

Subjectification in the Development of Clitic Doubling: A Diachronic Study of Romanian and Spanish

OANA A. DAVID
University of California, Berkeley

1 Introduction

This work focuses on a phenomenon commonly occurring in several Western Romance languages, Semitic languages, and the Balkan Sprachbund languages (Slavic, Greek, Macedonian, Albanian and Eastern Romance). The direct object of a simple transitive clause can be doubled by a coreferential pronominal clitic attaching to the verb, as illustrated by the following sentences:

- (1) Romanian
- a. *Președintele a convocat toți delegații.*
president.DEF has assembled all delegate.PL.DEF
- b. *Președintele i- a convocat pe toți delegații.*
president.DEF CL.3PL.MASC has assembled DOM all delegate.PL.DEF
'The president assembled all the delegates.'
- (2) Albanian
- a. *Ana lexoi librin.*
Ana read book.DEF
'Ana read the book.'
- b. *Ana e lexoi librin.*
Ana CL.3SG.MASC read book.DEF
'Ana read the book.'
- (Kallulli 2008:230)
- (3) Rioplatense Spanish
- La vas a llamar a Marta?*
CL.3SG.FEM will call DOM Marta?
'Are you going to call Marta?'

No, la voy a llamar a esta mina
 no CL.3SG.FEM will call DOM that girl
de la que hablábamos anoche.
 we were talking about last night

‘No, I’m going to call this girl we were talking about last night.’

(Estigarribia 2006)

In standard Romanian, as illustrated in (1b), as well as in Spanish (3b), a prepositional accusative differentially marks the instantiated direct object while a pronominal clitic that is coreferential with the direct object is also present. The data in (1-3) show that among the languages possessing some form of clitic doubling (hereafter CD) and differential object marking (hereafter DOM), Spanish in all its varieties is one of the few that, like Romanian, shows a juxtaposition of the two constructions. Further, Spanish is the language in which this pattern behaves most like that in Romanian, both in terms of form and function. On the other hand, there are languages like Albanian that have CD and no DOM. Many more languages have DOM and no CD, as Bossong has thoroughly documented (1985, 1991, 1998), but so far very few are found that have both.

The analysis pursued here operates on the assumption that prepositional accusative marking (or differential object marking) and clitic doubling should be seen as a single clause-level construction. This construction brings its own contribution to the clause independent of, yet semantically inheriting from its component parts. In Romanian, these two components gravitated towards each other based on their common high accessibility and their grounding in core transitivity. That is, *pe* is used with direct objects most likely to be prototypical patients, and follows the general rules of prototypical transitivity as posited in Hopper and Thompson (1980) and Næss (2007). von Heusinger and Kaiser (2010) find that the general properties of prototypical transitivity are also true of *a* in Spanish.

Therefore, in the current work, I extend my analysis of the evolution of this construction in Romanian to the study of a comparable Romance language, Spanish, concluding with a claim about the construction in all languages that have it: the intersection of the core transitivity scenario brought about by DOM, and the high accessibility scenario brought about by CD will result in this composite construction coming about. Nevertheless, the construction will not be functionally identical cross-linguistically. Further, Romanian CD-DOM is different from the CD-only construction in Romanian (and in Spanish), in addition to there being a marked functional difference between the two languages in the usage of the CD-DOM construction.

Taking a diachronic perspective, I will argue that subjectification is the driving force behind the grammaticalization of this construction; subjectification is the kind of semantic change that has been observed for the development of modals and causal connectives over time, as introduced by Traugott (1989, 1995, 1997, 2005), and explored by Sweetser (1990) and Sanders and Sweetser (2009),

Subjectification in the development of clitic doubling

among others. Specifically, during grammaticalization, there is a progression from more propositional/ideational content, to more textual/discourse-structuring functions and meanings and finally to expressive/pragmatic functions and meanings associated with a given form. That is to say, grammatical and lexical items come to be expressions of *epistemicity* (Traugott 1989:32), reflecting speaker's internal states, including beliefs, attitudes, etc. In addressing subjectification as key to the development of CD-DOM, I would like to further suggest that this sequence of semantic change also holds of constructionalization (Bybee 2003, Noël 2007), whereby constructions that are more expressive of propositional content unify over time to yield constructions that are more expressive of epistemic content.

In that vein, in Romanian there has been constructionalization further along a subjectification cline than in Spanish, resulting in a fully pragmatic construction in the modern language, while in Spanish the subjectification of the construction is slower, revealing an occasional glimpse of the pragmatic construction only in some newer varieties (such as is illustrated in Rioplatense Spanish in 3).

By observing the diachronic development and synchronic status of this construction in Romanian and Spanish, I will take this opportunity to challenge existing assumptions in modern syntax regarding the similar underlying structures of the construction cross-linguistically, (a common assumption when establishing similar formal syntactic projections). I will aim to show that while the construction is syntactically and morphologically similar in the two languages, it arose from two completely different diachronic constructionalization processes rooted in different lexical origins (specifically pertaining to the semantics of the prepositions *pe* and *a*), while maintaining similarity in function. I hope this opens the door to further detailed historical semantic reconstruction of the joint clitic doubling-differential marking construction in the few languages that possess it, such as Catalan, and to consider the typology of differential object marking from a semantic-pragmatic perspective.

The Romanian data is gathered from modern Romanian news sources, from examples found in the existing literature, and from a corpus compiled by the author¹. Spanish examples come from Spanish popular media sources and from the existing literature (covering several Spanish varieties); the study also references one of the oldest manuscripts in Continental Spanish, the *Cantar de mio Çid*.

