

Data, Theory, and Explanation: The View from Romance

Adam Ledgeway and Martin Maiden

1.1 Introduction

This is a book about doing linguistics by using data, comparative and historical, from the Romance languages. It explores what we can learn about linguistics from the study of Romance, rather than taking the more traditional approach of asking what we can learn about the structure and history of the Romance languages through the application of general linguistic principles and assumptions. In short, it asks not what linguistics can do for Romance, but, rather, what Romance can do for linguistics.

The Romance languages are among the most widely studied and researched language families in modern linguistics. Data from Romance have always been prominent in the linguistic literature and have contributed extensively to our current empirical and theoretical understanding of phonetics, phonology, morphology, syntax, semantics, pragmatics, sociolinguistics, and historical linguistics. Their prominence reflects the richly documented diachronic variation exhibited by the Romance family, which, coupled with our extensive knowledge and abundant textual documentation of the ancestral language, Latin, offers insights into a range of variation through time and space certainly unparalleled for any other Western languages. In short, the Romance languages and dialects constitute a treasure house of linguistic data of profound interest and importance not merely for Romance linguists, but for linguists generally. Indeed, this perennially fertile and still underutilized linguistic testing ground has a central role to play in challenging linguistic orthodoxies and shaping and informing new ideas and perspectives about language change, structure, and variation. This book takes seriously the idea that our knowledge and understanding of the many fields of linguistics have been and continue to be considerably enhanced – but in many cases shaped – by investigations of the Romance

data. It is therefore meant not for the exclusive use and interest of Romance linguists,¹ but for general linguists interested in the insights that a knowledge of the Romance evidence can provide for general issues in linguistic theory.

By exploring a range of comparative Romance data, this book contributes to a series of core questions and issues in linguistics, namely *I. What Is a Language?*; *II. Phonetics and Phonology*; *III. Morphology*; *IV. Syntax*; *V. Semantics and Pragmatics*; *VI. Language, Society, and the Individual*. The 30 chapters have been written, often collaboratively, by 50 internationally recognized Romance linguists, who were invited to contribute in these areas both on the basis of their expertise in specific fields of linguistics and for their expert knowledge of the relevant comparative Romance data. They have been encouraged to take a personal view of the principles and areas that have been influential in a particular subarea, bringing to bear the results of their own recent research wherever appropriate.

What follows in this introductory chapter is also a ‘personal view’ of Romance linguistics, but one that adopts a slightly different perspective from that taken in the rest of the book. At first sight, what we do in the remainder of this chapter may appear quirky, incoherent, perhaps even self-indulgent. Rather than addressing a particular topic in linguistic theory from a Romance perspective, we have chosen to explore our own, personal, experiences of doing Romance linguistics, and of how working with data from the Romance languages has made us reflect on wider issues in general linguistics. Recurrent themes in our work have been, respectively, morpho-syntactic change (Ledgeway) and sound change and its morphological consequences (Maiden). Within those areas, however, we have each concentrated here on a particular aspect, Ledgeway on the grammatical expression of functional categories and Maiden on Romance palatalization and its consequences. Now these two topics may seem to be the most curious of bedfellows, and indeed there is probably no significant overlap between them whatever. Moreover, each of these topics has led us along a number of different, and perhaps unexpected, sidetracks and byways but not, we think, dead ends! The result may seem eclectic and diffuse, but that is not

¹ There are numerous valuable manuals and handbooks, including classic comparative-historical and massively-detailed encyclopaedic treatments such as Meyer-Lübke (1890–1902), Lausberg (1965–66), Holtus, Metzeltin, and Schmitt (1988–2005), and Ernst, Gießgen, Schmitt, and Schweickard (2003–08), and the three volumes co-edited by the current editors (viz. Maiden, Smith, and Ledgeway 2011; 2013; Ledgeway and Maiden 2016), as well as a new De Gruyter series *Manuals in Romance Linguistics* (general editors: Günter Holtus and Fernando Sánchez-Miret) with a projected 30 or so volumes dedicated to individual Romance varieties, sub-branches of Romance, and specific Romance phenomena and themes. Then there are the many very useful smaller-scale works on comparative Romance such as Hall (1974), Elcock (1960; 1975), Harris (1978), Harris and Vincent (1988), Posner (1996), Alkire and Rosen (2010), Ledgeway (2012a), as well as detailed structural treatments of some of the better-known individual Romance languages (e.g., Maiden 1995; Penny 2000, 2002; Azevedo 2005; Fagyal, Kibbee, and Jenkins 2006; Pană Dindelegan 2013, 2016; Maiden et al. 2021).

the point. These topics are simply two representative fragments of the vast intellectual enterprise of Romance linguistics, and we believe that they have led us to the kind of conclusions that would also emerge if Romance linguists working in any other subdomains were invited to reflect on their personal experience of doing Romance linguistics. And what are those conclusions? That the comparative-historical study of the Romance languages can most effectively illuminate our understanding of human language, and particularly of language change, if it seeks to explain, rather than merely to describe, linguistic facts; that such explanation should be informed by, and can in turn illuminate and refine, general linguistic theory; but above all, and most fundamentally, that Romance linguistics can make its most powerful contributions to general linguistics when Romance linguists exploit to the maximum the extraordinary wealth of historical and comparative *data* which the Romance languages offer them.

1.2 The View from Morphosyntax and the Case of Functional Categories

1.2.1 From Latin to Romance: The Rise of Functional Categories

One of the most striking morphosyntactic differences between Latin and Romance has traditionally been taken to involve a distinction between morphology and syntax:² whereas Latin predominantly makes recourse to synthetic structures, Romance makes greater use of analytic structures, a development often interpreted as the surface reflex of a change in the basic ordering of head and dependency according to a well-known typological distinction from which many other basic properties are said to follow (Greenberg 1966; Lehmann 1974; Harris 1978: 4–6; Bauer 1995: 13).³ By way of illustration, consider Table 1.1, where we see that, in contrast to Romance, Latin lacks functional categories, in that none of the core grammatical categories such as subordination, tense, aspect, mood, transitivity, or definiteness is expressed analytically (cf. Ledgeway 2012a: ch. 4). At the same time, there is significant synchronic variation across Romance as to which of the functional categories are lexicalized and the distinctions they overtly mark. For instance, only French lexicalizes all the available heads of the functional projections in Table 1.1, including an overt transitive/causative light *v*(erb) *fait* ‘made’, whereas Italian only optionally encodes the partitive distinction through an overt DET(erminer) *del* ‘of.the (= some)’ (cf. Stark 2008). By contrast, Romanian fails to overtly lexicalize either of

² See, among others, von Schlegel (1818), Bourciez (1956: 23), Harris (1978: 15f.), Schwegler (1990), Posner (1996: 156f.), Vincent (1997a), Ledgeway (2011b: 383–87; 2012a: ch. 2; 2017a).

³ Harris (1978: 16), Vincent (1988: 55f., 62f.; 1997b: 166), Bauer (1995), Oniga (2004: 52), Ledgeway (2011b: §5; 2012a: ch. 5; 2014b; 2018a).

Table 1.1. *Synthetic vs analytic marking of core grammatical categories in Latin and Romance*

		COMP		Infl		V		DET	
Lat.	Dico/Uolo	Ø		eum	Ø	Ø	coxisse	Ø	panem.
Fr.	Je dis/veux	qu'		il	a/ait	fait	cuire	du	pain.
It.	Dico/Voglio	che			ha/abbia	Ø	cotto	(del)	pane.
Ro.	Spun/Vreau	că/să			a/fi	Ø	copt	Ø	pâine.
	I.say/want	that _(REALIS/IRREALIS)	him/he	has _{IND} /(be) _{SBJV}		made	bake(d)	some	bread
	'I say that he has/I want him to have baked some bread.'								

these functional categories, but uniquely displays robust marking on the COMP(lementizer) *că/să* ‘that’ for the realis/irrealis opposition (Gheorghe 2013b: 468–70), otherwise paralleled in the indicative/subjunctive distinction realized through the clausal INFL(exion) on the perfective auxiliary *a/ait* and *ha/abbia*, in turn further distinguished by way of a HAVE/BE split (viz. *a/fi*) in Romanian (Ledgeway 2014a). In short, we observe minimal differences among otherwise highly homogenous systems which can be read both vertically (Latin ⇒ Romance) and horizontally (French ⇒ Italian ⇒ Romanian) as cases of diachronic and synchronic/diatopic microvariation, respectively.

We thus conclude that marking of clausal boundaries, various verb-related grammatical categories, and definiteness and quantification in Romance is lexicalized by functional markers belonging to the categories of COMP(lementizer), AUX(iliary), light v(erb), and DET(erminer). In current theory, grammatical elements of this type are generally considered to head their own functional projections CP, I(nfl)P, vP, and DP which provide the locus of grammatical information for the clausal, sentential, verbal, and nominal groups, respectively. On this view, one of the most significant generalizations of the traditional synthesis-analysis approach can now be recast in terms of the emergence of these functional categories (Vincent 1997a: 105; 1997b: 149; Lyons 1999: 322f.) which, at least according to one view (though cf. Horrocks 2011; Ledgeway 2012a: ch. 5), were either entirely absent from Latin or only present in incipient form.

Although a consideration of the lexicalization or otherwise of the head positions made available by a universal structure of functional projections provides an elegant way of drawing a morphosyntactic typological distinction between Latin and Romance, it does not offer any further insight into the thorny question of how Romance can be distinctively and exhaustively defined purely on linguistic grounds (Section 1.3.6). Clearly, there are many other language families and areal groupings that equally show extensive evidence for the use of functional categories in similar ways to the Romance languages. Nonetheless, detailed study of Romance functional categories constitutes a fruitful and insightful area of investigation which can both

throw light on the comparative history of Romance and offer us important lessons in general linguistic theory. Indeed, differences in functional categories are best studied comparatively within a single family of languages where dimensions of variation between otherwise highly homogeneous linguistic systems of the family are often minimal, thereby allowing us to pinpoint what precisely may vary and the linguistic mechanisms underpinning such variation. In this respect, the richly documented diachronic and synchronic variation exhibited by the Romance family (cf. Section 22.2) offers privileged access to a range of variation through time and space unparalleled for other Western languages.

The Romance languages therefore offer us a valuable experimental testbed to investigate the ways in which current theories claim that it is possible for the morphosyntax of languages to vary. Building on the insights of the Borer–Chomsky Conjecture (cf. Baker 2008: 353), the relevant dimensions of Romance microvariation can be taken to lie in the functional lexicon and, in particular, in the overt lexicalization of specific formal feature values of individual functional heads and the functional categories that realize them (Borer 1984; Chomsky 1995). These feature values are not set in isolation, inasmuch as dimensions of variation ostensibly form an interrelated network of implicational relationships whereby the given value of a particular functional category may, in turn, entail the concomitant activation of associated lower-order grammatical choices, whose potential surface effects may consequently become entirely predictable, or indeed rule out other morphosyntactic properties. In what follows, we therefore consider a selection of representative case studies of comparative morphosyntactic variation which highlight a number of significant differences in the featural make-up of the functional heads C-T-v-D and their associated domains – the left periphery, the inflexional core of the sentence, the verb phrase, and the nominal group – and the parametric options they instantiate. By marrying, on the one hand, traditional Romance philological and dialectological scholarship through the study of syntactic microvariation across time and space with, on the other, the insights of recent syntactic theory, we show how a detailed, expert knowledge of the full extent of the Romance evidence can both test and challenge our theories of morphosyntax and expand the empirical linguistic data on which they are based. Unfortunately, non-standard Romance varieties are too often overlooked in this respect, even though they offer fertile, and frequently uncharted, territory in which to study microvariation. Such microvariation frequently reveals significant differences of real theoretical significance which would not otherwise be visible by simply comparing the grammars of the standard Romance languages (cf. the discussion of gender and number in Section 2.2).

Following a brief introduction in Section 1.2.2 to morphosyntactic variation across Romance in relation to parameters (Section 1.2.2.1), universals

(Section 1.2.2.2), language typology (Section 1.2.2.3), and the interfaces (Section 1.2.2.4), in Section 1.2.3 some case studies of microvariation across Romance are explored which highlight what Romance can do for syntactic theory by way of testing, challenging, and expanding our theory of language and the empirical base. By the same token, the tools and insights of current theories of syntax can also be profitably used to throw light on many of the otherwise apparently inexplicable facts of Romance microvariation, the topic of Section 1.2.4 where the role of syntactic theory for Romance is explored through the exploration of a number of Romance case studies which have traditionally proven, at the very least, extremely difficult to interpret in a unitary and satisfactory fashion.

1.2.2 Linguistic Variation

1.2.2.1 Parameters

One area where research into Romance functional categories has proven particularly influential is the investigation of linguistic parameters, those dimensions of linguistic variation along which natural languages are said to vary (for in-depth discussion, see Chapters 4 and 21, this volume, and Roberts 2019: §1.2; Ledgeway 2020b).⁴ Linguistic variation is not free or wild, but is subject to specific structural conditions which restrict the possible limits of variation of all natural languages. To cite just one simple example, it is well known (cf. Cheng 1997; Roberts 2019: ch. 7) that languages vary according to whether *wh*-interrogatives must be fronted to the C-domain, as in most Romance varieties (1a–b), or whether they must remain *in situ* as in Chinese (1c). Yet, in other languages *wh*-fronting is not so systematic, but shows a mixed distribution. This is the case in Brazilian Portuguese, colloquial French, and many dialects of north-(eastern) Italy, where the fronting or otherwise of *wh*-interrogatives variously depends on their phonosyntactic and discourse-pragmatic status (cf. also Section 20.4.3; Section 24.2.2). For instance, in the north-eastern Italian dialect of Lamon clitic and tonic variants of the *wh*-interrogative *WHAT* occur in fronted (2a) and *in situ* (2b) positions, respectively, and can even co-occur (2c), whereas discourse-pragmatically marked interrogatives such as D(iscourse)-linked complex *wh*-phrases (2d) invariably undergo fronting (De Cia 2018: 22f., 118).⁵

⁴ For examples and discussion of a phonological parameter, see Section 5.7.

⁵ See Munaro (1998), Ambar et al. (2001), Munaro, Poletto, and Pollock (2001), Munaro and Poletto (2002), Benincà and Poletto (2005), Kato and Miotto (2005), Manzini and Savoia (2011), Kato (2013), Bonan (2019), De Cia (2019). Note, however, that many of these analyses maintain that insitueness is only apparent, with the *wh*-interrogative raising to the lower or higher left periphery, variously accompanied by remnant movement. If correct, then the relevant fronting parameter displays a uniform behaviour across Romance.

- (1) a. **Cine** crede John că **eine** a cumpărat cărțile? (Ro.)
 b. ¿**Quién** cree John que **quién** ha comprado los libros? (Sp.)
 who believe.PRS.3SG John that have.PRS.3SG bought.PTCP the books(.DEF)
 c. John xiangzin **shei** mai-le shu? (Ch.)
 John believe who buy-ASP book
 'Who does John believe has bought the books?'
- (2) a. **Sa-** g- a -li dat a Simon? (Lamon)
 b. G- a -li dat **che** a Simon? (Lamon)
 c. **Sa-** g- a -li dat **che** a Simon? (Lamon)
 what= DAT.3= have.PRS.3 =SCL.3MPL give.PTCP what to Simon
 'What did they give Simon?'
- d. **Che** casa a -lo fat su Toni? (Lamon)
 what house have.PRS.3 =SCL.MSG do.PTCP up Toni
 'Which house did Toni build?'

Among those varieties which display overt fronting of *wh*-interrogatives it is possible to further distinguish between those which allow multiple fronting and those that do not (Bošković 2002): at first blush Slavonic (3a) belongs to the former group, whereas Romance (3b) appears to belong to the latter group (cf. Giurgea and Remberger 2016: 870). However, a more extensive examination of the Romance facts reveals a more nuanced picture in that, unlike other Romance varieties, Romanian (3c) requires multiple fronting (Rudin 1988).

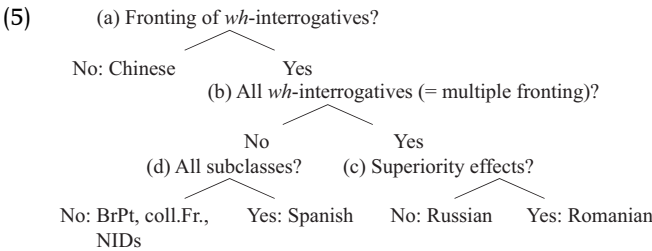
- (3) a. **Kto** čto kto kupil čto? (Ru.)
 b. ¿**Quién** (**qué) **quién** ha comprado **qué**? (Sp.)
 c. **Cine** ce **eine** a cumpărat ee? (Ro.)
 who what bought what
 'Who has bought what?'

In this respect the behaviour of Romanian appears to parallel that of Slavonic, hardly a surprising result given the widespread borrowing, not just of lexical, but also of functional features across languages of the so-called Balkan Sprachbund. Nonetheless, a closer look at the Romanian facts reveals that the features of the C-head which license multiple *wh*-fronting also impose ordering restrictions absent in Slavonic (Gheorghe 2013a). More specifically, in contrast to most Slavonic varieties where the order of multiple fronted *wh*-constituents is generally unconstrained (cf. 3a, 4a), in Romanian their order shows a sensitivity to superiority effects such that, for example, the subject must precede the object (cf. 3c vs 4b) and arguments must precede, in turn, all adjuncts (4c).

- (4) a. Čto kto kto kupil čto? (Ru.)
 b. ****Ce** cine **eine** a cumpărat ee? (Ro.)
 what who bought
 'Who bought what?'

- c. (**Când) Cine când ~~eine~~ a cumpărat-o ~~când?~~ (Ro.)
 when who when have.PRS.3SG buy.PTCP=it
 ‘Who bought it when?’

Consequently, the evidence of Romanian – today still too often overlooked in so-called comparative overviews of Romance – is fundamental for the study of the parameters and sub-parameters involved in the licensing of *wh*-fronting, since it exceptionally presents a mixture of both typical Romance and non-Romance options yielding apparently hybrid grammatical choices. By comparing in this way Romanian not only with other Romance languages, but also with the neighbouring Slavonic varieties it has come into contact with over time (cf. also Chapter 28, this volume), it is possible to isolate the properties of individual functional heads of the C-domain responsible for the fronting of *wh*-interrogatives and model the internal hierarchical organization of the options they instantiate. For example, keeping technical details to a minimum, the formal structural characterization of the variation observed so far in the licensing of the *wh*-interrogatives can be captured by way of (5).



Conceived along the lines of (5), parametric variation can be interpreted in a scalar fashion and modelled in terms of a series of hierarchical and implicational relationships (for further discussion, see Section 21.1). The simplest and least marked options that uniformly apply to all functional heads, are placed at the very top of the hierarchy, but, as one moves downwards, variation becomes progressively more restricted with choices becoming progressively more limited to smaller and smaller proper subsets of features and contexts. This gradual cascading effect produced by the options presented in (5) highlights how variation in relation to the ability of the C-domain to attract *wh*-interrogatives is not uniform but, rather, licenses differing degrees of surface variation in accordance with the growing markedness conditions that accompany the available parametric options as one moves down the hierarchy.

The simplest and least constrained option (viz. 5a) is exemplified by Chinese where all *wh*-interrogatives simply remain in their base

positions in all cases, since the C-head is inert and hence unable to license *wh*-fronting to the clausal left periphery. In all other varieties, by contrast, the relevant parameter shows a more marked setting, in that the C-head requires some degree of *wh*-fronting. The least marked option (viz. 5b) among these varieties is instantiated by languages where C indiscriminately attracts all *wh*-interrogatives giving rise to multiple fronting, inasmuch as the effects of the parameter are uniform since the 'rule' affects all *wh*-interrogatives without exception. In this respect, languages such as Chinese, on the one hand, and multiple-fronting languages, on the other, represent simpler and comparatively unmarked options, in that the C-head in these varieties either indiscriminately fails to attract any *wh*-interrogative or, on the contrary, systematically attracts all *wh*-interrogatives. However, as we have seen, within the subclass of languages specified positively for the option of multiple *wh*-fronting there is an additional split which introduces a further restriction in relation to the linear order of fronted *wh*-interrogatives (viz. 5c). While in Slavonic languages such as Russian the order of fronted constituents is largely unconstrained, in Romanian their order falls under specific structural conditions constrained by superiority effects. Finally, option (5d) identifies those varieties where the C-head licenses a more restricted type of fronting limited to a maximum of just one *wh*-interrogative. Such varieties do not, however, form a homogeneous grouping but can be further divided into at least two further subclasses according to whether *wh*-fronting displays a uniform or mixed behaviour. In languages such as Spanish and most other Romance varieties all *wh*-interrogatives may be fronted without exception, whereas in varieties such as Brazilian Portuguese, colloquial French, and many north(eastern) Italian dialects fronting only applies to specific subclasses of *wh*-interrogative.

Over recent years the significance of Romance dialects for the study of parametric variation has also been increasingly recognized. These prove particularly insightful since, although neighbouring dialects tend to be closely related to each other displaying in most cases a high degree of structural homogeneity, they often diverge minimally in significant ways which allow the linguist to identify and observe what lies behind surface differences in particular parametric settings across a range of otherwise highly homogenized grammars. By drawing on such microvariation, it is possible to determine which phenomena are correlated with particular parametric options and how such relationships are mapped onto the syntax. By way of example, consider the so-called dative shift construction, a phenomenon attested in a number of Germanic languages whereby an underlying indirect object such as the RECIPIENT *to Mary* in (6a) can be

reanalysed and promoted to direct object. Consequently, in the double object variant in (6b) *Mary* now occurs without the dative marker *to* and precedes the THEME *a book*. Furthermore, it has been claimed that the possibility of dative shift is linked to another structural property, that of stranding prepositions in *wh*-questions and relative clauses, as demonstrated in (6c).

- (6) a. John gave a book [to Mary].
 b. John gave [Mary] a book.
 c. [Who] did John give a book [_{PP} to [_{DP} *wh*ø]]?

