-1.SG-DAT several child enter-PAST
 ‘Several children entered my room’
İki kız-ı tanı-yor-du-m
 two girl-ACC know-PROG-PAST-1.SG
 ‘I knew two girls.’ glosses by von Heusinger and Kornfilt (2004)

In the first sentence a set of children is introduced and the accusative marker in the second sentence indicates that the two girls are part of that set of children that was introduced before. Thus the expression ‘two girls’ with the case suffix is considered to be specific by virtue of partitivity in this sentence. The same continuation with the direct object without case suffix is understood to refer to two girls not included in the set of several children already introduced.

In Romanian, partitivity is mostly explicitly expressed by the numeral *un* suffixed by the definite article *-a* or *-l* in singular, or *-i* or *-le* in plural, and the partitive preposition *dintre/din* (‘of’). This type of determiners can project only partitive noun phrases, i.e. if the determiner is immediately followed by a noun N, N is not interpreted as denoting the set of N’s in the model, but an already given, non-empty subset of N (Comorovski 1995). If a partitive noun phrase is a direct object, *pe*-marking is obligatory if the direct object refers to some human referent:

- (33) *Am luat -o pe una din ele de mână,*
 have taken CL.3.SG DOM one.DEF from they of hand
și ne-am pierdut în mulțime.¹¹
 and CL.1.PL have lost in crowd.
 ‘I took one of them (girls previously mentioned) by the hand and we got lost in the crowd.’

The notion of referentially anchored or **relative specificity** refers to such indefinite NPs that neither have wide scope nor are directly referential, but are still specific in some sense. Higginbotham (1987: 64) illustrates this with a well-known example:

¹¹ <http://elianferghete.blogspot.com/2007/07/din-lumea-celor-ce-nu-cuvanta.html>

In typical cases specific uses are said to involve a referent that the speaker ‘has in mind.’ But this condition seems much too strong. Suppose my friend George says to me, ‘I met with a certain student of mine today.’ Then I can report the encounter to a third party by saying, ‘George said that he met with a certain student of his today,’ and the ‘specificity’ effect is felt, although I am in no position to say which student George met with.

In such cases we can also find *pe* in Romanian, though as generally with specificity (except for some cases of partitivity) it is not obligatory.

- (34) *Ion spune că l- a întâlnit pe un student*
 John says that CL.3.SG has met DOM a student
de-al lui, dar nu știu pe care.
 of-POSS.ART his, but not know DOM which.
 ‘John says that he met a student of his, but I don’t know which one.’

Hence, as shown by the examples above, every subclass of specificity can be marked with *pe* and in all of these cases *pe* successfully disambiguates the indefinite NP for the specific reading.

At this level, however, the notion of specificity is treated as a general label for different, partly overlapping phenomena, which are evidently related even though the nature of this relation is not very clear. In the next step of our inquiry we shall attempt to use the notion of **anchored specificity** presented in von Heusinger and Kornfilt (2004) and von Heusinger (2007) based on ideas of Enç (1991), which captures an essential aspect of specificity and allows for a finer differentiation, in order to capture possible semantic differences between different types of indefinite direct objects in the diachronic evolution resulting in the described regress in DOM for indefinites between the 19th and the 20th century.

The notion of anchored specificity is probably the definition of specificity that covers the widest range of cases. It also has the formal means to distinguish and to relate to each other several subtypes of specificity. In this paper we shall not repeat the formal approach in von Heusinger (2007) but constrain ourselves to an intuitive explanation of the different types of anchoring.

The general idea of anchoring specificity is that similar to functional definite NPs, specific indefinites are referentially anchored to other discourse referents. Referential anchoring means that there is a salient function (or relation) such that the referent introduced by the indefinite NP can be linked by means of this function (or relation) to some present discourse referent: this discourse referent is called the referential anchor. Possible anchors are individuals introduced into the discourse, the speaker or even non-atomic discourse referents. The contextually salient function linking the indefinite to an anchor reaches from ‘being able to identify x’ and ‘being in the same location with x at a certain time’ to ‘feeling pity for x’, etc. Note, that the discourse position of the anchor evidently determines the maximal scope of the indefinite, in other words, the anchor must be visible for the discourse referent introduced by the indefinite NP.