2 Two types of subjectification

There are two main views on subjectivity and subjectification: a conceptualist approach and a pragmatic approach (Nuyts 2001, Narrog 2012). The conceptualist

¹ The corpus contains 300,000 words, and includes multiple fiction and non-fiction works by Constantin Cantacuzino (1700s), Ion Creangă (late 1800s), and Mircea Eliade (early 1900s).

approach represents the interpretation of Langacker (1987, 2002, 2006), which defines subjectivity in terms of grounding and construal, with a more subjective stance consisting of an offstage construal of the conceptualizer and a more objective perspective placing the conceptualizer onstage. Thus, deictics such as *I* and *you*, are expressions that maximally objectify the grounded discourse participants, placing them within the immediate scope of predication (Langacker 2007). On the other hand the pragmatic approach, in the spirit of Traugott and Dasher (2002) and the aforementioned earlier work by Traugott, holds that subjectification is a tendency towards more expressive meanings from more propositional meanings, and thus Traugott's approach is explicitly diachronically oriented, succeeding in putting forth generalizations about semantic change over time.

In this work, I utilize both notions of subjectification, but with a strong focus on Traugott's version. Langacker's notion explains the attenuation or bleaching that the component constructions undergo, while Traugott's notion explains the epistemic transformation of the resulting construction. Further, subjectification in the sense of Traugott in this case results in deobjectivization in the sense of Langacker (2006). More precisely, the component construction CD is highly grounded in the speech event, and as such its meaning is highly objective in the Langackarian sense. As deictic and referential expressions grammaticize into a construction whose use is shaped by discourse management, information structure, and epistemic stance, those lexical expressions forming part of that construction lose their relation to the ground. Thus, in the case of the doubling construction with object marking, deobjectivization in Langackarian terms, (that is, the loss of onstage status) goes hand in hand with subjectification in the Traugott sense. In short, deobjectivization is precisely what happens to pronouns when they become weak clitics in the doubling construction, as well as in general².

We can illustrate this process with an example from Romanian:

- (4) a. *Îl iubește.*
 CL.3SG.MASC loves
 '(She) loves him.'
- b. *Îl iubește pe un băiat cu ochii negri.*
 CL.3SG.MASC loves DOM a boy with eyes black
 '(She) loves a boy with dark eyes.'
- c. *Iubește un băiat cu ochii negri.*
 loves a boy with eyes black
 '(She) loves a boy with dark eyes.'

² Note that becoming less objective does not automatically mean they are becoming more subjective. The loss of objective status for deictics does not in and of itself result in achieving subjective status. Subjectification is achieved by virtue of the construction being more pragmatically meaningful over time.

Subjectification in the development of clitic doubling

In (4a), the referent, referred to deictically with a pronoun, is in the immediate scope of predication, and the meaning of the pronoun is referential. In (4b), however, the pronoun is no longer referential, but acts to reinforce the direct object, which is referential. The construction in (4b) incorporating both DOM and CD is distinct from (4c) in enhancing prominence. Prominence, in the definition adopted by Hopper and Thompson (1980), Goldberg (2006), and others, is a general form of informational salience, triggered by some morphological or syntactic form, in the mind of the addressee. As Goldberg (2006:136) points out, both topical and focal constructions are prominence-enhancing, because they force to the foreground information, whether new or old, and distinguish it from presupposed or backgrounded information. Accordingly, CD-DOM performs both topical and focal functions, as well as myriad other prominence-enhancing functions.

In the following section, I will outline the semantics of the component constructions CD and DOM in Romanian and Spanish, and illustrate their subjectification as they unite into a single construction.

3 Constructional components in historical perspective

The two main components contributing to CD-DOM are discussed in this section, namely the Canonical Pronominal Construction (CPC) and Differential Object Marking (DOM).

Like other Romance languages, Romanian and Spanish transitive clauses can contain direct objects instantiated purely by means of a verb-adjoined weak pronoun (Gerlach 2002, Lambrecht 2004), which I will call here the canonical pronominal construction (CPC), as is illustrated in (4a). This trait is attested in Romanian from the earliest periods. Referents referred to by a clitic pronoun alone are highly accessible (Ariel 1988, 1990, 2001, *inter alia*), being placed very high on the accessibility scale relative to other forms of nominal forms (Ariel 1990:73). In Accessibility Theory, referring expressions are instructions for recoverability of referents, and this recoverability lies within the scope of mental accessibility of the particular form, whether a full noun phrase, a pronoun, or even a null expression. This way of categorizing referring expressions lines up well with Langacker's view of onstage/offstage predication and grounding. Namely, clitics are highly grounded in the moment of discourse, as pronouns in general are argued to be (Langacker 2007). Thus, any construction involving only a clitic pronoun as a main argument is a construction whose meaning is lower in subjectivity: the referent, when expressed deictically, is maximally grounded in the external situation while providing minimal semantic information about the nature of the referent.

Also, like many languages with eroded case systems, Romanian retains a selective distribution of some case marking with certain direct objects. Namely, the dummy preposition *pe* (and its prior reflex *pre*), is used as the accusative case

marker for direct objects. From the earliest attestations, *pe* is used to mark direct objects highest in animacy, namely human nouns, deictic pronouns and anaphors. However, it is unclear when and why *pe* assumed this function. As the earliest written documents available with sufficient data illustrating DOM, the Sibiu manuscripts provide attestations of CD and DOM from 1551-53 (data from Hill & Tasmowski 2008). As a few sentences illustrate, at this historical juncture there is still a mix of *pe*- and non-*pe*-marked direct objects.

- (5) *Cine va primi pre voi*_[+pe], *mine*_[-pe] *primște*,
 who will receive pre you.pl me.acc receives
*și cine pre mine*_[+pe] *primește*
 and who pre me.acc receives
 ‘He who receives you receives me, and he who receives me...’

Alone, *pe* was and continues to be semantically a licenser of animate direct objects in Romanian, a generalization thought to hold for many DOM languages. However, there was a period during which *pe*-marked and non-*pe*-marked human direct objects, including pronouns, were in mixed usage, as illustrated even within a single sentence in (5) above.