Romance, by contrast, has been claimed to display neither dative shift nor preposition stranding (Kayne 1984; Larson 1988: 378; Holmberg and Platzack 1995), as the sharp ungrammaticality of the Portuguese examples in (7b–c) demonstrates:⁶

- (7) a. O João deu um livro [à Maria]. (Pt.)
 the João give.PST.PFV.3SG a book to.the Maria
 b. **O João deu [a Maria] um livro. (Pt.)
 the João give.PST.PFV.3SG the Maria a book
 c. **[Quem] deu o João um livro [_{PP} a [_{DP} *quem*]]? (Pt.)
 who give.PST.PFV.3SG the João a book to

However, this apparent Germanic-Romance parametric contrast, ultimately related to properties of the light *v* head and its extended projection, is contradicted by a number of Romance dialects where something very similar, if not identical, to dative shift, is found (Demonte 1995; Sornicola 1997: 35f.; Ledgeway 2009a: 844–47; cf. also Section 16.4), witness the representative Neapolitan examples in (8).

- (8) a. Giuanne nce rette nu libro [a Maria]. (Nap.)
 Gianni DAT.3= give.PST.PFV.3SG a book to Maria
 ‘Gianni gave a book to Maria.’
 b. Giuanne a rette [a Maria] nu libro. (Nap.)
 Gianni ACC.3FSG= give.PST.PFV.3SG DOM Maria a book
 ‘Gianni gave Maria a book.’
 c. **[Chi] rette nu libro [_{PP} a [_{DP} *chi*]]? (Nap.)
 who give.PST.PFV.3SG a book to
 ‘Who did he give a book to?’

The RECIPIENT argument *a Maria* ‘to Maria’, the underlying indirect object in (8a), has been promoted to direct object in (8b) where *a* is no longer the indirect object marker but, rather, the differential object marker

⁶ Note the orthographic and phonetic distinction in (7a–b) between the articulated preposition *à* [a] ‘to the’ (< *a* ‘to’ + *a* [e] ‘the.FSG’) and the feminine singular definite article *a* [e] ‘the.FSG’.

of specific animate objects, as shown by: (i) the position of *a Maria* in front of the THEME *nu libbro*; and (ii) by the fact that *a Maria* is now referenced by an accusative clitic pronoun *a* ‘her’ on the verb, and not the third person dative pronoun *nce* in (8a). Other neighbouring southern dialects, by contrast, such as Cosentino (cf. Ledgeway 2000: 46–52) exhibit a more constrained type of dative shift, inasmuch as RECIPIENT arguments (cf. *cci* ‘to her’ in 9a) may only surface as direct objects (cf. *a* ‘her’ in 9b) in monotransitive clauses. A similar dative–accusative alternation is also found in a number of non-standard Ibero-Romance varieties (Pineda 2016; 2020), including (central) Catalan (10a), various Spanish varieties (10b), and Asturian (10c).

- (9) a. Gianni **cci** scriva (na littera). (Cos.)
 Gianni DAT.3=write.PRS.3SG a letter
 b. Gianni **a** scriva (**na littera). (Cos.)
 Gianni ACC.3FSG=write.PRS.3SG a letter
 ‘Gianni will write her (a letter).’
- (10) a. En Joan (**li**) **la** telefona. (non-standard Cat.)
 the Joan DAT.3= ACC.3FSG= phone.PRS.3SG
 ‘Joan rings her.’
 b. Al día siguiente **la** telefoneó para invitarla **al**
 at.the day following ACC.3FSG= phone.PST.PFV.3SG for invite.INF=ACC.3FSG to.the
 cine. (non-standard Sp.)
 cinema
 ‘The following day he phoned her to invite her to the cinema.’
 c. Telefoneélu (*lʲ-y*). (Ast.)
 phone.PST.PFV.1SG=ACC.3MSG (≠DAT.3)
 ‘I rang him.’

These dialectal varieties reveal therefore three important things. First, dative shift is not a simple Germanic vs Romance parametric option (see also Pineda and Mateu 2020: iv–vii). Second, the supposed link between dative shift and preposition stranding, argued to be derivable from a single parametric option (cf. 6c vs 7c), does not hold. Rather, the presence of both phenomena in languages like English simply represents a fortuitous combination rather than the principled outcome of a particular parametric setting, since preposition stranding is not found in those same Romance varieties which license dative shift (cf. 8c). This conclusion is further confirmed by Romanian which, in contrast to other standard Romance languages but in line with neighbouring Slavonic patterns, also displays a number of lexicalized double object constructions (Pană Dindelegan 1968; 2010; 2013: 65–72) in which the RECIPIENT is marked accusative even in the presence of a THEME argument (11a–c), but which once again fails to license preposition stranding (11d).

- (11) a. *Imprudența aceasta l-a costat viața.* (Ro.)
 imprudence.DEF this ACC.3MSG=have.PRS.3SG cost.PTCP life.DEF
 ‘This act of folly cost him his life.’
- b. *Am întrebat-o ceva.* (Ro.)
 have.PRS.1SG ask.PTCP=ACC.3FSG something
 ‘I asked her something.’
- c. *Îi învățau limba engleză.* (Ro.)
 ACC.3PL= teach.PST.IPFV.3PL language.DEF English
 ‘They were teaching them English.’
- d. ***Cine vorbeau despre [DP cine]?* (Ro.)
 who speak.PST.IPFV.3PL about
 ‘Who were they speaking about?’

Third, it is incorrect to subsume all instances of accusative marking of RECIPIENT arguments under the generic heading of dative shift, since some Romance dialects prove sensitive to the mono- vs ditransitive distinction. It follows therefore that what might otherwise be taken to represent the surface reflexes of a single parametric setting in dialects like Neapolitan, namely the accusative marking of all RECIPIENT arguments irrespective of whether they occur in mono- or ditransitive clauses, turns out in fact to conceal two distinct structural operations in the light of evidence gleaned from Cosentino, non-standard (central) Catalan and Spanish varieties, and Asturian.

1.2.2.2 Language Universals

Romance also has much to contribute in the area of so-called universal principles of language, essentially a system of rules forming part of the genetic endowment known as Universal Grammar which is believed to hold of all human languages (cf. also Section 21.1). A good illustration of the valuable role that Romance can play in testing linguistic universals concerns the licensing of nominative Case. Within current theory, it is assumed that Infl, the locus of verbal inflexion, may be specified as [\pm Agr], featural specifications which in turn are argued to correlate with the verb’s ability or otherwise to license a nominative Case-marked subject. This [\pm Agr] distinction is supported by the evidence of many of the world’s languages, including French, where finite verbs license nominative subjects (12a), but non-finite verbs such as infinitives and gerunds, which lack overt morphological agreement, only allow null (Caseless) PRO subjects (12b):

- (12) a. *Vous rentrez à la maison.* (Fr.)
 you return.PRS.IND.2PL to the house
 ‘You return home.’
- b. *Avant de Ø_i Jean rentrer, vous_i avez téléphoné.* (Fr.)
 before of Ø Jean return.INF you have.PRS.2PL phone.PTCP
 ‘Before going home, you rang.’

Yet, the evidence of Romance dialects reveals that the supposed universal correlation between the specification of Infl and the availability of nominative Case is entirely spurious (Ledgeway 1998; 2000: ch. 4; Mensching 2000). In particular, dialects from the length and breadth of the Romance-speaking world demonstrate an abundant use of overt nominative subjects in conjunction with infinitival verbs:

- (13) a. L' üsu l' è d' acatâ tütu u padrùn. (Lig.)
 the usage SCL be.PRS.IND.3SG of buy.INF all the boss
 'It is customary for the boss to buy everything.'
- b. Nun sgarrava mai l' uri soi, senza dàricci corda nuddu. (Sic.)
 NEG ert.PST.IPFV.3SG never the hours its without give.INF=DAT.3 wind nobody
 'It always kept time without anyone ever having to wind it up.'
- c. E' di prima à mettasi idda in vinochju, altari ùn si ni
 and of before to put.INF=self she in knee altar NEG self= of.it=
 fighjulaia più (Crs.)
 look.PST.IPFV.3SG more
 'Until she knelt down, you couldn't look at the altar.'
- d. Su postinu est colatu prima de arrivare jeo. (Srd.)
 the postman be.PRS.IND.3SG pass.PTCP before of arrive.INF I
 'The postman passed by before I arrived.'
- e. Eu trabalha per elo s' amusa. (Occ., Gvd.)
 I work.PST.IPFV.1SG for she self= enjoy.INF
 'I would work so that she could enjoy herself.'
- f. ¿Qué tú me recomiendas para yo entender la lingüística? (Cuban Sp.)
 what you me=recommend.PRS.IND.2SG for I understand.INF the linguistic
 'What do you recommend for me so that I can get an understanding of linguistics?'

Examples such as (13a–f) which illustrate the so-called 'personal' infinitive (cf. Ledgeway 1998; 2000: ch. 4; Bentley 2014) highlight how traditional interpretations of finiteness in terms of a binary distinction are untenable, inasmuch as apparently non-finite forms function to all intents and purposes on a par with finite forms (cf. also Section 2.6.2). Indeed, we are led to conclude that finiteness cannot be understood in terms of superficial morphological marking alone, but, rather, must be interpreted as a scalar notion also including reference to semantico-syntactic criteria. Framing these observations in terms of the features T(ense) and Agr(eement) of the functional category Infl, the traditional binary opposition can be recast in accordance with the typological options in (14a–d) illustrated here by Romanian examples.

- (14) a. [+T, +Agr] ⇒ indicative
 Ion se trezește / trezea / trezi / trezise. (Ro.)
 Ion self= wake.PRS.IND.3SG wake.PST.IPFV.3SG wake.PST.PFV.3SG wake.PLPFV.3SG
 'Ion wakes/was waking/woke/had woken up.'
- b. [+T, –Agr] ⇒ personal infinitive
 Am plecat / Plec / Voi pleca înainte de a
 have.PRS.1SG leave.PTCP leave.PRS.1SG AUX.FUT.1SG leave.INF before of to

- se trezi Ion. (Ro.)
 self= wake.INF Ion
 'I (have) left/leave/will leave before Ion wakes up.'
- c. [-T, +Agr] ⇒ subjunctive
 Vreau / Voiam să se trezească Ion. (Ro.)
 want.PRS.1SG want.PST.IPFV.1SG PRT_{IRREALIS} self=wake.SBJV.3 Ion
 'I want/wanted Ion to wake up.'
- d. [-T, -Agr] ⇒ canonical (control) infinitive
 Ion_i dorește a Ø_i se trezi. (Ro.)
 Ion wish.PRS.IND.3SG to self= wake.INF
 'Ion wishes to wake up.'

Appealing to ideas developed in Ledgeway (1998; 2007), to which the reader is referred for a detailed treatment, the featural specification in (14a) represents canonical finite indicative verb forms which, from a semantico-syntactic perspective, show the greatest degree of temporal and referential autonomy, in turn, reflected morphologically by explicit marking for person, number, and tense. At the other extreme, the specification in (14d) identifies canonical control infinitives which lack any semantico-syntactic autonomy, as reflected by the complete absence of any morphological marking (cf. Sections 2.6.2, 20.5.2). Consequently, in contrast to indicative forms, such infinitival forms are limited to embedded contexts, where they depend on their associated matrix verb to supply their personal and temporal reference. In particular, the identity of the null infinitival subject is 'controlled' by a matrix argument (cf. Williams 1980; Hornstein 1999; Landau 2013), whereas the temporal frame of the infinitival clause is invariably interpreted as posterior to the matrix reference time, thereby giving rising to the obligatory irrealis unrealized future reading of the infinitival tense (Stowell 1982).

Specifications (14b–c), by contrast, represent intermediate points between (14a,d) on a scale of finiteness. The former identifies the personal infinitive which, despite lacking morphological marking for person, is nonetheless specified positively for the feature tense (cf. Ledgeway 2000: §4.4.2). Unlike control infinitives, the personal infinitive does not receive an obligatory irrealis interpretation but, rather, is referentially free in that its temporal frame can be located as simultaneous (13a–b), anterior (13c–d), or posterior (13e–f) to the matrix reference time. As a consequence, the personal infinitive can be taken to express a non-specific tense relative to the matrix predicate, thereby capturing the fact that the tense of the personal infinitive is always free though interpretatively bound by that of the matrix. Finally, the specification in (14c) provides an accurate description of the formal properties that distinguish subjunctive verb forms. Although endowed with overt marking for person and number (hence the featural

specification +Agr), subjunctive verb forms lack the referential temporal freedom of the personal infinitive (cf. Maiden 2016a: 111) but, rather, typically yield an irrealis interpretation (Stowell 1982; Ledgeway 2000: §4.4.2) similar to that observed with the canonical control infinitive (hence the featural specification –T).⁷

1.2.2.3 Typological Variation

Data like those exemplified in (13a–f) also illustrate how investigations of Romance and especially its dialects frequently reveal that the extent of typological variation within Romance, and indeed even within and beyond Indo-European, can prove to be considerably greater than is traditionally assumed and often of a typologically ‘exotic’ nature (for some phonological examples, see Chapter 9, this volume). In this respect, one only has to think of such examples as the Romance inflected infinitives (15a), gerunds (15b), and participles (15c),⁸ which, as hybrid categories, present further problems for the traditional binary interpretations of finiteness reviewed in Section 1.2.2.2 (cf. Ledgeway 2007c):

- (15) a. É doado supoñer-en as cousas. (Glc.)
 be.PRS.IND.3SG easy suppose.INF-AGR.3PL the things
 ‘It is easy for them to assume things.’
- b. Não saíndo-mos de casa, morrâmos à fome. (EuPt., Ervedosa do Douro)
 not leave.GER-AGR.1PL of home die.PRS.1PL at.the hunger
 ‘If we don’t leave the house, we’ll starve to death.’
- c. dato-no-sse insembla salute como convenne (ONap.)
 give.PTCP-AGR.3PL=self together greeting as require.PST.PFV.3SG
 ‘after having greeted one another as was customary’

Another acute example comes from the unique ‘infectious’ development of inflexion across all functional categories in the Marchigiano dialect of Ripatransone.⁹ Simplifying somewhat, in addition to some expected person/number agreement, the Ripano finite verb simultaneously displays masculine/feminine gender agreement with the subject, not to mention the possibility of agreement with third person so-called neuter subjects, by means of final inflexional vowel contrasts which appear quite remarkably to have

⁷ From the perspective of Romanian and the dialects of the extreme south of Italy this conclusion is not at all surprising, inasmuch as the distribution of subjunctive clauses in these varieties largely coincides with the use of the canonical control infinitive in other Romance varieties. Note finally that, alongside the subjunctive, the specification in (14c) also picks out the so-called inflected infinitive discussed in Section 1.2.2.3, underlining the intuition that the subjunctive is little more than an infinitive endowed with overt marking for person.

⁸ See, among others, Maurer (1968), Carballo Calero (1981), Loporcaro (1986), Jones (1993: 78–82), Longa (1994), Vincent (1996; 1998), Ledgeway (1998: 41–46; 2000: 109–14; 2009a: 585–90; 2012a: 293–95), Mensching (2000), Lobo (2001).

⁹ See Parrino (1967), Lüdtkke (1974; 1976), Mancini (1993), Harder (1998), Ledgeway (2012a: §6.3.4), Paciaroni and Loporcaro (2018), D’Alessandro (2020).

Table 1.2. *Ripano nominal and present indicative paradigms*

Nominal				Verbal			
	Masculine	Feminine	Neuter	Person	Masculine	Feminine	Neuter
Singular	<i>fijju</i>	<i>fijje</i>	<i>prəfuttə</i>	1	<i>magnu</i>	<i>magne</i>	
				2			
				3			<i>piove</i>
Plural	<i>fijji</i>	<i>fijja</i>		1	<i>magnemi</i>	<i>magnema</i>	
				2	<i>magneti</i>	<i>magneta</i>	
				3	<i>magni</i>	<i>magna</i>	

been grafted onto the verbal paradigm from the nominal paradigm. This is shown in Table 1.2 by a comparison of the nominal and verbal paradigms of *fijj* ‘son, daughter’ and *prəfuttə* ‘ham’, on the one hand, and *magnà* ‘eat’ and *piovərə* ‘rain’ on the other.

Far from being limited to finite verbs, such a rich and complex system of agreement has come to permeate all instantiations of Infl, including so-called non-finite verb forms such as perfective participles (16a), gerunds (16b), and (rhizotonic) infinitives (16c).

- (16)
- a.

Lu

frəki

/

le

frəkine

è

ddərmitu

/

ddərmitə

(Rip.)

the

boy

the

girl

be.PRS.3SG

sleep.PTPC.MSG

sleep.PTCP.FSG

‘The boy / the girl has been sleeping.’
- b.

Mamme

stieve

cucənnə

(Rip.)

mum

stand.PST.IPFV.FSG

cook.GER.FSG

‘Mum was cooking.’
- c.

Sai

skrivu

/

skrive

?

(Rip.)

know.PRS.2SG

write.INF.MSG

write.INF.FSG

‘Do you know how to write?’

However, these agreement features are not limited to Infl in Ripano, but are extended to all functional heads and the categories they host (Ledgeway 2012a: 308–10).¹⁰ As such, overt agreement with the clausal subject is found on parts of speech including in the C-domain *wh*-interrogatives (17a), in the *v*-domain predicative nominal complements (17b), and quite exceptionally within the nominal domain prepositions (17c), a behaviour which aligns Ripano with languages such as Welsh.

- (17)
- a.

C'aveti

peuri

/

C'aveta

peura

(Rip.)

have.PRS.2MPL

fright.MPL

have.PRS.2FPL

fright.FPL

‘You are afraid.’

¹⁰ For a full discussion of the phonological, morphological, and syntactic conditions operating on the distribution of Ripano agreement, see Paciaroni and Loporcaro (2018).

- b. **quannu** *passu* / **quanne** *passe* *lloka* (Rip.)
 when.MSG pass.MSG when.FSG pass.FSG there
 ‘whenever I/you(s)he pass(es) by there’
- c. **văcinu** *lu* *mara* / **văcine** *le* *case* (Rip.)
 near.MSG the.MSG sea.MSG near.FSG the.FSG house.FSG
 ‘near the sea/house’

We turn finally to a brief examination of the Romanian imperative. In contrast to other Romance and Indo-European languages,¹¹ the Romanian second person singular positive imperative frequently (though not systematically) marks, outside of the first conjugation, a transitivity distinction through an inflexional alternation between the desinences *-i* and *-e* which respectively encode the intransitive and transitive nature of the verb (Pîrvulescu and Roberge 2000; Zafiu 2013: 36f.; Maiden 2016a: 108; Maiden et al. 2021: 304f.).

- (18) a. **Fierbi!** (Ro.)
 boil.IMP.2SG
 ‘Boil!’
- a’. **Fierbe** *macaroanele!* (Ro.)
 boil.IMP.2SG *macaroni.DEF*
 ‘Boil the macaroni!’
- b. **Arzi!** (Ro.)
 burn.IMP.2SG
 ‘Burn!’
- b’. **Arde** *toate* *documentele!* (Ro.)
 burn.IMP.2SG *all.FPL* *documents.DEF*
 ‘Burn all the documents!’
- c. **Adormi!** (Ro.)
 fall.asleep.IMP.2SG
 ‘Go to sleep!’
- c’. **Adoarme-!** (Ro.)
 put.to.sleep.IMP.2SG=ACC.3MSG
 ‘Put him to sleep!’

Data like these, which reflect a transitivity feature encoded in the verb (presumably a formal feature of the light *v* functional head), prove truly remarkable. In Romance finite verbs display inflexional marking for the

¹¹ Even outside the imperative, formal marking of transitivity proves extremely rare cross-linguistically, as highlighted by Trudgill’s (2011: 73f., 105f.) remarks about the exceptional formal distinction between transitive and intransitive infinitives (the latter marked by final *-y*) in some English dialects of the south-west of England, e.g., Dorset *he can’t hit it* vs *he can’t swimmy*.

subject, not the object.¹² Inflexional marking on the verb referencing the object, by contrast, is restricted in Romance to non-finite verb forms such as the perfective participle in accordance with well-documented diachronic and diatopic variation (Loporcaro 1998; 2016). The Romanian examples in (18) highlight therefore two important typological considerations. First, the Romanian positive imperative shows an inflexional distinction for transitivity which is otherwise unparalleled in most languages of the world (cf. n. 11). Second, it is a frequent traditional observation that one of the most notable characteristics of the imperative is its lack of any inflexional marking or, at the very least, very minimal inflexional marking in accordance with a widespread cross-linguistic tendency (Pott 1859: 613; Bybee 1985: 173; Maiden 2006; 2007; Floricic 2008: 10; Ledgeway 2014c). Yet, the Romanian facts do not readily fit with this picture in that Romanian not only presents an inflexional alternation in contrast to the otherwise robust cross-linguistic tendency towards inflexional invariance, but also a minimal inflexion which encodes a typologically unusual distinction concerning the object.

1.2.2.4 The Interfaces

Functional categories also have an important role to play in understanding those linguistic phenomena whose surface form and formal licensing represent the convergence of two or more areas of the grammar. Indeed, investigation of the interfaces has figured prominently in much recent generative research thanks to developments such as phase theory (cf. Chomsky 2001; 2008), which postulate through the cyclical Spell-Out operation a direct mapping between narrow syntax and the conceptual-intentional and sensorimotor interfaces, and the cartographic enterprise (cf. Rizzi 1997; 2004; Cinque 1999; 2002; 2006; 2010; Belletti 2004; Benincà and Munaro 2010), which attempts to build semantic and pragmatic representations into the formal morphosyntactic architecture of the clause.¹³ In both cases functional categories and their associated projections prove fundamental: phases represent autonomous phonological and semantico-syntactic derivational domains defined in terms of the functional heads C, v, and D and their extended projections, whereas cartographic analyses assume richly articulated functional structures composed of a universally

¹² A notable exception is the dialect of Ripatransone where, under specific conditions, the finite verb may also encode the features of the object in conjunction with those of the subject (cf. Ledgeway 2012a: §6.3.4).

