

Romanian Object Clitics: Grammaticalization, agreement and lexical splits

Roxana-Maria Barbu
Carleton University

Ida Toivonen
Carleton University

Proceedings of the LFG'18 Conference

University of Vienna

Miriam Butt, Tracy Holloway King (Editors)

2018

CSLI Publications

pages 67–87

<http://csli-publications.stanford.edu/LFG/2018>

Keywords: Romanian, clitics, agreement marking, pro-drop

Barbu, Roxana-Maria, & Toivonen, Ida. 2018. Romanian Object Clitics: Grammaticalization, agreement and lexical splits. In Butt, Miriam, & King, Tracy Holloway (Eds.), *Proceedings of the LFG'18 Conference, University of Vienna*, 67–87. Stanford, CA: CSLI Publications.



Abstract

Direct object clitics in Modern Standard Romanian display different properties depending on whether or not they double an object. We propose a dual analysis for the clitics: they function as agreement makers when they double an object and as pronouns when they do not. Furthermore, the lexical entries differ beyond the presence or absence of pronominal referential features, and this accounts for the split behavior. The analysis is placed in its historical context and extended to other varieties of Romanian. Finally, we argue that the Romanian lexical split is not an isolated phenomenon: multiple similar splits can be found in the typology of agreement marking.

1 Introduction

Romanian object clitics can occur with or without an object double.¹ This is illustrated with the third person masculine clitic *-l-* in (1–3):²

- (1) L-am văzut.
3SG.M.ACC-have.1SG seen
'I saw him/it.'
- (2) L-am văzut pe el.
3SG.M.ACC-have.1SG seen ACC him
'I saw him.'
- (3) L-am văzut pe băiat.
3SG.M.ACC-have.1SG seen ACC boy
'I saw the boy.'

Example (1) marks the object with a clitic only; there is no independent object nominal. In (2), the clitic doubles a pronominal object, and in (3), it doubles a non-pronominal noun.

This phenomenon is often referred to as “*pro*-drop” in the literature: the overt pronoun is “dropped” or phonologically unrealized. We will make use of the traditional term *pro*-drop, but we use it in a theory-neutral way to refer to the basic data pattern, and not as a term that implies that something has actually been dropped.

The analysis of Romanian object clitics presented in this paper builds on the standard LFG analysis of *pro*-drop, as spelled out in Fassi Fehri (1984); Bresnan & Mchombo (1987); Bresnan et al. (2016, Chapter 8), and elsewhere. In line with previous LFG analyses, we propose that the Romanian clitics have a dual nature: they are ambiguous between agreement markers and referential pronouns.

¹We want to thank the audience at LFG18 for insightful questions and comments. We would also like to thank Raj Singh for his feedback. We are very grateful to Aurelia Barbu, Octavian Barbu, and Lenuța Focșa for discussing the Romanian data with us and providing acceptability judgements. This paper has been greatly improved thanks to comments from the editors and an anonymous reviewer.

²The following abbreviations are used in this paper: SG = singular, PL = plural, M = masculine, F = feminine, ACC = accusative, PTCPL = participle, OBJ = object, NUM = number, GEND = gender, PERS = person, DEF = definite, PST = past, and IMP = imperative.

In examples such as (1), where the clitic is not accompanied by an independent object, the clitic is a pronoun. In (2) and (3), the clitic is an agreement marker.

A uniform analysis where the clitic is consistently either an agreement marker or a pronoun may seem like a better analysis *a priori*, but there are empirical arguments against this position. The argumentation builds on the fact that the agreement-marking clitic is more restricted in its distribution than the pronominal clitic. The clitics thus do not only differ in pronominal status. In LFG terms, they differ beyond the presence or absence of the [PRED ‘pro’] feature.

The proposed analysis of Romanian clitics will be placed in a historical context. The “lexical split” in Romanian clitics might seem like an unusual quirk, but it in fact follows naturally from commonly assumed grammaticalization processes that such a split would occur. We will also consider dialectal variation within Romanian. The paper finally mentions a number of examples from a variety of languages that illustrate that the Romanian clitic system is not so exotic after all.

2 Object clitics in Modern Standard Romanian

2.1 A brief introduction to the clitics

The object clitic forms in Modern Standard Romanian (MSR)³ are given in (4):

(4) Direct object clitic pronouns

	SG	PL
1	<i>mă/m</i>	<i>ne</i>
2	<i>te</i>	<i>vă/v</i>
3M	<i>îl/l</i>	<i>îi/i</i>
3F	<i>o</i>	<i>le</i>

The morphophonological status of Romanian clitics is controversial (Dobrovie-Sorin, 1994; Monachesi, 1998; Popescu, 2000; Luís, 2004): Are they clitics or bound morphemes? This paper will follow the majority view and treat them as clitics (i.e., non-projecting, phonologically dependent words) and not bound morphemes. However, nothing in our analysis hinges on this decision. Since the LFG architecture allows for ‘mismatches’ between levels of grammatical structure, the morphophonological status of the clitics does not dictate whether they are agreement markers or pronouns. We return to this later; for now it suffices to note that our analysis can be translated into one that treats the clitics as bound morphemes.⁴

The phrase-structural realization of the clitics is not central to our analysis, but we outline our assumptions here for concreteness. The \hat{D} is a non-projecting

³There is variation in Romanian regarding the distribution of object clitics. Some of that variation will be discussed in Section 5. The dialect we present in this section and the next is widespread and consistent with most grammar books. Still, MSR might not be the best label to use, and we do not mean for it to have any special significance.

⁴In fact, we expect that a careful analysis of different varieties of Romanian might reveal that the morphemes are phonologically bound in some dialects and true clitics in others.