The most likely genesis for the *pe*-marking construction was proposed by Hatcher (1942), who suggests that *pe* marking arose out of intransitive constructions with verbs of attacking or violence, with the preposition phrase coupled with these verbs being reanalyzed as markers and the content of the prepositional phrase being reanalyzed as the direct object of a transitive sentence. If *pe* marking arose in this manner, with intransitives being reanalyzed as transitives starting with a small cluster of verbs that bring about sudden affect in their objects and spreading to other verbs, then the types of direct objects that are *pe*-marked are the type that tend to be affectees with those types of verbs. Lyons (1990) echoes this view of the origins of prepositionally marked direct objects as a form of adjunction, interpreting the Spanish analogue to *pe*, *a*, and the NP it introduces as an adjunct that reinforces or clarifies the reference of a null *pro* in the object position. However, this analogy with the Romanian construction may be only superficial in nature; although similar in discourse function, particularly in combination with clitic doubling, DOM in Romanian and in Spanish diverge when considering the diachronic dimension.

One important difference between DOM in Romanian and Spanish is its function in indirect object marking. That is, Spanish *a*-marking applies both to direct and indirect objects, while in Romanian *pe*-marking applies only to direct objects. This difference is owed in great part to the fact that Romanian is subject to grammatical changes typical of Balkan Sprachbund languages, while Spanish is not. It is a common feature of Balkan Sprachbund languages for there to be genitive-dative synchrony, such that the dative case comes to look more like the genitive case. This can result in ambiguity in Romanian (6a), whereas no such ambiguity exists in Spanish (6b-c):

Subjectification in the development of clitic doubling

- (6) a. *I-am aruncat mingia lui Mario.*
'I threw the ball to Mario. / I threw Mario's ball.'
- b. *Lancé la pelota a Mario.*
'I threw the ball to Mario.'
- c. *Lancé la pelota de Mario.*
'I threw Mario's ball.'

This similarity of form and function in Spanish direct and indirect object marking is not a matter of coincidence, and may be a clue to the origins of the DOM construction in Spanish, namely as an extension to direct object marking from indirect object marking. This hypothesis is tenable when considering the semantics of the prepositions involved. Semantically, *a* comes from a preposition of directed motion (meaning 'to' or 'toward') while *pe* comes from a preposition that captures trajector-landmark relations, namely 'above', 'on', or 'onto' (originally from *asupra* (Rom.), from Latin *super*). These prepositions, with these meanings are still used in these languages, their meanings not having been bleached. When taking into account the meaning of the prepositions themselves, *a* clearly expresses directed action or motion, rendering its object a recipient, while *pe* expresses contact, rendering its object an affected entity. In this way, thinking of *a* as originally an indirect object marker is plausible, given that indirect objects are usually recipients or goals of directed action and directed motion.

Thus, an extension to usage with direct objects in Spanish could be due to the construal of direct objects of transitive constructions as recipients of the action affecting them. On the other hand, even if Romanian had not undergone genitive-dative synchrony, it is unlikely that *pe* would have ever shown up as an indirect object marker, given the semantics of prepositional *pe* detailed above. What's more, *a* appears as both a direct and indirect object marker much earlier in Spanish than *pe* appears in Romanian for direct objects³, signaling no plausible common source of this type of marking for the two Romance languages (if one wanted to argue a common influence in developing DOM, say from Latin or some common vulgar ancestor).

Finally, when used independently of a clitic doubling construction, *a* has much broader scope than *pe*, being used with all types of direct objects, often

³ The earliest written Spanish is from the *Cantar de mio Çid*, from sometime between 1195 and 1207, which is full of both direct and indirect object marking uses of *a*. On the other hand, the earliest written Romanian is from 1521 ("Neacșu's letter"), which unfortunately does not happen to contain any tokens of *pe* or any instances of doubling. Thus, it would be impossible to state with certainty that Romanian had a Balkan type of unmarked doubling prior to 1521, or when marked doubling started to appear. However, it is likely that it had at least unmarked doubling given the status of other Balkan languages around that time, and given the shape of other contemporary Romanian languages, such as Istroromanian (cf. Zegrean 2012).

including inanimate ones. *A* without clitic doubling is also much more widespread in modern Spanish than *pe* without clitic doubling is in Romanian, a fact that is indicative of the different levels of entrenchment of the construction in the two languages.

Nevertheless, there are also many similarities between the two markers. Both *a* and *pe* tend to be used with animate, mostly human, direct objects, although there are, as mentioned, notable instances of non-human direct objects in Spanish. The exceptions, however, can be explained in accordance with principles of metonymy. For instance, both Spanish and Romanian allow DOM with animate referents other than humans, such as animals. This can be seen as an extension of DOM along the animacy scale to include non-human animate referents, while human referents remain the prototype, revealing an instance of category metonymy (Dancygier and Sweetser 2014). Further, Spanish, but not Romanian, also allows DOM with human-occupied regions, such as in the following examples:

- (7) a. *Debemos invadir a Namibia.*
'We should invade Namibia.'
- b. *Tuvieron que tomar a la ciudad.*
'They had to take the city.'
- (Kliffner 1984:196)
- c. *Yo lo quiero a mi país.*
'I love my country.'
- (Estigarribia 2006:124)

In these cases, the direct object metonymically refers to human referents. That is, when one invades a country or takes a city, one is in fact conquering the human inhabitants of those locations. Romanian does not allow this type of metonymic extension of *pe*-marked direct objects, and the Romanian versions of (7a-c) would not be grammatical; however, as mentioned, *pe* in Romanian is much younger than *a* in Spanish, and thus we can see this extension in Spanish as having had more time to develop. Further, given that in these cases the direct object is a location, this goes well with the directional motion sense of *a*, giving the sense that these locations are the endpoints of motion.