¹³ See, among others, Burkhardt (2005), Späth (2007), Grohmann (2009), Ramchand and Reiss (2012), Rothman and Slabakova (2011). Studies of the interfaces in relation to Romance data include Rizzi and Savoia (1993), Zubizarreta (1998), Elordieta et al. (2003), D'Imperio et al. (2005), Rao (2008), Silvestri (2009), Folli and Ulbrich (2010), Scheer (2011), Cruschina (2012), D'Alessandro and Scheer (2015), Fischer and Gabriel (2016), Manzini and Savoia (2016), Ledgeway (2018b; 2021a), Cruschina, Ledgeway, and Remberger (2019).

fixed order of projections which transparently encode and license various grammatical distinctions.

In this respect, Romance functional categories present the linguist with many valuable opportunities to study the linguistic interfaces, offering numerous insights into how different components of the linguistic system – syntactic, phonological, morphological, semantic, and pragmatic – do not necessarily operate in isolation but, rather, interact to license phenomena whose nature and distribution can only be fully understood in terms of the formal mapping between the interfaces. Consider, for instance, *rafforzamento fonosintattico* or *raddoppiamento fonosintattico* (RF) ‘phonosyntactic reinforcement/doubling’, a phonological process of sandhi assimilation in the dialects of southern Italy whereby a class of words that historically ended in a final consonant (e.g., *PLUS* > Cos. *cchiù* ‘more’) cause the lengthening or strengthening of the initial consonant of the following word in word₁ + word₂ sequences (e.g., Cos. *cchiù* [**m**:]ele ‘more honey’).¹⁴ It is not, however, a simple phonological rule whose application is based on pure linear adjacency, but shows sensitivity to structural configuration (Ledgeway 2009a: 46f.; 2018b). By way of illustration, consider the possible effects of the third person singular finite verb *vena* ‘comes’ (cf. < Lat. *UENIT*) on the form of the feminine singular definite article (< (I)LLAM) in the Cosentino near-minimal pairs in (19a–b), where under RF the vibrant represents the restoration of an underlying word-initial Latin long lateral (viz., [ll]- > [dd]- > [dd]- (> [j]-) > [r/r]-).

- (19) a. Quannu vena (**r)a primavera, mi sientu ggìa miegliu. (Cos.)
 when come.PRS.3SG the spring me= feel.PRS.1SG already better
 ‘When(ever) spring comes, I already start to feel better.’
 b. Quannu vena ** (r)a primavera, m’ affittu na casa a ru mare. (Cos.)
 when come.PRS.3SG the spring me=rent.PRS.1SG a house at the sea
 ‘When the (= this) spring comes, I’ll rent a house by the sea.’

In (19a), but not in (19b), the definite article cannot occur in its ‘reinforced’ variant with the initial vibrant despite immediately following the third person singular finite verb, but assumes its simple vocalic realization. As a result, the semantico-pragmatic reading of the two immediately postverbal subjects is not the same in (19a–b). In the former, the subject, although definite, is not referential but, rather, receives a generic interpretation, hence the unbounded reading of *quannu* ‘whenever’. In (19b), by contrast, *quannu* has its bounded interpretation and the postverbal definite subject, now marked by RF, is concomitantly fully referential, identifying a specific and known referent salient in the discourse or the extra-linguistic

¹⁴ Cf. Rohlfs (1966: 235–38), Loporcaro (1988; 1997b), Maiden (1995: 72–76), Fanciullo (1997), Sampson (2016: 675f.). For further discussion in this volume, see also Sections 6.4.3.2 and 7.4.2.

context which we can characterize as topical (hence the reading ‘the/this spring’).

The distribution of RF observed in these examples therefore highlights how Cosentino formally distinguishes between postverbal non-referential definite subjects and their referential variants. Given the assumption that for RF to take place Word₁ and Word₂ must surface in the same phasal domain (cf. Ledgeway 2018b; 2021a), there emerges a principled explanation for the facts in (19a–b). In particular, adopting Belletti’s (2004; 2005) seminal idea that the edge of the *v*-domain makes available a lower left periphery with dedicated Topic and Focus positions (cf. also Section 26.2.3), we can assume a direct mapping between syntax and pragmatic-semantic interpretation such that all referential constituents, when not raised to the higher left periphery within the C-domain, target a Topic or Focus position within the lower left periphery, whereas all non-referential constituents remain *in situ* within the lexical VP (cf. also Diesing’s 1992 Mapping Hypothesis). Consequently, we can associate the minimal pair in (19a–b) with the structural representations in (20a–b), where the presence of RF on the postverbal definite subject in (20b) signals a referential reading of the subject raised to SpecTop, namely, ‘When the (= this) spring comes’, whereas its absence in (20a) correlates with a non-referential interpretation of the definite subject *in situ*, namely, ‘Whenever spring comes’.

- (20) a. Quannu vena [TopP _____ [vP ~~vena~~ [vP ~~vena~~ a primavera]], ...
 b. Quannu vena [TopP [Spec *ra* primavera] [vP ~~vena~~ [vP ~~vena~~ a primavera]], ...

Following Ledgeway and Lombardi (2005), the finite verb in Cosentino targets a low functional head situated above the *v*-VP complex. It therefore follows that RF is licensed with referential postverbal subjects such as (20b) where the finite verb (viz. Word₁) and the immediately postverbal constituent (viz. Word₂) are transferred to the phonological component of the grammar in the same higher phasal cycle, since the postverbal subject surfaces in the left edge of the lower *v*P phase from where, in accordance with the Phase Impenetrability Condition, it remains accessible to phonosyntactic processes of the higher CP phase. In (20a), by contrast, the postverbal subject from its *in situ* position remains inaccessible to the potential RF effects of the third person singular finite verb, since it is contained within the lower *v*P phase from where it is sent to the phonological component in the lower cycle before the spell-out of the RF trigger in the higher phasal cycle.

We therefore see that the distribution of Cosentino RF involves an isomorphic mapping of syntax and phonology at the interfaces, with phonological domains aligning with syntactic domains to externalize in the phonological component syntactic information which, in turn, may spell out key semantico-pragmatic distinctions such as referentiality and topicality. In particular, the licensing of RF is constrained by specific

locality conditions which can be exhaustively computed and modelled in terms of a phase-theoretic cartographic approach, providing new and interesting data to further test the nature and computation of phasal domains ultimately defined in terms of specific functional projections. At the same time, we have seen that these same structural representations explicitly encode semantico-pragmatic information through the activation or otherwise of (lower) left-peripheral positions which, though not necessarily linearly distinguished on the surface, witness the immediately postverbal position of both subjects in (19a–b), nonetheless leave their mark in the phonological component which reads and externalizes these postverbal positions in distinct phasal cycles. We thus witness in these examples the output of an interaction of the syntactic, phonological, semantic, and pragmatic components of the grammar which contrive to derive strings which can be read at each of the interfaces.

1.2.2.5 Interim Conclusions

From the preceding introductory discussion, it is clear that the wealth of Romance standard and especially dialectal data, although frequently overlooked, have a great deal to contribute to research into such areas as parametric variation, linguistic universals, typological variation, and the interfaces. Nonetheless, the syntax of the dialects still represents a relatively poorly understood area of Romance linguistics, to the extent that there still remains a considerable amount of fieldwork to be done in recording and cataloguing the linguistic diversity within the România, as well as in bringing such facts to the attention of the wider linguistic community. With this in mind and keeping the technical detail to a minimum, in what follows a number of issues relating to the syntax of Romance will be discussed under the two broad headings of what Romance can do for syntactic theory and what syntactic theory can do for Romance. Under the former heading a number of assumptions will be reviewed about language structure and variation that have been proposed in the literature, demonstrating how in the specific cases examined the Romance data contradict such principles and parameters, rendering them either invalid or in need of further elaboration. Under the latter heading, by contrast, some of the less familiar and more problematic aspects of Romance syntax will be brought to light which can be shown to find an enlightening interpretation in light of current theoretical assumptions.

1.2.3 What Romance Can Do for Syntactic Theory

1.2.3.1 Pro-drop Parameter

One of the best known and most widely studied parameters is the so-called pro-drop (or null subject) parameter (for recent overviews, see Biberauer,

Holmberg, Roberts, and Sheehan 2009; Koeneman and Zeijlstra 2021; and Roberts 2019: ch. 3). Limiting our attention to Romance and Germanic, it is traditionally claimed that, with the exception of modern French, morphological Agr(eement) for person and number on the Romance verb is sufficiently rich to license a null subject (21a), whereas in such languages as English Agr is so impoverished that that it is unable to recover the identity of a null pronominal subject which must instead be phonologically expressed (22a; cf. also Section 17.5.1). By the same token, it is also assumed (cf. Chomsky 1981: 28; Rizzi 1986: 410; Haider 2001: 285; Koeneman and Zeijlstra 2021: §2.2.3) that expletive (or non-referential) pronouns are null in the former (21b) but overt in the latter (22b).

- (21) a. Ø (/Él) llo^{ra}. (Sp.)
 Ø he cry.PRS.IND.3SG
 b. Ø / **Él / **Ello llue^{ve}. (Sp.)
 Ø he it rain.PRS.IND.3SG
- (22) a. He/ **Ø cries.
 b. It Ø rains.

On a par with others, Rizzi (1982: 143) derives this supposed universal distinction from the pro-drop parameter, which he argues yields the four language types illustrated in Table 1.3.

Table 1.3. *Typology of null subjects*

Subject pronoun	Type 1 English	Type 2 Spanish	Type 3 German	Type 4 ?
null referential	–	+	–	+
null expletive	–	+	+	–
	non-pro-drop language	consistent pro-drop language	partial pro-drop language	

Language types 1 and 2 are exemplified by English and Spanish, respectively. In Spanish both null expletives and null referential pronouns are licensed, an example of a so-called consistent pro-drop language (Holmberg and Roberts 2010), whereas in English both types of null pronoun are excluded. Type 3 is argued to characterize German, an example of a so-called partial pro-drop language where, in contrast to referential pronouns which are invariably overt (23a), (certain classes of) overt expletive pronouns are only licensed when they occur in clause-initial position to satisfy the surface V2 requirement (23b–c):

- (23) a. Er / **Ø weint. (Ger.)
 He Ø cry.PRS.IND.3SG
- b. Es wird heute getanzt. (Ger.)
 it become.PRS.IND.3SG today dance.PTCP
- c. Heute wird Ø / **es getanzt. (Ger.)
 today become.PRS.IND.3SG Ø it dance.PTCP
 ‘There is dancing going on today.’

On the other hand, Rizzi (1982: 143) explicitly argues that type 4 languages with overt expletive subjects but null referential subjects are ‘excluded for intrinsic reasons’. However, the evidence of a number of non-standard Romance varieties demonstrates that type 4 languages do indeed exist. For example, although Neapolitan is a pro-drop language (24a), it also displays structures such as those in (24b) where the subject position is filled by the overt expletive *chello* ‘that’ (Sornicola 1996; Ledgeway 2009a: 290–94; 2010a), a pattern replicated by a number of other non-standard Romance varieties (25a–g).

- (24) a. Ø / Isso chiagne. (Nap.)
 Ø he cry.PRS.3SG
- b. Ø / Chello chiove. (Nap.)
 Ø that rain.PRS.3SG
- (25) a. El sera verdade que o centro fala e a periferia non
 it be.FUT.3SG truth that the centre speak.PRS.3SG and the periphery NEG
 responde? (Glc.)
 reply.PRS.IND.3SG
 ‘Is it true that the centre speaks while the periphery fails to reply?’
- b. Ele era umas dores de cólica enorme! (EuPt., Unhais da Serra)
 it be.PST.IPFV.3SG some pains of colic enormous
 ‘It was a horrible case of stomach cramps!’
- c. ¡Ello hay un búho en el techo! (Dominican Sp.)
 it have.PRS.IND.3SG an owl in the rood
 ‘There’s an owl on the roof!’
- d. Ell és veritat! (Bal.Cat.)
 It be.PRS.IND.3SG true
 ‘It is true!’
- e. Iddu cchi mm’ importa? (Ctz.)
 it what me= matter.PRS.3SG
 ‘What does it matter to me?!’
- f. Ma iddu chi cc’ è cosa? (Plm.)
 but it what there= be.PRS.3SG what
 ‘What is it?!’
- g. kə pløj. (Corrèze, Nocc.)
 that rain.PRS.IND.3SG
 ‘It is raining.’

With the exception of the northern Occitan example in (25g), for which see Oliviéri (2009) and Kaiser, Oliviéri, and Palasis (2013), in most cases the use of the overt expletive in such examples is associated with specific pragmatic functions (Sornicola 1996: 325f.; Ledgeway 2003; 2009a: 290f.; Hinzelin 2009; Corr 2015; 2016), typically marking the illocutionary force of the clause as exclamative (cf. 25b,c,d) or interrogative, often with rhetorical overtones (cf. 25a,e,f), a usage which still requires much more detailed investigation. Summing up, the data considered here lead us to conclude, with Kaiser, Oliviéri, and Palasis (2013), that the ability to drop referential pronouns and the licensing of overt expletives are not necessarily mutually exclusive or, for that matter, two interrelated properties of a single pro-drop parameter (cf. also Roberts 2019: ch. 3). At the same time, however, as argued in Ledgeway (2010a), the overt expletives of most of these Romance dialects cannot be equated *tout court* with those of languages like English and French in view of the marked pragmatic functions of the former and the purely syntactic nature of the latter.

The wealth and extent of the Romance evidence is such that our typology of null subjects does not, however, stop here. Turning our attention to medieval Romance, it has long been noted that null subjects in these varieties display an asymmetric distribution,¹⁵ a pattern more robustly represented in Gallo-Romance (French, Occitan, and northern Italian dialects) than either in Ibero-Romance or central-southern Italo-Romance (Wolfe 2018). In particular, whereas in root clauses null subjects are freely licensed (26a–c), in subordinate clauses pronominal subjects must usually be phonologically expressed (27a–c), although not interpreted as emphatic or contrastively focused. Illustrative in this respect are the old French and old Umbro-Tuscan examples in (27a,c), where, despite the coreference of main and embedded clause subjects, the latter is overtly realized yielding a structure which would be judged ungrammatical, for example, in modern Spanish or Italian where the presence of an overt pronoun in the same context would typically signal switch reference.

- (26) a. Et l'ors vint Ø en sa chambre (OFr., *Histoire ancienne* 604.16)
and then come.PST.PFV.3SG Ø in her room
'And then she came into her room'
- b. Del cor Ø sospir e dels olhs Ø plor (OPrv., *Era-m cossehlitz, senhor* 19)
of.the heart Ø sigh.PRS.IND.1SG and of.the eyes Ø cry.PRS.IND.1SG
'I sigh from the heart and cry from the eyes'
- c. manifestamente Ø l' hoe veduto nelle cose [...] (OTsc., *Novellino* 3)
manifestly Ø it=have seen in.the things
'I have seen this clearly in those things [...].'

¹⁵ Adams (1987), Hirschbühler and Junker (1988), Dupuis (1989), Roberts (1993: §3.2), Benincà (1994; 2006; 2010: §3.2.1), Salvi (2004: 16f., 26–31), Dufier (2010), Ledgeway (2012a: 74f.; 2021b: §§2.2.3–4).

- (27) a. Por ce qu' ele_i estoit tostans en doute qu' ele_i ne perdist ce
 for this that she_i be.PST.IPFV.3SG always in doubt that she NEG=lose.PST.SBJV.3SG this
 qu' ele_i trop amoit (OFr., *Histoire ancienne* 604.7)
 that she too love.PST.IPFV.3SG
 'Because she was always in doubt that she would lose what she so dearly loved'
- b. Non es meravelha s' eu chan / Melhs de nul autre
 not be.PRS.IND.3SG wonder if I sing.PRS.IND.1SG better of no other
 chantador (OPrv., *Era-m cossehlitz, senhor* 1–2)
 singer
 'It is no wonder if I am a better singer than all others'
- c. Elli_i conosce certamente ch' elli_i avea ucciso lo migliore
 he know.PRS.IND.3SG certainly that he have.PST.IPFV.3SG kill.PTCP the best
 cavaliere del mondo (OUmb.-Tsc., *Tristano Riccardiano* 238.6–7)
 knight of.the world
 'He certainly knows that he had killed the best knight in the land'

This asymmetrical distribution leads us to conclude that null subjects in medieval Romance were not licensed exclusively, if at all, by rich verb agreement for person and number, insofar as the featural specification of Infl is equally rich in both root and subordinated clauses, but by a structural property (to be discussed in Section 1.2.4.2) which aligns the verb with distinct functional positions in both clause types. Indeed, the unreliability of verb morphology as a diagnostic for correctly predicting the distribution of null subjects is further evidenced by Ripano, a pro-drop variety, where we have seen that finite verbs (cf. Table 1.2) privilege the marking of gender over that of person, witness the three singular persons of the paradigm *magnu* (MSG) vs *magne* (FSG), despite traditional claims that the licensing of null subjects is directly linked to (the overt marking of) the person feature.¹⁶ In summary, the null pronominal types seen in medieval Romance and Ripano cannot be readily accommodated in terms of current theories of a binary null subject parameter (cf. also Oliviéri 2009; Kaiser, Oliviéri, and Palasis 2013) and, in particular, the typology of null subjects predicted by such models as that observed in Table 1.3 which ultimately reduce the distribution of null pronominals to the availability of rich morphological Agr for person on the functional head Infl lexicalized by the finite verb.

1.2.3.2 Verb Positions

Exploiting the fixed positions of VP-adverbs like ALWAYS as a diagnostic indicator of the left edge of the v-VP complex, it is possible to distinguish between overt verb-raising languages like French (28a), where the finite

¹⁶ Cf. Koenenman and Zeijlstra (2021: §2.1), who claim that in consistent pro-drop languages the minimal requirement for the licensing of null subjects is that the verb morphology express at the very least person and number.

verb raises to the Infl position to the left of VP-adverbs, and languages like English (28b), where the verb remains *in situ* to the right of such VP-adverbs and the Infl position is not overtly lexicalized in the syntax, a difference traditionally retraced to the respective richness of verbal inflexion in the two languages (Emonds 1978; Pollock 1989; Belletti 1990).

- (28)
- | | | | | | | |
|----|------|-------------------|-------------|-------------------|----------------|-----------------------|
| a. | Jean | [_{Infl} | skie | [_{v-VP} | <u>souvent</u> | skie]] (Fr.) |
| b. | John | [_{Infl} | Ø | [_{v-VP} | <u>often</u> | skis]] (Eng.) |

Nonetheless, recent research has revealed a much more nuanced interpretation of Romance verb movement than these familiar broad-brush treatments which classify Romance *tout court* as having overt verb movement (for in-depth discussion, see Sections 17.3, 20.2.1, 21.2.3; and Schifano 2018). Following the seminal work of Cinque (1999), Infl is now commonly interpreted as a general label for the rich inflexional area of the clause (the I-domain) made up of a series of distinct functional projections dedicated to marking various temporal, aspectual, modal, and voice distinctions ranging over the lexical verb, its arguments, and possible adjuncts which can also be identified by the semantically corresponding adverbial modifiers they host (cf. also Cinque 2002; 2006; Belletti 2004; Rizzi 2004). Armed with these assumptions about a universal fixed hierarchy of adverb positions and corresponding functional projections, it is now possible to construct a fine-grained typology of Romance varieties along the lines of (29):

- (29)
- | | | | | | | |
|----|---------------------------------------|---------------------|-----------------|-----------------------|-----------------|--------------------------------|
| | | [_{IP} ... | | [_{v-VP} ... | |]] |
| a. | | <u>normalmente</u> | | <u>todavía</u> | veo | <u>todo</u> veo (Sp.) |
| b. | | <u>di solito</u> | vedo | <u>ancora</u> | vedo | <u>tutto</u> vedo (It.) |
| c. | je vois | <u>d'habitude</u> | vois | <u>encore</u> | vois | <u>tout</u> vois (Fr.) |
| | I see.PRS.IND.1SG | usually | see.PRS.IND.1SG | still | see.PRS.IND.1SG | all |
| | 'I can usually still see everything.' | | | | | |

Although in all three varieties exemplified in (29) the finite lexical verb invariably leaves its base position to vacate the verb phrase, witness its position to the left of the completive adverb ALL/EVERYTHING immediately adjacent to the v-VP, it raises to different functional projections within the I-domain as illustrated by its differential position with respect to different adverb classes. For example, in Spanish (29a) the finite verb raises to the head position of the continuative aspectual projection immediately below the adverb STILL, whereas in Italian (29b) it raises slightly higher to the head position of the habitual aspectual projection below the adverb USUALLY, and in French (29c) it raises to the highest available position above all adverb classes.

On the basis of evidence like this, a number of surface differences across Romance can be interpreted in terms of the varying extent of verb movement around different adverb classes. Traditionally there have been

many attempts to relate the extent of movement to the richness or otherwise of the verb's inflexional features (Roberts 1985; Lightfoot and Hornstein 1994; DeGraff 1997; Rohrbacher 1997; Vikner 1997; Biberauer and Roberts 2010; Zwart 2020). In essence, approaches of this type attempt to derive syntactic operations from cross-linguistic morphological differences in individual languages (cf. also the discussion of the *Mirror Principle* in Section 17.3). Admittedly, this view finds some initial support in the Romance vs Germanic contrast in (28a–b) where the Romance verb form, undoubtedly the inflexionally richest of the two, raises the highest. However, a brief comparison of the results reported in (29a–c) above suffices to dispel such an approach, inasmuch as all the Romance varieties exemplified are what may be termed inflexionally rich languages, yet they display some quite marked differences in the extent of finite lexical verb movement. This conclusion is further substantiated by the observation that much of the rich inflexion of the modern French verb, unlike that of Spanish or Italian, is predominantly orthographic, yet it shows higher verb movement of finite lexical verbs than the other two varieties.

Thus, to conclude, the rich comparative evidence of multiple verb positions observed across Romance forces us to postulate a richer functional structure for the clause than has traditionally been assumed. More generally, it has been established that from an empirical and a theoretical perspective broad-brush characterizations of Romance as invariably involving overt V-raising prove neither descriptively nor explanatorily adequate, inasmuch as a more nuanced picture has to be recognized. At the same time, this same evidence has highlighted the danger of assuming a direct correlation between the richness of inflexion and the extent of verb movement.

