ON THE (NON)ISOMORPHISM OF ROMANCE DEFINITE ARTICLES AND CLITIC PRONOUNS: A DIACHRONIC PERSPECTIVE*

Judy B. Bernstein
William Paterson University

Francisco Ordóñez Stony Brook University

Francesc Roca
Universitat de Girona

Abstract The observed isomorphism of Romance definite articles and $3^{\rm rd}$ person clitic pronouns, both typically l-forms descended from Latin ILLE, suggests a shared syntactic structure (Uriagereka 1995, Roca 1996). Less-studied Romance varieties such as Balearic Catalan, which displays s-articles descended from Latin IPSE as well as ILLE-derived $3^{\rm rd}$ person clitics, invite further consideration. The historical development, distribution, and function of the forms are accommodated via a multilayered structure for DP, one that encodes different features in a higher DP projection (for IPSE-derived s-articles) than in a lower one (for ILLE-derived l-clitics). Properties of these forms in the modern languages can be traced back to their Latin ancestors (Vincent 1997, Ledgeway 2012). In the uniform Romance languages (l-forms for articles and clitics), the same multilayered approach applies, though the evidence is masked.

^{*} The authors thank two anonymous reviewers for their insightful and constructive comments and one of the reviewers for help with the Old French, Old Sardinian and Old Catalan translations. We also thank Franck Floricic for help with the Old Sardinian translations. We are grateful for feedback from audiences at these conferences: 23rd Diachronic Generative Syntax (DiGS) at New York University (2022), 31st Colloquium on Generative Grammar (CGG) at Universitat de les Illes Balears (2022), and XXXe Congrés International de Linguistique et de Philologie Romanes (CILPR) at Universidad de La Laguna (2022). This research is partly funded by the research projects PID2021-123617NB-C42 and 2021SGR00483 in the case of the third author.

French

1 Introduction: Romance articles and clitic pronouns

Romance definite articles and accusative 3^{rd} person clitic pronouns (both descended from Latin ILLE) are typically isomorphic, as shown in the paradigms in (1) and the examples in (2)–(3). Examples (2 a) and (3 a) illustrate definite articles and (2 b) and (3 b) illustrate the isomorphic clitics.

(1)	definite articles		clitic pronouns	
	el, la, los, las	\longleftrightarrow	lo, la, los, las	Spanish
	el/lo, la, els/los, les	\longleftrightarrow	el/lo, la, els/los, les	Catalan
	le, la, les	\longleftrightarrow	le, la, les	French
	o, a, os, as	\longleftrightarrow	0, a, os, as	Portuguese

- (2) a. *La fille est arrivée tard*. '**The** girl arrived late.'
 - b. Je *la* vois. 'I see **her**.'
- (3) a. *Els* nois *han arribat tard*. Catalan 'The boys arrived late.'
 - b. *Els* hem vist. 'We saw **them**.'

This isomorphism is no accident, and several authors have proposed that the correspondence points to a shared syntactic structure, both elements corresponding to the same D head (e.g., Uriagereka 1995, Roca 1996).

In this paper we focus on Romance varieties that do not display the usual isomorphism of definite articles and clitic pronouns. As it turns out, all Romance languages we are aware of display clitics that developed from Latin ILLE. The same is not true for the definite articles. For example, Sardinian displays definite articles descended from Latin IPSE. The definite article and clitic paradigms for both languages are shown in (4); examples $(5\,a)$ and $(6\,a)$ illustrate definite articles in their syntactic contexts and $(5\,b)$ and $(6\,b)$ clitic pronouns in theirs.

- (4) <u>definite articles</u> <u>clitic pronouns</u> su, sa, sos, sas \longleftrightarrow <u>lu, la, los, las</u> Sardinian $es/so, sa, es/sos, ses^1$ Balearic Catalan el/lo, la, els/los, les \longleftrightarrow el/lo, la, els/los, les
- (5) a. *A connoskes su babbu de Maria*? Sardinian 'You know **the** father of Maria' (Jones 1993: 34)
 - b. **Lu** piket! 'Take it!' (Jones 1993: 28)
- (6) a. *Sa muntanya* vs. *la muntanya*. Balearic Catalan '**the** mountain' (specific) vs. '**the** mountain' (non-specific)
 - b. L'he vista.
 'I saw her.'

In this paper we explore the isomorphism, as well as its absence, between Romance definite articles and clitic pronouns. In section 2, we trace the historical development of definite articles and accusative clitic pronouns in Old French, Old Sardinian and Old Catalan and compare these with the patterns in the present-day varieties. In section 3, we take up the issue of the relationship between Latin demonstrative ille and the clitics and definite articles that developed from it, a path that is consistent across the Romance languages. We compare this to the relationship between ipse and the definite article that developed from it. Building on the articulated DP structure of Bernstein, Ordóñez & Roca (2021), we consider interpretive differences between forms in both the diachronic and synchronic data. We adopt an articulated DP structure that accommodates the range of synchronic and diachronic facts we have examined. We offer some concluding remarks in section 4.

2 Determiners and clitics and their evolution from Latin to Romance

Clitic pronouns and definite articles are two of the most important grammatical features that distinguish Latin from Romance languages. Latin lacked these two grammatical forms that are now present in all Romance languages.

¹ The masculine forms *so* and *sos* are only used after the preposition *amb* 'with'): *amb so càvec* ('with the hoe'), *amb sos cavalls* ('with the horses'). These forms are not general to Balearic Catalan: they are used mainly in Ibiza and in some areas of Mallorca, though not in Palma (see Veny 1999: 71–72; Veny & Massanell 2015: 193).

The definite article forms and the non-reflexive third person clitic pronouns have in common the fact that they originate from Latin pronouns. The distal demonstrative pronoun ILLE yielded the clitic pronouns and the definite *l*-article (*el*, *la*, etc.) of most Romance languages; the originally emphatic pronoun IPSE yielded the definite *s*-article (*es*, *su*, *sa*, etc.) that we find today in Sardinian and in some Catalan dialects, and that formerly was also present in linguistic varieties (e.g., Occitan, Gascon, Sicilian) of the southern Romania (see Vincent 1997, Ledgeway 2012, Aebischer 1948, among others).