Accordingly, if an indefinite NP introduces for instance a student into the discourse as in (35), then one possible way of anchoring this student would be that the speaker, James, knows who the student is. In this case the student is referentially anchored to the speaker, i.e. there is a function such that the new discourse referent is the value of the function with the speaker as an argument. Eventually, however, it is not the speaker but some discourse referent, say George, who the speaker is talking about, who knows who the introduced student is. Accordingly, the student is referentially anchored to George:

- (35) James: “George met [a certain student of his]”
 reading (i): [a certain student of his] = f(James) anchored to the speaker
 reading (ii): [a certain student of his] = f(George) anchored to the subject

Different types of anchors have very different effects on the referential properties (and scope properties) of indefinites, which may hence differ with respect to their referential stability (see Farkas and von Stechow 2003). The types of specificity distinguished above can be modeled based on the referential anchors as follows:

Indefinites **referentially anchored on the speaker** are such that the referent of the indefinite is known to the speaker or identifiable through the speaker. This type of anchored specificity generally yields the widest possible scope, because the anchor is fixed outside of the range of any sentential operator. This also applies for scope-islands like conditionals. Generally we can consider referential anchoring to the speaker as a concept very similar to the epistemic specificity, however its range is somewhat more extended, since anchoring functions are not necessarily (although often) restricted to a function that would spell out something like “the speaker has in mind x”.

As explained above, following Enç (1991)’s proposal we accept that partitives are specific. **Partitive anchoring** can be viewed as a specific relation between the anchor and the referent of the indefinite such that the referent of the indefinite is part of the anchor-set. Note that partitive indefinites are anchored by means of a relation and not a function, since in this case there is no clear way of distinguishing the exact referents except by virtue of the fact that they are members of a set. It is not necessarily clear which members of the set are meant.

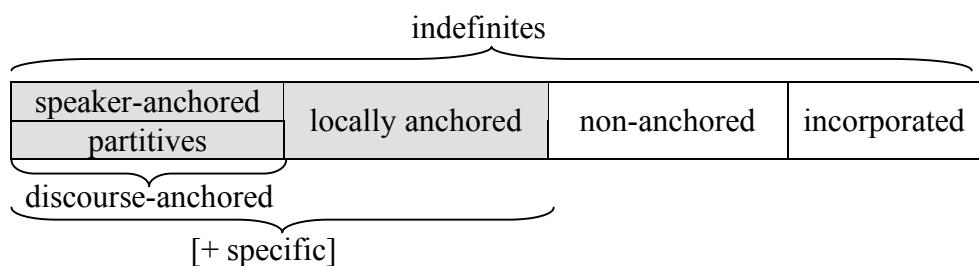
Indefinites referentially anchored on some referent of the current sentence can be called **locally anchored indefinites**. Relative specificity is a typical example for this subclass. In principle, local anchoring means that the indefinite cannot get wide scope over some operator that has scope over its anchor, but it is still specific as anchored to a given discourse referent. Locally anchored indefinites may have both narrow and intermediate scope.

Finally **non-anchored indefinites** are such that they generally take narrow scope and are merely existential. As compared to locally anchored indefinites, non-anchored indefinites are not specific: in intensional contexts and under quantifiers they take on narrow scope, while in neutral contexts they do not have any properties related to specificity. As opposed to **semantically incorporated** indefinites, however, they may act as antecedents for anaphoric pronouns.

The notion of specificity can be related to the notion of referential anchoring as follows: Any referentially anchored indefinite is specific. In this light, specificity turns out to be a fine-structured domain: depending on the anchor types, different kinds of specificity can be distinguished. We shall model the different types of specificity on a scale which we shall include in the definiteness scale as an extension for indefinites. The basic idea of this scale is that elements on the left have properties that make them more referential, while elements on the right are less referential, and therefore we expect phenomena like DOM to spread to indefinites along the specificity scale. The ordering of the scale is clear-cut between speaker-anchored, locally anchored and non-anchored indefinites and between partitives, locally anchored indefinites and non-anchored indefinites:

$$\left. \begin{array}{l} \text{speaker anchored indefinites} \\ \text{partitives} \end{array} \right\} < \text{locally anchored indefinites} < \text{non-anchored indefinites}$$

The ordering between speaker-anchored indefinites and partitives is not straightforward, as both of them show a high degree of specificity. While speaker-anchored indefinites can be considered as referential, partitives share the fact that they have a discourse antecedent with the definite NPs. For this reason we shall not order these subcategories at all and we extend a section of the definiteness scale such that we regard partitives and speaker-anchored indefinites as parallel categories. Speaker-anchored indefinites and partitives build the group of discourse-anchored indefinites together.