D node (Toivonen, 2003). Its distribution is restricted by the Romanian phrase structure rules in (5) and (7) below:

$$(5) \quad I^o \longrightarrow \begin{array}{c} \hat{D} \quad I^o \\ (\uparrow\text{OBJ})=\downarrow \quad \uparrow=\downarrow \end{array}$$

The c-structure for (1–3) is given in Figure 1:

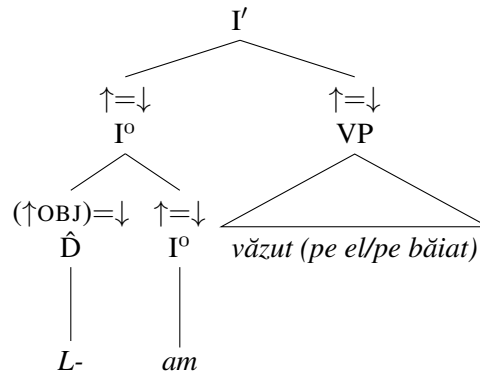


Figure 1: C-structure

The rule in (5) accounts for all clitics except the third person singular feminine clitic *-o*, which can cliticize to past participles as in (6):

- (6) Ai pǎcălit-o.
 have.2SG tricked-3SG.F
 ‘You tricked her.’

Examples such as (6) are allowed by the following rule:

$$(7) \quad V^o \longrightarrow \begin{array}{c} V^o \quad \quad \quad \hat{D} \\ \uparrow=\downarrow \quad \quad \quad (\uparrow\text{OBJ})=\downarrow \\ (\uparrow\text{PTCPL})=\text{PAST} \quad (\uparrow\text{OBJ NUM})=\text{SG} \\ \quad \quad \quad \quad \quad (\uparrow\text{OBJ GEND})=\text{FEM} \\ \quad \quad \quad \quad \quad (\uparrow\text{OBJ PERS})=3 \end{array}$$

We assume that all Modern Romanian object clitics are of the category \hat{D} , and the distribution of the clitics is accounted for by the phrase structure rules that govern the distribution of \hat{D} .⁵ The word order facts are intriguing, and rules beyond (5) and (7) are needed to account for the full distribution. For example, it does not follow from the two rules above that the feminine clitic *-o* cannot precede the auxiliary in examples like (6).⁶ However, we set the details of Romanian word order aside, since they are not directly relevant to our main focus here: the Romanian direct object clitics’ status as agreement markers and incorporated pronouns.

⁵There might be other Romanian clitics that are of the category \hat{D} ; for example, the indirect object pronouns. If so, it is important to ensure that the right forms map onto the right \hat{D} nodes. This can be done with case specifications on the entries and rules, or by “constructive case” (Nordlinger, 1998).

⁶The placement of the third person feminine clitic is intricate. Zafiu et al. (2016, 2.2.3.7) describe

2.2 Uniform hypothesis 1: The clitics are agreement markers

Before the Romanian object clitics are given a formal analysis, two alternative hypotheses will be considered and rejected. The first hypothesis (presented in this subsection) is that the clitics are agreement markers whether or not an independent NP object is (overtly) present. The second hypothesis (presented in Section 2.3) is that the clitics are pronouns whether or not an independent NP object is present. Both of these hypotheses are uniform: the clitics are *not* ambiguous between agreement markers and pronouns.

The analysis that is most commonly associated with the term *pro*-drop assumes the presence of a phonologically empty pronoun ('little *pro*') in non-doubling cases. On this hypothesis, the third person plural masculine pronoun *i-* agrees with *ei* or *băieți* in (8), and it agrees with an empty pronoun *pro* in (9):

- (8) I-am văzut pe ei / pe băieți.
3PL.M-have.1SG seen ACC them.M / ACC boys
'I saw the boys.'
- (9) I-am văzut *pro*.
3PL.M-have.1SG seen
'I saw them.'

Analyzing the clitics uniformly as agreement markers is empirically problematic: the doubling clitics are governed by different restrictions than the non-doubling clitics. In MSR, clitic doubling occurs only with objects that are human, definite and marked with the preposition *pe*. The object-marking preposition *pe* can be compared to the Spanish *a*, and Romanian clitic doubling falls under "Kayne's generalization" (Kayne 1975; Aoun 1981, 275; Jaeggli 1981, 39), which states that clitics can only double prepositionally marked phrases. The non-doubling clitics refer more freely. These generalizations are illustrated below.

Non-doubling clitics can refer to animals and inanimates:

- (10) Cățelul a furat cârnatul. L-am prins cu el în
dog.DEF has stolen sausage.DEF 3SG.M-have.1SG caught with it in
gură.
mouth
'The dog stole a sausage. I caught it with it in its mouth.'
- (11) Tabloul este unicat. L-am cumpărat la o licitație.
painting.DEF is unique 3SG.M-have.1SG bought at an auction
'The painting is unique. I bought it at an auction.'

the distribution as follows: "In MR [Modern Romanian], *o* is always enclitic with the compound past tense, the analytic pluperfect (dialectal), the conditional, and gerundial periphrases, and proclitic with the *voi* infinitival future and with the future perfect. It can be either proclitic or enclitic with the perfect subjunctive."

However, note that the object *cârnatul* ‘sausage’ is not doubled by a clitic in (10). Although non-doubling clitics can refer to non-humans, clitics cannot double non-human NPs. This is further illustrated in (12):

- (12) a. Am văzut melcul.
 have.1SG seen snail.DEF
 ‘I saw the snail.’
 b. *L-am văzut (pe) melc.
 3SG.M-have.1SG seen ACC snail

Clitics also cannot double indefinites:

- (13) a. Am văzut un băiat.
 have.1SG seen a boy.
 ‘I saw a boy.’
 b. *L-am văzut (pe) un băiat.
 3SG.M-have.1SG seen ACC a boy

Finally, (14) illustrates that clitics do not double objects that are not *pe*-marked:

- (14) a. Am văzut (*pe) băiatul.
 have.1SG seen ACC boy.DEF
 ‘I saw the boy.’
 b. *L-am văzut băiatul.
 3SG.M-have.1SG seen boy.DEF

In sum, the doubling clitic (which will be analyzed here as an agreement marker) is restricted in ways that the non-doubling clitic (here, a pronoun) is not.

2.3 Uniform hypothesis 2: The clitics are pronouns

A second potential uniform analysis of the clitic is that it is uniformly a pronoun (cf. Alexopoulou’s 1999 on Greek clitics, and Aoun 1981 on Romanian, Hebrew and Lebanese Arabic clitics). Compare *o-* in example (6) above to example (15):

- (15) Ai păcălit-o pe Ioana.
 have.2SG tricked-3SG.F ACC Joanna
 ‘You tricked Joanna.’