1.2.3.3 Mapping the Left Periphery of the Clause

As seen in the previous section, one area of considerable interest in much recent syntactic research has been the role of functional categories in throwing light on the fine structure of the clause (for an overview, see Cruschina and Ledgeway 2016). Standardly, the structure of the clause in a typical SVO language, of which all the modern Romance varieties are arguably examples, has been taken to present (at least) the positions indicated in the linear template in (30a), exemplified from Italian in (30b):

- (30) a. Subject Aux Adverb Verb Object Adjunct(s)
 b. Ugo ha sempre preparato la pasta a cena. (It.)
 Ugo have.PRS.IND.3SG always prepare.PTCP the pasta at dinner
 'Ugo has always made pasta for dinner.'

The linear arrangement in (30a–b) highlights how the confines of the sentential core can be identified with the preverbal subject position situated

at the left edge of IP and the complement or adjunct position situated at the right edge of *v*-VP (31a). However, following the seminal work of Rizzi (1997), in recent years research within generative syntax has increasingly focused on the investigation of the C-related functional categories and the positions they lexicalize within the left periphery (cf. also Benincà and Munaro 2010), the syntactic space immediately to the left of the sentential core (31b).

- (31) a. [IP S Aux [_{v-VP} V O (X)]]
 b. [Periphery ... [IP S Aux [_{v-VP} V O (X)]]]

Unsurprisingly, a considerable amount of work on the split C-domain has been conducted on the basis of the rich (dialectal) microvariation offered by Romance varieties (for an overview and relevant bibliography, see Ledgeway 2012a: 154–71, and Section 20.2.1), which in many cases provide invaluable overt evidence with which to map the fine structural organization of the left periphery. In particular, the left periphery, traditionally defined in terms of CP and its associated Spec(ifier) and head positions hosting *wh*-operators and complementizers (Chomsky 1986: §1; cf. also Section 1.2.2.1), respectively, is now conceived as a split domain, hierarchically articulated into several fields and associated projections (for an in-depth overview, see Chapter 26, this volume). Revealing in this respect are Italian topicalization examples such as (32a–c), where one of the constituents of the sentential core in (30b) conveying old/given information has been fronted to the left periphery and picked up, where available, by a resumptive clitic pronoun on the verb in the sentential core, a case of (clitic) left-dislocation.¹⁷ Yet, this does not exhaust all available structural possibilities, as demonstrated by the examples in (33a–d) where as many as two constituents have been fronted under (clitic) left-dislocation, illustrating the fact that topic is a recursive category capable of multiple realizations within the same utterance.

- (32) [Topic_{OLD} [S V Adv O (X)]]
 a. [La pasta_i, [Ugo la_i prepara sempre a cena.]] (It.)
 the pasta Ugo it= prepare.PRS.IND.3SG always at dinner
 b. [Ugo_i, [lui_i prepara sempre la pasta a cena.]] (It.)
 Ugo he prepare.PRS.IND.3SG always the pasta at dinner
 c. [A cena, [Ugo prepara sempre la pasta.]] (It.)
 at dinner Ugo prepare.PRS.IND.3SG always the pasta

¹⁷ In these and following examples we indicate topicalized constituents by underlining, contrastively focalized constituents by small caps, and informationally focalized constituents with bold.

- (33)
- | | | | |
|----|---|---|--|
| | [Topic _{OLD} | [S V Adv O (X)] | |
| a. | [Ugo _i , <u>la pasta</u> _j ,
Ugo the pasta | [lui _i la _j prepara
he it= prepare.PRS.IND.3SG | sempre a cena.]] (It.)
always at dinner |
| b. | [<u>La pasta</u> _j , Ugo _i
the pasta Ugo _i | [lui _i la _j prepara
he it= prepare.PRS.IND.3SG | sempre a cena.]] (It.)
always at dinner |
| c. | [Ugo _i , <u>a cena</u>
Ugo at dinner | [lui _i prepara
he prepare.PRS.IND.3SG | sempre la pasta.]] (It.)
always the pasta |
| d. | [<u>A cena</u> , <u>la pasta</u> _i
at dinner the pasta | [Ugo la _i prepara
Ugo it= prepare.PRS.IND.3SG | sempre.]] (It.)
always |

Further support for the richly articulated structure of the left periphery comes from a consideration of focus structures such as the Italian examples in (34), where a constituent of the sentential core in (30b) is fronted under corrective focus to correct a previous assertion (cf. 34a–c). Significantly, however, fronted focused constituents are not in complementary distribution with fronted topicalized constituents, as the traditional simplex CP structure with its single left-peripheral position presupposes, but can co-occur with left-dislocated topics in the strict order Topic + Focus, as shown in (35a–d).

- (34)
- | | | | |
|----|------------------------|---|---|
| | [Focus _{NEW} | [S V O (X)] | |
| a. | [SEMPRE
always | [Ugo prepara
Ugo prepare.PRS.IND.3SG | la pasta a cena.]] (It.)
the pasta at dinner |
| b. | [A CENA
at dinner | [Ugo prepara
Ugo prepare.PRS.IND.3SG | sempre la pasta.]] (It.)
always the pasta |
| c. | [LA PASTA
the pasta | [Ugo prepara
Ugo prepare.PRS.IND.3SG | sempre a cena.]]
always at dinner |
- (35)
- | | | | |
|------|---|---|---|
| | [Topic + Focus | [S V O (X)] | |
| a. | [Ugo _i , SEMPRE
Ugo always | [lui _i prepara
he prepare.PRS.IND.3SG | la pasta a cena.]] (It.)
the pasta at dinner |
| b. | [Ugo _i , A CENA
Ugo at dinner | [lui _i prepara
he prepare.PRS.IND.3SG | sempre la pasta.]] (It.)
always the pasta |
| c. | [<u>La pasta</u> _i , SEMPRE
the pasta always | [Ugo la _i prepara
Ugo it= prepare.PRS.IND.3SG | a cena.]] (It.)
at dinner |
| d.** | [SEMPRE <u>la pasta</u> _i
always the Pasta | [Ugo la _i prepara
Ugo it= prepare.PRS.IND.3SG | a cena.]] (It.)
at dinner |

Evidence like this from Italian and other Romance varieties highlights that the existence of a single left-peripheral position is empirically inadequate. Rather, the relevant left-peripheral positions must be reconceived as distinct pragmatic-syntactic spaces along the lines of Benincà and Poletto (2004), according to which we can identify from left to right at least two fields termed Topic and Focus, respectively (see also Cruschina 2012; Cruschina and Ledgeway 2016: §31.3.4). Not only is this demarcation between both fields justified at a pragmatic-semantic level, in that

elements appearing in the Topic field are generally interpreted as ‘old’ or ‘given’ information whereas the Focus field is typically associated with informationally ‘new’ elements, but it also finds confirmation at the syntactic level. For instance, in contrast to elements appearing within the Topic field, which often call for a resumptive pronominal (clitic) where available (cf. 32a–b, 33a–d), those appearing within Focus (cf. 34c) prove incompatible with a pronominal copy. Moreover, it has already been observed that topic is a recursive syntactic category allowing several reiterations within a given utterance, whereas focus is restricted to just one occurrence per utterance.

Robust evidence like this demonstrates that topicalized and focused constituents target distinct spaces within the left periphery, forcing us to recognize a representation of the C-domain along the lines of (36).

(36) [_{CP} Comp [_{TopP} Topic^{REC} [_{FocP} FOCUS [_{IP} ...]]]]

1.2.3.3.1 Complementizer Positions

In addition to the Topic and Focus fields highlighted above, the left periphery of the clause also hosts subordinators or complementizers. Alongside finite complementizers derived from QUOD/QUID ‘that’ and QU(I)A ‘(because >) that’ which introduce tensed clauses (cf. (a) examples below), Romance varieties also present a series of non-finite complementizers derived from the prepositions AD ‘to’ and DE ‘of, from’ to introduce infinitival clauses (cf. (b) examples below), which to all intents and purposes parallel the use of their finite counterparts.

- (37) a. Et gouyat_i qu’ aymeré *que* [_{IP} Ø_{iij} dansèsses dap ére.] (Gsc.)
 the young.man that like.COND.3SG that Ø dance.PST.SBJV.2SG with her
 ‘The young man would like you to dance with her.’
 b. Et gouyat_i qu’ aymeré *a* [_{IP} Ø_{iij} dansa dap ére.] (Gsc.)
 the young.man that like.COND.3SG to Ø dance.INF with her
 ‘The young man would like to dance with her.’

- (38) a. Dignes-li_i *que* [_{IP} Ø_{iij} vingui!] (Cat.)
 tell.PRS.SBJV.2SG=DAT.3 that Ø come.PRS.SBJV.3SG
 b. Dignes-li_i *de* [_{IP} Ø_{iij} venir!] (Cat.)
 tell.PRS.SBJV.2SG=DAT.3 of Ø come.INF
 ‘Tell him to come!’

Despite appearances, the presumed parallelism between finite and non-finite complementizers is not, however, perfect, as revealed by their respective positions in relation to topics and foci: whereas finite complementizers precede topics and foci (39a), non-finite complementizers invariably follow both types of fronted constituent (39b).

- (39) a. So *che, la data*, [_{IP} l' ho sbagliata]. (It.)
 know.PRS.IND.1SG that the date it= have.PRS.IND.1SG mistake.PTCP
 b. So, *la data, di* [_{IP} averla sbagliata]. (It.)
 know.PRS.IND.1SG the date of have.INF=it mistake.PTCP
 'I know (that), the date, I got (it) wrong.'

Romance evidence like this forces us to assume that the Topic and Focus fields outlined above are, in turn, closed off upwards by a higher complementizer position (termed Force) marking the illocutionary force of the clause and hosting such items as the Romance finite declarative complementizer *quel/che* 'that', and downwards by a complementizer position (termed Fin(iteness)) specifying the modality and/or finiteness of the clause and hosting such items as the Romance infinitival complementizers *de/di* 'of', as schematicized in (40).

- (40) [_{Periphery} *quel/che*_{Force} + Topic^{REC} + FOCUS + *de/di*_{Fin} [_{IP}...]]

In fact, some Romance varieties present dual finite complementizer systems which appear to exploit both the higher and lower complementizer positions within the left periphery (cf. also Sections 2.6.2, 21.3.1). This is the case in Romanian and many southern Italian dialects,¹⁸ which contrast an indicative/realis complementizer (QU(1)A 'because' > *că/ca*) which lexicalizes the higher complementizer position, and therefore precedes topics and foci (41a), and a subjunctive/irrealis complementizer (SI 'if' > Ro. *să*, QUOD/QUID 'that' > USID *col/che*, QUO(MODO), (QUO)MODO 'how' > Sal. *cu*, Cal./NESic. *mu/mi/ma*) that lexicalizes the lower complementizer position, and therefore follows topics and foci (41b):

- (41) a. Cred [_{CP} *că* MÂINE [_{IP} merg la teatru.]] (Ro.)
 believe.PRS.IND.1SG that_{REALIS} tomorrow go.PRS.1SG to theatre
 'I believe that I'm going to the theatre TOMORROW.'
 b. Vreau [_{CP} MÂINE *să* [_{IP} merg la teatru.]] (Ro.)
 want.PRS.1SG tomorrow that_{IRREALIS} go.PRS.1SG to theatre
 'I want to go to the theatre TOMORROW.'

Additional compelling evidence for these two complementizer positions comes from those varieties which allow the simultaneous lexicalization of both positions around a fronted topic or focus constituent. These so-called 'recomplementation' structures are found across a wide range of Romance varieties (Ledgeway 2012a: §4.4.2.2; cf. also Section 2.6.3), including many early Romance texts (42a–b; Wanner 1998; Paoli 2003a; Ledgeway 2004:

¹⁸ For Romanian see, among others, Dobrovie-Sorin (1994: 93–111), Alboiu and Motapanyane (2000: §4.2), and for southern Italy Calabrese (1993), Lombardi (1997; 1998), Ledgeway (1998; 2004; 2005; 2006; 2007c; 2009b; 2012b; 2016: §63.3), Manzini and Savoia (2005, I: 455–501, 650–76), Ledgeway and Lombardi (2014), Colasanti (2018), Groothuis (2019), Andriani, Groothuis, and Silvestri (2020: §3).

§4.3.2.2; 2005: 380–89; Vincent 2006; Munaro 2016), and a number of modern Italo-Romance dialects (42c; Paoli 2002; 2003a; 2003b; 2005; D’Alessandro and Ledgeway 2010) and modern Ibero-Romance varieties (42d; Demonte and Fernández-Soriano 2009; Gupton 2010: 227–34; Villa-García 2010; 2015; Martínez Vera 2020).

- (42) a. Onde *dize* Josepho *que en casa de so padre que le*
 where say.PRS.IND.3SG Josepho that in house of his father that him=
llamaron *primera* *mientre* *Ciro* (OSp.)
 call.PST.PFV.3PL first -ly *Ciro*
 ‘Where upon Josepho said that in his father’s house he was originally called *Ciro*’
- b. *èy* *manifesta cosa che homo che se ave* *a defendere*
 be.PRS.IND.3SG obvious thing that man that self= have.PRS.IND.3SG to defend.INF
a la patria soa intre li amici e li canussienti suoy cha
 to the fatherland his among the friends and the acquaintances his that
ave *a chesta parte gran prerogativa e gran avantayo* (ONap.)
 have.PRS.IND.3SG at this part big prerogative and big advantage
 ‘it is abundantly clear that, a man who has to defend himself in his own country among his friends and acquaintances, has in this respect considerable privilege and advantage’
- c. A Teeja a *credda* *che a Maria ch’ a parta*
 the Teresa SCL believe.PRS.IND.3SG that the Mary that SCL leave.PRS.SBJV.3SG
duman. (Lig.)
 tomorrow
 ‘Teresa believes that Mary will leave tomorrow’
- d. *Dixéronme que a esse rapaz que o coñecemos na*
 say.PST.PFV.3PL=me that DOM that boy that him= know.PST.PFV.1PL in.the
festa. (Glc.)
 party
 ‘They told me that we met that guy at the party.’

Finally, unique within Romance is the situation found in Béarnais Gascon since around the sixteenth century, where the [+finite] feature of affirmative root clauses is exceptionally spelt out in the systematic lexicalization of the lower complementizer position C_{Fin} through the complementizer *que* ‘that’ (see also Section 21.3.2).¹⁹ Cross-linguistically, explicit typing of declarative force in this way represents an extremely rare option (cf. Lyons 1968: 307; Bybee 1985: 147; Cinque 1999: 130; Franco 2013; see also Section 24.2.1), yet Gascon highlights how models of language need to accommodate it. Firm proof that root-clause *que* spells out the lower complementizer head is provided by the observation that, apart from object clitics, nothing can intervene between *que* and the finite verb so that all preverbal lexical subjects must occur to the left of *que*. This latter observation highlights the fact that, unlike in other Romance varieties, there is

¹⁹ For extensive discussion and analysis, see Ledgeway (2020a) and sources cited therein.

apparently no preverbal subject position within the sentential core in Gascon, such that all subjects have to be syntactically fronted to a topicalized (typically if old and definite; 43a) or focalized (typically if new and indefinite; 43b) position within the left periphery. These facts find an immediate explanation if we assume that *que* lexicalizes the C_{Fin} position, since finiteness is standardly assumed to license nominative Case, hence Gascon would appear to have grammaticalized the locus of finiteness and, by implication, the licensing of nominative subjects within the left periphery, as happens in Verb Second (V2) varieties, rather than in the sentential core. Indeed, Ledgeway (2020a) shows how this exceptional typological development is externally motivated, being ultimately the result of contact with a non-Romance language, Basque, and can be formally integrated into the V2 Parameter if the latter can be satisfied by direct insertion ('external merge') of *que* in the C_{Fin} head rather than V-movement to the same (cf. also Ledgeway 2008).

- (43) a. [_{TopP} La Maria [_{FinP} *qu'* [_{IP} *apara las pomas.*]]] (Gsc.)
 the Maria that collect.PRS.IND.3SG the apples
 'Maria gathers up the apples.'
- b. [_{FocP} **Quauque** **trufandèr** [_{FinP} *que* [_{IP} *vos dirà. .*]]] (Gsc.)
 whatever joker that you= say.FUT.3SG
 'Any joker will tell you . . .'

This analysis is further supported by the behaviour of embedded declarative clauses such as (44) where both complementizer positions are simultaneously realized: whereas *que* continues to lexicalize C_{Fin} on a par with root declaratives, from where it follows fronted topics and foci, subordination is marked in the higher C-position Force, also lexicalized by *que* 'that' but which precedes fronted topics and foci.

- (44) *Que* *sémbre* [_{ForceP} *que* [_{FocP} *TOUTS* [_{FinP} *que* [_{IP} *pouderém séné*
 that seem.PRS.IND.3SG that all that can.COND.1PL without
 pénes debisa parié.]]]]] (Gsc.)
 difficulties speak.INF equally
 'It seems that we can all speak equally without problems.'

On the basis of the Romance evidence reviewed in the preceding sections, the fine structure of the left periphery can be summarized as in (45):

- (45) [_{ForceP} *que/che* [_{TopP} Topic^{REC} [_{FocP} *FOCUS* [_{FinP} *de/di/che/să/cu/mu* [_{IP} . .]]]]]]

1.2.3.4 Interim Conclusions

The discussion so far has highlighted the importance of Romance data, and especially those from the still underutilized non-standard varieties and dialects, in making us rethink some of our most basic assumptions

1.2.4 What Linguistic Theory Can Do for Romance

Undoubtedly one of the most striking differences between medieval and modern Romance varieties is manifested in the often radically differing word order patterns they permit in root clauses. By way of illustration, consider the old Neapolitan sentences in (46a–b) and their modern Italian translations in (47a–b):

- (46a–b) illustrate a frequent early Romance structure in which the preverbal position is occupied by some constituent other than the subject, namely the direct object (*viechy*) and a non-subcategorized adverbial phrase (*de poy queste parole*), respectively. In the former case, the fronted rhematic

direct object, which conveys new information, constitutes an example of informational focus (Lambrecht 1994: ch. 5; Cruschina 2012), and contrasts sharply with many modern Romance varieties such as Italian (cf. 34–35), where preposing of rhematic constituents is only found under quite restrictive pragmatic conditions to license corrective focus, insofar as rhematic objects conveying informational focus canonically occur in postverbal position (cf. 47a). Similarly, example (46b) demonstrates how when the preverbal position is occupied by a constituent distinct from the subject, the latter, whenever overtly realized, is generally required to follow the verb, giving rise to an apparent case of verb-subject inversion. Significantly, in this and similar examples of inversion the subject does not simply follow the verb but also precedes any other sentential constituents (subcategorized or otherwise), witness the order subject (*lo messayo*) + direct object (*licencia*) in (46b). In modern Italian, by contrast, postverbal subjects generally follow their associated objects and other sentential constituents, and in such cases are typically associated with rhematic interpretations, whereas the postverbal subject in (46b) is clearly thematic. Consequently, in the modern Italian translation in (47b) the thematic subject obligatorily occurs in preverbal position.

Word order patterns like these, which can be readily replicated for other early Romance varieties, have led a number of linguists to argue that medieval Romance word order is characterized by a V2 constraint, the origins of which can be retraced to late Latin (Ledgeway 2017b).²⁰ During this V2 stage sentences consist therefore of two principal parts (48a), a sentential core (IP) with fixed S V O Adv order on a par with what we have already witnessed in Section 1.2.3.3 for modern Romance, and a richly articulated left periphery (CP) along the lines of (45) to whose lowest C position (Fin) the finite verb raises in root clauses, where it is preceded by one or more elements fronted from the sentential core to the Topic and Focus fields to be assigned a pragmatically salient reading. In embedded clauses, by contrast, the left periphery generally hosts an overt Comp(lementizer) and the finite verb is consequently forced to remain within the sentential core, yielding the order S+V+O+Adv (48b). Thus, as the following representative early Romance examples demonstrate, alongside S+V+X (48c) we also frequently find in main clauses O+V(S) (48d), IO+V(S) (48e), O_{pp}+V(S) (48f), and Adv+V(S) (48g), whereas embedded clauses invariably display rigid S+V+O+Adv (48h):

²⁰ See, among others, Skårup (1975), Vanelli, Renzi, and Benincà (1985), Vanelli (1986; 1999), Adams (1987), Fontana (1993; 1997), Roberts (1993), Benincà (1994; 1995; 2006; 2013), Lemieux and Dupuis (1995), Ribeiro (1995), Vance (1995; 1997), Lombardi and Middleton (2004), Salvi (2004; 2012; 2016: 1005–09), Labelle (2007), Ledgeway (2007b; 2008; 2017b; 2021b), Lombardi (2007), Radwan (2011), Salvesen (2013; 2014), Bech and Salvesen (2014), Poletto (2014), Steiner (2014), Wolfe (2015a; 2015b; 2015c; 2015d; 2018a), Cruschina and Ledgeway (2016: 571f). For an analysis of V2 in old Romanian, see Nicolae (2015: 155–98), Nicolae and Niculescu (2015), and Dragomirescu and Nicolae (2015).

- (48) a. [LeftPeriphery_{Top} / Foc V [SententialCore S ∇ O Adv]]
 b. ... [LeftPeriphery Comp [SententialCore S V O Adv]]
 c. [CP Lo cavaliere prese [IP ~~lo cavaliere prese~~ i marchi]] (OTsc., *Novellino* 4)
 the knight take.PST.PFV.3SG the marks
 ‘The knight took the marks’
 d. [CP **Grande duelo avien** [IP las yentes
 great sorrow have.PST.IPFV.3PL the peoples
 cristianas ~~avien grande duelo~~] (OSp., *Mío Cid* 29–30)
 Christian
 ‘The Christian peoples felt great sorrow’
 e. [CP A ceste paroles respont [IP la reine ~~respont a ceste paroles~~] (OFr., *M. Artu* 59, 84)
 to these words reply.PRS.IND.3SG the queen
 ‘The queen replied to these words’
 f. [CP D’ALGUÑAS COUSAS *me calarei* [IP ~~me calarei d’algũas cousas~~]
 of some things myself=fall.silent.FUT.1SG
 (Opt., *Diálogos de S. Gregório* 1.5.25)
 ‘I shall remain silent about certain matters’
 g. [CP **Luenh es** [IP lo castelhs e la
 far be.PRS.IND.3SG the castle and the
 tors ~~es luenh~~] (OOcc., Rudel, *Pro ai del chan essenhadors* 3.1)
 tower
 ‘Far is the castle, far is the tower’
 h. la honret [CP q’ [IP **ella fetz** so q’ el
 her=honour.PST.PFV.3SG that she do.PST.PFV.3SG that that he
 volc]] (OPrv., *Era-m cossehlatz, senhor* 30,1)
 want.PST.PFV.3SG
 ‘And he honoured her such that she did what he wanted’

In conclusion, it has been seen that the fine structure of the sentential core (Section 1.2.3.3) and the left periphery (Section 1.2.3.3.1) independently established above on the basis of modern Romance data provide us with the necessary pragmatico-syntactic tools to interpret the facts of medieval Romance word order. However, above it was simply assumed that the finite verb in root clauses raises to the lowest C-head Fin, although the fine structure of the C-domain in (45) makes available two C-head positions. This has led some to argue that under V2 syntax the finite verb may also target the highest position C_{Force}. Following Roberts (2012; 2015), Wolfe (2015d; 2018), and Dadan (2019: ch. 3), it is thus possible to recognize a typology, according to which V2 grammars differ in terms of whether the finite verb targets a low position C_{Fin} (49a) or a high position C_{Force} (49b) within the C-domain.