It has been argued that personal *a* acts as a marker of individuation, casting its function more in terms of construal than in terms of pragmatics (Kliffner 1984). Individuation, in Hopper and Thompson's terms (1980) is defined as the conceptual distinctness of the object from the subject and from the background. Individuation also singles out animacy as a parameter, among several, with animate direct objects being more highly individuated than inanimate ones. If we take as a basis the hypothesis that DOM in Spanish spread from indirect object marking, then the mandatory humanness of the direct object comes naturally, since this is a general requirement of indirect objects. On this basis, though, the

path by which the humanness requirement came about is very different for Spanish and for Romanian: for Spanish, the prototypical object to be marked is the indirect object (necessarily human), and there was a spread to direct objects, while in Romanian the humanness requirement was imposed by the verb, since *pe*-marking started out as an adjoined prepositional phrase with intransitive sentences featuring verbs of high-affectedness pertaining specifically to human interaction.

4 Propositional functions

The similarities and differences detailed above between the two languages with respect to DOM are important to note when considering the pragmatics of the CD-DOM construction (which we call CD-*pe* in Romanian and CD-*a* in Spanish, and throughout CD-DOM to refer to both). Individually, in both languages, the CPC and the DOM and their respective semantic specifications discussed above would constitute, in Traugott's subjectification transitions, the more propositional meanings of these constructions. On the other hand, the information-structuring and discourse-structuring functions, such as prominence and topicality (Kallulli 2008), referential anchoring (von Heusinger 2002), referential persistence (Chiriacescu & von Heusinger 2009), and high accessibility (Ariel 1988) all come to make up its textual and expressive meanings.

Although CD-*pe* has come to acquire expressive functions, nevertheless there are several contexts in which, I will argue, the CD-*pe* construction can still be considered as fulfilling a propositional function. These are the cases in which CD-*pe* is mandatory and, no non-CD-*pe* variant is available. When the direct object is a pronoun, a bare noun, a proper name, a bare demonstrative, a demonstrative adjective, a bare superlative, or a bare quantifier, the CD-*pe* construction is mandatory, and there is no variant available without CD-*pe*:

- (8)
- a. Pronouns
*Președintele *(m-)a salutat *(pe) mine.*
'The president greeted me.'
 - b. Bare nouns
*Președintele *(l-)a salutat *(pe) delegat.*
'The president greeted the delegate.'
 - c. Proper names
*Președintele *(l-)a salutat *(pe) Ion.*
'The president greeted Ion.'
 - d. Bare demonstratives
*Președintele *(l-)a salutat *(pe) acela.*
'The president greeted that one.'

- Partitive pronouns + adjectives
- e. *Președintele *(l-)a salutat *(pe) cel înalt.*
'The president greeted the tall one.'
- Bare superlatives
- f. *Președintele *(l-)a salutat *(pe) cel mai înalt.*
'The president greeted the tallest one.'
- Bare quantifiers
- g. *Președintele *(l-)a salutat *(pe) fiecare.*
'The president greeted each (of them).'

This is strikingly different from Spanish, in which by and large the CD-*a* variant alternates with the non-CD-*a* variant, notwithstanding regional and dialectal differences in distribution as well as variations in perceived degrees of acceptability. In this regard, for the specific set of direct objects detailed in (8), CD-*pe* has a grammaticalized non-pragmatic form (alongside the pragmatic usage to be detailed below in (9)). On the other hand, in Spanish use of CD-*a* with all non-pronominal direct objects is pragmatically-motivated (although the range of direct object types is more limited); this means that in Spanish there is no direct object type for which CD-*a* is obligatory, and that those direct objects with which it can occur can have alternatives, either as *a*-only or with no marking at all.⁴

Additionally, CD-*pe* also has a pragmatic usage: when the direct object in Romanian is a modified definite noun, an indefinite noun⁵, a proper name, a demonstrative NP, a superlative NP, or a quantified NP, the speaker has a choice between a CD-*pe* and a non-CD-*pe* variant on the same utterance, with different pragmatic effects:

- Modified definites
- (9) a. *Președintele (l-)a salutat (pe) delegatul înalt.*
'The president greeted the tall delegate.'

⁴ With the striking exception, cross-dialectally, of use with pronouns, with which CD-*a* is mandatory, and no *a*-only variant or zero-marking variant is available. This goes well with the hypothesis that CD-DOM has its beginnings in a high accessibility scenario, which pronouns fulfill because they are deictic, and thus maximally accessible. However, this obligatory use is isolated to personal pronouns, not all deictics, thus making it a constraint specific to pronouns.

⁵ By default an indefinite noun in a CD-*pe* clause takes a specific reading, or as others like to argue, a referential reading (Anagnostopoulou & Giannakidou 1995). The specificity disambiguation effects of CD with indefinites is thoroughly discussed in von Heusinger (2002) and Leonetti (2003), among others.

Subjectification in the development of clitic doubling

- Indefinites (with and without modification)
- b. *Președintele (l-)a salutată (pe) un delegat înalt.*
'The president greeted a tall delegate.'
- Demonstrative nouns
- c. *Președintele (l-)a salutată (pe) acel delegat.*
'The president greeted that delegate.'
- Superlative NPs
- d. *Președintele (l-)a salutată (pe) cel mai înalt delegat.*
'The president greeted the tallest delegate.'
- Quantified NPs
- e. *Președintele (l-)a salutată (pe) fiecare delegat.*
'The president greeted each delegate.'

In Spanish, the same alternations are available for this class of direct objects for CD-*a*, although their proportion in spoken language is quite lower than in Romanian.