According to this analysis, *-o* is uniformly a pronoun, and *pe Ioana* in (15) would be an adjunct, likely a right-dislocated adjunct. The reading of the example would be something like “You tricked her, Joanna” (with *her* and *Joanna* co-referring).

However, this hypothesis is problematic. First, the restrictions on doubling are equally unexpected under a uniform pronoun hypothesis as under a uniform agreement marker hypothesis (Section 2.2). For example, it is unclear why it would in principle be impossible to right dislocate an inanimate object.

Second, on the uniform pronoun hypothesis, *pe Ioana* is an adjunct in (15). However, proper names (like *Ioana*) *must* be doubled by a clitic:

- (16) * Ai păcălit (pe) Ioana.
 have.2SG tricked ACC Joanna
 ‘You tricked Joanna.’

Since the clitic is obligatory here, the only way to express proper names as objects would be indirectly through a clitic (the “real object”) on this hypothesis. Categorically disallowing proper name objects seems like an unusual constraint.

Third, the example in (15) does not display the typical characteristics of right dislocation. There is no intonational break before the object, and the object is not stressed. The sentences in (17–18) are actual right dislocation examples:

- (17) Ai păcălit-o pe ea, Ioana.
 have.2SG tricked-3SG.F ACC her Joanna
 ‘You tricked her, Joanna.’
- (18) L-am văzut pe el, proful.
 3SG.M-have.1SG seen ACC him prof.DEF
 ‘I saw him, the prof.’

For these reasons, we reject the hypothesis that the clitics are uniformly pronouns.

3 A lexical split hypothesis

In LFG, *pro*-drop is commonly formalized with an optional pronominal PRED feature in the relevant lexical entries. The lexical entry for the third person masculine singular clitic would then look something like (19) (this entry will be revised):

- (19) *i/l-*: ((↑ PRED) = ‘pro’)
 (↑ PERS) = 3
 (↑ NUM) = SG
 (↑ GEND) = MASC

Each PRED feature value is unique, and PRED features can therefore not unify, unlike other features. This means that the clitic cannot co-occur with an independent object (with its own PRED feature) when the PRED feature is present. This holds whether the object is pronominal or not, since pronominal PRED features are also unique. When PRED is absent, the clitic can co-occur with an object on the condition that the other features match. It then functions as a regular agreement marker.⁷ When the PRED feature is present, the clitic *is* the object pronoun, and its PRED feature contributes to the completeness of the f-structure.

The optionality of PRED means that there are in effect two lexical entries, one with a PRED feature and one without:

⁷Except it is a slightly unusual agreement marker, since agreement markers are more typically bound morphemes than clitics.

(20)	Pronoun:		Agreement:	
	(↑ PRED)	= 'pro'	(↑ PERS)	= 3
	(↑ PERS)	= 3	(↑ NUM)	= SG
	(↑ NUM)	= SG	(↑ GEND)	= MASC
	(↑ GEND)	= MASC		

This “lexical split” opens up the possibility for further differences between the entries, and the Romanian clitics indeed differ beyond the PRED feature.

The observant reader will notice that the lexical entries in (20) do not account for all the generalizations listed in Section 2: doubling is only possible with *pe*-marked, animate, definite objects. Following Cornilescu (2000) and others, *pe* is analyzed here as an accusative case marker. The requirement that agreement-marking clitics only double *pe*-marked nouns is captured here with a *constraining equation* demanding accusative case: (↑ CASE) =_c ACC. We assume that the pronominal clitic (the clitic with a PRED feature) is also specified for case, since the object clitics differ in form from other clitics (e.g., dative clitics). However, the pronominal clitic directly contributes the case, so the accusative feature is introduced with a regular defining equation: (↑ CASE) = ACC.

The *pe*-morpheme is an instance of *differential object marking*. There are two ways to express definite and animate NP objects in Romanian, through *pe*-marking (21) and through *-ul*-marking (22):⁸

(21) L-am văzut pe băiat.
 3SG.M-have.1SG seen ACC boy
 ‘I saw the boy.’

(22) Am văzut băiatul.
 have.1SG seen boy.DEF
 ‘I saw the boy.’

The *pe*-marker is obligatory with personal pronouns and proper names with human referents:

(23) L-am vizitat pe el / pe Ion.
 3SG.M-have.1SG visited ACC him / ACC John
 ‘I visited him/John.’

(24) * Am vizitat el / Ion.
 have.1SG visited him / John

Inanimate objects and objects with an indefinite article are not *pe*-marked: examples (12–13) above cannot take *pe*. Unmodified *-ul*-marked direct objects are not *pe*-marked either:

(25) * L-am vizitat pe băiatul.
 3SG.M-have.1SG visited ACC boy

⁸The morphological definiteness marker is referred to as *-ul* here, to distinguish it from *pe* which is also only compatible with definites. However, *-ul* has several allomorphs.

Onea & Hole (2017) propose that *pe*-marking is restricted to strong definites, but this proposal is difficult to reconcile with the fact that proper names are obligatorily *pe*-marked. An alternative account is provided by von Heusinger & Chiriacescu (2013), who suggest that *pe* marks discourse structuring potential; and further proposals are provided by Hill (2013); Ticio & Avram (2015), and Tigău (2015). There is no consensus in the literature about the conditions for the different types of objects, and we will not try to adjudicate between the proposals here.

Although *pe*-marking is not the topic of this paper, it is closely tied to clitic-doubling and we therefore cannot completely set it aside. Clitic doubling and *pe*-marking almost always co-occur, but there is nevertheless some evidence that the animacy requirement is tied to the *pe*-marker and not (necessarily) to the clitic itself. Clitic doubling only occurs with *pe*, but *pe* can occur without a clitic. In examples with *pe* and without a clitic, the animacy requirement remains:

- (26) Petru n-a văzut pe nimeni.
Peter not-have.3SG seen ACC nobody
'Peter didn't see anybody.'
- (27) Petru a văzut pe cineva / pe careva.
Peter have.3SG seen ACC somebody / ACC somebody
'Peter saw somebody.'
- (28) Pe cine ai văzut?
ACC who have.2SG seen
'Who did you see?'
- (29) Petru a văzut ceva. / * Petru a văzut pe ceva.
Peter have.3SG seen something
'Peter saw something.'