The theoretical expectation bound to this specificity scale is that different languages may cut this scale at different points (if at all). We do not expect languages to necessarily distinguish between all of these fine-structured types of specificity. They may choose different sections of different length for overt specification, however these sections are expected to start on the left. For example, Russian does make such a distinction: as shown in Haspelmath (1997) and implemented by Geist (2008), the indefinite system of Russian distinguishes between speaker-anchored and non-speaker-anchored indefinites as shown in the contrast between (36a) and (36b):

- (36) a. *Igor vstretil koe-kakuju studentku.*
 Igor met koe-wh' student
 Igor met a student. [speaker knows whom]
- b. *Igor vstretil kakuju-to studentku.*
 Igor met wh-to student
 Igor met a student. [speaker does not know whom]

5. DOM, clitic doubling and the fine structure of specificity

In this chapter we attempt to give an explanatory account for the observed regress of *pe*-marked indefinite direct objects between the 19th and the 20th century.

Based on the collected data the exact conditions of DOM for indefinite human direct objects in the 19th century cannot be given with certainty. In Table 14 we present the distribution of the indefinite direct objects and DOM in the 19th century in a more detailed perspective. This time generics, counted simply as non specific in Table 5, are counted separately:

19 th century indefinites DOs	specific	non-specific	generics
[+pe]	6	1	2
[-pe]	0	6	1
total	6	7	3

Table 14 Distribution of DOM for indefinite direct objects in the 19th century.

The data presented above show that DOM was preferred with anchored indefinite direct objects and strongly not preferred with non-anchored indefinites. Hereby characteristic generics are marked with *pe*, just like other anchored indefinites. Accordingly, only one non-specific indefinite direct object gets *pe*-marked (cf. (4)). Hence we can assume that the semantic import of DOM with regard to the referential properties of the direct object can be regarded as the marking of referential anchoring. Why exactly one non-specific direct object has been marked is not clear, however one may argue that example (4) gets *pe* marking for reasons that may be related to the quantifier *mulți* ('many'), or that in fact (4) shows that the process leading to differential marking of specific indefinites continues and already allows the marking of non-specific but still argumental indefinites as well. This development of DOM indicating the argumental status of the direct object has also taken place in Spanish (see Leonetti 2004).

The generic indefinites marked with *pe* in the 19th century, which can be observed similarly in present-day Romanian as well, constitute a problem that we will not discuss here in

detail. We only sketch a solution here that needs to be developed by further research. According to the main view (Krifka et al. 1995), there are two main types of generics: kind and characteristic generic uses. While kind generics tend to be expressed by definite NPs in Romanian, characteristic generics can be expressed by specific indefinites as well. It can be argued that characteristic generics involve a generic operator, and for Romanian we assume that generic operators have widest possible scope and have a high enough discourse position to grant wide scope for indefinites. Accordingly, generic indefinites need to be anchored and thus they may be marked with *pe*.

The diachronic fact obviously correlated with the regress in the marking of indefinite direct objects after the 19th century is the spreading of clitic doubling to the nominal domain. As our data have shown in the first half of the 19th century or the second half of the 18th century, clitic doubling with full noun phrases is unlikely, however in the second half of the 19th century clitic doubling becomes possible (even if not necessarily preferred) first with *pe*-marked definite, and after some probably minimal delay, for *pe*-marked indefinite direct objects as well. Again our data are much too restricted to account for the semantic conditions of clitic doubling in late 19th century Romanian. Some observations are, however, still possible:

Clitic doubling strictly follows the evolution of DOM. This evolution is only related to pronominal direct objects for quite a time: as shown in Table 6, it took more than 300 years for weak pronouns to systematically occur with strong pronouns and another century to systematically replace these. Obviously, the spreading of clitic doubling only starts towards the domain of full noun phrases after clitic doubling is obligatory with personal pronouns and there is the (preferred) possibility to replace them totally. From there on clitic doubling spreads to definite NPs. Even though we do not have enough data or other reliable information about the exact distribution of clitic doubling with definite NPs in the late 19th century, we hypothesize that the functional load of clitic doubling was to express familiarity. First, this corresponds to the bleached function of pronouns, and second, this function is assumed for clitic doubling in other languages, such as Greek. Alexiadou (2006) assumes that clitic doubling of definite direct object NPs in Greek marks familiarity, while the definite article marks uniqueness.¹² Leonetti (this volume) discusses the interaction of DOM and clitic doubling in Spanish. In section 3.4 he discusses the reinterpretation of the +definite feature of clitic pronouns as discourse anchored or in his own words: “The [+definite] feature in the clitic finds its corresponding feature in the indefinite DP because the indefinite determiner is assumed to operate on a contextually given set that is inferentially retrieved.” (Leonetti this volume).

The starting point for our detailed investigation of the interrelation between DOM and clitic doubling of indefinite direct objects in Romanian is that DOM develops into the domain of indefinite NPs, triggered by specificity and we might also assume that the development was to continue to non-specific indefinite NPs triggered by the argumental status of such NPs. Such an evolution has also taken place in Spanish (cf. Leonetti 2004). At this point clitic doubling is also possible for indefinite NPs in Romanian – here we assume that clitic doubling expresses a referential property and we hypothesize that this property must be closely related to familiarity (the relevant property for clitic doubling with definites). Since indefiniteness is not compatible with familiarity, the latter notion is reanalyzed as discourse linking, and we

¹² Alexiadou (2006, 3) follows Alexiadou and Anagnostopoulou (1997) and Anagnostopoulou (1999) and states “clitic doubling in Greek has a clear semantic contribution where only familiar definite NPs can be clitic doubled; in (8) [= (i)] below, the DP that is clitic-doubled involves reference to a particular entity in a particular context:”

(i) *tin ida* *ti gata*
her saw-1.SG the cat

Anagnostopoulou (1999, 783) also argues for different subclasses of specific indefinites, such as referential indefinites (our specific or anchored indefinites) and partitive indefinites (our discourse or speaker anchored indefinites), in order to explain the variation between the conditions for clitic doubling with indefinites in Greek, (Porteño) Spanish and Romanian

get the option that indefinite direct objects can appear in three different environments: Type 1: [+clitic doubling] [+pe]; Type 2: [-clitic doubling] [+pe]; and Type 3 [-clitic doubling] [-pe]. These types have already been illustrated in Table 12. The three types combine two distinct phenomena, namely clitic doubling and DOM, both having their own semantic import. While the semantic import of *pe* for indefinite direct objects seems to be that the direct object is referentially anchored, the semantic import of clitic doubling involves discourse linking (in this case: anchoring to the speaker or to a set in the discourse, i.e. partitivity). This yields a clear correlation between the syntactic realization of indefinite direct objects and the fine-structured specificity discussed above. In a final step, this instable three-way distinction is reduced or simplified to an opposition with a very close correlation of DOM and clitic doubling, leaving Type 2 as a very marginal option if at all. At the same time, the potential further development of DOM is stopped and first steps are taken backwards since it is so closely related to clitic doubling. This system is what we find in contemporary Romanian. We illustrate the evolution in Table 15.

Starting point: indefinite NPs are *pe*-marked, the triggering condition is \pm specificity and further development with the triggering condition \pm argumental

definite NPs	indefinite specific (= anchored) NPs		non specific indefinite NPs	incorporated nouns
+pe	+pe		-pe (\pm pe)	-pe

Additional referential marker: CL with functional load: familiarity; reanalysis to discourse anchoring, three way system (instable), spreading of *pe* into \pm argumental stops, regression to +anchored

definite NPs	indefinite specific (= anchored) NPs		non specific indefinite NPs	incorporated nouns
	discourse anchored	locally anchored		
+pe +CL	+pe +CL	+pe -CL	-pe -CL	-pe -CL