Considering the data in (8) and (9) above, we are left to wonder what natural categories these two types of direct objects fall into. The types in (8) have in common the fact that most are fairly high on the accessibility scale by being deictic expressions. Especially for (8d-g), one needs to be present in the context of conversation to know who is being referred to. The types in (9) have in common several features: they are lower on the accessibility scale, they represent nominal expressions that are more semantically robust, and they can reconstruct referents in the minds of addressees without recourse to context. For this reason, the CD-*pe* construction is mandatory with the former and optional with the latter. This optionality allows the specialization of CD-*pe* as a construction that, when chosen by speakers, expresses more than the content of the sentence. In short, CD-*pe* has developed expressive functions.

5 Expressive functions

I define CD-*pe*'s expressive function primarily in terms of information structuring effects. In Traugott's terms, propositional meanings are those based in the external described situation, and when the meaning of a lexeme or construction changes, it changes from this latter type of meaning to one based on the internal described situation (evaluative/perceptual/cognitive) (1989:34). Traditionally, the focus has been on the meaning change of lexical items, such as modal auxiliaries and modal adverbs, and the findings reveal change in the degree of epistemicity, as well as shifts from the sociophysical domain to the emotional and psychological domain (Sweetser 1984, 1990).

In this work, I am exploring the semantic change involved in the development of larger multiword constructions with loci across the clause. Additionally, I would like to suggest that constructions whose sole purpose is to create informational asymmetry in the mind of the addressee, i.e., information-structure constructions, are constructions that by definition serve expressive purposes. When creating focus, for instance, the construction captures the speaker's relative prioritization of that piece of information over all others in the utterance. The CD-*pe* and CD-*a* constructions also bring with the additional speaker-internal meanings, such as the higher degree to which the speaker construes the direct object to be affected by the subject, the higher degree of importance the affectedness of the direct object has to the outcome of things, and the higher degree to which the speaker believes the content of the sentence matters to the rest of the conversation, or to the immediate stretch of talk. These are not properties of the lexical verbs, or of the compositional semantics of the sentence independent of the CD-DOM; rather, CD-DOM alone brings about these effects.

The sentences in (10) and the subsequent discussion illustrate the difference between CD-*pe* constructions and its alternatives:

- (10) a. *Președintele a convocat toți delegații.*
 president.DEF has assembled all delegates.DEF
- b. *Președintele i-a convocat pe toți delegații.*
 president.DEF CL.3PL.MASC-has assembled DOM all delegates.DEF
- c. *Președintele a convocat pe toți delegații.*
 president.DEF has assembled DOM all delegates.DEF
 'The president assembled all the delegates.'
- d. *Președintele i-a convocat.*
 president.DEF CL.3PL.MASC-has assembled
 'The president assembled them.'
- e. **Președintele a convocat îi /-i.*
 president.DEF has assembled pro.3PL.MASC / CL.3PL.MASC
 (As 'The president assembled them.')

(10b) has a prominence-raising effect relative to (10a). Thus, the relationship between (10a) and (10b) is one of constructional alternation: they are in complementary distribution with respect to pragmatic force. The difference between (10b) and (10c), on the other hand, is more subtle. The latter represents a remnant from a previous constructional stage and occurs less frequently than CD-*pe*. Its equivalent in Spanish, however, is much more common, because CD-*a* has not taken hold quite as firmly in Spanish, and thus sentences such as (10c) are

Subjectification in the development of clitic doubling

more often found in Spanish where in Romanian formulations such as in (10b) would be preferred where allowed.

The current analysis, which assumes the merger of CPC with DOM, suggests that the evolution of (10b) out of the merger of (10d) with (10c) represents the leap from the mere proposition-expressing status of *pe*-only constructions to the expressive status of CD-*pe*. In uttering (10c), a speaker is not necessarily providing an addressee with any information as to his relative prioritization of information, or the degree of importance the proposition may have, either in general, or to the future of the conversation. It also does not provide much in the way of indicating high affectedness on the direct object. This is because *pe*-marking that occurs with human direct objects is unsurprising and an unmarked form of expression (human objects are regularly *pe*-marked). In uttering the CD-*pe* version in (10b), however, a speaker has already given consideration to the relative degree of accessibility of the referent in the mind of the speaker. The additional doubling with a clitic goes above the expected differential marking of human objects, and represents a marked form.

From the point of view of the epistemic state of the speaker, when one uses CD-*pe*, one is also delivering information with much more certainty than otherwise, as one is relying on both one's own knowledge and the knowledge the addressee is assumed to share. Thus, there is certainty not about the truth of the state of affairs described, but about the degree to which the addressee is familiar with it, signaling a certain degree of intersubjectivity. This comes about most strikingly with specific indefinite noun phrases. *Pe*-marking has been documented most frequently in the literature using evidence from this type of direct object (Farkas 1978, Dobrovie-Sorin 1994, von Heusinger 2002, 2008). However, in the bigger picture presented in the current work, after contextualizing indefinites in a broader range of direct objects as laid out in (8) and (9), we see that indefinites achieve high specificity when coupled with CD-*pe* precisely because indefinites are naturally low on the accessibility scale, and CD-*pe* has the effect of rendering them more prominent, and hence more accessible. The semantic contrast between the accessibility level of indefinites and that of CD-*pe* is more easily noticed than it is when nominal expressions are higher up on the accessibility scale. Thus, indefinites achieve a specific reading with CD-*pe*, since specificity is nothing more than the certainty of the speaker that he and the addressee have mental access to the same referent (or at least, in producing the utterance, the speaker is achieving drawing the attention of the addressee to the referent). Thus, we explain specificity not with a specificity operator or feature present on indefinite DPs, but rather as a pragmatic effect that results when CD-*pe*, itself a meaningful grammatical construction imbued with a semantics and pragmatics, couples with a nominal expression low on the accessibility scale. The striking contrast between the high accessibility of one versus the low accessibility of another is what yields, in our eyes, the (epi)phenomenon that is specificity.