The objects in (26–28) are all specified [HUMAN +], and *pe* is obligatory. The object *ceva* in (29) is [HUMAN –] and cannot be *pe*-marked.

The *pe*-marked examples in (26–28) are human, but they are not definite. Clitic doubling only occurs on *pe*-marked definite objects. We propose that *pe* is specified for animacy but not definiteness, and the agreement clitic is specified for definiteness but not animacy. The revised lexical entries for the pronominal and agreement-marking clitics are given in (30):

- | | | | | |
|------|-----------------|------------|-------------------|--------------------|
| (30) | Pronoun: | | Agreement: | |
| | (↑ PRED) | = 'pro' | (↑ PERS) | = α |
| | (↑ PERS) | = α | (↑ NUM) | = β |
| | (↑ NUM) | = β | (↑ GEND) | = γ |
| | (↑ GEND) | = γ | (↑ CASE) | = _c ACC |
| | (↑ CASE) | = ACC | (↑ DEF) | = + |
| | (↑ DEF) | = + | | |

It is crucial for our analysis that the agreement marker is specified for definiteness. We assume that the pronominal clitic is also specified for definiteness, as personal pronouns typically have a definite interpretation.

Let us briefly return to the issue of human referents. The revised agreement-marking entry does not include a HUMAN feature, but it is in fact difficult to determine whether it should. The clitic only agrees with objects with human referents. However, this might be an indirect effect of the requirement that it can only agree with *pe*-marked NPs, and we have independent evidence that *pe*-marking is restricted to objects with human referents. We therefore take the more conservative view that the agreement clitic is not specified for animacy at all, since a HUMAN specification on the clitic is formally unnecessary. On the other hand, Hill (2013) and David (2015) show that the animacy requirement developed independently on the clitic and *pe*, and the clitics never double inanimates. If we were to assume that the agreement marker has an additional specification (\uparrow HUMAN) = +, there would be no empirical consequences. This would then be a further difference between the pronominal clitic and the agreement-marking clitic, as the pronominal clitic can refer freely to animals and inanimates (10–11).

The data presented so far suggest an analysis where the *pe* is specified with the feature [HUMAN +] and can therefore only mark accusative case on [HUMAN +] nouns. However, the data are a bit more complicated, as illustrated by (31):

- (31) Am probat șapte paltoane. L-am cumpărat pe cel
 have.1SG tried seven trench coats. 3SG.M-have.1SG bought ACC the
 mai frumos.
 nicest
 ‘I tried seven trench coats. I bought the nicest.’

Example (31) has clitic doubling and *pe*, even though the superlative (without an overt head noun) refers back to a noun that is not [HUMAN +]. Whether we assume that *cel mai frumos* is nominalized (a deadjectival noun) or it contains an empty *one*-type pronominal, it is clear that it does not have the feature [HUMAN +] in (31). *Cel mai frumos* can refer back to human nouns in other examples, so the simplest analysis is that *cel mai frumos* is unspecified for HUMAN; that is, it has no HUMAN feature at all. If this is the case, then *pe* cannot be specified with a [HUMAN +] feature. Since *cel mai frumos* is unspecified for HUMAN, the [HUMAN +] feature from *pe* would simply be added to the feature structure in *pe cel mai frumos*, which does not work for (31). Instead, we propose that *pe* is lexically specified as follows:

- (32) *pe*: (\uparrow CASE) = ACC
 (\uparrow HUMAN) \neq -

The equation (\uparrow HUMAN) \neq - ensures that *pe* does not unify with nouns specified as [HUMAN -]. Nothing prevents it from unifying with [HUMAN +] nouns, or with nouns which are unspecified for the [HUMAN] feature.

Let us summarize the main points of this section. Following previous work in LFG on “*pro*-drop”, the Romanian object clitic is analyzed here as ambiguous between a pronoun (with [PRED ‘pro’]) and an agreement marker (without [PRED ‘pro’]). The agreement marker is more restricted than the pronoun and co-occurs only with definite, human, *pe*-marked objects. These generalizations are captured by the lexical entries in (30) and (32).

4 Diachronic variation

A wealth of research on the grammaticalization of pronouns and agreement marking (Givón & Li, 1976; Mithun, 1988; Hopper & Traugott, 1993) has shown that the diachronic development typically follows the path in (33):

- (33) independent pronoun > weak pronoun > clitic pronoun > agreement affix
> fused agreement marker

Numerous previous analyses of *pro*-drop have noted that it is quite unsurprising that pronoun/agreement ambiguities should emerge given the grammaticalization cline in (33) (e.g., Fassi Fehri 1984; Bresnan & Mchombo 1987; Toivonen 2001; Morimoto 2002; Butt 2007; Coppock & Wechsler 2010). When pronouns transition into agreement affixes, it seems natural that there could (at least in some cases) be a stage where the forms are not immediately reanalyzed as wholesale agreement, but instead are agreement markers when they double an NP and pronouns when they do not.

Since the mappings between the constituent structure, the feature structure, the lexicon, and the prosodic structure are quite flexible, the ambiguous stage between pronoun and agreement marker is not tied to only one kind of morphosyntactic realization. A linguistic element can in principle be ambiguous between a pronoun and an agreement marker regardless of its status as an independent word, a clitic, a bound agglutinative morpheme, or a fused morpheme. The grammaticalization path in (33) thus conflates two common sequences of changes that are often parallel. The first sequence concerns prosody and c-structural realization:

- (34) independent word > small word > true clitic > affix > fused affix

A “small word” can be a word that does not project a phrase but is phonologically independent *or* a word that projects a phrase but is phonologically dependent on a syntactic host. By ‘true clitic’, we mean a form that does not project a phrase and is phonologically dependent on a host, but is not a bound morpheme. The table in (35) from Toivonen (2003, 45) serves to illustrate the distinctions:

(35)		NON-PROJECTING	PROJECTING
PHON. DEP.		French ‘clitic’ pronouns Serbo-Croatian pronouns and auxiliaries	Kwakwala & Yagua determiners English reduced auxiliaries Swedish genitive marker Finnish, Russian and Bulgarian question particles
PHON. INDEP.		Swedish verbal particles Yoruba weak pronouns	English lexical verbs Yoruba strong pronouns Estonian question particles

The second scale has to do with referential capacity:

(36) noun > pronoun > agreement marker

This scale does not concern the prosodic or phrase-structural realization of a linguistic entity. Nouns, pronouns and agreement markers typically have similar features (e.g., person, animacy, and case), but they differ in their PRED features. Nouns have a contentful nominal PRED feature, pronouns have the PRED feature ‘pro’, and agreement markers have no PRED feature at all.

