

Evidence for an articulated DP structure from the history and diversity of Romance languages

We examine definite articles and accusative pronominal clitics in Romance languages and observe that their frequent isomorphism in the modern languages masks an asymmetrical historical development, one that calls into question the idea that both types of elements correspond to the same D head. A detailed analysis of the diachronic patterns combined with a comparison of non-mainstream varieties shows that, a) clitics develop earlier than certain semantic classes of definite articles; and b) the facts support a more articulated structure for DP.

The parallelism between Romance definite articles and clitic pronouns has been observed and capitalized on in the literature. We find overlap in both form (see (1)) and etymology, both sets of elements having developed from the Latin demonstrative ILLE (see (2)).

(1)	<u>definite articles</u>	<u>clitic pronouns</u>	
	<i>el, la, los, las</i>	<i>lo, la, los, las</i>	Spanish
	<i>el/lo, la, els/los, les</i>	<i>el/lo, la, els/los, les</i>	Catalan
	<i>le, la, les</i>	<i>le, la, les</i>	French
	<i>o, a, os, oas</i>	<i>o, a, os, as</i>	Portuguese
(2)	<u>definite article</u>	<u>clitic pronoun</u>	
	ILLAM > <i>la</i>	ILLAM > <i>la</i>	Spanish/Catalan/French

This overlap, combined with the [+definite] nature of both sets of forms, has led to proposals that both the definite articles and clitic pronouns correspond to D (e.g., Uriagereka 1995, Roca 1996).

Data from non-mainstream Romance varieties call for a more nuanced approach.

Sardinian and some varieties of Catalan display a different definite article, an *s-* form, one that developed from the Latin emphatic pronoun IPSE: *sa baca* ‘the cow’ (Sardinian). Some varieties of Picard display a *ch-* form that developed from the Latin emphatic interjection ECCE: *che fu* ‘the fire’. Yet despite the diversity in the forms and sources of definite articles across Romance, only ILLE gave rise to pronominal clitics (Ledgeway & Maiden 2016):

(3)	<u>clitic pronouns</u>	
a.	<i>lu, la, los, las</i>	Sardinian
b.	<i>el/lo, la, els/los, les</i>	Balearic Catalan
c.	<i>li, les/lus</i>	Picard (Boulogne-sur-Mer; Haigneré 1901)

This asymmetry suggests that the grammatical nature of the forms derived from ILLE and those derived from IPSE or ECCE cannot be entirely identical and furthermore, that the development of the definite article and the clitic pronouns was not simultaneous.

Although Sardinian displays only an *s-* definite article, Balearic Catalan and varieties of French Picard display two different forms, *s-* and *l-* articles and *ch-* and *l-* articles, respectively. The choice between the two articles depends on syntactic and semantic properties of the DP: a DP with deictic or anaphoric function will display *s-* (Balearic Catalan) or *ch-* (Picard) and a DP expressing uniqueness, genericity or inherent reference (e.g., proper names, pronouns) will display *l-*. The distribution corresponds to a uniqueness hierarchy (Löbner 1985, 2011; Ortmann 2014) that makes a prediction about the likelihood of an overt D. That is, a nominal expression

with anaphoric reference is more likely to display a definite article than a nominal expression with inherent reference. It also makes a prediction about chronology: D1 before D2.

Bernstein et al. (2021) expresses the uniqueness hierarchy as an articulated DP structure:

(4) [DP₁ D1 ... [DP₂ D2 ... [NP N]]]

Each D head correlates with different syntactic and semantic properties: D1 with those properties higher on the uniqueness hierarchy (argument status, deictic/anaphoric function), D2 with those properties lower on the hierarchy (uniqueness/genericity, inherent reference). Applied to the Romance facts discussed above, the *s-* and *ch-* articles of Balearic Catalan and French Picard correspond to D1, the *l-* articles to D2. Sardinian *s-* can fill either D1 or D2, depending on syntactic/semantic function. Similarly, *l-* can fill either D1 or D2 in Romance languages with only *l-* articles. Romance pronominal clitics, which are *l-* forms and lack deictic function, naturally correspond to D2.

Diachronically, we find support for this articulated DP structure. Specifically, we see that definite articles develop earlier in some syntactic/semantic contexts than others. Dufresne et al. (2018; adapting material from Boucher 2005) identifies a hierarchy of contexts favoring overt determiners in Old French:

(5) demonstratives/possessives/quantifiers (obligatory) > definite specific reference (optional) > indefinite specific reference (optional) > generic/non-specific/abstract (bare) > mass Ns (bare)

Over time, definite articles will appear obligatorily in all of these contexts, the pattern we find in modern French. This shows that D1 is filled with definite articles before D2 in the history of French. We find evidence that this progression applied more generally across Romance.

In Old Catalan (of Mallorca) we find multiple instances of the same noun, its modification or absence thereof determined by semantic context. In (6), we see *rey* ('king') introduced with the definite article *lo* (*l-* articles predominate in the formal register), indicating that this is the existing king, the one currently in power and being discussed. The subsequent instances of *rey*, uttered by an interlocutor, are bare, as they indicate the needs of kings in general. These facts comport with the hierarchy of overt definite articles in (5). Furthermore, the examples serve to illustrate an overt D1 definite article (*lo rey*) and simultaneous absence of D2 definite articles (*rey*), contexts that would require overt articles in modern Catalan. Despite unfilled D2 in (6), we find D2 filled with a clitic pronoun in (7) (italicized *la*).

(6) Qual és la raó per la qual *lo rey* fa mostrar al fill major philosophia e al fill manor fa mostrar d'armes? Respós lo pastor e dix que *rey* ha major necessitat aver saviea natural que sciencia d'armes; cor per sciencia natural pot *rey* haver conexença de Deu e de sa persona.

[Llull, *Llibre de meravelles*; 14th c.]

'What is the reason for which the king has his eldest son teach philosophy and has his youngest son teach about weapons? The shepherd replied and said that kings have more need of (natural) wisdom than of weapons because kings can get knowledge about God and about themselves through (natural) science.'

(7) cor l'aer dona humiditat escalfada al aygua e la aygua *la reeb*

[Llull, *Llibre de meravelles*; 14th c.]

'because air gives warm moisture to water and water receives it'

Our examination of DP structure in Romance varieties like Sardinian, Balearic Catalan, and Picard reveals an absence of complete isomorphism between definite articles and clitic pronouns. The more articulated DP that we adopt can account for these facts, as well as for the sequencing of definite articles (D1 > D2) relative to clitic pronouns (D2). Homogenous varieties that display *l-* uniformly (e.g., Spanish, French) don't easily reveal the whole story.