Because much of the time CD-*pe* prominence manifests as topicality, it carries with it the implication that more information is to follow pertaining to

some portion of the CD-*pe* sentence (which is not always necessarily the direct object referent, but may be the predicate, or the subject, as the textual analysis below will show). This is a feature of the spoken language, since the speaker has live feedback from the addressee as to their knowledge state, and as well is likely to know the addressee, and hence have access to his knowledge state. In the written language, these knowledge-structuring functions manifest as text-structuring functions, which will be detailed below.

5 Textual Functions

In this section I will detail the range of textual functions of the construction, which build on the expressive functions, using data from the modern language.

CD-*pe* is very common in news headlines. This trend is consonant with the function of CD-*pe* as signaling referential persistence, which is defined as “the number of occurrences of co-referential expression in the following text (Chiriacescu and von Heusinger 2009:2).” This idea is based on the notion of topic continuity by Givón (1981), in which the notion of topic is extended beyond the sentence and into the larger discourse. With this notion of topic continuity, we observe that news headlines serve to give a summary of the article to come, and pique the interest of the reader enough to want to read more. Thus, the CD-*pe* construction in headlines serves to establish the referent that is expected to persist in the body of the article.

- (11) *Culoarea apei din Marea Neagra se schimba. Fenomenul care i-a surprins pe specialiștii NASA.*

‘The color of the Black Sea is changing. The phenomenon that surprised NASA specialists.’

- (12) *Gluma hilară care l-a facut pe un bărbat din California să înoate disperat către țarm.⁶*

‘The hilarious joke that made a California man swim frantically towards the shore.’

These headlines establish topics to be continued in the body of the text. Chiriacescu (2010) also finds with experimental data that, when posed with a piece of text introducing a topic with CD-*pe*, that topic is more likely to be re-introduced several sentences later. Referential persistence is best illustrated in longer stretches of text. Below is a fragment from a news article from a Romanian news source describing a local murder case. The function of CD-*pe* as a text

⁶ <http://stirileprotv.ro/show-buzz/fun/gluma-hilara-care-l-a-facut-pe-un-barbat-din-california-sa-inoate-disperat-catre-tarm-video.html>

Subjectification in the development of clitic doubling

structuring device, introducing and sustaining discourse topics, is illustrated with the example in Text 1⁷:

Text 1 News story

- | | | | |
|---|--|--|---|
| 1 | <i>Un tânăr de 30 de ani din localitate <u>l-a omorât pe un bărbat de 48 de ani</u> chiar într-unul din cimitirele din Axente Sever.</i> | <i>A young man of 30 from that town <u>killed a 48-year-old man</u> right in one of the cemeteries of Axente Sever.</i> | CD- <i>pe</i> establishes direct object as discourse topic, likely to become subsequent subject |
| 2 | <i><u>Bărbatul care a fost omorât</u> era din Țapu și venise la Axente Sever. Criminalul se întorcea acasă, iar la un moment dat a izbucnit un conflict între cei doi.</i> | <i><u>The man who was killed</u> was from Tapu and had come to Axente Sever. The killer was returning home, and suddenly conflict erupted between the two men.</i> | Previous direct object now subject |
| 3 | <i>Tânărul de 30 de ani i-a aplicat mai multe lovituri, <u>l-a înjunghiat, după care l-a târât în cimitir și l-a lovit cu capul de una dintre cruci.</u></i> | <i>The 30-year-old man delivered several blows (to him), stabbed him, and dragged him into the cemetery and hit his head against a stone cross.</i> | Referential persistence: topic sustained with CPC |
| 4 | <i>Chiar mama criminalului a fost cea care <u>l-a descoperit pe bărbatul omorât și a sunat de urgență la Poliție.</u></i> | <i>It was the killer's mother who <u>discovered the killed man and called the police.</u></i> | Topic refreshed with additional use of CD- <i>pe</i> |

The discourse structuring effects achieved by the ‘CD-*pe* sandwich’ in the text above are typical of the news genre. Authors of news stories make initial assumptions about the knowledge status of the reader, and take this assumption as a starting point for the presentation of the content of the text. The use of reference strategies that manipulate accessibility is one way to achieve this strategic presentation. The reader is not assumed to know anything about the topic beforehand (a striking difference from CD-*pe* use in conversation), but the first CD-*pe* is establishing the topic and promising relevance later in the story.

In this sample text, it is CD-*pe* which is introducing and more importantly sustaining the discourse topic. This matches the findings of Chiriacescu (2010), who finds that topics are more likely to be sustained when introduced with CD-*pe*. Here, we see that CD-*pe* has both topic-establishing and topic-sustaining functions. As such, CD-*pe* is fulfilling the textual functions predicted by the subjectification cline, as proposed in the current work.

⁷ <http://www.tribuna.ro/stiri/eveniment/11-ani-de-inchisoare-pentru-criminalul-din-cimitirul-de-la-axente-sever-86901.html>

6 Romanian and Spanish CD-DOM in comparative perspective

The resulting construction from the unification of CPC and DOM into CD-DOM, while present in both Romanian and Spanish, is quite different functionally and distributionally. Firstly, *CD-pe* is much more common in Romanian than *CD-a* is in Spanish, regardless of dialect or genre. When they do occur in Spanish, however, the pragmatic implications detailed above for Romanian are also present. In Spanish also, the diachronic conditions were ripe for the formation of a *CD-a* construction once *a* hypothetically spread from indirect to direct objects. While it developed along the same subjectification cline, it never reached the full textual structuring function it did in Romanian; that is, the same referential persistence phenomenon cannot be achieved with *CD-a* as was detailed for *CD-pe* in Section 5. *CD-a* use in Spanish is relatively more unpredictable (no texts of the kind illustrated for Romanian can be found as easily, although abundant in Romanian), and much of the use of *CD-a* is still subject to dialectal variability in the many Spanish varieties and to acceptability judgments. When it appears, however, (usually in spoken conversation), it serves the same expressive functions as in Romanian: an indication of high degree of shared accessibility to the referent between speaker and addressee, a raised degree of affectedness on the direct object, and a raised degree of importance or salience attributed to the event expressed by the sentence. This suggests that, although older, *CD-a* has not yet come to acquire the same textual function in Spanish.