Even though changes along the cline in (34) tend to be closely tied to changes along (36), the two scales are not intrinsically connected. This disconnect is carefully investigated in van Rijn (2016), who concludes: “loss of referentiality correlates with a loss in form, but in a relative rather than an absolute sense [...] function and form evolve in the same direction, but need not evolve at the same pace”.

Before presenting some historical data from Romanian, we want to clarify two points. First, this paper focuses on the scale in (36), and not (34). As mentioned above, we assume that the relevant elements are true clitics, but they may in fact be bound affixes. It is also possible that some of the Romanian forms are clitics and some are affixes, and there might well be differences between dialects concerning the prosodic and morphological status of the “clitics”. We think this question is important, but the LFG architecture allows us to focus on the pronoun/agreement marker status without taking a stand on the clitic/bound morpheme status.

Second, although we agree with the claims from the historical linguistics literature that certain kinds of changes often follow specific grammaticalization scales, we do not argue that change *must* happen this way. We furthermore do not claim that pronouns changing into agreement markers *must* go through a stage of optionality exactly as proposed here. The point is rather that the optionality of Romanian object clitics seems quite natural when considering the grammaticalization path that has been argued to be the origin of agreement marking in many different languages.

We propose that Romanian object clitics are following a succession of stages as outlined in (37):

(37)	clitic	NP
STAGE 1:	pronoun	adjunct
STAGE 2:	pronoun	
	or agreement marker	object
STAGE 3:	agreement marker	object

The current stage is Stage 2. During a previous stage, the clitics were unambiguous pronouns. When doubled, the doubling NP was an adjunct, not an object. The next natural stage would be Stage 3, where the clitics are unambiguous agreement markers. It is of course not possible to say whether the clitics will reach that stage. Again, certain changes along the scale in (34) tend to occur more or less concurrently with the pronoun to agreement marker shifts, but we set this aside here.

On this view, the clitic would have been optional during Stage 1, with doubling under certain discourse conditions. Until recently, the clitic was indeed optional (see, e.g., *The Grammar of the Romanian Academy* 1963; Zafiu et al. 2016, Section 2.4.2.5), and examples such as (38–39) (with no object clitic) were possible:

(38) Pre tine vădzuu.
ACC you.SG seen.1SG
'I saw you.' (Zafiu et al. 2016, Section 2.4.2.5, [1683])

(39) Domnul Domnezeu făcu pre om.
Lord.DEF God made ACC man
'The Lord God made the man.' (Zafiu et al. 2016, Section 2.4.2.2, [1582])

Equivalent examples in MSR are unacceptable. In (38–39), *pre tine* and *pre om* (*pre* is an old form of *pe*) are regular objects, and not adjuncts added as afterthoughts.

The fact that the clitic used to be optional is consistent with our proposal. Another important part of the puzzle would be to show that the NPs were adjuncts in doubling examples. However, distinguishing arguments from adjuncts is in general not easy (see, e.g., Whaley 1993; Needham & Toivonen 2011), and it is especially difficult when the evidence is restricted to written records. Evidence from intonation and native speakers' grammaticality judgements are not available.

One potential source of evidence for adjuncthood is word order: arguments tend to be more restricted in terms of phrase-structural position, and adjuncts are often ordered after arguments (Jackendoff, 1977; Pollard & Sag, 1987, i.a.). However, word order in Old Romanian is quite free, even freer than in Modern Romanian (Nicolae, 2016), so it is difficult to draw any conclusions based on ordering.⁹

Another indication that the doubled NPs were indeed adjuncts in Old Romanian is that doubling was much less restricted than it is now. In fact, it seems like it was syntactically quite unrestricted: Zafiu et al. (2016, Section 2.4.2.5) claim that there

⁹Clitics have a higher degree of freedom in their placement relative to the verb in Old Romanian than in Modern Romanian (Zafiu et al., 2016, Section 2.2.2.1). This indicates that the Old Romanian "clitics" were in fact free-standing words. Since the changes in (34) and (36) tend to go hand-in-hand, it is not surprising to find indications that the clitics were until recently independent words.

were “no proper rules of clitic doubling” in Old Romanian. Repetition of discourse participants through adjunction (in phenomena such as right or left dislocation) is restricted through discourse considerations such as emphasis and afterthoughts. The restrictions on true agreement are clearer and easier to detect in a text, since the restrictions are based on the syntax within the clause. In order to understand discourse constraints, we need access to larger corpora, and to a certain extent speaker/writer intentions. It seems that there were no *syntactic constraints* on clitic doubling in Old Romanian: clitic doubling was possible with non-human objects, *-ul*-marked objects, and it did not have to co-occur with *p(r)e*. For example, the object *flămânzii* in (40) is *-ul*-marked and there is no *p(r)e*:

- (40) flămânzii săturați-i
 hungry.PL.DEF.ACC feed.IMP.2PL-3PL.M
 ‘feed the hungry’ (Zafiu et al. 2016, Section 2.4.2.5, [1601])

The lack of syntactic constraints on clitic doubling in Old Romanian is consistent with an adjunction hypothesis and problematic on an agreement hypothesis.

We suggest that the first change was that the PRED feature became optional:

- | | | | | |
|------|-----------------|------------|-----------------|------------|
| (41) | Stage 1: | | Stage 2: | |
| | (↑ PRED) | = ‘pro’ | ((↑ PRED) | = ‘pro’) |
| | (↑ PERS) | = α | (↑ PERS) | = α |
| | (↑ NUM) | = β | (↑ NUM) | = β |
| | (↑ GEND) | = γ | (↑ GEND) | = γ |
| | (↑ CASE) | = ACC | (↑ CASE) | = ACC |
| | (↑ DEF) | = + | (↑ DEF) | = + |

Note that “Stage 2” in (41) is very close to our proposed lexical entry for MSR. One further change occurred that led to the current stage: the accusative case feature morphed into a requirement for an overtly case marked (*pe*-marked) NP object. This change was formalized above as a constraining equation, and it occurred in the agreement-marking clitic but not the pronominal clitic.