- (49) a. [FrameP [Spec] [ForceP [Spec] [TopP^{REC} [Spec] [FocP [Spec] [FinP V_{Fin} [TP V_{Finr} ...]]]]]
 b. [FrameP [Spec] [ForceP [Spec] V_{Fin} [TopP [FocP [FinP [TP V_{Finr} ...]]]]]

The representations in (49) make some precise and testable predictions about the types of structures that can be generated in both types of V2 grammars. In a V-in-Fin V2 grammar (cf. 49a) the verb sits in the lowest head position C_{Fin} from which it can be preceded, potentially

simultaneously, by a focus (SpecFocP), one or more topics (SpecTopP^{REC}), and a frame element (SpecFrameP).²¹ In a V-in-Force grammar (cf. 49b), by contrast, the possibilities are greatly reduced since the verb targets the higher C_{Force} head, from where it can only be preceded by a fronted focalized or topicalized constituent in SpecForce (> V2) and additionally by a frame element in SpecFrameP (> V3). Grammars of the former type have been argued to characterize, among others, medieval southern Italo-Romance varieties (cf. Ledgeway 2007b; 2008) as well as late Latin (Ledgeway 2017b), where we witness V-to-Fin raising with optional fronting of one or more constituents to the left periphery (> SpecFocP, SpecTopP^{REC}, SpecFrameP). Consequently, it is correctly predicted that in these V2 varieties, alongside frequent linear V2 orders, V1 and V3* orders are not only possible but also not insignificant. On the other hand, V2 grammars of the latter V-in-Force type have been argued to characterize medieval Gallo-Romance (cf. Wolfe 2018; Ledgeway 2021b), where V-to-Force movement is accompanied by obligatory fronting to SpecForceP, be that of a focus or a topic. This correctly predicts a much stricter surface V2 linearization where V1 and V4 orders are extremely rare, if not entirely absent/impossible, and where V3 orders are attested, but are qualitatively constrained since the only position above SpecForceP is SpecFrameP such that V3 orders necessarily instantiate Frame+Topic/Focus sequences.

1.2.4.2 Pro-drop Parameter Revisited

Returning to our observation regarding the distribution of null subjects in medieval Romance (cf. examples 26–27 in Section 1.2.3.1), this same generalized raising of the finite verb to C(omp) under V2 also provides us with an elegant and highly natural explanation for the asymmetric distribution of null subjects in early Romance. In particular, when the verb raises to the vacant C position, null subjects are freely licensed (50a), whereas in subordinate clauses, where the finite verb is forced to remain *in situ* within the sentential core, pronominal subjects must be phonologically expressed (50b), although not interpreted as emphatic or contrastively focused.

- (50) a. Et lors vint Ø en sa chambre (OFr.)
and then come.PST.PFV.3SG Ø in her room
'And then she came into her room'
- b. Por ce qu' ele_i estoit tostans en doute qu' ele_i ne perdist ce
for this that she be.PST.IPFV.3SG always in doubt that she NEG= lose.PST.SBJV.3SG this
qu' ele_i trop amoit (OFr.)
that she too love.PST.IPFV.3SG
'Because she was always in doubt that she would lose what she so dearly loved'

²¹ As documented in the literature (cf. Benincà and Poletto 2004; Ledgeway 2010b), frame elements sit outside the clause proper above ForceP and introduce a dislocated topical constituent, typically represented either by a (hanging) topic or by a scene-setting/circumstantial adverb(ial) that spells out the spatial-temporal coordinates of the utterance (cf. also Section 26.3.3).

This asymmetrical distribution leads us to conclude that null subjects in medieval Romance were not licensed exclusively, if at all, by rich verb inflexion for person and number, but, by a property which the finite verb uniquely acquires by raising to the vacant C position, presumably the locus of finiteness in medieval Romance (cf. Ledgeway 2012a: §3.4.1). Now we have just seen in Section 1.2.4.1 how medieval Romance V2 can be classified according to a V-in-Fin and V-in-Force typology, the latter type characterizing medieval Gallo-Romance varieties where the root-embedded asymmetry in the distribution of overt and null pronominal subjects is strongest, if not systematic. Superficially, this asymmetrical distribution might lead us to hypothesize that the functional head responsible for the licensing of pro-drop in medieval Gallo-Romance is C_{Force} , since null subjects occur in root V2 contexts where the finite verb targets C_{Force} , but not in embedded contexts where the verb only raises to Infl and its associated subject can only be realized by an overt pronominal. Less frequently, however, we also find in medieval Gallo-Romance examples of embedded V2 (cf. Salvi 2004: ch. 1; Benincà 2006: 24), in which the finite verb exceptionally raises to C_{Fin} with concomitant fronting of one or more constituents to the left periphery, since the C_{Force} position is already lexicalized by the subordinating complementizer (51a). Such cases can be unambiguously identified in that they deviate from the otherwise predominant embedded SVO pattern (viz. V-in-Infl; cf. 48b), displaying an immediately preverbal constituent other than the subject (cf. Salvesen and Walkden 2017), and the subject, if pronominal, can be null (51b) as in canonical root V2 contexts.

- (51) a. ... [C_{Force} *que* [$C_{\text{Top/FocP}}$ *Topic/FOCUS* [C_{FinP} *V* [IP *pro* \forall]]]
- b. Tantost com ele sot [...] *que* ·VII· NÉS i avoit Ø
 soon as she know.PST.PFV.3SG that 7 ships there= have.PST.IPFV.3SG Ø
 charges (OFr., *Histoire ancienne* 598.8)
 load.PTCP
 ‘As soon as she found out [...] that he had 7 ships fully laden there’

Significantly, this observation that null subjects are also possible in cases of embedded V2 involving V-in-Fin forces us to identify the licensing of pro-drop with C_{Fin} , through which the finite verb passes *en route* to C_{Force} in root V2 contexts, since if the relevant head were C_{Force} , then this would incorrectly predict the ungrammaticality of null subjects in embedded V2. This is a natural conclusion given the traditional strong association of Fin(iteness) with inflexional reflexes of number and person which presumably license and spell out the relevant phi-feature specification of the null subject.

Finally, it is interesting to note how this archaic pattern in the asymmetrical distribution of null subjects has been exceptionally retained in modern Corsican (Marchetti 1974: 25, 51, 85, 94, 119). In particular, Corsican behaves as a canonical Romance null subject variety in root clauses (52a–b), but in embedded contexts requires referential (52a–c) and non-referential (52d) pronominal subjects to be overtly realized, albeit in

reduced clitic form (cf. tonic/clitic *io/e*, *o* (1SG), *tù/tu* (2SG), *ellu/ella* / *(e)llu/(e)lla* (3M/FSG), *noi/no* (1PL), *voi/vo* (2PL), *elli/elle* / *(e)lli/(e)lle* (3M/FPL)).

- (52) a. Ø avemu da fà e divuzioni à tutti i santi chè no
(we) have.PRS.IND.1PL from do.INF the devotions to all the saints that we=
truvemu. (Crs.)
find.PRS.IND.1PL
'We'll pray to all the saints we meet.'
- b. Ø e cumprereu à u scagnu quand'è vo falate. (Crs.)
(you) them= buy.FUT.2PL a the office when you= descend.PRS.IND.2PL
'You buy them at the office when you go down.'
- c. Eo ogni volta ch' o tornu in Cervioni, è ch' o u sentu
I each time that I=return.PRS.IND.1SG in Cervione and that I= it= hear.PRS.IND.1SG
u nostru chjoccu mi mette sottusopra. (Crs.)
the our chiming me= put.PRS.IND.3SG under.over
'Each time I go back to Cervione and hear our chiming bell I get overcome.'
- d. Ch' ellu un ci sia troppu rumore! (Crs.)
that it= not there= be.PRS.SBJV.3SG too.much noise
'Make sure that there isn't too much noise!'

This distribution of null and overt pronouns is however only superficially similar to medieval Romance, in that modern Corsican is not a V2 variety with the finite verb occurring in the sentential core in all cases, irrespective of the realization of the pronominal subject. However, when the finite verb does raise to the C-domain, as happens in partial interrogatives (53), realization of the clitic subject pronoun is blocked in a similar fashion to what was seen for medieval Romance (cf. 50a).

- (53) [_{CP} QUANDU falate [_{IP} Ø falate quandu?]] (Crs.)
when descend.PRS.IND.2PL Ø
'When are you coming down?'

In conclusion, these modern Corsican data highlight the force of reanalysis in syntactic change, whereby the original asymmetrical distribution of null and overt pronominal subjects, the surface output of an underlying V2 grammar, has been subject to reanalysis and realigned with a non-V2 grammar in which their distribution is no longer licensed by specific functional heads targeted by the finite verb (viz. $C_{Fin} \Rightarrow$ null/overt vs $Infl \Rightarrow$ overt), but, presumably, by a [\pm root] distinction borne by the finite verb raised to a position within the I-domain. This example shows, in turn, how the cluster of properties associated with a given parameter, V2 of the V-in-Force type in this particular case, are not intrinsically bound together but, rather, represent separate micro-parameters which can also be licensed individually under other scenarios. For example, formal registers of modern French and many dialects of north(east-ern) Italy continue to license raising of the finite verb to the C-domain in specific marked non-veridical contexts (cf. Ledgeway 2015a; Section 21.3.2), but have lost focus-fronting to the left periphery (54a). This contrasts with most

Although a very short text, the *Placito capuano* is not without its problems, raising a number of qualitative and interpretive issues for the philologist (cf. Section 3.2.3). In particular, given the extremely brief and formulaic nature of the text, it is legitimate to question what is the value, if any, of the linguistic evidence that such a short piece can genuinely offer the historian of the language. Indeed, this is a problem which arises with many of the earliest attestations of the Romance vernacular including, for example, the earliest Romance text, the *Strasbourg Oaths* of 842, another short sworn oath produced in an early French dialect of disputed origin.²³

Another unresolved issue thrown up by the *Placito capuano* concerns the correct reconstruction of the pragmatico-semantic interpretation of the fronted constituents stacked up at the beginning of the embedded clause (namely, *kelle terre, per kelle fini . . . , trenta anni*), our reading of which is without doubt greatly hindered by the limited nature of our textual evidence. Again this is a frequent problem faced by philologists reading early texts, which in many cases only offer a rather brief glimpse of the language, especially when they only exist in fragmentary form, and whose linguistic physiognomy is often deliberately limited by the specific style and register of the text type.

A final observation concerns the appearance of the complementizer *ko* ‘that’ in (55). Contrary to expectations in light of the discussion of dual complementizer systems above (cf. Section 1.2.3.3.1), the epistemic main clause predicate *sao* ‘I know’ selects for an indicative clausal complement headed, not by the expected indicative/realis complementizer *ca*, but by what appears to be the subjunctive/irrealis complementizer *ko* (< QUOD), a variant of *che* (< QUID). How are we then to interpret the appearance of *ko* in this instance? Is it an example of a scribal error, or should it be taken at face value? As we shall see below, a consideration of this question in light of our preceding discussion of the fine structure of the left periphery provides an illuminating solution.

Putting together the results of the discussions of the previous sections regarding the fine structure of the left periphery of the clause (Section 1.2.3.3) and the philological evidence of the *Placito capuano*, it becomes clear that a knowledge of the relevant linguistic and philological facts can profitably complement one another. We begin by observing how, despite only providing a glimpse of the early vernacular, the *Placito capuano* is of immense interest both to the historian of the language and the historical linguist since, although a very short text, it is nonetheless astonishingly rich in linguistic evidence. More specifically, it is quite remarkable that such a short text, and the first one from the Italian Peninsula no less,

²³ See Tabachovitz (1932), Ewert (1935), Lüdtkke (1963; 1966), Castellani (1969; 1978), López (1994), Ayres-Bennett (1996: 16–30), Balibar (1997: 61–68), Beck (2014).

should provide such extensive early evidence of the fine structure of the left periphery and, above all, incontrovertible proof for the existence of the two left-peripheral fields postulated in Section 1.2.3.3, which, in turn, hold the key to a proper pragmatico-semantic interpretation of the fronted constituents stacked up at the beginning of the embedded clause. In particular, the rigid ordering of the Topic + Focus fields postulated in Section 1.2.3.3, together with the observations regarding the potential recursion of topics but not foci, which is limited to a single occurrence, allows us to infer that the left periphery of the embedded clause hosts two thematicizations within the Topic field, namely, *kelle terre* ‘those lands’ (picked up by the resumptive clitic pronoun *le* ‘them’) and *per kelle fini que ki contene* ‘with those borders which are contained here’, and a contrastive focus *trenta anni* ‘(for) thirty years’ within the Focus field, as represented in (56).

- (56) Sao [ForceP ko [TopP kelle terre, per kelle fini que ki contene,
 know.PRS.IND.1SG that those lands for those confines that here contain.PRS.IND.3SG
 [FocP TRENTA ANNI [IP le possette parte sancti Benedicti.]]]
 thirty years them= possess.PST.PFV.3SG party saint.OBL Benedict.OBL
 ‘I know that, those lands, within those borders which are contained here [in the document/map
 before me], have belonged for thirty years to the part [= monastery] of St. Benedict [of
 Montecassino].’

Turning now to the unexpected use of the complementizer *ko* (< QUOD) rather than *ca*, this too finds a principled explanation in terms of the structural organization of the complement clause. As demonstrated in Ledgeway (2004; 2005), in the early dialects of southern Italy, including those of Campania, the distribution of the two complementizers *co/che* (< QUOD/QUID) and *ca* (< QU(I)A) is not quite as neat as the traditional descriptions reviewed in Section 1.2.3.3.1 would lead us to expect. Whereas all types of subjunctive/irrealis clause are introduced by *che*, indicative/realis complement clauses are headed either by *ca* or *che*. Simplifying the facts somewhat (for detailed discussion, see Ledgeway 2005: §3), it will suffice to note here that either *ca* (57a) or *che* (57b) are employed when the left periphery of an embedded indicative/realis clause does not contain any topics or foci, whereas *che* alone is found in the presence of fronted topics/foci (57c), as witnessed by the following old Neapolitan examples:

- (57) a. Homero [...] dice a li suoy libri *ca* [IP foro nave
 Homer say.PRS.IND.3SG to the his books that be.PST.PFV.3PL ships
 MCLXXXVI] (ONap.)
 1186
 ‘Homer [...] says in his books that there were 1186 ships’
 b. Purriase ben dicere *che* [IP fo causa multo legiere] (ONap.)
 can.COND.3SG=self well say.INF that be.PST.PFV.3SG cause very light
 ‘it could indeed be said that there was very little reason for it’

- c. Considerava *che* [_{TopP} a quista insula de Colcosa [...], [_{IP} non se nce
 consider.PST.IPFV.3SG that to this island of Colchis NEG self= there=
 poteva gire se non per mare]] (ONap.)
 can.PST.IPFV.3SG go.INF if not by sea
 'He thought that, this island of Colchis [...], was only accessible by sea'

Given these distributional facts, it is possible to argue that old Neapolitan had just one indicative complementizer generated in the lowest complementizer position (C_{Fin}) as *ca*: whenever raised to the higher complementizer position (C_{Force}), as proves obligatory whenever topics or foci are present, it surfaces in the form *che* (namely, ca_{C-Fin} vs $che_{C-Force}$). In short, the overt form assumed by indicative uses of *col/che* is interpreted as nothing more than the surface morphological reflex of raising *ca* from its base position to the higher complementizer position within the left periphery. Indeed, this analysis is directly supported by old Neapolitan recomplementation examples such as (42b) above where, crucially, the higher complementizer invariably surfaces in the morphological form *col/che* but the lower position is always spelt out as *c(h)a*, and never *vice versa*.

In this light, we can now return to the initially mysterious selection of *ko* in the *Placito* in (55). It turns out after all that the use of *ko* here is not a scribal error, but reflects the availability and use of two different complementizer positions determined by the informational structure of the embedded clause (cf. 58a) which, as noted in (56), contains a contrastively focused constituent preceded by two topicalized constituents. It is the activation of these Topic-Focus fields in the embedded left periphery which is directly responsible for the presence of *ko*, inasmuch as the complementizer is forced to surface in the higher complementizer position in the presence of fronted topics or foci where it is spelt out as *ko* (58b). It now comes as no surprise therefore that the complementizer *ca* (or its graphic variants *cha*, *ka*) should not be employed in (55).

- (58) a. ... [_{CP} *ko/che* Topic + Focus *ka/c(h)a* [_{IP} ...]]
 b. Sao [_{CP} ***ko***_i *kelle terre*, *per kelle* ... *TRENTA ANNI* ***ka***_i [_{IP} ...]]

This example also highlights the dangers of dismissing too hastily the linguistic attestations offered to us by even the most meagre of philological evidence. In this particular case, the *Placito capuano* provides an invaluable early example of the *ca* vs *col/che* alternation, which is not attested again in the textual record for the Campania region for at least another 350 years. Thanks however to this early attestation, it is possible to conclude with confidence that the relevant complementizer alternation and associated positions licensed by the absence/presence of fronted topics and foci, otherwise richly attested in texts from the fourteenth century onwards

(Ledgeway 2004; 2005; Vincent 2006), must date back to at least the tenth century. Of course, it might be objected that the linguistic evidence of the texts is so formulaic that its value for the linguist is questionable on a number of counts. Indeed, a comparison of the *Placito capuano* with the other three *Placiti cassinesi* produced three years later in the nearby localities of Sessa Aurunca (59a) and Teano (59b–c) reveals such an extraordinarily high level of structural, discourse, and lexical uniformity across all four texts that it would seem naïve to imagine that all four sworn oaths faithfully reproduce authentic, spontaneous tokens of the spoken vernacular of the time.

- (59) a. Sao cco kelle terre, per kelle fini que tebe mostrai
 know.PRS.IND.1SG that those lands for those confines that you.DAT show.PST.PFV.1SG
 Pergoaldi foro, que ki contene, et trenta anni
 Pergoaldi.OBL be.PST.PFV.3PL that here contain.PRS.IND.3SG and thirty years
 le possette.
 them= possess.PST.PFV.3SG
 'I know that, those lands, within those borders which I have shown you, and which are
 contained here [in the document/map before me], belonged to Pergoaldi [the abbot of the
 monastery of San Salvatore] and that he owned them for thirty years.'
- b. Kella terra, per kelle fini que bobe mostrai, sancte Marie
 that land for those confines that you.DAT show.PST.PFV.1SG saint.OBL Maria.OBL
 è, et trenta anni la possette parte sancte Marie.
 be.PRS.IND.3SG and thirty years it= possess.PST.PFV.3SG part saint.OBL Maria.OBL
 'That land, within those borders which I have shown you [in the document/map before me],
 belong to [the monastery of] Santa Maria [di Cengla], and were possessed by the part
 [= monastery] of Santa Maria [di Cengla] for thirty years.'
- c. Sao cco kelle terre, per kelle fini que tebe mostrai,
 know.PRS.IND.1SG that those lands for those confines that you.OBL show.PST.PFV.1SG
 trenta anni le possette parte sancte Marie.
 thirty years them= possess.PST.PFV.3SG part saint.OBL Maria.OBL
 'I know that those lands, within those borders which I have shown you [in the document/
 map before me], have belonged for thirty years to the part [= monastery] of Santa Maria [of
 Cengla].'

On the contrary, given the nature of the four oaths, which all had the specific purpose of persuading the court that a set of disputed lands had been in the possession of a given monastery for thirty years and consequently a legitimate part of the latter's estate, it is more than likely that the individuals enlisted by the Church to serve as 'independent' witnesses, presumably under the promise of personal financial reward, were given very precise instructions regarding what they were required to swear under oath. Nonetheless, it would be rash to disregard the evidence of these four short, highly formulaic written testimonies on these grounds alone; rather, given the Church's deliberate efforts to place presumably authentic-sounding words and structures of the vernacular in the mouths of their witnesses, it is still possible to see in the language of these four texts a

deliberate hypercharacterization of some of the most salient traits of the spoken language of the time such as the fronting of topicalized and focused constituents, which still constitutes to this day a characteristic feature of the spoken, rather than written, registers of Romance (cf. Duranti and Ochs 1979; Schweickard 1986; Rossi 1999; Cruschina 2012). In short, the *Placiti* presumably reflect competence rather than performance.

To sum up, the discussion in this section has demonstrated that, when theory and philological evidence are considered together, the results of traditional philological and linguistic scholarship can be considerably enhanced (cf. Section 2.7). In particular, current theoretical assumptions about the design and architecture of the left periphery of the clause provide us with some novel and powerful tools to shed light on the interpretation and linguistic choices of one of the earliest Romance texts. At the same time, the *Placiti cassinesi* provide independent and robust evidence for these same syntactic assumptions. In short, and as abundantly demonstrated in this case study, linguistics and philology should complement each other to produce enlightening results, rather than be seen as alternatives to be pursued in isolation from each other.