Both the accusative clitic pronoun (*l*-forms) and the definite determiner (*l*-forms and *s*-forms) were present in early stages of Romance languages. However, the development in the use of each element presents an asymmetry: while the use of clitic pronouns has remained essentially the same from an interpretative point of view, the use of the definite article has undergone a remarkable evolution, progressively extending to new contexts (Roberts 2019) and, consequently, encoding different semantic and pragmatic values. In addition, while definite determiners had two sources in Latin, ILLE and IPSE, Romance clitics derive only from ILLE.

In section 2.1 we discuss the development of the determiners and in section 2.2 we discuss the development of the clitic pronouns.

2.1 ILLE- and IPSE-derived determiners

ILLE in classical Latin was a demonstrative and IPSE was an emphatic pronoun. Both ILLE and IPSE underwent a process of grammaticalization (van Gelderen 2011, among others) involving semantic, morphological, and phonological simplification and reduction that led to the modern l-articles in (7) and s-articles in (8):

- (7) a. ille > l-(pronoun) (definite article) b. el niño Spanish the.м boy
- (8) a. ipse > s-(emphatic pronoun) (definite article) b. su babbu de Maria the.м father of Maria

Examples of ILLE and IPSE can be found in Late Latin in the *Peregrinatio Aetheriae*. In this text ILLE and IPSE appear internal to a DP. In the example below, Ledgeway (2012: 92) points out that *illam montis* is most likely translated with

a demonstrative, 'that mount', and *ipsa summitate* is most likely translated as 'that-very summit', with an emphatic value:

```
ergo quarta pervenimus in summitatem illam
                                                           Late Latin
thus fourth arrived.1PL in summit
                                           that.F
montis Dei sancti Syna, ubi
                                    data
                                           est lex
                                                    (\dots)
mount God saint Sinai where given is
                                               law
persubissemus in ipsa
                              summitate
reach.1pl
               in that-very. F summit
'The fourth day we arrived at that summit of Mount Sinai, where law
is given (...) we reached that very summit.'
                                     (Peregrinatio Aetheriae; 4<sup>th</sup>-5<sup>th</sup> c.)
```

We assume that the examples in which IPSE and ILLE modify a noun, as in (9) above, represent an intermediate stage of these demonstrative pronouns on the path to becoming determiners. This is schematized as follows:

The path to definite article status included a change in the featural character of the forms. While IPSE lost its emphatic value, ILLE lost its deictic one in the transition to Medieval Romance. Some Medieval Romance languages followed the path of an ILLE-derived article, such as French, Italian and Spanish; others followed the path of an IPSE-derived article, such as Sardinian; and others maintained a mixed system, like Medieval Catalan. Both ILLE and IPSE forms retained their anaphoric value and were well-suited to becoming determiners. The development of the forms is schematized here in terms of formal features:

As a result of this evolution and as we show in (11)–(13) below, French became a uniform l-article language, Sardinian a uniform s-article language, and Catalan a language where s-articles and l-articles competed during the medieval period and still survive with different distributions in some dialects

(see Ledgeway 2012: 100–105):²

- (11) a. Elle no'nt eskoltet les mals conseliers Old French she not listened the.m.pl bad.pl counselors 'She didn't listen to the villains who advised her.'

 (Cantilène de Sainte-Eulalie, 5; 9th c.)
 - non la b. Niule cose pouret omque pleier / la polle none thing not her could ever bend the.F girl sempre non amast menestier lo Deo always not love-subj3sg the.м God.gen service 'Nothing could ever deter the girl from always loving the (Cantilène de Sainte-Eulalie, 9-10; 9th c.) service of God.' 'Nothing could ever persuade the girl not to continually love the service of God.' [Ayres-Bennett (1996: 32) translation]
- (12) a. et levate vois sas filias: Old Sardinian and take.2pl you.2pl the.f.pl daughters

 Furata integra et anbos sos filios

 Furata whole and both.m.pl the.m.pl sons

 'And you, take his daughters: Furata entirely and both her sons.'

 (CSNT, 164.4; 12th c.)
 - b. ca cun liuertu suo l'ockisit su seruu since with freedman his him-killed.3sc the.m servant uostru su seruu meu your the.m servant mine 'since with his freedman my servant killed your servant' (CSPS, 110; 11th-13th c.)
- (13) a. *Met.se en convenenza ab zos cavalers* Old Catalan put.refl in agreement with the.m.pl knights *e ab los homes* and with the.m.pl men

 'He gets into an agreement with the knights and the men.'

 (*Greuges dels homes de Sant Pere Graudescale*; 12th c.)

² The Old Sardinian examples are taken from Cennamo, Ciconte & Andriani (2022) and belong to three documents of two different varieties: Condaghe di San Pietro di Silki (CSPS) from the 11th–13th c. and Condaghe di San Nicola di Trullas (CSNT) fom the 12th c. in the case of Old Logudorese, and Condaghe di Santa Maria di Bonarcado (CSMB) from the 12th c. in the case of Old Arborense. All the Old Catalan examples are from documents contained in the Corpus Informatitzat del Català Antic (CICA).

b. des mas d'en Ramon de sa Scara sa of.the.м farm of-ART Ramon of the. F Scara the. F mas de Pere de sa (...) del Guardia of.the.м farm of Pere of the.F Guardia plow.duty sa jova the.f plow.duty 'from Ramon de sa Scara's farm, the plow duty (...) from Pere

de sa Guardia's farm, the duty of plowing'

(Capbreu de Castellbisbal; 12th c.)