The literature on the historical development of clitic doubling and *pe* sheds light on the emergence of the agreement marker’s *pe* requirement. Clitic doubling and *pe*-marking developed separately, and the doubling developed after *pe* (Chiriacescu, 2007; von Heusinger & Onea Gáspár, 2008; Hill, 2013; Tigău, 2014; David, 2015). Even though clitic doubling and *pe*-marking are not intrinsically connected, several authors have argued that they serve overlapping (though distinct) discourse functions (Leonetti, 2008; Hill, 2013; Tigău, 2014). It is still unclear exactly what those discourse functions are, but it seems that discourse prominence and high referentiality are relevant notions (Leonetti, 2008; Hill, 2013; Tigău, 2014). Since the functions are overlapping, *pe* and doubling often co-occurred even before this was a grammatical requirement. We propose that the frequent co-occurrence of the forms led to a reanalysis where the co-occurrence became a morphosyntactic requirement: Modern Romanian agreement clitics require *pe*.

5 Synchronic variation

This section will present two additional documented varieties of current Romanian. The variation can be modelled quite readily with the lexical entries in (30) above as a starting point. The varieties involve only minor featural differences.

The first variety is presented by Tigău (2010, 2014). Tigău reports that some speakers of Romanian allow clitic doubling with indefinites:

- (42) Petru (l-)a vizitat pe un prieten.
 Peter 3SG.M-have.3SG visited ACC a friend
 ‘Peter visited a friend.’

Even the speakers who allow doubling with indefinite objects allow it only sometimes. Tigău (2010, 2014) argues that doubled indefinite objects get a *specific* interpretation (see also Aoun 1981, Chapter 3).

The difference between MSR and the indefinite-doubling dialect described by Tigău is captured with minimally different lexical entries. Recall that both the pronominal clitic and the agreement marker are specified as [DEFINITE +] in MSR (30). In the dialect considered here, the pronoun is the same as in MSR, but the agreement marker differs by being marked for specificity instead of definiteness:

- | | | | | |
|------|-----------------|---|-------------------|-----------------------------|
| (43) | Pronoun: | | Agreement: | |
| | (↑ PRED) | = | ‘pro’ | |
| | (↑ PERS) | = | α | (↑ PERS) = α |
| | (↑ NUM) | = | β | (↑ NUM) = β |
| | (↑ GEND) | = | γ | (↑ GEND) = γ |
| | (↑ CASE) | = | ACC | (↑ CASE) = _c ACC |
| | (↑ DEF) | = | + | (↑ SPECIFIC) = + |

In both dialects, clitics double only *pe*-marked objects, hence the constraining equation for accusative case in the agreement marker (which is the version of the clitic that doubles the object).

The Aromanian dialect (AR; spoken in Albania, Macedonia, Romania, Bulgaria, Serbia and Croatia) and the Megleno-Romanian dialect (MR; spoken in Greece and Macedonia) differ from the varieties discussed above. In AR and MR, all and only definite objects are doubled (Tomić 2006, Chapter 4; Tomić 2008, 84; Hill 2013). There is no *pe* requirement; in fact, AR and MR do not have accusative *pe*-marking at all. There is also no animacy requirement, which follows if the [HUMAN] requirement is associated with the *pe*.

Sentence (44) is an MR example. The clitic *l-* doubles the object *filmu*, which is inanimate and carries definiteness morphology, but is not *pe*-marked.

- (44) L-am vizut filmu.
 3SG.M-have.1SG seen film.DEF
 ‘I saw the film.’ (Tomić, 2006, 295)

The lexical entries for the pronominal and agreement-marking clitics in AR and MR (given in (45)) are identical except for the PRED feature:

(45)	Pronoun:		Agreement:	
	(↑ PRED)	= 'pro'		
	(↑ PERS)	= α	(↑ PERS)	= α
	(↑ NUM)	= β	(↑ NUM)	= β
	(↑ GEND)	= γ	(↑ GEND)	= γ
	(↑ CASE)	= ACC	(↑ CASE)	= ACC
	(↑ DEF)	= +	(↑ DEF)	= +

The AR/MR lexical entry for the pronoun is the same as the pronoun entry in the other dialects. However, the AR/MR agreement-marking clitic differs from both of the other dialects in that the equation for CASE is not a constraining equation. It further differs from the indefinite-doubling dialect in that it is marked for definiteness but not for specificity.

This section has extended the analysis proposed for MSR in Section 2 to other dialects of Romanian. The lexical split analysis makes it possible to consider the agreement-marking clitics separately from the pronominal clitics. The dialectal data show no indication that the pronominal clitics differ across dialects. However, the agreement markers display differences in their lexical features. These differences, which are formally minimal, have easily observable empirical effects.

6 Lexical splits cross-linguistically

The analysis of Romanian object clitics presented above adopts a lexical split hypothesis: each clitic form is associated with two lexical entries. The fact that the agreement markers and pronouns display differences beyond the PRED feature supports the analysis. Adopting standard assumptions about the grammaticalization of pronouns, it is not surprising that lexical splits like the Romanian one should emerge. If these claims are correct, then we should expect lexical splits to be quite common in agreement systems cross-linguistically. This section presents a list of examples of agreement systems with forms that seem to be ambiguous between agreement markers and pronouns, and where the difference goes beyond mere referential status.

Subject agreement in Modern Standard Arabic. Fassi Fehri (1984, 1988, 1993) carefully analyzes subject agreement in Modern Standard Arabic. He proposes that “some affixes have two different lexical entries” (Fassi Fehri, 1988, 119). The kinds of splits he describes look strikingly similar to the Romanian clitics. For example, he provides different lexical entries for the feminine affix *at*. Two entries are pronominal but differ in GENDER, NUMBER, PERSON and HUMAN features. A third entry is an agreement marker and has only one feature: [GENDER FEMININE].