Stabilization and simplification of the three way system into a two way system (strong correlation between *pe* and CL), no further development of *pe* (still some synchronic reflexes of the old system)

definite NPs	indefinite specific (= anchored) NPs discourse anchored & locally anchored		non specific indefinite NPs	incorporated nouns
+pe+ CL	+pe+ CL	(+pe -CL)	-pe -CL	-pe -CL

Table 15 Transitional distinction and regression effect for indefinite NPs depending on CL doubling

If this hypothesis is correct, we expect Type 1 indefinite direct objects to be referentially anchored to the speaker or to some discourse antecedent set (partitives), since they inherit the semantic import of clitic doubling, which is familiarity. Type 2 indefinite direct objects are only marked for referential anchoring, which means that they can refer to speaker anchored indefinites, partitives and locally anchored indefinites. And finally Type 3 indefinite direct objects are not marked for anchoring, hence they may be anchored or not.

However this system has, upon a closer look, some very striking peculiarities which can be shown in the contrast between (37a) (37b) and (37c):

- (37) a. *Poliția l -a arestat pe un elev.*
 Police CL.3.SG has arrested DOM a student.
 The police arrested a student.
- b. *Poliția a arestat pe un elev.*
 Police has arrested DOM a student.
 The police arrested a student.
- c. *Poliția a arestat un elev.*
 Police has arrested a student.
 The police arrested a student.

The semantic analysis of (37a) yields that the direct object “a student” must be referentially anchored due to DOM and it must be anchored to the speaker or some discourse antecedent set (partitive), due to the presence of clitic doubling. In effect, the stronger condition must be met, thus we either have a partitive reading or the indefinite is anchored to the speaker. In (37c) nothing is being marked about the referential anchoring of the direct object, since unmarked indefinite direct objects may be both specific and non-specific. (37b), on the other hand, yields the information that the direct object is referentially anchored, however not necessarily to the speaker or a discourse antecedent.

Since (37b) does contain information about referential anchoring but does not explicitly say that the direct object is partitively anchored or anchored to the speaker, a standard implicature arises, that this is not the case, i.e. the implicature is that the referent is locally anchored, since otherwise the stronger form would have been used.

Under this analysis, the use of structures like (37b) is very restrained, namely for locally anchored specific indefinites. For locally anchored indefinites two alternatives are available, namely the lack of both DOM and clitic-doubling (Type 3) and the presence of DOM and the lack of clitic doubling (Type 2). Type 2 standardly implicates that a higher degree of specificity is excluded, while Type 3 does not implicate anything.

Hence, in neutral contexts Type 3 indefinite direct objects seem to be the unmarked and probably the preferred alternative. This way, the situation captured by our 19th century Bible seems to change: while *pe*-marked indefinites are preferred in the early 19th century (53%), the overlapping effect with clitic doubling results in the situation in which *pe*-marking is only preferred together with clitic doubling. This change of preference may explain the regress in differentially object marked indefinites, namely a regress in Type 2 indefinite direct objects.

This analysis stipulates a semantic distinction between Type 1 and Type 2 indefinite direct objects for the end of the 19th century and the first half of the 20th century. For most speakers of present-day Romanian this distinction is not accessible anymore, since they reject Type 2 indefinite direct objects generally. However, the synchronic data presented in Table 13, repeated here in a slightly different form as Table 16, suggests that some correlation still exists in present-day Romanian:

verb/subject	Type 1 [+pe] [+ clitic doubling]	Type 2 [+pe] [- clitic doubling]	Type 3 [-pe] [- clitic doubling]
to call 1	5%	0%	95%
to phone 1	27%	2%	71%
to ask 1	27%	5%	68%
to beat up 1	17%	0%	83%
to help 1	10%	4%	86%
Sum 1:	17%	2%	81%
to call 3pl	1%	2%	97%
to phone 3pl	13%	12%	75%
to ask 3pl	17%	12%	71%
to beat up 3pl	11%	3%	86%
to help 3pl	2%	6%	94%
Sum 3pl:	9%	7%	85%

Table 16 Distribution of the three types of indefinite direct objects depending on person/number of the subject.