In Old Romance the definite article was not systematically present in all the contexts where it is used today. Vincent (1997: 161-162) pointed out that, initially, the use of the article was favored in contexts such as the subject position and in establishing anaphoric relations (repeated items, discourse continuity). Dufresne, Tremblay & Déchaine (2018: 27) identify additional factors favoring bare nouns Old French: singularity, mass/abstract nouns, generic nouns and indefiniteness. In the above examples the definite article displays a more advanced stage where it can introduce nominal expressions that perform other syntactic functions (direct object, complement of preposition) and denote entities interpreted as definite in the context (with no need of an anaphoric relation), but it is not until the end of the medieval period that the definite article is consolidated in all the syntactic positions and semantic functions that characterize it now. The following examples show that the definite article was absent in contexts where it would generally be required today:³

- (14) a. Voldrent faire diaule servir Old French wanted.3pl her make devil 'They wanted to make her serve the devil.' (Cantilène de Sainte-Eulalie, 4; 9th c.)
 - b. verité que ama tant nëis aus the king liked so-much truth that not-even the Sarrazins...

Sarrazins...

'The king liked the truth so much that not-even the Sarrazins...' (Joinville; 13th c.)

Soz ciel n'at home qui puisset conforter. c. under heaven not-has man that can comfort 'There is no man under heaven who can comfort.'

(Alexis; 11th c.)

³ The original source of the Old French examples in (14b,c) was Anglade (1958: 159,177).

- (15) a. Ego deili ad iss ccoda de Old Sardinian а gave-him to him dom canyon of a mimi a isse deitimi balle Savina church and he ром valley Savina gave-me to me 'I gave him the canyon of the church and he gave me the Savina valley.' (*CSPS*, 276.1-3; 11th-13th c.)
 - b. Judice levedi a Ciçella et clesia ad Stephane judge took.3sg dom Ciçella and Church dom Stephane 'And the judge took Ciçella and the Church (took) Stephane.'

 (CSMB, 89.7; 12th)
 - c. Posit ad su prideru d'Arriora a kertare,
 put.pst.3sg dom the priest of-Arriora to quarrel.inf
 donnu Trogodori Coco, in corona de sinodu
 sir Trogodori Coco in assembly of synod

 '(On his behalf he) put the priest of Arriora, sir Trogodori Coco,
 to (lead the) quarrel at the assembly of the church council.'

 (CSMB, 178; 12th)
- (16) a. Totes questions que seran entre • ls Old Catalan issues all that be.fut.3pl between-the.м.pl habitadors de la ciutat sien ménades en lochs inhabitants of the.f city be.subj.3pl taken in places públics (...) e ve[n]dretsde cort по а casa public and not come.fut.2sg to house of court ne de battle nor of major

'All the disagreements between the inhabitants of the city should be dealt with in public places (...) and shall not come to the house of the court nor to the house of the major.'

(*Carta de franquesa*, 119; 13th c.)

b. ...ni quant o[m] póc creure que Déus nor however much man can believe that God fos home ni que verge aprés l'enfantament be.subj.3sg man nor that virgin after the-delivery estegués verge be.subj.3sg virgin

'...nor how many men can believe that God was a man nor that the Virgin remained a virgin after giving birth.'

(Vides de Sants Rosselloneses, 71; 13th c.)

c. *E adoncs Sen Gregori, per revelacion de Sant Esperit,* and then Saint Gregory by revelation of Holy Spirit conec que diable s' era mès en él knew that devil REFL-was put in him 'And, because of a revelation of the Holy Spirit, St. Gregory knew that the devil had gotten inside him.'

(Vides de Sants Rosselloneses, 307; 13th c.)

The kinds of nominal phrases that were interpreted as definite and that used to appear without a determiner include nouns denoting unique referents ((14 a,c) (15 a,b) (16 b,c)); abstract nouns ((14b)); nouns preceded by a universal quantifier ((16 a)); and nouns preceded by a preposition ((15 c)). The contrast between the older stages of the languages, where the definite article

contrast between the older stages of the languages, where the definite article was (or could be) absent, and the modern stages, where the definite article is required, can be observed by comparing (a) and (b) in the following examples:^{4,5}

- (17) a. *In figure de colomb volat a ciel* Old French in form of dove flew.3sg to sky/heaven

 'In the semblance of a dove, she flies to heaven.'

 (*Cantilène de Sainte-Eulalie*, 25; 9th c.)
 - b. Sous la forme d'une colombe, elle Modern French under the.f form of.a.f dove she s'envole au ciel REFL-is flown to-the sky/heaven 'In the semblance of a dove, she flew to heaven.'
- (18) a. Parthivi filos de Gavini Troccu Old Sardinian divided.1sg children of Gavini Troccu cun Jorgi de Tanar, a Maria et a Justa. with Jorgi from Tanar DOM Maria and DOM Justa 'I shared Gavinus Troccu's children, Maria and Justa, with Jorgi of Tanar.' (CSPS, 296. 1-2; 11th-13th c.)

⁴ Examples from the *Cantilène* are from Ayres-Bennett (1996). The Modern French translations of (17 a) and (17 b) are from de Nantes (1898).

⁵ We thank Franck Floricic for providing the Modern Sardinian example (18b).

- b. Paltei Modern Sardinian fidzas de sas divided.1sg the.f.pl children of Gavino Troccu, (a) Maria e Giusta, kin Giorgio Gavino Troccu dom Maria and Justa with Giorgio Tanar. from Tanar 'I shared Gavino Troccu's children, Maria and Justa, with Giorgio of Tanar.'
- (19) a. del remedi d'infern e de la Old Catalan of the avoidance of-hell and of the.f glòria de paradís glory of paradise

 'of the avoidance of hell and of the glory of paradise'

 (Homilies d'Organyà; 13th c.)
 - b. *del remei de l'infern e de* Modern Catalan of the avoidance of the-hell and of *la glòria del paradís* the.f glory of the paradise 'of the avoidance of hell and of the glory of paradise'

This means that the use of the definite article was progressively extended in Romance from its first uses – essentially linked by the discourse or pragmatic context to definite and specific identification – to its use with nouns denoting unique referents, as well as with abstract nouns. The latter types of nominals tended to appear, to a greater or lesser extent, without a determiner until the last medieval centuries in several Romance languages. There is, then, a hierarchy of contexts to which the definite article extended. Dufresne et al. (2018: 26–27), adapting previous work by Boucher (2005), gives the following hierarchy for the use of determiners in French:

(20) Demonstratives, possessives and quantifiers > Definite specific reference > Indefinite specific reference > Generic reference/non-specific/abstract nouns > Mass nouns

According to this hierarchy, at the very first stages (i.e., Gallo-Romance) the only nominal expressions obligatorily headed by an overt D element were those introduced by a demonstrative, a possessive, or a quantifier. In Old French the presence of determiners increased: nominal constructions with specific reference were introduced by a determiner, and generics, non-specific, abstract and mass nouns appeared systematically as bare NPs.