Possessive suffixes in Finnish. Pronominal possession in Standard Finnish can be marked with an independent pronoun and a suffix on the possessed noun, or

a suffix alone. In first and second person, the independent pronoun is optional and expressions with or without the independent possessor have the same literal meaning (they differ in emphasis). When a third person independent pronoun is “dropped” and possession is marked by just a suffix, the possessor is necessarily bound by a subject within the minimal finite clause. Conversely, when an independent pronoun is present, the possessor *cannot* be bound by a subject. In Toivonen’s (2000) analysis, the anaphoric suffix has a PRED feature and the suffix agreeing with a non-anaphoric independent pronoun does not. The entries also differ in that the agreement suffix is restricted to agreement with human personal pronouns.

Rioplantense Spanish object clitics. Varieties of Spanish display clitic systems very similar to that of Romanian (see, e.g., Mayer 2017). Andrews (1990) and Estigarribia (2013) analyze Rioplantense Spanish within an LFG framework. They both propose entries for pronominal clitics that differ from the agreement clitics beyond the PRED feature. In Estigarribia’s analysis, the agreement marker has the following feature that the pronominal clitic lacks (p. 300): $\neg(\downarrow\text{SPECIFIC})$ –.

Pakin Lukunosh Mortlockese object suffixes. Odango (2014) argues that the object markers in the Micronesian language Pakin Lukunosh Mortlockese show a split. Most of the suffixes exclusively function as incorporated pronouns. However, when an independent object is present, the third person singular object suffix behaves like a general transitivity marker. Object markers often grammaticalize into transitivity markers (Lehmann, 2002; Mayer, 2017). A natural grammaticalization path is: independent object pronoun > incorporated pronoun > agreement marker > transitivity marker.

7 Conclusion

Romanian object clitics can double direct objects, provided that the objects are animate, definite and *pe*-marked. The sections above described the clitic doubling in Modern Standard Romanian and provided an analysis in Lexical Functional Grammar. According to our analysis, the Romanian clitics are ambiguous between pronouns and agreement markers.

Baker & Kramer (2018) write: “Controversy and uncertainty have plagued the question of whether ‘object markers’ (OMs) are object pronouns cliticized to the verb or realizations of object agreement.” They also note that “[i]t is an awkward fact that generative linguistics has had a hard time distinguishing reliably between pure agreement and clitic doubling (CD).” We believe that this awkwardness can be explained by the special challenges that pronouns and agreement marking pose for syntactic theory. Three challenges have been addressed in this paper.

First, as argued in this paper following Bresnan & Mchombo (1987) and others, it is not uncommon for the same forms to be ambiguous between agreement markers and pronouns. Second, it is often difficult to determine the morphological status of these elements because they display characteristics that are typical neither for regular, free standing, fully projecting words, nor for bound morphemes. In

other words, these elements often involve unusual mappings between p-structure and c-structure (see Section 4, and also Spencer & Luís 2012; Bögel 2015; Lowe 2016). Section 4 also mentioned the third challenge, which concerns the alignment between c-structure and f-structure. Prototypically, pronouns are some kind of “small words” or clitics, and agreement markers are bound morphemes. However, mismatches are possible and atypical alignment can occur: although it is possible to describe the typical c-structural expression of pronouns and agreement markers, these generalizations are not universal principles of grammar. Clitics can function as agreement markers, and pronouns can be morphologically incorporated.

References

- Alexopoulou, Theodora. 1999. *The syntax of discourse functions in Greek: A non-configurational approach*. Edinburgh: University of Edinburgh dissertation.
- Andrews, Avery. 1990. Unification and morphological blocking. *Natural Language and Linguistic Theory* 8(4). 507–558.
- Aoun, Joseph. 1981. *The formal nature of anaphoric relations*. Cambridge: Massachusetts Institute of Technology dissertation.
- Baker, Mark & Ruth Kramer. 2018. Doubled clitics are pronouns: Amharic objects (and beyond). *Natural Language and Linguistic Theory* 36(4). 1035–1088.
- Bögel, Tina. 2015. *The syntax-prosody interface in Lexical Functional Grammar*. Konstanz: University of Konstanz dissertation.
- Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler. 2016. *Lexical-functional syntax, second edition*. Hoboken, NJ: Wiley-Blackwell.
- Bresnan, Joan & Sam Mchombo. 1987. Topic, pronoun, and agreement in Chicheŵa. *Language* 63(4). 741–782.
- Butt, Miriam. 2007. The role of pronominal suffixes in Punjabi. In Annie Zaenen, Jane Simpson, Tracy Holloway King, Jane Grimshaw, Joan Maling & Christopher Manning (eds.), *Architectures, rules, and preferences: Variations on themes by Joan Bresnan*, Stanford: CSLI.
- Chiriacescu, Sofiana. 2007. Pe-Markierung und Diskurz-Prominenz im Rumänischen. Magisterarbeit, University of Stuttgart.
- Coppock, Elizabeth & Stephen Wechsler. 2010. Less-travelled paths from pronoun to agreement: The case of the Uralic objective conjugations. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG11 conference*, 165–185. Stanford: CSLI.