The data above show very clearly that the use of Type 2 indefinite direct objects is not simply marginal but is highly related to different verb types and the person and number of the

clause subject. We can explain these correlations as traces of a nearly finished diachronic change process involving the semantic differentiation described above:

The surprising correlation between structures with first person subject vs. third person plural subject and the three different types of indefinite direct objects can be explained in terms of referential anchoring. Since generally subjects can be used as referential anchors, both specificity in general and more specifically the referential anchoring to the speaker are more likely to occur if the speaker is the sentence subject at the same time. Hence, if the subject is in the first person, Type 1 indefinite direct objects are more likely to occur (17% Type 1 vs. 2% Type 2). If however the sentence has a third person subject, referential anchoring on this subject yields local anchoring: in these cases Type 2 is more prominent (9% Type 1 vs. 7% Type 2). Moreover, the distribution of Type 2 indefinite direct objects is highly sensitive to verbal semantics. Generally speaking, if the event described by the verb is such that the agent and the patient are strongly related, such as ‘to beat up’ or ‘to phone’, Type 2 is nearly excluded with first person subjects, since in these cases it does not make sense to anchor the patient to something other than the agent/speaker. For communication verbs, on the other hand, Type 2 direct objects are likely if the subject is in the third person. A detailed analysis of these dependencies must, however, remain subject to future investigations.

According to our analysis the diachronic regress in DOM with indefinite direct objects can be explained as a pragmatically triggered preference shift. In a first step, local anchoring and speaker anchoring/partitivity are distinguished, such that the former category is marked by *pe* alone while the latter one is additionally marked by clitic doubling. This distinction leads however to a significant regress in DOM without clitic doubling due to an implicature. Moreover, the variant without any direct object marking at all gets preferred for locally anchored indefinites. This leads to a general regress in DOM.

However, in the past decades Type 2 structures are being perceived as archaic and systematically are tending to disappear. Hence a neutralization process is taking place again: in present-day Romanian all specific indefinites can be marked with the object marker accompanied by clitic doubling, but there still seem to be statistical correlations showing some stronger preference for clitic doubling if the direct object is higher on the proposed specificity scale. Accordingly, the semantic differentiation that led the way for the spreading of clitic doubling to indefinite direct objects is to a certain extent neutralized: clitic doubling does not have more restrained conditions than DOM in present-day Romanian. This, however, has not (yet) led to the situation observed in the 19th century, because due to the overlapping effect with clitic doubling a preference shift has taken place in the 20th century. Today, clitic doubling and *pe*-marking for indefinites are strongly correlated but occur less frequently than in the 19th century.

6. Summary and open questions

In this paper we have provided evidence for some theoretical assumptions about the nature of the diachronic development of DOM in Romance and argued that subtle semantic distinctions involving the fine structure of specificity are relevant for the diachronic development of DOM in Romanian.

In the first line we have shown that the diachronic evolution of DOM in Romanian can be captured as a gradual spreading from higher to lower referential expressions. The evolution involves transitory steps consisting in a finer semantic differentiation, thus initially only a certain part of the category DOM spreads to receive *pe*. As topicality and specificity have been shown to be such relevant transitory distinctions for definite and indefinite direct objects specifically in Spanish (cf. von Heusinger and Kaiser 2005), in Romanian the distinction between anaphoric and non-anaphoric definite expressions seems to have been the trigger for the differential marking of definite direct objects and specificity for the marking of indefinites.

However, a finer structured semantic differentiation may as well block and reverse the spreading of DOM. Such was the case in 19th century Romanian, as clitic doubling overlapped with DOM. Clitic doubling, being a later development for nominal direct objects and having obviously stronger, i.e. more restrained conditions than DOM, combines with DOM and leads to a re-interpretation of the semantic import of the latter in the form of a pragmatically triggered process. The pragmatic trigger assumed here is a standard implicature.

In the last step this semantic differentiation triggered by the spreading of clitic doubling has been widely abandoned such that clitic doubling mostly co-occurs with post-verbal DOM-marked direct objects, however at the point being, DOM still did not reach the level of the 19th century Bible translation. This, of course, may change in the following decades.

There are, however, important open questions: in the first line, a systematic analysis of the conditions of DOM and clitic doubling with indefinites in synchronic Romanian should take place, especially taking into account the effect of verb semantics onto the distribution of DOM. Further, the system briefly developed here should be extended to deal with generic uses, which are possible in Romanian with *pe*-marked indefinites and clitic doubling, and our results should also interface with current analyses of DOM with definite NPs in Romanian (eg. Dobrovie Sorin 2007), which we did not discuss at all in this paper. Finally, if the analysis presented here is correct, we would expect a general cross linguistic relevance of the sketched pragmatic mechanism. We can only leave this as a subject of further research for the time being.