Finally, in Middle and Classical French an overt determiner is required with many types of nominals.^{6,7} A similar characterization is sketched by Ortiz Ciscomani (2009: 367–375) in relation to the definite article in Spanish: a) the definite article was initially used pragmatically to identify or individualize specific referents in the context; b) it started to introduce generic, abstract and mass nouns in the 12th c. and, primarily, from the 13th c. on; and c) it was generalized by the 15th c. Over time, overt articles will appear obligatorily in all of these contexts. This progression applied more generally across Romance, though there is some cross-linguistic variation in the endpoint of the progression.

Ortmann (2014), building on Löbner (1985, 2011; see also Schwarz 2013), characterizes such a hierarchy in terms of uniqueness and distinguishes 'pragmatic uniqueness' and 'semantic uniqueness', the former more likely to appear with overt definite articles than the latter. For Ortmann, in contexts of pragmatic uniqueness (when a definite article is more likely) a nominal expression's reference is unambiguous due to the linguistic 'context of utterance' (e.g., deictic or anaphoric nominal expressions). In contexts of semantic uniqueness (when a definite article is less likely), a nominal expression's reference is unambiguous due to the lexical semantics of the noun (e.g., unique nouns like 'the sun'). Ortmann's 'uniqueness hierarchy', as well as the hierarchy of Dufresne et al. (2018) above, are consistent with both the synchronic and diachronic facts concerning the appearance of overt definite articles in Romance languages. In other words, these hierarchies predict the presence vs. absence of a definite article at a particular stage of the language, as well as predict in which direction a language will change over time.

In summary, the emergence of determiners was made possible by the loss of deictic and emphatic features of ILLE and IPSE respectively. Determiners did not appear in all contexts systematically and were subject to various semantic and pragmatic conditions that differ from the modern uses. The emergence and evolution of clitic pronouns was very different, as we will see in the next section.

⁶ Dufresne et al. (2018: 42–43) shows that in 12th century Old French the frequency of object bare mass nouns remains more or less constant over time, while that of object bare abstract nouns actually increases.

⁷ Mathieu (2009: 152) ties the increase in frequency of determiners in Old French to the loss of overt interpretable φ -features of nouns.

⁸ See Heim (2011) for a historical overview for the semantics of definite articles. Heim assumes that semantic uniques like *the hospital* and *the sun* should be treated as having idiom-like properties. See Ludlow & Segal (2004) and Carlson, Sussman, Klein & Tanenhaus (2006).

2.2 Only ille-derived clitics

As mentioned above, none of the Romance languages developed an IPSE-derived object clitic. Even languages that had an IPSE-derived determiner, like Balearic Catalan and Sardinian, developed an ILLE-derived clitic. In the schema above in (10) IPSE has an emphatic feature that ILLE lacks. We take this feature of IPSE to be the reason for why no Romance language developed an IPSE clitic pronoun. Vincent (1997: 162) also points out that there is an underlying asymmetry between Latin ILLE and IPSE, the latter associated with second mention, topichood, and contrast, and the former with repeated mention and items "taken as a given." Along the same lines, Ledgeway (2012: 90) states that "IPSE, unlike ILLE, performed a topic-marking function (...) only picking out informationally prominent second-mention items (and hence an unsuitable candidate for the object clitic paradigm)." This difference in feature make-up is crucial to understanding the lack of IPSE clitics.

The following data from Old French, Old Sardinian, and Old Catalan show that ILLE-derived accusative clitics were already present in relatively early stages of these languages and that they served to identify, usually through an anaphoric relation, the referent of the direct object:

- (21) Buona pulcella fut Eulalia (...) Old French good nice-girl was Eulalia

 Voldrent la veintre li Deo inimi

 wanted.3PL her beat the.PL God enemies

 'Eulalia was a good girl (...) the enemies of God wanted to conquer her.'

 (Cantilène de Sainte-Eulalie, 1,3; 9th c.)
- (22) Nârrunmi « Homines amus Old Sardinian told.3pl-me that men have.1pL intro in Bulluso: a Iorgi de Contra et fugitos escaped into in Bulluso ром Iorgi of Contra and the. F muliere et filios suos. Iohanne Susanna. wife and two children their Iohanne and Susanna Batugitendelos!». Εt ego batusindelos. bring.2pl-loc-them and I brought.1sg-Loc-them 'They told me that: "we have people who run away, inside of Bulluso: Iorgi de Contra and his wife and their two children, Iohanne and Susanna. Bring them here!". And I did that.' (CSNT, 208; 12th c.)

⁹ The special nature of accusative clitic pronouns (morphological, syntactic, semantic and prosodic) is also captured in the tripartition of pronouns detailed in Cardinaletti & Starke (1999).

Old Catalan (23) et quod una fèmina l'avia fuit dit and was said that a woman him-had.3sg fuit presa. ad femina et Εt mort (...) et anaren killed and went.3PL to woman and was captured and mès-la in sua domo (...) e ipse P. de P. de Lobeira put.3sg-her in his house and that P. de Lobeira dis la trairia que ja по jutgament, а Lobeira said.3sg that yet not her bring.3sg to judgment que malauta era de febres. that ill was of fever

'and people said that a woman had killed him (...) and they went to the woman and captured her. And P. de Lobeira put her in his house (...) and P. de Lobeira said that he wouldn't bring her for trial yet because she was ill.'

(Abusos comesos a Lleida per Petrus de Lobeira, 91; 12th c.)

In all these examples, the accusative pronoun refers back to a discourse antecedent: in the Old French example, the antecedent is the proper name *Eulalia*, which denotes an inherently specific referent; in the Old Sardinian one, the clitic refers to the people who have just been mentioned (*Iorgi de Contra*, *issa muliere*, and *ii duos filios*, *Iohanne* et *Susanna*); finally, in the Catalan example, the referent is introduced into the discourse by means of the indefinite *una* (*una fémina* 'a woman') and, then, it is resumed in the discourse by the accusative clitic.