1.2.4.4 Dual Complementizer Systems

In Section 1.2.3.3.1 it was observed how Romanian and a number of southern Italian dialects employ a dual complementizer system which distinguishes between a realis complementizer derived from *QUIA* ‘because’ (> Ro. *că*, SItR. *ca*) and an irrealis complementizer derived from *QUID/QUOD* ‘that, which’ (> USID *che*, *chi/co*), *QUO(MODO)/(QUO)MODO* ‘how’ (> Sal. *cu* / SCal. *mu*, *ma*, *mi*) or *si* ‘if’ (> Ro. *să*): while the former heads clauses selected by declarative and epistemic predicates typically marked by the indicative (60a), the latter is employed after predicates such as volitionals that characterize the state or events of their complements as unrealized at the time of speaking typically marked by the subjunctive when available (60b).

- (60) a. Su’ ssicuru c’ a ttie nu tte tice none. (Lec.)
 be.PRS.IND.1SG sure that_{REALIS} to you NEG you= say.PRS.IND.3SG no
 ‘I’m sure that he won’t tell you no.’
- b. Spettamu lu miètecu nesciu cu nni fazza la lizzetta. (Lec.)
 wait.PRS.1PL the doctor our that_{IRREALIS} DAT.3= do.PRS.SBJV.3SG the prescription
 ‘We’re waiting for our doctor to give him the prescription.’

Furthermore, it was noted above how both complementizers differ with respect to their relative positions in conjunction with fronted topicalized and focused constituents. For instance, as the preceding Leccese examples demonstrate, while *ca* precedes all such fronted elements (cf. 60a), *cu* must follow (cf. 60b). Facts like these led us to propose in terms of Rizzi’s (1997) split CP perspective that the realis complementizer (e.g., *ca*) lexicalizes the

higher C_{Force} head and the irrealis complementizer (e.g., *cu*) the lower C_{Fin} head, as sketched in (61):

- (61) [_{ForceP} *ca*_{REALIS} [_{TopP} Top [_{FocP} Foc [_{FinP} *cu*_{IRREALIS} [_{IP} . . .]]]]]

A further distinction between the two complementizers which, at first sight, does not seem to immediately follow from (61) concerns their phonological realization (Calabrese 1993; Terzi 1994; 1996; Ledgeway 2013; 2015b). For example, while in Salentino *ca* must always be pronounced (62a), *cu* may optionally remain unpronounced (62b), a case of C-drop.

- (62) a. Te prumettu *ca* tornu. (Sal.)
 you= promise.PRS.1SG that_{REALIS} return.PRS.1SG
 ‘I promise you that I will return.’
 b. Ogghiu (*cu*) llū faci stare cittu. (Sal.)
 want.PRS.1SG that_{IRREALIS} it= make.PRS.2SG stand.INF quiet
 ‘I want you to shut him up.’

Strikingly, even in those Romance varieties which do not display a dual complementizer system it is not infrequent for the finite complementizer *que/che* to remain phonologically unpronounced, on condition that it introduces an irrealis complement (63b), but not, crucially, when it introduces a realis complement (63a).²⁴

- (63) a. Sabem *que* esperes. (Cat.)
 know.PRS.IND.1PL that wait.PRS.IND.2SG
 b. Preguem (*que*) esperis. (Cat.)
 pray.PRS.IND.1PL that wait.PRS.SBJV.2SG
 ‘We know/ask that you wait.’

In view of the superficial similarity in the conditions regulating the distribution of C-drop in varieties with dual complementizer systems and languages such as Catalan with apparently only one complementizer, it is tempting to extend the analysis of dual complementizer systems to languages of the latter type. In particular, we have established on the basis of dual complementizer systems that C-drop is a property uniquely licensed by the lowest C-related position, presumably involving a residual V2 effect (cf. Section 1.2.4.2) with raising of the modally-marked finite verb to C_{Fin} in the absence of the irrealis complementizer (Poletto 2001), but not by the realis complementizer which lexicalizes the highest C-related position C_{Force} and hence unavailable to finite verb-raising under V2. By the same token, in its irrealis uses standard Romance *que/che* must also lexicalize C_{Fin} , thereby explaining its complementary distribution with finite verb-raising under

²⁴ See further the discussion in Section 21.3.2, as well as Ledgeway (2016: 1021f.) and references cited therein.

C-drop, in contrast to its realis uses where it lexicalizes the highest position C_{Force} .

In short, we are led to conclude that the dual complementizer system explicitly attested for such varieties as Romanian and the dialects of southern Italy must also be assumed to hold more generally for Romance where, despite the two complementizers not being lexically distinguished (cf. also Ledgeway 2009b; Ledgeway and Lombardi 2014), the relevant distinction between the two homophonous complementizers *que/che* is marked indirectly by their differential positions within the left periphery and their compatibility with C-drop (64a–c).

(64)	a.	[ForceP	Ø	[Top/FocP	[Fin	<i>que/che</i> _{IRREALIS}	[IP V _{SBJV}]]]]
	b.	[ForceP	Ø	[Top/FocP	[Fin	V _{SBJV}	[IP V _{SBJV}]]]]
	c.	[ForceP	<i>que/che</i> _{REALIS}	[Top/FocP	[Fin	Ø	[IP V _{IND}]]]]

1.2.4.5 Nominal Functional Structure

1.2.4.5.1 Articles

An area of spectacular diachronic and synchronic microvariation in Romance regards the numerous dimensions of variation characterizing the Romance nominal group,²⁵ many of which can be captured in terms of the functional categories and associated functional structure that make up the DP. An obvious place to start is Quintilian's oft-quoted remark 'noster sermo articulos non desiderat' ('our language has no need of articles'), highlighting a salient typological difference between Latin and Greek in nominal functional structure of a parametric nature. Effectively, Quintilian's observation distinguishes between languages that lack articles such as Latin, which fail to grammaticalize definiteness overtly in the syntax through the lexicalization of the D(eterminer) position with a definite article (cf. Bošković 2005a; 2005b; 2008; Ledgeway 2012a: §4.2.2.1), and varieties like Romance, which from around the eighth century (Ledgeway 2012a: 96) grammaticalized the marking of definiteness on D through a weakened form of the Latin distal demonstrative *ILLE* 'that' or, less frequently, the Latin intensifier *IPSE* '-self, same' (> Bal./Costa Brava Cat. *es/sa*, Srd. *su/sa*). In accordance with the cross-linguistic generalization that marking of indefiniteness is dependent on the prior availability of marking for definiteness (Longobardi and Guradiano 2009; Keenan 2011; Longobardi

²⁵ See also Longobardi (1994; 2012), Guardiano and Longobardi (2005; 2017a; 2017b), Longobardi and Guardiano (2009), Longobardi, Guardiano, Silvestri, Boattini, and Ceolin (2013), Ledgeway (2007a; 2015a), Longobardi, Ghirotto, Guardiano, Tassi, Benazzo, Ceolin, and Barbujani (2015), Longobardi, Ceolin, Ecay, Ghirotto, Guardiano, Irimia, Michelioudakis, Radkevic, Luiselli, Pettener, and Barbujani (2016), and Giusti (2016).

2012: 308–15), it is possible to further derive the fact that, despite presenting a definite article, the earliest Romance varieties often still fail to grammaticalize the [\pm count] distinction on D and hence lack an indefinite article in their earliest attestations. Indeed, systematic usage of the indefinite article, which continues a weakened form of the Latin numeral for ‘one’ (< UNUM/-AM), does not become established until around the fourteenth century (Pozas Loyo 2010: ch. 5; Maiden 1995: 121; Ledgeway 2012a: §4.2.1). Before then bare DPs are employed for non-particularized referents (65a), whereas the indefinite article is reserved for particularized new referents (65b), presumably a residue of its numeral origin (Parry and Lombardi 2007: 91f.). In the modern languages, by contrast, indefinite DPs, whether particularized or not, require the article (65c):

- (65) a. **Enfant** nos done qui seit a ton talent (OFr., *Vie de Saint Alexis* 5.5)
child us= give.IMP.2SG who be.PRS.SBJV.3SG to your wish
‘Give us a child of your pleasing’
- b. **Un fi** lor donet (OFr., *Vie de Saint Alexis* 6.3.)
a son to.them= give.PRS.IND.3SG
‘He gave them a son’
- c. **Busco una minyona que em neteja/netegi la casa.** (Cat.)
search.PRS.IND.1SG a maid that me= clean.PRS.IND/SBJV.3SG the house
‘I’m looking for a maid who is cleaning/to clean my house.’

Although in later stages of Romance that grammaticalize both the definite and indefinite articles the definite article displays considerable attenuation of its original deictic force, increasingly coming to mark shared cognition between speaker(s) and addressee(s), it still retained considerable identifying force, witness its exclusion in early texts with unique, abstract, and generic referents (66a; cf. Parry and Lombardi 2007: 83f.; Renzi 2010: 318f., 329f., 332–37), a usage often fossilized in modern proverbs and set expressions (66b–c). In the modern languages, by contrast, shared cognition between speaker(s) and addressee(s) assumes increasing importance, such that the article is now generally required with unique, abstract, and generic referents (66d).

- (66) a. **leichatz estar ypocresie** (OGsc., *Disciplina clericalis* 1v.32–33)
let.IMP.2PL be.INF hypocrisy
‘let hypocrisy be’
- b. **Parar/desparar taula** (Cat.)
lay.INF/clear.INF table
‘To lay/clear the table’
- c. **Noblesse oblige** (Fr.)
nobility oblige.PRS.IND.3SG
- d. **Dreptatea este lumina vieții.** (Ro.)
justice.DEF be.PRS.IND.3SG light.DEF life.GEN.DEF.SG
‘Justice is the light of life.’

This difference between earlier and later stages of Romance can be captured through a microparametric opposition between weak and strong D languages (Guardiano and Longobardi 2005). Languages of the former group include early Romance varieties which do not require overt association in the syntax between N and D, hence kind-reference is not explicitly lexicalized on D, witness the absence of the article in old Galician examples such as *e les quede **libertad** de poder cobrar de lor herederos* (*Historia de la Santa A. M. Iglesia de Santiago de Compostela* 73) ‘and may they retain (the) **freedom** to be able to inherit’. In strong D languages such as modern Romance varieties, by contrast, kind-reference is licensed through explicit association of N and D in the syntax, witness the obligatory use of an expletive article in such modern Galician examples as ***(a) escravitude é todo o contrario da liberdade* ‘(the) slavery is the complete opposite of (the) freedom’. In this respect, many Balearic Catalan varieties and, to a lesser extent Catalan dialects spoken along the Costa Brava, prove particularly revealing in that they show a further dimension of synchronic microvariation on ‘strong’ D which explicitly marks a lexical distinction between the deictic and expletive functions of the article through the opposition between IPSE-derived (67a) and ILLE-derived (67b) articles, respectively (Ledgeway 2012a: 100–03).

- (67) a. Sa mort d’ en Joan (MajCat.)
 the death of the Joan
 ‘Joan’s death’
 b. Pensam en la mort. (MajCat.)
 think.PRS.IND.1PL in the death
 ‘We think about death.’

1.2.4.5.2 Noun Positions

Among the strong Romance D varieties we can further observe a parametric distinction between those that exhibit N(-to-D)-raising and those that do not. Particularly instructive in this respect are adjective-noun orders (for an overview and relevant bibliography, see Ledgeway 2012a: 50–57). Simplifying somewhat, prenominal and postnominal adjective positions typically correlate with the following respective interpretations in the modern standard varieties of Romance:²⁶ (i) inherent/non-inherent (68a); (ii) descriptive/distinguishing (68b); (iii) subjective/objective (68c); and (iv) figurative/literal (68d).

²⁶ See further Arnholdt (1916), Sandfeld and Olsen (1960: 98–114), Alisova (1967), Reiner (1968), Lapesa (1975), Vincent (1986; 2007: 57–61), Bernstein (1993), Giorgi and Longobardi (1991), Bosque (1996), Berruto (1998), Pountain (1998), Demonte (1999), Abeillé and Godard (1999), Radatz (2001), Cinque (2010), Gonzaga (2004), Ledgeway (2007a; 2012a: §3.2.2.1), Andriani (2015; 2018), Giusti (2016: 545–49).

- (68) a. un àrid desert vs una regió àrida (Cat.)
 a arid desert a region arid
 'an arid desert' vs 'an arid region'
- b. unas interesantes películas vs unas películas interesantes (Sp.)
 some interesting films some films interesting
 'some interesting films' vs 'some films which are interesting'
- c. une (splendide) maison (splendide) (Fr.)
 a splendid house splendid
 'a splendid house'
- d. nu vècchià chəmbagnə vs nu chəmbagnə vècchià (Bar.)
 a old companion a companion old
 'an old (= of long standing) friend' vs 'an old (= in years) friend'

In early Romance, by contrast, the distribution of adjectives was quite different, insofar as contrastive readings were not necessarily associated with the postnominal position as in modern Romance, but could equally be licensed in the prenominal position (Ledgeway 2007a; 2009a: 214f; Vincent 2007; Thiella 2008; Giusti 2010: 599–609; Brăescu and Dragomirescu 2014). This is illustrated in the following near-minimal pairs, where the adjective in each case invariably gives rise to a contrastive reading irrespective of its pre- or postnominal position.

- (69) a. pro christian poblo vs lo nom christiien (OFr., *Strasbourg Oaths* / *Sequence of S. Eulalia*)
 for Christian people the name Christian
 'for the Christian people' vs 'the Christian name'
- b. La carnal amor del spiritu et el desseo carnal (OSp., *Libro de actoridades*)
 the carnal love of the spirit and the desire carnal
 'Carnal love of the spirit and carnal desire'
- c. li spangnoli soldati vs le compangnie spangnole (ONap., *Cronaca dei tumulti* 34–35 / 65)
 the Spanish soldiers the companies Spanish
 'the Spanish soldiers' vs 'the Spanish troops'

Strikingly similar is the situation found in modern Wallon (Bernstein 1991; 1993: ch. 4; Bouchard 2002), where all adjectival classes, apart from those expressing nationality (70d), occur in prenominal position (70a–c), perhaps representing a conservative feature, although we cannot exclude the (reinforcing) influence of neighbouring Flemish varieties. Language contact with Croatian presumably also explains the preferred preposing of adjectives in Istro-Romanian (Zegrean 2012: 91–96), as illustrated in (71a–c). Once again, the exception is represented by nationality adjectives, at least the class in *-an*, witness the contrast in (71d).

- (70) a. on neûr tchapê (Wal.)
 'a black hat'
- b. du l' corante êwe (Wal.)
 of the running water
 'some running water'

- c. lès cûts pans (Wal.)
‘the cooked loaves’
- d. le peûpe italyin (Wal.)
the people Italian
‘the Italian people’
- (71) a. (negre) pâre (negre) (IRo.)
black bread black
‘brown bread’
- b. doi (otrovni) šarpel’i (otrovni) (IRo.)
two poisonous snakes poisonous
‘two poisonous snakes’
- c. (uscâte) lemne (uscâte)
dried wood dried
‘dried wood’
- d. ur (taljanksi/**taljan) fečor (taljan/taljanksi) (IRo.)
a Italian boy Italian
‘an Italian boy’

By contrast, in a number of, especially non-standard, Romance varieties including Occitan (72a; cf. Wheeler 1988: 268), Sardinian (72b; cf. Jones 1993: 42), and central-southern Italian dialects (72c; cf. Rohlfs 1969: 330; Ledgeway 2007a; Andriani 2015; 2018), the prenominal adjectival position is extremely restricted and generally replaced by the postnominal position, which is neutral to the contrastive vs non-contrastive distinction.²⁷

- (72) a. lo (**vièlh) pònt vièlh d’ Avinhon (Lgd.)
the old bridge old of Avignon
‘the old Avignon bridge’
- b. na (**piccerella) maruzza piccerella (Nap.)
a small snail small
‘a small snail’
- c. una (**nova) mákkina nova (Srd.)
a new car new
‘a new car’

Exploiting the analysis in Section 1.2.3.2 of variable verb positions around different adverbial classes distributed across the clause, it is possible to make sense of the variation witnessed in (69)–(72) along parallel lines in terms of the varying extent of N(oun)-movement in relation to different adjectival classes (cf. also Section 20.2.2). In particular, different adjectival positions can be reinterpreted as the surface reflex of the head noun

²⁷ Cf. the postnominal position of the adjectives in (69a–c) with the corresponding prenominal position in their French (le *vieux* pont d’Avignon) and Italian (una *piccola* lumaca; una *nuova* macchina) translations.

variously moving across a universally fixed series of at least seven different adjectival classes situated immediately above the NP within the functional structure of the DP (cf. Cinque 1994; 2010; Giusti 2016: 545), as summarized in (73):

- (73) [DP (N) [AP₁ Adj_{Quantity} (N) [AP₂ Adj_{Quality} (N) [AP₃ Adj_{Size} (N) [AP₄ Adj_{Age} (N) [AP₅ Adj_{Form} (N) [AP₆ Adj_{Colour} (N) [AP₇ Adj_{Origin} [NP N]]]]]]]]]]]

Assuming the much-simplified structure in (74a) in which AP₁ and AP₂ can be broadly understood as the functional ‘areas’ in which the various subclasses of non-contrastive and contrastive adjectives, respectively, are generated, we can formally capture in a highly simple manner the differences between the non-standard varieties in (72a–c) on the one hand and standard Romance varieties (68a–d) on the other: in the former the nominal head typically raises to the highest available position above the highest adjectival projection (AP₁), which hosts non-contrastive adjectives from where it precedes both non-contrastive and contrastive adjectives (74b), whereas in the latter the noun only targets the higher adjectival field (viz. AP₁), from where it precedes contrastive adjectives but follows those with a non-contrastive reading (74c):

- (74) a. ... (N) [AP₁ Adj_{Contr.} (N) [AP₂ Adj_{Non-contr.} [NP N]]]
 b. ...lo **pònt** [AP₁ [vièlh] **pònt** [AP₂ [NP **pònt**]]] (Lgd.)
 c. ...le [AP₁ [vieux] **pont** [AP₂ [NP **pont**]]] (Fr.)

By the same line of reasoning, we can explain the frequent prenominal position of contrastive adjectives in early Romance (cf. 69a–c) by assuming that N-raising is only optional in the early varieties (75a–b), since, as noted in Section 1.2.4.5.1, these were originally weak D languages which initially did not require overt association in the syntax between N and D, namely, N-(to-D)-raising, independently yielding the observed adjective-noun order. As observed above, this archaic pattern is still preserved to the present-day in Wallon and Istro-Romanian, where the nominal head only moves at most to the left of the subclass of contrastive adjectives encoding nationality/origin (76a–b). Nonetheless, these are both strong D languages since they require the use of the article with, for example, generic reference (77a–b), highlighting how low N-raising is compatible with both weak (e.g., early Romance) and strong (e.g., Wallon, Istro-Romanian) D grammars, although in the latter case the relevant cluster of properties appears to be the output of adstratal contact with Flemish and Croatian.

- (75) a. li [AP₁ [AP₂ [spangnoli] [NP **soldati**]]] (ONap.)
 the Spanish soldiers
 b. le [AP₁ **compagnie** [AP₂ [spangnole] [NP **compagnie**]]] (ONap.)
 the companies Spanish

- (76) a. lès [AP1 ... [AP2 [cuts] pans... [NP pans]]] (Wal.)
 the cooked loaves
- b. le [AP1 ... [AP2 ... peúpe [AP2-Origin italyin] [NP peúpe]]] (Wal.)
 the people Italian
- (77) a. Les omes sont pus sovint bilingues ki les femreyes. (Wal.)
 the men be.PRS.IND.3PL more often bilingual that the women
 'Men are more often bilingual than women.'
- b. Căprele mârâncu iârba. (IRo)
 goats.DEF.FPL eat.PRS.IND.3PL grass.DEF.FSG
 'Goats eat grass.'

In conclusion, there are truly striking parallels between the nominal and verbal domains and the functional structures associated with these, as revealed by the fixed positions of distinct classes of adjectival and adverbial modifiers, respectively. In particular, we have seen how different diachronic and diatopic varieties of Romance provide clear evidence for some considerable microvariation in terms of the extent of N-movement and its effects on semantico-pragmatic interpretation at the interfaces. Among the various patterns observed, we have also identified a strong diachronic tendency for Romance nouns to climb progressively higher within the available nominal functional structure, the end result of which has given rise in a number of non-standard varieties to a syntactic neutralization of the erstwhile interpretive difference between contrastive and non-contrastive adjectival readings.

1.2.4.5.3 Expletive Articles

Although the preceding discussion forces us to conclude that, with the exception of Wallon and Istro-Romanian, D and the functional field above NP uniformly attracts N in modern Romance to yield the typical noun-adjective_{Contrastive} order, further fine-grained differentiation of this particular microparameter is required to understand the observed split among Romance varieties in relation to the licensing of proper names through the use or otherwise of an expletive article (Longobardi 1994). The relevant difference is to be interpreted in terms of which types of N can be attracted by D in individual Romance varieties. The least marked option is that which characterizes varieties such as standard Spanish, where D indiscriminately attracts all types of N, including proper names which overtly raise to D and therefore prove incompatible with the definite article, e.g., (**el) Juan/(**la) Juana '(the) John/(the) Jane'. The more marked and restrictive option is exemplified by varieties such as European Portuguese where D fails to attract proper names, a small and lexically definable subclass of nominals, which, by virtue of the strong D setting, can only be rescued through merger of an expletive article in D, e.g., o João/a Joana 'the John/the Jane'. Catalan varieties have moved the furthest in this direction (Wheeler, Yates,

and Dols 1999: 67f.), developing a specialized paradigm for proper names based on clitic reflexes of DOMINUS/-A ‘master/mistress’ > *en/na* in Balearic Catalan (e.g., *en Joan/na Joana*) and on a blend of DOMINUS- and ILLE-derived forms in the standard language (e.g., *en Joan/la Joana*).