A comparative dialectal study of Spanish is needed, with access to both spoken and written texts, to fully understand the breadth of pragmatic development of the construction. Another important phenomenon of interest, which occurs in some varieties of Spanish, but not in Continental Spanish, nor in Romanian, is that of unmarked clitic doubling, such as the following:

(13) *Yo lo voy a comprar el diario justo antes de subir.*

‘I will buy the newspaper just before coming upstairs.’

(Porteño Spanish (Argentina), Belloro 2007:16, from Suñer 1988)

This type of doubling seems to occur only with inanimate direct objects, keeping *a*-marking exclusively for animate objects. This construction, however, may have arisen independently, and possibly stems from an object right-dislocation construction instead of an object-marking strategy. More investigation is needed to see if the intonational contour is consistent with that of right-dislocation, and also to see just how widespread this construction is. It is worth tracing the origins of this construction as well, since in most discussions this construction is simply thought of as a variant of *CD-a*.

7 Conclusions and future directions

We have explored Traugott’s concept of subjectification in the process of historical construction formation in Romanian, while also considering an

Subjectification in the development of clitic doubling

analogous construction in Spanish. The study tracked the make-up of a new construction, which developed expressive and textual functions from two component constructions, CPC and DOM, each of which continue to exist in the languages and are grounded in propositional, content-expressing functions. An argument was made for the consideration of information structure constructions as innately expressive, given that they reflect the inner world of the speaker, namely, the speaker's relative prioritization of information for the outcome of the events described. For future research, a broad corpus study encompassing both spoken and written data is called for in order to discern the precise degree of entrenchment of CD-*a*. A comparative historical study would be highly valuable, considering that the literature offers little treatment of the semantics and pragmatics of this construction, and still treats it as surface manifestations of the same functional syntactic projections of clitics. This exploration of CD-DOM constructions has given a glimpse into the subtler semantic and apparent distributional differences in the usage of the construction in two Romance languages.

References

- Ariel, Mira (2001). Accessibility theory: An overview. In T. Sanders, J. Schilperoord & W. Spooren (Eds.), *Text representation: Linguistic and psycholinguistic aspects* (29-87). Amsterdam/Philadelphia: John Benjamins.
- Ariel, Mira (1990). *Accessing noun-phrase antecedents*. London: Routledge.
- Ariel, Mira (1988). Referring and accessibility. *Journal of Linguistics*, 24-1.65-87.
- Belloro, Valeria A. (2007). *Spanish clitic doubling: A study of the syntax-pragmatics interface*. PhD dissertation, State University of New York, Buffalo.
- Bossong, Georg (1998). Le marquage de l'expérient dans les langues d'Europe. In J. Feuillet (Ed.), *Actance et valence dans les langues de l'Europe* (259-294). Berlin/New York: Mouton de Gruyter.
- Bossong, Georg (1991). Differential object marking in Romance and beyond. In D. Kibbee, & D. Wanner (Eds.), *New analyses in Romance linguistics: Selected papers from the Linguistic Symposium on Romance Languages XVIII, Urbana-Champaign, April 7-9* (143-170). Amsterdam/Philadelphia: John Benjamins.
- Bossong, Georg (1985). Markierung von Aktantenfunktionen im Guaraní: Zur Frage der differentiellen Objektmarkierung in nicht-akkusativischen Sprachen. In F. Plank (Ed.), *Relational typology* (1-29). Berlin: Mouton.
- Bybee, Joan (2003). Cognitive processes in grammaticalization. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure vol. 2* (145-167). Psychology Press.

- Chiriacescu, Sofia (2010). Topic shift and *pe*-marking in Romanian. Handout, *Specificity from empirical and theoretical points of view*, University of Stuttgart.
- Chiriacescu, Sofia & Klaus von Heusinger (2009). *Pe*-marking and referential persistence in Romanian. In A. Riester & E. Onea (Eds.), *Focus at the syntax-semantics interface, Working papers of the SFB 732, Vol. 3*. University of Stuttgart.
- Dancygier, Barbara & Eve Sweetser (2014). *Figurative language*. Cambridge: Cambridge University Press.
- Dobrovie-Sorin, Carmen. 1994. *The Syntax of Romanian*. Berlin: Mouton de Gruyter.
- Estigarribia, Bruno (2006). Why clitic doubling? A functional analysis for Rioplatense Spanish. In T. Face & C. Klee (Eds.), *Selected Proceedings of the 8th Hispanic Linguistics Symposium* (123-136). Somerville: Cascadilla.
- Farkas, Donka (1978). Direct and indirect object reduplication in Romanian. In *Papers from the 14th Regional Meeting of the Chicago Linguistics Society (CLS)*, 14.88-97.
- Gabriel, Christoph, & Esther Rinke (2010). Information packaging and the rise of clitic-doubling in the history of Spanish. In G. Ferraresi & R. Lühr (Eds.), *Diachronic studies on information structure: Language acquisition and change* (63-86). Berlin: Mouton de Gruyter.
- Gerlach, Birgit. (2002). *Clitics between syntax and lexicon (No. 51)*. John Benjamins.
- Givón, Talmy. (1981). On the Development of the Numeral 'one' as an indefinite Marker. *Folia Linguistica Historia*, 2.35-53.
- Goldberg, Adele (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.
- Hill, Virginia & Liliane Tasmowski (2008). Romanian clitic doubling: A view from pragmatics-semantics and diachrony. In D. Kallulli & L. Tasmowski (Eds.), *Clitic doubling in the Balkan languages* (227-255). Amsterdam/Philadelphia: John Benjamins.
- Hopper, Paul J. & Sandra A. Thompson (1980). Transitivity in grammar and discourse. *Language*, 56-2.251-299.
- Kallulli, Dalina (2008). Clitic doubling, agreement and information structure: The case of Albanian. In D. Kallulli & L. Tasmowski (Eds.), *Clitic doubling in the Balkan languages* (227-255). Amsterdam/Philadelphia: John Benjamins.