Significantly, clitic pronouns are also found referring back to bare NPs, i.e., referring to nominal expressions that are not introduced by any determiner, as in (24)–(25):

(24) *Iudicarunmi ad testimonios et ego* Old Sardinian ordered.3pl-me dom witnesses and I batusilos et vinkilu brought.1sg-them and defeated.1sg-him

'And they ordered me to provide witnesses and I brought them and defeated him.'

(CSNT, 211.8-9; 12th c.)

(25) Et de les chasteles che Old Catalan vós mi and of the.pl castles that you me comendaretes, potestate no vosende vetaré, ni authority not you from-it deprive.fut.1sg nor entrust vetare non **la** vos ende faré, ni de dia deprive not it you from-it do.fut.1sg nor of day nor of nuit night

'And I won't deprive you of your authority on the castles that you entrust to me during day or night, nor will I deprive you of it.'

(Jurament feudal de fidelitat; 11th c.)

This indicates that the appearance of the accusative clitic pronoun did not correlate with that of the definite article: while the article follows a progression with various steps sensitive to the content of the nominal expression, the clitic pronoun is employed as an identifier of a grammatical function (direct object) regardless of whether or not the noun phrase with which it is co-referent is introduced by a determiner.

The fact that in languages with *s*-articles, such as Sardinian and Catalan, no cases of clitic pronouns with *s*- are documented supports the hypothesis that the introduction of accusative clitic pronouns and definite articles (and also strong pronouns) followed different patterns. The following examples show clitic pronouns in contexts with several different forms of definite articles and strong pronouns:

- (26) a. a llarga la levait su servo tuo Old Sardinian away her took.3sg the.m servant your sa coliberta mea the.f servant mine

 'Your servant took away my servant.' (CSPS, 111; 11th-13th c.)
 - b. ...cun ken la aveat ad issa scu. Petru with whom her had.3sg ром she saint Petru '...with whom saint Peter had her' (Lit.: with whom her had her saint Petrus) (CSPS, 28.4; 11th-13th c.)
- (27) Et monstre Remon **ipso** termine ke Old Catalan and shows Remon the limit that proclame per Ribeles, que est de Cion ad enlà. claims for Ribelles that is from Sió towards there so ample lo monstre de là (...) ad riba de Scio. (...) E wide it shows from there to bank of Sió and

ke sia assemblament de ipso epíscopo si en so that be.subj.3sg in the opinion of the bishop if ipsos omines quin faciant **ipsa** batalla, dicat ipso say.subj.3sg the who make.subj.3pl the battle el primers ipso fromiment (...) o.l epíscopo si pharà п if do.fut.3sg in the first.pl bishop the supply or-it rechulirà de Remon Gonball get.fut.3sg from Remon Gonball 'And Remon shows the limit that he claims for Ribelles, which is from Sió on, so wide does he show it as from there (...) to Sió river. (...) And let it be in the bishop's opinion if (...) the men who fight in the battle, let the bishop decide if he will get the supply from the first

ones (...) or he will get it from Remon Gonball.'
(Disputació referent als límits entre Guissona i Ribelles, 65; 11th c.)

In the Old Sardinian examples the use of the clitic pronoun is closely related to the DP *sa coliberta mea* ('my servant') in (26 a) and to the strong pronoun *issa* ('she') in (26 b) in a kind of clitic doubling construction with right dislocation of the DP, according to Cennamo et al. (2022). If the evolution of the pronoun IPSE in this language had been parallel to that of clitic pronouns, definite articles and strong pronouns, it might be expected that in cases such as these the three entities would be isomorphic. In the Old Catalan example there are plenty of determiners derived from IPSE (most in a Latin form *ipso/ipsa*, but also in the reduced form *so* of the Catalan article) and one instance from ILLE (*el primers* 'the first ones'), but the clitic pronoun is consistently realized as an *l*-form in this excerpt and throughout the whole document.

To sum up, in this section we have provided evidence of definite articles and accusative clitics in three Romance languages of the Middle Ages (Old French, Old Sardinian, and Old Catalan). The clitic pronoun follows paths different from those of the definite article. Whereas the definite article is sensitive to the nature of the noun it introduces, the semantic interpretation of the nominal expression, and the syntactic function, the clitic pronoun seems to act instead as a function marker linked to resumption of an argument in direct object position (see Vincent 1997: 161).

3 Romance DPs: the syntax and a semantic mapping

In this section, we consider the structural implications of the differences between definite articles and accusative clitic pronouns. To that end, we adopt Bernstein et al.'s (2021) complex DP structure and map the different functions and interpretations onto it, and then we extend this analysis to incorporate the clitic pronouns.

3.1 An articulated structure for DPs: definite articles vs. clitic pronouns

Present-day Balearic Catalan has both the *s*-definite article (from Latin IPSE) and the *l*-definite article (from Latin ILLE). The first one is found in anaphoric and deictic contexts and with relative clauses, contexts associated with topic-hood (Vincent 1997: 162), and the second one with uniques and what Lyons (1999: 162) describes as "identifiable in the wider situation." In the following examples we show that the *s*-article introduces a noun that heads a relative clause, as in (28 a), its unambiguous reference related to its syntactic structure due to the 'context of utterance' (i.e., its syntactic or semantic context; Ortmann 2014: 294); a noun phrase interpreted anaphorically with the indefinite *un rem* ('an oar'), as in (28 b); and a noun phrase interpreted deictically, as in (28 c).

- (28) a. Sa muntanya que veus és molt alta. Balearic Catalan the.f mountain that see.2psg is very tall.'
 - b. Això és un rem (...) I seguí camí envant this is a.m oar and continued.3sg path ahead amb es rem... with the.m oar

 'This is an oar (...) and he went ahead with the oar...'

(Veny 1999: 156)

c. Veus sa casa d'allà? see.2sg the.f house of-there 'Do you see the house over there?'