- Cornilescu, Alexandra. 2000. Notes on the interpretation of the prepositional accusative. *Bucharest Working Papers in Linguistics* 2. 91–106.
- David, Oana. 2015. Clitic doubling and differential object marking: A study in diachronic construction grammar. *Constructions and Frames* 7(1). 103–135.
- Dobrovie-Sorin, Carmen. 1994. *The syntax of Romanian: Comparative studies in Romance*. Berlin/New York: Walter de Gruyter.
- Estigarribia, Bruno. 2013. Rioplatense Spanish clitic doubling and “tripling” in Lexical-Functional Grammar. In Chad Howe, Sarah A. Blackwell & Margaret Lubbers Quesada (eds.), *Selected proceedings of the 15th Hispanic Linguistics Symposium*, 297–309. Somerville, MA: Cascadilla Proceedings Project.
- Fassi Fehri, Abdelkader. 1984. Agreement in Arabic, binding and coherence. In *Conference on agreement in natural language*, Stanford University.
- Fassi Fehri, Abdelkader. 1988. Agreement in Arabic, binding and coherence: Approaches, theories, descriptions. In Michael Barlow & Charles Ferguson (eds.), *Agreement in natural language*, 107–158. Stanford: CSLI.
- Fassi Fehri, Abdelkader. 1993. *Issues in the structure of Arabic clauses and words*. Dordrecht: Kluwer.
- Givón, Talmy & Charles Li. 1976. Topic, pronoun and grammatical agreement in subject and topic. In Charles Li (ed.), *Subject and topic*, 1551–188. New York: Academic Press.
- von Heusinger, Klaus & Edgar Onea Gáspár. 2008. Triggering and blocking effects in the diachronic development of DOM in Romanian. *PROBUS* 20(1). 67–110.
- Hill, Virginia. 2013. The direct object marker in Romanian: a historical perspective. *Australian Journal of Linguistics* 33(2). 140–151.
- Hopper, Paul & Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge, UK: Cambridge University Press.
- Jackendoff, Ray. 1977. *X-bar syntax: A study of phrase structure*. Cambridge, MA: MIT Press.
- Jaeggli, Osvaldo. 1981. *On some phonologically null elements in syntax*. Cambridge: Massachusetts Institute of Technology dissertation.
- Kayne, Richard. 1975. *French syntax: The transformational cycle*. Cambridge, MA: MIT Press.
- Lehmann, Christian. 2002. *Thoughts on grammaticalization, 2nd edition*. Erfurt: Arbeitspapiere des Seminars für Sprachwissenschaft der Universität Erfurt.

- Leonetti, Manuel. 2008. Specificity in clitic doubling and in differential object marking. *PROBUS* 20(1). 33–66.
- Lowe, John. 2016. Clitics: Separating syntax and prosody. *Journal of Linguistics* 52(2). 375–419.
- Luís, Ana Alexandra Ribeiro. 2004. *Clitics as morphology*. Essex: University of Essex dissertation.
- Mayer, Elisabeth. 2017. *Spanish clitics on the move: Variation in time and space*. Berlin: De Gruyter Mouton.
- Mithun, Marianne. 1988. The development of bound pronominal paradigms. In Winfred P. Lehman & Helen-Jo Jakusz Hewitt (eds.), *Language typology: Typological models in reconstruction*, 85–104. Amsterdam: John Benjamins.
- Monachesi, Paola. 1998. The morphosyntax of Romanian cliticization. In P. Copen, H. van Halteren & L. Teunissen (eds.), *Proceedings of Computational Linguistics in the Netherlands 1997*, 99–118.
- Morimoto, Yukiko. 2002. Prominence mismatches and differential object marking in Bantu. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG02 conference*, 292–314. Stanford: CSLI.
- Needham, Stephanie & Ida Toivonen. 2011. Derived arguments. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG11 conference*, 401–421. Stanford: CSLI.
- Nicolae, Alexandru. 2016. Word order and configurationality. In Gabriela Pană Dindelegan (ed.), *The syntax of old Romanian*, Chapter 10. Oxford: Oxford Scholarship Online.
- Nordlinger, Rachel. 1998. *Constructive case: Evidence from Australian languages*. Stanford: CSLI.
- Odango, Emerson Lopez. 2014. The status of subject and object markets in pakin lukunosh mortlockese. *Oceanic Linguistics* 53(1). 110–135.
- Onea, Edgar & Daniel Hole. 2017. Differential object marking of human definite direct objects in Romanian. *Revue roumaine de linguistique* 62(4). 359–376.
- Pollard, Carl & Ivan Sag. 1987. *Information-based syntax and semantics*. Stanford: CSLI.
- Popescu, Alexandra. 2000. The morphophonology of the Romanian clitic sequence. *Lingua* 110(1). 773–799.
- Spencer, Andrew & Anna R. Luís. 2012. *Clitics: An introduction*. Cambridge: Cambridge University Press.

- Ticio, Emma & Larisa Avram. 2015. The acquisition of differential object marking in Spanish and Romanian: Semantic scales or semantic features? *Revue roumaine de linguistique* 60(4). 383–401.
- Tigău, Alina Mihaela. 2010. Towards an account of differential object marking in Romanian. *Bucharest Working Papers in Linguistics* 12(1). 137–158.
- Tigău, Alina Mihaela. 2014. Specificity effects with clitic doubling and *pe* marking. *Bucharest Working Papers in Linguistics* 16(1). 43–62.
- Tigău, Alina Mihaela. 2015. Romanian differential object marking and presupposition. *Cahiers de linguistique française* 32. 215–227.
- Toivonen, Ida. 2000. The morphosyntax of Finnish possessives. *Natural Language and Linguistic Theory* 18(3). 579–609.
- Toivonen, Ida. 2001. Language change, lexical features and Finnish possessives. In Miriam Butt & Tracy Holloway King (eds.), *Time over matter: Diachronic perspectives on morphosyntax*, 209–225. Stanford: CSLI.
- Toivonen, Ida. 2003. *The phrase structure of non-projecting words: A case study of Swedish particles*. Dordrecht: Kluwer.
- Tomić, Olga Mišeska. 2006. *Balkan Sprachbund morphosyntactic features*. Dordrecht: Springer.
- Tomić, Olga Mišeska. 2008. Towards grammaticalization of clitic doubling: Clitic doubling in Macedonian and neighbouring languages. In Dalina Kallulli & Liliane Tasmowski (eds.), *Clitic doubling in the Balkan languages*, 65–87. Amsterdam/Philadelphia: John Benjamins.
- van Rijn, Marlou. 2016. The grammaticalization of possessive person marking: A typological approach. *Transactions of the Philological Society* 114(2). 233–276.
- von Heusinger, Klaus & Sofiana Chiriacescu. 2013. The discourse structuring potential of differential object marking: the case of indefinite and definite indirect objects in Romanian. *Revue roumaine de linguistique* 58(4). 439–456.
- Whaley, Lindsay. 1993. *The status of obliques in linguistic theory*. Buffalo: State University of New York dissertation.
- Zafiu, Rodica, Alexandru Nicolae, Dana Niculescu, Gabriela Pană Dindelegan, Irina Nicula Paraschiv, Adina Dragomirescu & Andra Vasilescu. 2016. The verb and its arguments: the root clause. In Gabriela Pană Dindelegan (ed.), *The syntax of Old Romanian*, Chapter 2. Oxford: Oxford Scholarship Online.