Subjectification in the development of clitic doubling

- Kliffner, M. D. (1984). Personal *a*, kinesic and individuation. In *Papers from the XIIIth Linguistic Symposium on Romance Languages (195-216)*. Amsterdam: Benjamins.
- Lambrecht, Knud (2004). On the interaction of information structure and formal structure in constructions: The case of French right-detached *comme-N*. In M. Fried & J.-O. Östman (Eds.), *Construction Grammar in a cross-language perspective*. Amsterdam/Philadelphia: John Benjamins.
- Langacker, Ronald W. (2007). Constructing the meanings of personal pronouns. In G. Radden (Ed.), *Aspects of meaning construction (171-187)*. Amsterdam/Philadelphia: John Benjamins.
- Langacker, Ronald W. (2006). Subjectification, grammaticization, and conceptual archetypes Ronald W. Langacker. In, A. Athanasiadou (Ed.), *Subjectification: Various paths to subjectivity (17-40)*. Berlin/New York: Mouton de Gruyter.
- Langacker, Ronald W. (2002). *Concept, image, and symbol*. Mouton de Gruyter.
- Langacker, Ronald W. (1996). Conceptual grouping and pronominal anaphora. *Typological Studies in Language*, 33.333-378.
- Langacker, Ronald W. (1990). Subjectification. *Cognitive Linguistics*, 1-1.5-38.
- Langacker, Ronald W. (1987). *Foundations of Cognitive Grammar: Theoretical prerequisites, Volume 1*. Stanford University Press.
- Lyons, Christopher (1990). An agreement approach to clitic doubling. *Transactions of the Philological Society*, 88-1.1-57.
- Lyons, (1999). *Definiteness*. Cambridge: Cambridge University Press.
- Næss, Åshild (2007). *Prototypical transitivity (Vol. 72)*. John Benjamins.
- Narrog, Heiko. (2012). *Modality, subjectivity, and semantic change: A Cross-linguistic perspective*. Oxford University Press.
- Noël, Dirk (2007). Diachronic construction grammar and grammaticalization theory. *Functions of Language*, 14-2.177-202.
- Nuyts, Jan. (2001). *Epistemic modality, language, and conceptualization: A cognitive-pragmatic perspective*. Amsterdam/Philadelphia: John Benjamins.
- Sanders, Ted & Eve Sweetser (Eds.) (2009). *Causal categories in discourse and cognition (Vol. 44)*. Walter de Gruyter.
- Sweetser, Eve (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure (Vol. 54)*. Cambridge: Cambridge University Press.
- Sweetser, Eve (1984). *Semantic structure and semantic change: a cognitive linguistic study of modality, perception, speech acts, and logical relations*. PhD dissertation. University of California, Berkeley.

- Traugott, E. C. (1997). Subjectification and the development of epistemic meaning: The case of promise and threaten. In T. Swan (Ed.), *Modality in Germanic languages* (185-210). Berlin/New York: Mouton de Gruyter.
- Traugott, Elizabeth C. (1995). Subjectification in grammaticalization. In D. Stein & S. Wright (Eds.), *Subjectivity and subjectivisation: Linguistic perspectives* (31-54). Cambridge: Cambridge University Press.
- Traugott, Elizabeth C. (1989). On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language*, 65-1.31-55.
- Traugott, Elizabeth C. (1982). From propositional to textual and expressive meanings: Some semantic-pragmatic aspects of grammaticalization. In W. P. Lehmann & Y. Malkiel (Eds.), *Perspectives on historical linguistics: Papers from a conference held at the meeting of the Language Theory Division, Modern Language Assn., San Francisco, 27-30 December 1979* (245-271). Amsterdam/Philadelphia: John Benjamins.
- Traugott, Elizabeth C. & Richard B. Dasher (2002). *Regularity in semantic change*. Cambridge: Cambridge University Press.
- von Heusinger, Klaus (2002). Specificity and definiteness in sentence and discourse structure. *Journal of Semantics*, 19-3.245-274.
- von Heusinger, klaus & Georg A. Kaiser (2010). Affectedness and differential object marking in Spanish. *Morphology*, 21.3-4. 593-617.
- von Heusinger, Klaus & Georg A. Kaiser (2005). The evolution of differential object marking in Spanish. In K. von Heusinger, G.A. Kaiser & E. Stark (Eds.), *Proceedings of the Workshop Specificity and the evolution/emergence of nominal determination in Romance, Arbeitspapier 119* (33-69). Fachbereich Sprachwissenschaft, Universität Konstanz.
- von Heusinger, Klaus & Edgar Onea-Gáspár (2008). Triggering and blocking effects in the diachronic development of DOM in Romanian. *Probus*, 20.67-110.
- Zegrean, I.-G. (2012). *Balkan Romance: Aspects on the syntax of Istro-Romanian*. Unpublished PhD dissertation, Università Ca' Foscari Venezia.

Oana David
University of California, Berkeley
Department of Linguistics
1203 Dwinelle Hall
Berkeley, CA 94720

oanadavid@gmail.com