In contrast, *l*-articles introduce uniques, their reference understood between speaker and hearer due to shared experience, as is the case of nouns like those in (29 a) and *la mar* ('the sea') in (29 b):

- (29) a. *el cel, la terra, la muntanya, l'hospital, el re* Balearic Catalan 'the sky', 'the earth', 'the mountain', 'the hospital' 'the king'
 - b. *No en sabrà res de la mar* no of-it know.fut.3sg nothing of the.f sea 'S/he won't know anything about the sea.'

Thus, the Balearic Catalan *l*-article is used when "the hearer's general knowledge or knowledge of the wider situation and of appropriate associations is sufficient to identify the referent" (Lyons 1999: 163, discussing the parallel case of Fering definite articles). These are the same contexts that

align with Ortmann's (2014) 'semantic uniqueness' (i.e., independent of pragmatic/discourse or syntactic configuration, which would instead align with 'pragmatic uniqueness'). An overview of syntactic or semantic contexts favoring one or the other form of definite article is provided in (30) and the structure adopted to accommodate the data is provided in (31) (additional data and discussion are found in Bernstein et al. 2021):¹⁰

- (30) a. *s*-articles: individuated common nouns, nouns modified with relative clauses or adjectives, prepositional complements
 - b. *l*-articles: uniques, 'shared experience' nouns, proper names, pronouns

(31)
$$[_{DP1}[_{D1} s-....[_{DP2}[_{D2} l-[_{NP} N]]]]]$$

In that work the authors pursue a multilayer approach to DPs, along the lines of Zamparelli (1995) and Ihsane (2008). However, Bernstein et al. add an extra layer, DP2, for definite articles corresponding to unique referents. Thus, it is claimed that the ILLE-derived article (i.e., the l-article) in Balearic Catalan is a suitable candidate for D2 and that the IPSE-derived article for D1. The evidence for the claims were based on syntactic distribution, (non)adjacency to the noun, interpretive patterns, as well as instances of co-occurrence of D1 and D2 elements. 11

How do clitics fit into the structure in (31)? We address this issue next.

3.2 The clitic pronoun in the DP system

On the basis of data from Old Catalan, Bernstein et al. (2021: 22) shows that in early Romance both IPSE- and ILLE-derived articles could occupy D1, but only ILLE-derived articles could occupy D2, and only to a limited extent. In other words, IPSE forms are found in D1 only; ILLE forms are found in D1 and to a lesser extent in D2, which was often null:

(32)
$$[_{DP1} [_{D1} s-/l-..... [_{DP2} [_{D2} (l-).....]]]]$$
 Early Romance

- (i) es conco en Toni
 the uncle ART Toni
 'Uncle Toni'

 Balearic Catalan
- (ii) es l'amo the the-boss

Moreover, they motivate movement of DP2 to DP1 when the nominal expression is an argument.

¹⁰ The contexts of $(30 \, a)$ would align with Ortmann's 'pragmatic uniqueness' and those of $(30 \, b)$ with his 'semantic uniqueness'.

¹¹ Bernstein et al. (2021: 7) provide evidence that D1 and D2 elements can co-occur:

In section 2 above, we observed these patterns and also provided examples (recall (21)–(23)) illustrating accusative l-clitics in early Romance. We will propose that these l-clitics occupy D2. The claim then is that D1 is contemporaneously filled with the definite article (ILLE- or IPSE-derived form) and D2 is filled with an ILLE-derived form, either an l-clitic or, to a lesser extent, an l-article. Interestingly, the D2 l-clitic has continued to function continuously up until the present time.

We interpret the continuity of the 3rd person accusative *l*-clitic as evidence of it being established early in Romance. The question we must consider is whether there is empirical evidence that connects D2 determiners and D2 clitics. We believe there are phonological and pragmatic reasons for assuming such a connection.

At the phonological level, Balearic Catalan D2 l-articles and clitics differ in important respects from D1 s-articles. Whereas D2 l-articles and clitics resist stress and emphasis, D1 s-articles may be stressed. Observe these contrasts between s-articles (33 a), and l-articles and clitics (33 b,c): 12

- (33) a. *SA casa* (D1 article; deictic/anaphoric) Balearic Catalan 'THE house'
 - b. **LA muntanya* (D2 article; unique) Balearic Catalan 'THE mountain'
 - c. **LA veuré* (D2 clitic) Balearic Catalan it.f will see

In other words, only the *s*-forms, as D1 elements may be stressed; the *l*-forms, whether definite articles or clitics, may not be stressed.

Semantically, we have already observed that clitics can refer back to nominal expressions that are full DPs or even bare NPs (recall (24)–(25)). This is because D2 clitics, like personal pronouns, are anaphoric by their nature and independent of the pragmatic context or syntactic configuration they appear in (Ortmann 2014, Löbner 2011). In other words, they embody Ortmann's no-

- (i) Je lui ai dit. (< je la lui ai dit) present-day French 'I to-him/her has said it.'
- (ii) Tient une chartre, mais ne li puis tolir. (< le lui) Old French 'He has a charter, but I cannot take it from him.' (*La Vie de St. Alexis*, 11th c.)

¹² Another relevant aspect of Romance clitics, arguably linked to their status as D2 elements, is the fact that the $3^{\rm rd}$ accusative l-clitics may fail to be expressed in contexts where they would co-occur with dative $3^{\rm rd}$ person clitics. We illustrate with present-day French and Old French (examples from Anglade 1958: 177):

tion of 'semantic uniqueness', akin to proper names. Across most of present-day Romance, both pragmatic uniqueness and semantic uniqueness are associated with l- for both definite articles and accusative clitics. Previous analyses appealed to this isomorphism to claim that both the l-articles and clitics correspond to a unique D head. However, Balearic Catalan shows that this cannot be the case. This is because the language makes use of a morphological distinction that coincides with a semantic distinction: the s-definite article signals pragmatic uniqueness (i.e., D1); l- signals semantic uniqueness (i.e., D2), either as a definite article or as a clitic pronoun. Thus, the articulated DP we adopt (recall (32)) more accurately captures the distribution of Romance articles and their relationship to accusative clitics.