7. References

- Aissen, Judith (2003). Differential Object Marking: Iconicity vs. Economy. *Natural Language and Linguistic Theory* 21: 435-483.
- Alexiadou, Artemis (2006). *On the cross-linguistic distribution of (in)definiteness spreading*. Handout ÖLT Syntax Workshop, Universität Klagenfurt, 8-9.12.2006. URL: <http://ifla.uni-stuttgart.de/~artemis/publications.html>
- Alexiadou, Artemis and Elena Anagnostopoulou (1997). Toward a Uniform Account of Scrambling and Clitic Doubling. In *German: Syntactic Problems, Problematic Syntax?*, Werner Abraham and Elly van Gelderen (eds.), 143-161. Tübingen: Max Niemeyer.
- Anagnostopoulou, Elena (1999). Conditions on Clitic Doubling in Greek. In *Clitics in the Languages of Europe*, Henk van Riemsdijk (ed.), 761-798. Berlin: Mouton De Gruyter.
- Bossong, Georg (1985). *Empirische Universalienforschung. Differentielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen: Narr.
- Bende Farkas, Ágnes (2002). *Specific Indefinites in Romanian and the Marker PE*. Paper presented at "Semantische und syntaktische Aspekte von Spezifität in Romanischen Sprachen", University of Konstanz.
- Chiriacescu, Sofiana (2007). *Pe-Markierung und Diskurs-Prominenz im Rumänischen*. Magisterarbeit. Universität Stuttgart.
- Comorovski, Ileana (1995): On Quantifier Strength and Partitive Noun Phrases. In *Quantification in Natural Languages*, Emmon Bach, Eloise Jelinek, Angelika Kratzer and Barbara H. Partee, eds., 145-177, Dordrecht: Kluwer.
- Comrie, Bernard (1975). Definite and Animate Direct Objects: A Natural Class. *Linguistica Silesiana* 3: 13-21.
- Croft, William (2003). *Typology and universals*. Cambridge: Cambridge University Press.
- Delille, Karl Heinz (1970). *Die geschichtliche Entwicklung des präpositionalen Akkusativs im Portugiesischen*. Bonn: Romanisches Seminar.
- Dobrovie-Sorin, Carmen (1994). *The Syntax of Romanian*. Berlin: Mouton de Gruyter.