Once again this is a simplification of the relevant Romance facts, in that there are further microparametric distinctions involved which, for space limitations, we cannot discuss in detail here. For example, while standard Italian appears to pattern with Spanish in all relevant respects with personal proper names, e.g., (***il*) *Gianni*/*(**la)* *Gianna*, the two languages differ with respect to proper names denoting large geographical expanses, e.g., It./Sp. *(**la)/(**la)* *Francia* ‘(the) France’. However, even in Spanish there are certain lexical exceptions where the article proves obligatory, e.g., *(**la)* *India* ‘(the) India’, *(**el)* *Reino Unido* ‘(the) United Kingdom’ (but cf. *(**la)* *Gran Bretaña* ‘(the) Great Britain’), or optional, e.g., *(la)* *Argentina* ‘(the) Argentina’, *(el)* *Canadá* ‘(the) Canada’. A similar distribution, though often with different class membership (cf. Sp./Pt. *(**la)* *España*/*(a)* *Espanha* ‘(the) Spain’, *(el)* *Brasil*/*(o)* *Brasil* ‘(the) Brazil’), also obtains in Portuguese where the article is generally excluded, e.g., *(**a)* *França* ‘(the) France’, *(**o)* *Portugal* ‘(the) Portugal’, but in some cases is obligatory, e.g., *(**a)* *Índia* ‘(the) India’, *(**o)* *Japão* ‘(the) Japan’, and in others is optional, e.g., *(a)* *Uganda* ‘(the) Uganda’, *(o)* *Timor Leste* ‘(the) East Timor’. Such unpredictable lexically-based variation, which is often subject at the same time to considerable diatopic and idiolectal variation, is indicative of nanoparametric variation. In a similar fashion, in Romanian and northern-central regional varieties of Italian, D attracts only masculine proper names, e.g., Ro. *Ion*/*(**ul)* ‘John(=DEF)’, It. *(**il)* *Gianni* ‘(the) John’, but not feminine proper names which must occur with an expletive article, namely, Ro. *Ioana* ‘Jane=DEF’ < *Ioană* + -a ‘Jane + =DEF’, NCI. *la* *Gianna* (cf. Cornilescu and Nicolae 2015; Ledgeway 2017c). To our knowledge, the reverse situation, namely, obligatory N-to-D raising with feminine proper names coupled with the obligatory use of expletive articles in conjunction with masculine proper names, is not found in any Romance variety in line with general assumptions regarding the relative markedness of gender categories (viz. masculine > feminine). In sum, data like these highlight once again the role of functional categories such as those associated with the D-domain in licensing extensive Romance microvariation.

1.3 The View from Romance Palatalization

1.3.1 Sketch of the Two Major Romance ‘Waves’ of Palatalization and Their Consequences

The history of the major Romance palatalizations can be thought-provoking in several quite different respects. Before we see what these are, we need a

preliminary sketch of the palatalizations in question. The phonological process of palatalization has occurred in various ways over the history of the Romance languages. It is a cross-linguistically common phenomenon which has a fairly simple articulatory explanation as an assimilation: see, e.g., Anderson (1973: 106–08); Trask (1996a: 60f.; 1996b: 254f.; 2000: 243). Many manifestations of the phenomenon in Romance are locally restricted and of relatively late date (see, e.g., Loporcaro 2011: 143–50; Repetti 2016), but the two palatalizations discussed in what follows are ‘Romance’ in the sense that their effects are part of the shared inheritance of all Romance languages, and this is because they operated at an early date. The first palatalization (see, e.g., Loporcaro 2011: 144; Repetti 2016: 658–62) is attested (from inscriptional evidence) in the second century, and involves the emergence of a novel series of palatal consonants (Latin had no palatals apart from the glide [j]). These arose by assimilation of certain consonants to an immediately following ‘yod’ (i.e., the glide [j]), often resulting in the ‘absorption’ of the yod into the preceding consonant.²⁸ As well as producing palatalization, [j] often caused affrication of preceding consonants, so that the first palatalization is also a major source of novel affricate consonants. The second palatalization (see, e.g., Loporcaro 2011: 147f.; Repetti 2016: 662–64) appears, on the basis of inscriptional evidence, to have occurred in or by the fifth century, and involves a change such that velar consonants [k] and [g] are palatalized (principally yielding the affricates [tʃ] and [dʒ], but with various subsequent developments according both to geographical area and to phonological environment). As we see later, the effects of this second palatalization are to be observed over a more restricted geographical area than those of the first palatalization.

Both palatalizations can be usefully illustrated from their effects on inflexional morphology. Indeed, these morphological consequences of the palatalizations will crop up repeatedly in what follows. In some Romance varieties (occasionally in Italo-Romance and very systematically in Romanian)²⁹ the second palatalization produced alternation between singular and plural in nouns and adjectives (e.g., SG PORCUS ‘pig’ ~ PL PORCI > It. *por*[k]o ~ *por*[tʃ]i, Ro. *por*[k] ~ *por*[tʃ]i; Ro. SG SICCUS ‘dry’ ~ PL SICCI > Ro. *se*[k] ~ *se*[tʃ]i),³⁰ but the most dramatic effects of both palatalizations, across the Romance languages, are to be seen in the verb. An important point in what follows is that, in many Romance languages, both the first and the second palatalization produced allomorphy in the final consonant of lexical

²⁸ There was a proliferation of new yods in late Latin because the Latin unstressed vowels /e/ and /i/, when immediately followed by a vowel, became [j].

²⁹ For discussions of why the phenomenon is rare in Italo-Romance, and why the (masculine) plural forms originate specifically in Latin nominative (rather than accusative) case-forms, see Maiden (1996; 2000).

³⁰ The morphology of nouns and adjectives was such that there happened to be no potential for the creation of alternations by the first palatalization.

roots of non-first conjugation verbs such that the first person singular present indicative,³¹ and some or all forms of the present subjunctive, come to share a distinctive root allomorph.³² That both types of palatalization create the same paradigmatic *pattern* of alternation is actually coincidental. The phonological environment for palatalization by yod was met just in the first person singular present and in throughout the present subjunctive; in contrast, the phonological environment for palatalization of velars was extensively met throughout the paradigm of the relevant verbs, *except* in first person singular present indicative, and throughout the present subjunctive. The two palatalizations effectively operated on complementary sets of paradigm cells.

A fine-grained account of how the effects of the palatalizations (and affrications) have played out in the historical morphology of the Romance languages is beyond our scope here (see also Section 10.5.1). There are many details and apparent or real exceptions that require special and lengthy explanations: for discussions, see Maiden (2010; 2011b; 2018: 122–48); Maiden et al. (2021: 350–56). Among these complications are the fact that in Romanian the first and second person subjunctive forms have been replaced by indicative forms, and that in Italian the expected alternant is now often absent in some verbs in the first and second persons plural of the present subjunctive. Nonetheless, the broad effects of the palatalizations can be illustrated by the verb forms in Tables 1.4–1.5. Subsequent sound changes in some Romance varieties have meant that the original palatal (or affricate) nature of the alternants may no longer be directly apparent as such, although Italian (Table 1.5) stays fairly close to what are likely to have been the origin products of the palatalizations. Yet the alternations illustrated in Tables 1.4–1.6 are all, in origin, the direct results of palatalization. The choice of ‘old’ Italian and ‘old’ Romanian in these examples is motivated by the fact that by the twentieth-century analogical levelling had rather obscured these original effects.

The most important point to note here is that the Romance languages, generally, have acquired, as a result of the palatalizations, a recurrent pattern of alternation such that forms of the present subjunctive share a distinctive root allomorph with the first person singular present indicative (and in some cases also with the third person plural present indicative). This is an observation to whose significance we will return in Sections 1.3.5 and 1.3.6 where, perhaps surprisingly, we will see that these facts have a

³¹ First conjugation verbs never presented the appropriate environment for creation of alternation caused by yod. These verbs were in principle exposed to the ‘second palatalization’ in their present subjunctive, but what actually happened here is discussed in Section 1.3.2.

³² In Italian and some other central Italo-Romance dialects these effects include the third person plural present indicative. In Romanian the effects of the second, but not the first, palatalization extend slightly further, in that they include the third person plural present indicative.

Table 1.4. (Old) Romanian

(a) Effect of palatalization by yod						
	1SG	2SG	3SG	1PL	2PL	3PL
PRS	<i>vă[ɖz]u</i> 'see'	<i>vedzi</i> ³³	<i>vede</i>	<i>vedem</i>	<i>vedeți</i>	<i>vădu</i>
SBJV	---	---	<i>va[ɖz]ă</i>	---	---	<i>va[ɖz]ă</i>
PRS	<i>au[ɖz]u</i> 'hear'	<i>audzi</i>	<i>aude</i>	<i>audzim</i>	<i>audziți</i>	<i>audu</i>
SBJV	---	---	<i>au[ɖz]ă</i>	---	---	<i>au[ɖz]ă</i>
PRS	<i>sa[j]u</i> 'jump'	<i>sari</i>	<i>sare</i>	<i>sărim</i>	<i>săriți</i>	<i>saru</i>
SBJV	---	---	<i>sa[j]e</i>	---	---	<i>sa[j]e</i>

(b) Effect of palatalization of velars						
	1SG	2SG	3SG	1PL	2PL	3PL
PRS	<i>zicu</i> 'say'	<i>zi[tʃ]i</i>	<i>zi[tʃ]e</i>	<i>zi[tʃ]em</i>	<i>zi[tʃ]eți</i>	<i>zicu</i>
SBJV	---	---	<i>zică</i>	---	---	<i>zică</i>
PRS	<i>împingu</i> 'push'	<i>împin[ɖʒ]i</i>	<i>împin[ɖʒ]e</i>	<i>împin[ɖʒ]em</i>	<i>împin[ɖʒ]eți</i>	<i>împingu</i>
SBJV	---	---	<i>împingă</i>	---	---	<i>împingă</i>

Table 1.5. Old Italian

(a) Effect of palatalization by yod						
	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	<i>va[ʎʎ]o</i> 'am worth'	<i>vali</i>	<i>vale</i>	<i>valemo</i>	<i>valetē</i>	<i>va[ʎʎ]ono</i>
PRS.SBJV	<i>va[ʎʎ]a</i>	<i>va[ʎʎ]i</i>	<i>va[ʎʎ]a</i>	<i>va[ʎʎ]amo</i>	<i>va[ʎʎ]ate</i>	<i>va[ʎʎ]ano</i>
PRS.IND	<i>te[ɲɲ]o</i> 'hold'	<i>tieni</i>	<i>tiene</i>	<i>tenemo</i>	<i>tenete</i>	<i>te[ɲɲ]ono</i>
PRS.SBJV	<i>te[ɲɲ]a</i>	<i>te[ɲɲ]i</i>	<i>te[ɲɲ]a</i>	<i>te[ɲɲ]amo</i>	<i>te[ɲɲ]ate</i>	<i>te[ɲɲ]ano</i>
PRS.IND	<i>ve[ɖɖʒ]o</i> 'see'	<i>vedi</i>	<i>vede</i>	<i>vedemo</i>	<i>vedete</i>	<i>ve[ɖɖʒ]ono</i>
PRS.SBJV	<i>ve[ɖɖʒ]a</i>	<i>ve[ɖɖʒ]i</i>	<i>ve[ɖɖʒ]a</i>	<i>ve[ɖɖʒ]amo</i>	<i>ve[ɖɖʒ]ate</i>	<i>ve[ɖɖʒ]ano</i>

(b) Effect of palatalization of velars						
	1SG	2SG	3SG	1PL	2PL	3PL
PRS	<i>dī[k]o</i> 'say'	<i>dī[tʃ]i</i>	<i>dī[tʃ]e</i>	<i>dī[tʃ]emo</i>	<i>(dite)</i>	<i>dī[k]ono</i>
SBJV	<i>dī[k]a</i>	<i>dī[k]i</i>	<i>dī[k]a</i>	<i>dī[tʃ]amo</i>	<i>dī[tʃ]ate</i>	<i>dī[k]ano</i>
PRS	<i>le[gg]o</i> 'say'	<i>le[ɖɖʒ]i</i>	<i>le[ɖɖʒ]e</i>	<i>le[ɖɖʒ]emo</i>	<i>le[ɖɖʒ]ete</i>	<i>le[gg]ono</i>
SBJV	<i>le[gg]a</i>	<i>le[gg]i</i>	<i>le[gg]a</i>	<i>le[ɖɖʒ]amo</i>	<i>le[ɖɖʒ]ate</i>	<i>le[gg]ano</i>

bearing on such issues as the tendency for linguists inappropriately to take the perspective of standard languages in their analysis of historical

³³ The [ɖʒ] found in the second person singular (*ve[ɖʒ]i*) and also in the first and second persons plural of some verbs (*au[ɖʒ]im*, *au[ɖʒ]iți*) is of different origin from and probably of later date than that found in first person singular and the subjunctive. See Maiden (2011b: 64).

Table 1.6. *Portuguese*

(a) Effect of palatalization by yod						
	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	<i>te[n]o</i> ‘have’	<i>tens</i>	<i>tem</i>	<i>temos</i>	<i>tendes</i>	<i>têm</i>
PRS.SBJV	<i>te[n]a</i>	<i>te[n]as</i>	<i>te[n]a</i>	<i>te[n]amos</i>	<i>te[n]ais</i>	<i>te[n]am</i>
PRS.IND	<i>ve[z]o</i> ‘see’	<i>vês</i>	<i>vê</i>	<i>vemos</i>	<i>vedes</i>	<i>vêem</i>
PRS.SBJV	<i>ve[z]a</i>	<i>ve[z]as</i>	<i>ve[z]a</i>	<i>ve[z]amos</i>	<i>ve[z]ais</i>	<i>ve[z]am</i>
PRS.IND	<i>fa[s]o</i> ‘do’	<i>fazes</i>	<i>faz</i>	<i>fazemos</i>	<i>fazeis</i>	<i>fazem</i>
PRS.SBJV	<i>fa[s]a</i>	<i>fa[s]as</i>	<i>fa[s]a</i>	<i>fa[s]amos</i>	<i>fa[s]ais</i>	<i>fa[s]am</i>
(b) Effect of palatalization of velars						
	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	<i>digo</i> ‘say’	<i>dizes</i>	<i>diz</i>	<i>dizemos</i>	<i>dizeis</i>	<i>dizem</i>
PRS.SBJV	<i>diga</i>	<i>digas</i>	<i>diga</i>	<i>digamos</i>	<i>digais</i>	<i>digam</i>

analyses, and on the fundamental question of how Romance languages are to be defined. We begin, however, by looking at how the palatalizations can make us reflect on the nature of sound change, and the relation between sound change and morphologization.

1.3.2 The Palatalization of the Velars and the Emergence of a Sound Change

The Romance palatalization of velars before front vowels (PVFV), illustrated from alternation in verbs (and some nouns and adjectives) in Section 1.3.1, operates across the board, wherever the relevant phonological environment is met. Some further examples, this time from non-alternating environments, are given in Table 1.7.³⁴

Romance PVFV seems to offer a classic scenario in which sound change occurs regularly and apparently without exception. Where PVFV does not occur, the reason is apparently geographical (the sound change simply failed historically to occur in a particular area, again across the board), or chronological (apparent counterexamples to PVFV simply postdate the historical period at which the sound change operated), or ‘analogical’ (morphological alternations originally triggered by the sound change have subsequently been removed by so-called analogical levelling, often attributed to a preference for a one-to-one correspondence between form and meaning).

³⁴ The Spanish forms given in Table 1.7 are in standard European Spanish. The consonant [θ] (elsewhere [s]) can be shown to go back to an affricated outcome of the velar before a front vowel.

Table 1.7. *Effects of palatalization of velars in some Romance languages*

Latin	Italian	Romanian	Spanish
CIUITATEM 'town'	[tʃ]ittà	[tʃ]etate	[e]iudad
GELU 'frost, ice'	[dʒ]elo	[dʒ]er	[j]elo
GENERUM 'son-in-law'	[dʒ]enero	[dʒ]inere	[j]erno
CENAM 'dinner'	[tʃ]ena	[tʃ]ină	[e]ena
PACEM 'peace'	pa[tʃ]e	pa[tʃ]e	pa[θ]
LEGES 'laws'	le[dʒ]i	le[dʒ]i	le[j]es
PISCEM 'fish'	pe[ʃ]e	pe[tʃ]e	pre[θ]

The best example of the 'geographical' exception comes from Sardinian (or at least central Sardinian dialects), which systematically lacks PVFV (cf. Lausberg 1966: 315f., 322f.). The physical and cultural isolation of Sardinia probably explains why the island was unaffected by PVFV – a phenomenon which is not attested in inscriptions until the fifth century and may not have reached all parts of a fracturing late Roman Empire. That geographically isolated areas within some linguistic domain may fail to participate in innovations that originate outside those areas is unremarkable.³⁵ Apparently all we would need to do, then, for Sardinian is to adjust the statement that the Romance languages underwent palatalization of velars before front vowels so as to exclude at least one, isolated, place, Sardinia. And this is what manuals of Romance linguistics regularly and understandably do. An example of the 'chronological' type of exception is Ro. *gem* 'I moan' [dʒem] < Lat. *GEMO*, Ro. *înțelege* 's/he understands' [înce'leʒe] < Lat. *INTELLEGIT*, with the expected palatalization, but *ghem* [gem] 'ball of wool',³⁶ or *intelighenție* [inteli'gentsije] 'intelligentsia', without it. These words fail to undergo PVFV because the former comes from late Latin *'glemu, and developed as [gem] only in relatively recent linguistic history, while the latter is a twentieth-century loanword from Russian. Both forms simply postdate the period at which PVFV operated. All that is needed here is to state a chronological limit after which the sound change ceased to operate and after which these words must have entered the language or developed velar + palatal sequences. Analogical levelling could be illustrated from Italian (see also Section 10.5.2). This language does have the

³⁵ This does not mean that isolated or peripheral areas are inherently linguistically archaic. There is simply no reason in principle why they should be, and it is perfectly possible for innovations to occur in them which do not occur elsewhere. But the language of isolated areas is liable to be *different* from the cognate languages from which they are separated. Nor should we ever assume that areas that seem 'isolated' today were necessarily so in the past; for a valuable object lesson on this point, see Varvaro (1984).

³⁶ More accurately, [g'em], since velars in modern Romanian automatically display a small degree of phonetic palatalization (but not affrication) in modern Romanian. For more discussion of types of palatalization at work in modern Romanian, see Section 9.4.3.

expected velar ~ palatal alternation between singular and plural in SG *porco* ['porko] 'pig' ~ PL *porci* ['pɔrʃi], or SG *amico* [a'miko] 'friend' ~ PL *amici* [a'miʃi], but not in SG *secco* ['sekko] 'dry' ~ PL *secchi* ['sekki], or SG *fungo* ['funʒo] 'mushroom' ~ PL *funghi* ['funʒi]. The latter non-palatalized cases can surely be assigned to analogical levelling: we may assume that the alternation was originally present (e.g., ['funʒo] ~ PL ['fundʒi]) as a result of regular sound change operating before the front vowel of the plural ending, but that it was then eliminated, quite plausibly on the model of countless other nouns and adjectives whose roots do not alternate for number, a development perhaps additionally motivated by a desire to unify form with meaning.³⁷ So far, so good. Most manuals of Romance historical linguistics indeed create a fairly tidy view in which this sound change occurred across the board, and if there are exceptions they can be neatly assigned to these well understood categories. Most of the time, this procedure is perfectly justified. But if we pay proper attention to *all* the historical-comparative information available to us, the picture becomes much more nuanced, fragmentary, and blurred.

Linguists should always ask themselves how plausible the changes they postulate are, in the light of what we know about the history of the people who speak the relevant languages. For many families of related languages we may have little or no knowledge of the broader historical context in which changes occur, but for Romance linguistics the possibilities are often different and Romance linguists should avoid too abruptly extirpating purely linguistic facts from the wider historical context in which they occurred. This is a point on which the Italian Romanist Alberto Varvaro repeatedly insisted, and it is one that he happens to have illustrated with reference to the Romance palatalization of velars. For example, Varvaro (2004a: 88) is critical of Wartburg (1967) for the 'summary and uninterested way' in which the latter treats some 'cases of regional conservatism in the development of Latin', including palatalization of velars. Varvaro observes that the modern extension of the phenomenon over almost the whole Romance domain (and its presence in all those Romance languages which have attained the status of a national or literary language, apart from some geographically marginal or apparently isolated areas), should not blind us to the fact that this sound change (like any other) was an innovation whose current geographical spread could not have been achieved overnight, and indeed that it was a late innovation which appears to postdate the fall of the western Roman Empire. In fact, failure to participate in PVFV is crucially not limited to Sardinian and this, in Varvaro's view, is significant. Varvaro (2004a: 88–90) observes that the (voiceless) velar is preserved intact in

³⁷ The apparent levelling is the overwhelming majority case in modern Italian. For a more nuanced account of the historical mechanisms involved, and the possible additional role of semantic factors, see Maiden (2000).

palatalizing contexts in loans from Latin into Germanic, Greek, Slavonic, and Berber (and, albeit partially, Maghreb Arabic): e.g., German *Keller* < CELLARIUS ‘cellar’, *Kiste* < CISTA ‘chest’, Berber *akiker* < CICER ‘chickpea’, *iger* < AGER ‘field’ (see, e.g., Tagliavini 1969: 173, 177). Varvaro also notes evidence that in the Romance of the Moselle area PVFV did not take place before the sixth century. All this must mean, Varvaro contends, that the current geographic spread of PVFV can date from no earlier than the end of the western Empire and that the phenomenon spread over almost the entire Latin area when political unity had already disintegrated, at a period of maximum fragmentation when new, smaller, political units were emerging. In short, just because a phenomenon is geographically widespread does not mean that it is ancient nor (we may add) that it emerged abruptly and spread easily and uniformly over the Empire. This perspective should make us wonder whether PVFV was as straightforward a process as simple ‘before and after’ comparisons between Latin and the Romance languages might suggest.

The sense that PVFV first spread in a tentative, non-uniform, way is augmented by a closer look at the comparative details. First, absence of PVFV turns out to involve not only isolated areas of the *Romània* (Sardinian) but also the far less isolated Vegliote Dalmatian – which is in a geographically intermediate position between Italo-Romance and Daco-Romance. There is indeed palatalization of velars before front vowels in Vegliote, but the phenomenon seems recent and in fact independent of the more widespread Romance examples of PVFV. In Vegliote, inherited velar consonants do not show palatalization before inherited mid front vowels (e.g., ‘*kaina* ‘dinner’ < CENAM, ge‘lut ‘cold’ < GELATUM). There is PVFV before high front vowels (Bartoli 1906: §425; Lausberg 1966: 316; Solta 1980: 148f.), but this development appears to be an independent, modern phenomenon, since some of the high front vowels in question have emerged in the recent history of the language (e.g., ʃol ‘arse’ < *kyl < CULUM; ʃil ‘sky’ < *kjelu < C(A)ELUM). In short, there is no good evidence that ‘Romance’ PVFV ever occurred in Vegliote either.