At this point one may wonder why IPSE was not suitable for D2 in the early stages of Romance. As indicated by Ledgeway (2012: 90) IPSE, contrary to ILLE, is emphatic and has a topic-marking function. Latin ILLE, on the other hand is not emphatic and lacks a topic-marking function. Instead, it "is used when an item is repeated in discourse" and "is taken as given" (Vincent 1997: 162). D2 does not support emphatic elements and therefore, no IPSE-derived form could emerge in D2 in early stages of Romance varieties. By the time IPSE lost its emphatic nature, the system had already been set for ILLE in all Romance languages. ILLE, crucially, was not emphatic.

Definite articles do not play a direct role in identifying arguments, but rather in pinpointing referential properties of a nominal expression. Clitic pronouns, in contrast, do identify arguments. Anaphoric properties of Latin ILLE inherited by the pronominal clitics need to be expressed overtly. We adopt Panagiotidis's (2002: 64) proposal that clitics are D elements that take *pro* as a complement (see (34 a)) and that determiners are D elements that take an overt complement (see (34 b)).

As we observed earlier, Romance 3rd person clitics derived from ILLE appeared in documents as early as the 9th c. in French and the 12th c. in Catalan and did not change substantially once established. This contrasts with the development of the determiner system across Romance languages. Definite articles appear early in contexts of pragmatic identification and specific reference, at approximately the same time as the clitic pronouns. However, unlike the clitic pronouns, their appearance extended to new contexts (e.g., uniques,

abstract nouns) over a period of time. In other words, new contexts requiring definite articles evolved relatively slowly, unlike the case of clitic pronouns. Moreover, the endpoint of the extension of definite articles varies cross-linguistically across Romance, with some languages (e.g., French) employing them in a greater variety of contexts than other languages (e.g., Italian). The clitic pronouns have been relatively stable and parallel across Romance.¹³

The observations of the last paragraph are nicely illustrated by the patterns in Sardinian, a language with ille-derived clitics and ipse-derived articles. As in other Romance languages, the pronominal l-clitics and s-articles appeared early ($11^{\rm th}$ c.). Over time, these D1 s-articles extended to D2 contexts (uniques, etc.). As the s-article extends to new contexts, it loses its emphatic feature. Once extended to D2 contexts, why doesn't the s-form displace the pronominal l-clitics? We conjecture that it is now too late to displace the pronominal l-clitics, which were well established in D2 across Romance from the outset. Put another way, the historical data suggest to us that there is a 'critical period' for establishing clitics after which they do not change substantially.

4 Conclusion

In this paper we have examined the development of Romance accusative clitic pronouns and definite articles. We aimed to shed light on a number of interrelated phenomena: a) the often-observed isomorphism between Romance definite articles and accusative pronominal clitics; b) the etymological origin of clitics; c) the functional and semantic asymmetries between Latin ILLE and IPSE; and d) the absence of IPSE-derived clitics.

The often-observed isomorphism of *l*-definite articles and *l*-clitic pronouns across Romance languages obscures subtle yet important syntactic and semantic distinctions between types of definite nominal expressions, and these are brought to light by languages with two-way definite article systems such as we have seen for Balearic Catalan. What Balearic Catalan reveals is that the actual alignment in Romance is between *l*-clitics and only certain types of nominal expressions such as those whose accompanying *l*-articles convey a redundancy of reference (e.g., when accompanying proper names) or the uniqueness of reference of certain common nouns. The remaining types of nominal expressions, distinguishable semantically and also syntactically across Romance, are further distinguished in Balearic

¹³ Brazilian Portuguese is interesting in that it has lost 3rd person clitic pronouns and allows bare nouns in contexts different from those found in other Romance languages.

Catalan by means of a different form, an *s*-article. The first set of forms, *l*-clitics and *l*-articles, are in D2 and the second set, *s*-articles, are in D1.

We have highlighted in this paper how the present-day patterns were set in motion in the transition from Latin to Romance. That is, semantic and functional properties of Latin IPSE (topic-marking function and focus value; Vincent (1997)), which rendered it unfit as a candidate for clitic (but not article), and of Latin ILLE (e.g., discourse repetition, absence of emphatic status) set the stage for the development of pronominal clitics as well as articles.

We have shown that the Romance clitic paradigm with ILLE is fully developed earlier than the ILLE determiner system, which generalizes over time. The more articulated structure for DPs of Bernstein, Ordóñez & Roca (2018, 2021), which we have adopted, accommodates the synchronic and diachronic patterns. We have claimed that the non-emphatic and function-oriented clitics correspond to the lower D2, along with the generalized definite articles, while D1 is reserved for articles with deictic or pragmatic function.

The frequent isomorphism of Romance definite articles and clitics is not to be ignored, only more deeply mined. Patterns displayed in understudied Romance varieties shed light on the syntax and semantics of definite articles and clitics and support distinguishing two different classes, which we have mapped onto an articulated DP structure.

REFERENCES

Aebischer, Paul. 1948. Contribution à la protohistoire des articles *ille* et *ipse* dans les langues romanes. *Cultura neolatina* 8. 181–203.

Anglade, Joseph. 1958. *Grammaire élémentaire de l'ancien français*. Paris: Librairie Armand Colin.

Ayres-Bennett, Wendy. 1996. A history of the French language through texts. New York: Routledge.

Bernstein, Judy B., Francisco Ordóñez & Francesc Roca. 2018. Sardinian descendants of IPSE in comparative relief. In Annamaria Chilà & Alessandro De Angelis (eds.), *Capitoli di morfosintassi delle varietà romanze d'Italia: teoria e dati empiric*, 97–116. Palermo: Centro di Studi Filologici e Linguistici Siciliani.

Bernstein, Judy B., Francisco Ordóñez & Francesc Roca. 2021. Emphatic elements and the development of definite articles: Evidence for a layered DP in early Romance. *Journal of Historical Syntax* 5. 1–32. doi:10.18148/hs/2021.v5i16-25.63.