- Dobrovie-Sorin, Carmen (2007). Article-drop in Romanian and extended heads. In *Pitar Mos: A Building with a View*, Gabriela Alboiu et. al. (eds.), 99-107. Bucharest: Bucharest University Press.
- Enç, Mürvet (1991). The Semantics of Specificity. *Linguistic Inquiry* 22: 1-25.
- Farkas, Donka (1978). Direct and indirect object reduplication in Rumanian. In *Papers from the Seventeenth Regional Meeting of the Chicago Linguistics Society* (CLS 14), 88-97. University of Chicago, Chicago.
- Farkas, Donka (2002). Specificity Distinction. *Journal of Semantics* 19: 213-243
- Farkas, Donka and Klaus von Heusinger (2003). *Stability of reference and object marking in Romanian*. Ms., Universität Stuttgart, (Paper given at the XV European Summerschool in Logic, Language and Information (ESSLLI) in Vienna).
- Fodor, Janet Dean and Ivan A. Sag (1982). Referential and quantificational indefinites. In: *Linguistics and Philosophy* 5: 355-398.
- Geist, Ljudmila (2008). Specificity as Referential Anchoring: Evidence from Russian Indefinite Pronouns. In *Proceedings of SuB 12*. Digital Publications at the University of Oslo and the Faculty of Humanities. (to appear)
- Grice, H. Paul (1989). Logic and Conversation. In *Studies in the Way of Words*, Paul Grice. Cambridge.
- Haspelmath, Martin (1997). *Indefinite Pronouns*. Oxford: Oxford University Press.
- von Heusinger, Klaus (2002). Specificity and Definiteness in Sentence and Discourse Structure. *Journal of Semantics* 19: 245-274.
- von Heusinger, Klaus (2007). Referentially anchored indefinite NPs. In *Existence: Syntax and Semantics*, Ileana Comorovski and Klaus von Heusinger (eds.), 273-292. Dordrecht: Springer.
- von Heusinger, Klaus (2008). *Verbal Semantics and the Diachronic Development of Differential Object Marking in Spanish* (in this volume).
- von Heusinger, Klaus and Georg A. Kaiser (2005). The evolution of Differential Object Marking in Spanish. In *Proceedings of the Workshop "Specificity and the Evolution/Emergence of Nominal Determination Systems in Romance"*, Klaus von Heusinger, Georg A. Kaiser and Elisabeth Stark (eds.), 33-69. Universität Konstanz: Fachbereich Sprachwissenschaft (Arbeitspapier 119).
- von Heusinger, Klaus and Georg A. Kaiser (2007). Differential object marking and verb classes in Spanish. In *Proceedings of the Workshop "Definiteness, Specificity and Animacy in Ibero-Romance Languages."*, Georg A. Kaiser and Manuel Leonetti (eds.), 83-109. Universität Konstanz: Fachbereich Sprachwissenschaft (Arbeitspapier 122).
- von Heusinger, Klaus and Jaklin Kornfilt (2005). The case of the direct object in Turkish: Semantics, Syntax and Morphology. *Turkic Languages* 9: 3-44.
- Higginbotham, James (1987). Indefinites and Predication. In *The Representation of (In)definiteness*, Eric Reuland and G. B. Alice ter Meulen (eds.), 43-70. Cambridge/Mass.: MIT Press.
- Hopper, Paul J. and Sandra A. Thompson (1980). Transitivity in grammar and discourse. *Language* 56: 251-299.
- Kamp, Hans and Agnes Bende Farkas (2006). *Specific Indefinites: Anchors and Functional Reading*. Ms. Universität Stuttgart.
- Krifka, Manfred et al. (1995). Genericity: An Introduction. In *The Generic Book*, Greg N. Carlson and Francis J. Pelletier (eds.), 1-124. Chicago: University of Chicago Press.
- Laca, Brenda (2002). Gramaticalización y variabilidad – propiedades inherentes y factores contextuales en la evolución del acusativo preposicional en español. In *Sprachgeschichte als Varietätengeschichte: Beiträge zur diachronen Varietätenlinguistik des Spanischen und anderer romanischer Sprachen. Festschrift für Jens Lüdtke zum 60. Geburtstag*, Andreas Wesch. (ed.), 195-303. Tübingen: Stauffenburg.

- Laca, Brenda (2006). El objeto directo. La marcación preposicional. In *Sintaxis histórica del español. Vol. 1: La frase verbal*, C. Company (ed.), 423-475. Mexico: Fondo de cultura económica / Universidad Nacional Autónoma de México.
- Leonetti, Manuel (2004). Specificity and differential object marking in Spanish. *Catalan Journal of Linguistics* 3: 75-114.
- Leonetti, Manuel (2008). *Specificity in Clitic Doubling and in Differential Object Marking* (in this volume).
- Lyons, Christopher (1999). *Definiteness*. (Cambridge Textbooks in Linguistics) Cambridge: Cambridge University Press.
- Mardale, Alexandru (2007). *Bare Plurals and Differential Object Marking in Romanian and Spanish*, Paper presented at the Workshop DOM in Romance, University of Stuttgart.
- Melis, Chantal (1995). El objeto directo personal en El Cantar de Mio Cid. Estudio sintáctico-pragmático. In *El complemento directo preposicional*, Carmen Pensado (ed), 133-163. Madrid: Visor.
- Milsark, Gary (1974). *Existential Sentences in English*. Ph.D. Diss. MIT, Cambridge/Mass.
- Popescu, Alexandra (1997). *Objektklitika und Argumentlinking im Rumänischen*. Magisterarbeit. Universität Düsseldorf.
- Schäfer-Prieß, Barbara. (2002). "O acusativo preposicional na história da língua portuguesa." In *História da Língua e História da Gramática. Actas do Encontro*, Brian Head, José Teixeira, Aida Sampaio Lemos et alii (eds.), 405-419. Braga: Universidade do Minho.