Boucher, Paul. 2005. Definite reference in Old and Modern French: The rise and fall of DP. In Montserrat Batllori, Maria-Lluïsa Hernanz, Carme Picallo & Francesc Roca (eds.), *Grammaticalization and*

- *parametric variation*, 95–108. New York: Oxford University Press. doi:10.1093/acprof:oso/9780199272129.003.0007.
- Cardinaletti, Anna & Michal Starke. 1999. The typology of structural deficiency: A case study of the three classes of pronouns. In Henk van Riemsdijk (ed.), *Clitics in the languages of Europe*, 145–233. Berlin, New York: Mouton de Gruyter. doi:10.1515/9783110804010.145.
- Carlson, Gregory, Rachel Sussman, Natalie Klein & Michael Tanenhaus. 2006. Weak definite noun phrases. In Christopher Davis, Amy Rose Deal & Youri Zabbal (eds.), *Proceedings of the North Eastern Linguistics Society* (= *NELS*) 3, 179–196. Amherst, MA: GLSA, University of Massachusetts.
- Cennamo, Michela, Francesco M. Ciconte & Luigi Andriani. 2022. Syntactic and semantic constraints on Differential Object Marking in Old Sardinian. University of Naples Federico II, University of Chieti-Pescara & University of Utrecht, Ms.
- CICA. n.d. Corpus informatitzat del català antic. http://www.cica.cat/.
- Dufresne, Monique, Mireille Tremblay & Rose-Marie Déchaine. 2018. The variable use of determiners in Old French and the argument DP hypothesis. *Linguistic Variation* 18(1). 23–48. doi:10.1075/lv.00003.duf.
- van Gelderen, Elly. 2011. *The linguistic cycle: Language change and the language faculty*. Oxford: Oxford University Press.
- Heim, Irene. 2011. Definiteness and indefiniteness. In Klaus von Heusinger, Claudia Maienborn & Paul Portner (eds.), *Semantics. An international handbook of natural language meaning*, vol. 2, 996–1025. Berlin, Boston: De Gruyter Mouton. doi:10.1515/9783110255072.996.
- Ihsane, Tabea. 2008. *Layered DP: Form and meaning of French indefinites*. Amsterdam, Philadelphia: John Benjamins. doi:10.1075/la.124.
- Jones, Michael Allen. 1993. Sardinian syntax. London: Routledge.
- Ledgeway, Adam. 2012. From Latin to Romance: Morphosyntactic typology and change. Oxford: Oxford University Press.
- Löbner, Sebastian. 1985. Definites. *Journal of Semantics* 4(4). 279–326. doi:10.1093/jos/4.4.279.
- Löbner, Sebastian. 2011. Concept types and determination. *Journal of Semantics* 28(3). 279–333. doi:10.1093/jos/ffq022.
- Ludlow, Peter & Gabriel Segal. 2004. On a unitary semantical analysis for definite and indefinite descriptions. In Marga Reimer & Anne Bezuidenhout (eds.), *Descriptions and beyond*, 420–436. Oxford: Oxford University Press.
- Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press.
- Mathieu, Éric. 2009. From local blocking to Cyclic Agree: The role and meaning of determiners in the history of French. In Jila Ghomeshi, Ileana Paul

- & Martina Wiltschko (eds.), *Determiners: Universals and variation*, 123–157. Amsterdam, Philadelphia: John Benjamins. doi:10.1075/la.147.04mat.
- de Nantes, J.M.J.A. 1898. *Littérature française depuis ses origines jusqu'à nos jours*. Paris: Ancienne Librairie Poussielgue.
- Ortiz Ciscomani, Rosa María. 2009. La creación y generalización del artículo definido. In Concepción Company Company (ed.), Sintaxis histórica de la lengua española. Segunda parte: La frase nominal, vol. 2, 271–386. Mexico: Fondo de Cultura Económica/Universidad Nacional Autónoma de México.
- Ortmann, Albert. 2014. Definite article asymmetries and concept types: Semantic and pragmatic uniqueness. In Thomas Gamerschlag, Doris Gerland, Rainer Osswald & Wiebke Petersen (eds.), *Frames and concept types: Applications in language and philosophy*, 293–321. Cham: Springer. doi:10.1007/978-3-319-01541-5_13.
- Panagiotidis, E. Phoevos. 2002. *Pronouns, clitics and empty nouns: 'pronominality' and licensing in syntax*. Amsterdam, Philadelphia: John Benjamins. doi:10.1075/la.46.
- Roberts, Ian. 2019. *Parameter hierarchies and Universal Grammar*. Oxford: Oxford University Press. doi:10.1093/oso/9780198804635.001.0001.
- Roca, Francesc. 1996. Morfemas objetivos y determinantes: los clíticos del español. *Verba* 23. 83–119.
- Schwarz, Florian. 2013. Two kinds of definites cross-linguistically. *Language and Linguistics Compass* 7(10). 534–559. doi:10.1111/lnc3.12048.
- Uriagereka, Juan. 1995. Aspects of the syntax of clitic placement in Western Romance. *Linguistic Inquiry* 26(1). 79–123.
- Veny, Joan. 1999. *Aproximació al dialecte eivissenc*. Palma (Mallorca): Editorial Moll.
- Veny, Joan & Mar Massanell. 2015. *Dialectologia catalana. Aproximació pràctica als parlars catalans*. Barcelona: Universitat de Barcelona Publicacions i Edicions.
- Vincent, Nigel. 1997. The emergence of the D-system in Romance. In Ans van Kemenade & Nigel Vincent (eds.), *Parameters of morphosyntactic change*, 149–169. Cambridge: Cambridge University Press.
- Zamparelli, Roberto. 1995. *Layers in the Determiner Phrase*. Rochester: University of Rochester Doctoral dissertation.

Bernstein, Ordóñez & Roca

Judy B. Bernstein Department of Language, Literature, Culture, Linguistics Department and Writing William Paterson University 300 Pompton Road Wayne, NJ 07470 USA bernsteinj@wpunj.edu

Francesc Roca Dept. Filologia i Comunicació Universitat de Girona Pça. Ferrater Mora, 1 17004 Girona Spain francesc.roca@udg.edu

Francisco Ordóñez Stony Brook University Stony Brook, NY 11794 USA

francisco.ordonez@stonybrook.edu