At first glance, PVFV in Romanian looks to be exactly the same historical phenomenon as PVFV in Italian, and many manuals of Romance linguistics (e.g., Lausberg 1966: 316, 323) assume or imply that they are indeed historically the ‘same’ phenomenon.³⁸ Closer scrutiny reveals some thought-provoking differences, however. In other Romance languages, Latin [kwi] or [kwe] become [ki] or [ke] but this change ‘counterfeeds’ PVFV, and is therefore clearly later than it, since the resulting [ki], [ke] do not undergo

³⁸ See also Varvaro’s remarks (Varvaro 2004a: 88f. n. 46) on similar assumptions made by Mihăescu (1983).

Table 1.8. *Differential treatment of reflexes of [kw] in Italian and Romanian*

Latin [kw]	Italian [k]	Romanian [ʃ]	
QUAERERE	[k]iedere	[ʃ]ere	‘ask for’
QUI	[k]i	[ʃ]ine	‘who(?)’
AQUILAM	(a[g]ila)	a[ʃ]eră	‘eagle’

Table 1.9. *Differential palatalizing effect of proto-Romance plural *-e in Italian and Romanian*

*vakke ‘cows’	> It. vakke (<i>vacche</i>)	Ro. *vaʃe > vaʃ (<i>vaci</i>)
*large ‘broad’ _{FPL}	> It. large (<i>larghe</i>)	Ro. *large > lardʒ (<i>largi</i>)

palatalization. In Romanian, in contrast, these same outcomes *do* undergo palatalization (Table 1.8).^{39,40}

Another respect in which Romanian differs from Italo-Romance in the range of ‘input’ to PVFV involves the desinence [-e] of feminine plural nouns and adjectives.⁴¹ In Italian this ending *never* triggers palatalization of preceding velars. Maiden (1996) argues that this is probably because the ending developed via a process [-as] > *[-ai] whose final stage, [-e], emerges only after PVFV has ceased to be productive in Italo-Romance. Yet the same ending, with the same origin, *always* triggers PVFV in Romanian (Table 1.9). Note that in Romanian feminine plural [-e] often subsequently became [-i] (then frequently losing its syllabic value altogether):

Finally, nearly all Romance languages systematically fail to present the expected palatalization of velars in the present subjunctive of first conjugation verbs, historically characterized by the desinence [-e]. For example It. [‘kariki] *carichi* ‘load’_{PRS.SBJV.SG} not **[‘kariʃi] *carici*, Sp. [‘karye] *cargue* not **[‘karʃe] *carge*, the non-occurring forms being the expected output if palatalization had originally happened. This is a fact to whose significance we return later, but we should note that, in contrast, Romanian systematically does present the expected palatalization in the first conjugation present subjunctive (e.g., third person [iŋ‘karʃe] *încarce* not **[iŋ‘karke] *încarche*).

A possible interpretation of the foregoing facts is that PVFV applied in Romanian at a later historical stage than it did in Italo-Romance. This interpretation is supported by a series of observations, made a century ago, which the major manuals of historical Romance linguistics have

³⁹ For the *-ne* at the end of the Romanian form of ‘who’ in Table 1.8, see Maiden et al. (2021: 177).
⁴⁰ The modern standard Italian form *aquila* ‘eagle’ in Table 1.8, retaining the pronunciation [kw], may be a learned borrowing from Latin. The old Tuscan form *aghila*, given here, happens to show voicing of the velar as well as loss of the labial glide.
⁴¹ In Romanian the feminine plural inflexional ending is also the ending of the feminine singular genitive-dative case-form.

frankly overlooked. In the 1920s Petar Skok (1926: 408) argued that a proper understanding of Balkan Romance palatalizations and their relation to PVFV elsewhere in the Romance world could only be gained by taking into account not only the purely linguistic details but also the historical circumstances in which the eastern Romance languages evolved, and the evidence of non-Romance languages with which they were in contact. Skok carefully documents Romance loans in south Slavonic, which took place during the period of Slav settlement along the Adriatic coast, and demonstrates the complete absence of palatalization in those loans. Skok's conclusion (as also suggested by the evidence of Dalmatian, discussed above) is that PVFV was absent on the Romance-speaking Adriatic littoral. After the arrival of the Slavs, according to Skok, the remaining nomadic Romance speakers of the interior were cut off from the urban linguistic models of the coast. Skok reaches the radical conclusion that Romanian palatalization is actually a post-sixth-century innovation *independent* of the PVFV that we see in 'western' Romance languages (Skok 1926: 409).

So was there one general Romance PVFV, or were there at least two independent PVFVs? Our purpose here is not to provide a 'right answer', but merely to show how our understanding of linguistic change (in this case, of a particular phonological change) may become far more nuanced if we carefully scrutinize not only the 'internal' linguistic details but also the 'external' historical facts. We do not have to accept either of the extreme positions (straightforward, monogenetic, almost pan-Romance, palatalization of velars vs fragmented, independent, polygenetic palatalizations occurring separately in the 'west' and the 'east'). One could perfectly well imagine a situation in which PVFV was, so to speak, 'in the air', with possibly more educated, and therefore perhaps linguistically more conservative, urban, speakers who tended not to palatalize velars before front vowels, and less educated and more rustic ones who tended to do so. In fact, a scenario in which palatalization of velars was originally 'latent', with some parts of the population tending to produce it more than others, would perhaps help us better understand how it was ever possible for palatalization to fail to occur in medieval or modern 'western' Romance present subjunctives, mentioned earlier. Some examples are given in Table 1.10.⁴²

In this case in 'western' Romance it is conventionally assumed (e.g., Penny 2002: 177 for Spanish) that regular PVFV at first took place, but that the resultant alternation was then removed from the present subjunctive as an effect of analogical levelling in favour of the velar alternant, the latter being by far predominant in the inflexional paradigm of the relevant verbs. Of course analogical levelling of the effects of palatal-velar alternations

⁴² The only exceptions of which we are aware in any Romance language outside Romanian occur in old French (see Fouché 1967: 202).

Table 1.10. *Failure of expected palatalization of velar consonants in ‘western’ Romance first conjugation present subjunctives vs expected palatalization in Romanian*

1SG	2SG	3SG	1PL	2PL	3PL
Portuguese <i>ro[g]e</i> ‘ask’	<i>ro[g]es</i>	<i>ro[g]e</i>	<i>ro[g]emos</i>	<i>ro[g]eis</i>	<i>ro[g]em</i>
Spanish <i>to[k]e</i> ‘touch’	<i>to[k]es</i>	<i>to[k]e</i>	<i>to[k]emos</i>	<i>to[k]éis</i>	<i>to[k]en</i>
Catalan <i>pa[g]i</i> ‘pay’	<i>pa[g]is</i>	<i>pa[g]i</i>	<i>pa[g]em</i>	<i>pa[g]eu</i>	<i>pa[g]in</i>
Languedocien <i>to[k]e</i> ‘touch’	<i>to[k]es</i>	<i>to[k]e</i>	<i>to[k]em</i>	<i>to[k]etz</i>	<i>to[k]en</i>
Italian <i>to[kk]i</i> ‘touch’	<i>to[kk]i</i>	<i>to[kk]i</i>	<i>to[kk]iamo</i>	<i>to[kk]iate</i>	<i>to[kk]ino</i>
Romanian <i>ro[g]</i> ‘ask’	<i>ro[dʒ]i</i>	<i>roa[dʒ]e</i>	<i>ru[g]ăm</i>	<i>ru[g]ați</i>	<i>roa[dʒ]e</i>
<i>to[k]</i> ‘chop’	<i>to[tʃ]i</i>	<i>toa[tʃ]e</i>	<i>to[k]ăm</i>	<i>to[k]ați</i>	<i>toa[tʃ]e</i>

(and of many other kinds of alternation) are commonplace in the history of the Romance verb, but such levellings simply do not show anything like the sheer, absolute, exceptionlessness we see in the first conjugation. So complete is the lack of evidence for any original palatalization of velars in Romance first conjugation present subjunctives that Maiden (1992: 305 n. 21) actually implies that PVFV simply *never happened* in the first conjugation. It is undoubtedly plausible that analogical pressure played some role: after all, the relevant first conjugation verbs show the velar alternant in all forty-odd cells of their inflexional paradigm with the exception of just the six cells of the present subjunctive. But the utterly exceptionless rejection of the expected palatalization would make much more sense if we assumed that they took place in a linguistic world where palatalization of velars was still latent, a *tendency* in the linguistic system, and where the ‘balance of power’ between phonological pressure for the assimilation to take place and morphological pressure to maintain a one-to-one relationship between lexical meaning and the forms of the lexical root was very far from settled. In a memorable metaphor, Varvaro (2004c: 41) once likened the widespread yet still curiously ‘patchy’ and geographically sporadic emergence of another sound change (so-called *yeísmo*, whereby original [ʎ] > [j], in different varieties of Spanish), to an ‘epidemic’ in which a ‘virus’ is present throughout a community but only develops full-blown

symptoms in some individuals (i.e., here and there, but not everywhere). Perhaps the initial situation with PVFV was the same. Let us put together all the comparative evidence from the history of PVFV: localized exceptions in areas that were by no means obviously isolated, evidence from loans that PVFV had not taken place in the early history of Balkan Romance; curious discrepancies between the input to PVFV in Romanian and other Romance languages; the remarkable wholesale failure of the expected PVFV to occur in one morphologically defined category. In this light, perhaps we should be thinking not of a discrete across-the-board sound change, but of a scenario much more similar to that postulated by Varvaro for Spanish *yeísmo*, such that in the early Romance world the sound change was possible, but nowhere firmly established and where norms lacking the palatalization still existed. The ‘fuller picture’, obtained by considering as many of the available facts as possible, certainly makes any notion of an abrupt and general transition from velars to their palatalized forms look simplistic.

We are hardly being original here in proposing what is, in effect, a ‘variationist’ approach to phonological change (or to language change more generally). The importance of such an approach in Romance linguistics is explored in far more detail in Section 30.2, to which readers are referred. The point here, rather, is that our minds should be open to the possibility that apparently very simple changes may have emerged in a situation of linguistic variation, and that careful scrutiny of the comparative-historical facts, even at a distance of centuries, may point us in just that direction.

1.3.3 When Does Phonological Conditioning of Morphological Alternation ‘Stop’? Comparative Romance Evidence

It is usually easy to demonstrate that some alternation originally triggered by sound change has become ‘morphologized’: for example, once the relevant pattern of alternation is analogically extended outside the original phonological conditioning environment, so that its occurrence cannot be analysed as a phonological effect of that environment, or where there is abundant other evidence that the original triggering sound change is long defunct in the language.⁴³ But even indisputably morphologized alternations can still display a surprising sensitivity to phonological environment. Comparative scrutiny of dialectal variation in Romanian verb morphology brings to light a surprising situation, in which a change which is indisputable evidence for the morphologization of an alternation type reveals a residual sensitivity to phonological environment. In the inflexional paradigm of Romanian verbs, the velar~palatal alternation in second and third conjugation verbs is distributed in such a way that the velar alternant

⁴³ For further discussion and references, see for example Bybee (2001: 55, 96–98).

Table 1.11. *The distribution of velar alternants in the Romanian verb*

infinitive	gerund		present	subjunctive
<i>lin</i> [ɖʒ]e	<i>lin</i> [g]ând	1SG	<i>lin</i> [g]	<i>lin</i> [g]
		2SG	<i>lin</i> [ɖʒ]i	<i>lin</i> [ɖʒ]i
		3SG	<i>lin</i> [ɖʒ]e	<i>lin</i> [g]ă
		1PL	<i>lin</i> [ɖʒ]em	<i>lin</i> [ɖʒ]em
		2PL	<i>lin</i> [ɖʒ]eți	<i>lin</i> [ɖʒ]eți
		3PL	<i>lin</i> [g]	<i>lin</i> [g]ă

Table 1.12. *The Romanian velar~palatal alternation unique to a fugi ‘flee’*

infinitive	gerund		present	subjunctive
<i>fu</i> [ɖʒ]i	<i>fu</i> [ɖʒ]ind	1SG	<i>fu</i> [g]	<i>fu</i> [g]
		2SG	<i>fu</i> [ɖʒ]i	<i>fu</i> [ɖʒ]i
		3SG	<i>fu</i> [ɖʒ]e	<i>fu</i> [g]ă
		1PL	<i>fu</i> [ɖʒ]im	<i>fu</i> [ɖʒ]im
		2PL	<i>fu</i> [ɖʒ]iți	<i>fu</i> [ɖʒ]iți
		3PL	<i>fu</i> [g]	<i>fu</i> [g]ă

occurs just in the first person singular and third person plural present tense, in the third person of the subjunctive, and in the gerund. The distributional pattern of the velar is exactly correlated with historically non-palatalizing environments (i.e., with an original following non-front vowel). We may take the example of the verb *a linge* ‘lick’ (Table 1.11). Note that the vowel spelled *â* in standard Romanian is the high central [ɨ]; crucially, it is not a front vowel.

Dozens of Romanian second and third conjugation verbs follow this pattern. However, there happens to be only one *fourth* conjugation verb,⁴⁴ namely *a fugi* ‘flee, run’, which also displays the velar~palatal alternation. This verb shows exactly the same inflexional distributional pattern as in second and third conjugation verbs, but deviates from it specifically in respect of the gerund. Here, in the standard language, we find not the velar alternant as in second and third conjugation verbs, but instead the palatal alternant. This fact has an entirely straightforward historical phonological explanation: the fourth conjugation gerund ending is *-ind*, and therefore contains the environment for PVFV (Table 1.12).

What is the wider dialectal picture for the same verb? The data offered by the Romanian linguistic atlases offer a kind of ‘chequer board’ image in which one may find in one place the gerund type [fu‘ɖʒind] with the

⁴⁴ That there is only one such fourth conjugation verb in modern Romanian is basically a historical accident.

expected palatal as in standard Romanian,⁴⁵ but in the next place the type [fu'gind] with the velar. That the velar alternant should appear in many dialects also in the gerund of a *fugi* is not surprising, given that, as we have seen, in all other Romanian verbs that have the velar~palatal alternation the velar also occurs in the gerund. But one thing is remarkably conspicuous by its absence: we find [fu'dʒind] and we find [fu'gind] but we absolutely *never* find a perfectly conceivable – indeed expected – third alternative, namely [fu'gind], in which the velar would have been analogically extended on the analogical model of second and third conjugation verbs, but the characteristic gerund ending of the fourth conjugation *-ind* were nonetheless preserved. So the analogical introduction of [g] always entails the parallel appearance of the ending *-ind*. There is absolutely no phonological impediment in Romanian – nor has there been for many centuries – to the existence of a potential *[fu'gind], in which the velar is followed by a front vowel: compare, for example, ModRo. *ghimp* [gimp] 'thorn', *ghete* ['gete] 'boots', [kin] 'torture'. Nor, it may be added, is there any constraint against the reverse situation, where such a palatal is not followed by a front vowel: e.g., *geam* [dʒam] 'window', *magiun* [ma'dʒun] 'plum jam', *ciori* [ʃor] 'crows', *deci* [deʃ] 'so' (the orthographic front vowels following the velars in these examples have a merely diacritic value). Yet, despite the lack of phonological obstacles to the contrary situation, the analogical generalization of [g] is always and unfailingly accompanied by the selection of a non-fourth-conjugation gerund ending containing a non-front vowel, *-ind* *-ând*.

The absolutely crucial point here is that the analogical change in the root allomorph *must* have chronologically preceded (or at least been simultaneous with) the replacement of the inflexional ending *-ind* by *-ând*. One's first assumption may be that first the ending *-ind* was somehow replaced by *-ând* and that only then was the velar alternant selected. This is essentially to say that the selection of the velar alternant is phonologically motivated, and it suggests that the alternation between the palatal affricate and the velar is actually conditioned by the phonological environment. But this analysis cannot be correct, since it would make a *fugi* the only fourth conjugation verb in which such a replacement of the gerund ending has ever taken place; no other fourth conjugation verb has a gerund ending *-ind* (spelled *-ând*), and postulating such a change just to account for *fu[g]ând* would be utterly ad hoc. In contrast, the analogical replacement of [dʒ] by [g] is very clearly motivated morphologically: it is beyond doubt that the gerund type *fu[g]ând* involves the analogical extension of a pattern of alternation whose current distribution is fundamentally *morphological*, not *phonological*, in nature. The extension is motivated by the fact that in all other

⁴⁵ The dialectal data may be directly observed in *ALRII* map 2153; *NALRO* Itenia map 924; *ALRRM* Muntenia Dobrogea map 543.

verbs with the relevant alternation, the velar alternant also occurs in the gerund. The change cannot be phonologically motivated because, *a priori*, fourth conjugation verbs have the ending *-ind*, and *a fugi* was, and firmly remains in all other respects, a fourth conjugation verb. Nonetheless, when the morphological change occurs, it actually brings with it the selection of a gerund ending containing the non-front vowel characteristic of non-fourth conjugation gerunds. Why? These facts suggest a residual sensitivity on the part of speakers to what is, in universal terms, clearly a more ‘natural’ environment for a velar alternant to occur in, given the availability of a choice between the alternants [dʒ] by [g]. The implication is that speakers prefer *fu[g]ând* over a potential and phonologically perfectly possible ***fu[g]ind* simply because the language provides two types of gerund ending and the former is simply the phonologically more natural possibility of the two. If this inference is correct, then an indisputably morphological phenomenon, while not phonologically triggered, has not entirely broken free from phonological conditions on the distribution of the alternation.

In fact, Romanian is not the only Romance variety in which the diachronic behaviour of velar~palatal alternations in the gerund shows sensitivity of a clearly morphologized alternation to phonological environment – this time of a negative kind. The Italian [g]~[dʒ] alternation is clearly no longer triggered by a following front vowel. This can easily be demonstrated by medieval and modern examples such as *lun[g]i ~ lun[g]e* (‘long’ M/F.PL), [g] *iro* ‘dormouse’, [k] *iedere* ‘ask’, *par[k]i* ‘parks’, *pes[k]e* ‘peaches’, and significantly by old Italian present subjunctive forms such as *le[g]gi* ‘read’, *cres[k]i* ‘grow’. The [g]~[dʒ] and [k]~[ʃ] alternations in old (and modern) Italian have a paradigmatic distribution such that the velar alternant occurs in the first person singular and third person plural present indicative and in most forms of the present subjunctive (see also Section 1.3.6). However, this distributional pattern is not limited to velar~palatal alternations, because other sets of alternants, of various origins, show the same alternation pattern. For example, in old Italian, the verbs *leggere* ‘read’, *morire* ‘die’, and *tenere* ‘hold’ (Table 1.13).

A much fuller account of the linguistic facts presented here will be found in Maiden (2018: 84–166). Note that the gerunds for the relevant class of verbs end in *-endo* and therefore contain a front vowel immediately following the root. The point of interest here is that in old Italian the three gerunds *faccendo* ‘doing’, *sappiendo* ‘knowing’, and *vegnendo* ‘coming’⁴⁶ happen to show the same root allomorph as the present subjunctive and (according to verb), the first person singular present indicative and third person plural present indicative: 1SG.PRS.IND *faccio*, 3SG.PRS.SBJV *faccia*, GER *faccendo*; [1SG.PRS.IND *so*], 3SG.PRS.SBJV *sappia*, GER *sappiendo*;

⁴⁶ The modern forms (which already existed as well in old Italian) are *facendo*, *sapendo*, *venendo*.

Table 1.13. *Recurrent alternation pattern in old Italian verbs*

infinitive	gerund		present	subjunctive
<i>le[ddʒ]ere</i>	<i>le[ddʒ]endo</i>	1SG	<i>le[gg]o</i>	<i>le[gg]a</i>
		2SG	<i>le[ddʒ]i</i>	<i>le[gg]i</i>
		3SG	<i>le[ddʒ]e</i>	<i>le[gg]a</i>
		3PL	<i>le[gg]ono</i>	<i>le[gg]ano</i>
infinitive	gerund		present	subjunctive
<i>mo[r]ire</i>	<i>mo[r]endo</i>	1SG	<i>muo[j]o</i>	<i>muo[j]a</i>
		2SG	<i>muo[r]i</i>	<i>muo[j]i</i>
		3SG	<i>muo[r]e</i>	<i>muo[j]a</i>
		3PL	<i>muo[j]ono</i>	<i>muo[j]ano</i>
infinitive	gerund		present	subjunctive
<i>te[n]ere</i>	<i>te[n]endo</i>	1SG	<i>te[ɲ]o</i>	<i>te[ɲ]a</i>
		2SG	<i>tie[n]i</i>	<i>te[ɲ]i</i>
		3SG	<i>tie[n]e</i>	<i>te[ɲ]a</i>
		3PL	<i>te[ɲ]ono</i>	<i>te[ɲ]ano</i>

Table 1.14. *Sporadic analogical extension of L-pattern alternants into the gerund in old Italian*

INF	1SG.PRS.IND	3SG.PRS.SBJV	GERUND	
<i>potere</i> 'be able'	<i>posso</i>	<i>possa</i>	<i>possendo</i>	or <i>potendo</i>
<i>vedere</i> 'see'	<i>veggo</i>	<i>veggia</i>	<i>veggendo</i>	or <i>vedendo</i>
<i>tenere</i> 'hold'	<i>tegno</i>	<i>tegna</i>	<i>tegnendo</i>	or <i>tenendo</i>
<i>piacere</i> 'please'	<i>piaccio</i>	<i>piaccia</i>	<i>piaccendo</i>	or <i>piacendo</i>
<i>avere</i> 'have'	<i>ho</i>	<i>abbia</i>	<i>abbiendo</i>	or <i>avendo</i>
<i>volere</i> 'want'	<i>voglio</i>	<i>voglia</i>	<i>vogliendo</i>	or <i>volendo</i>

1SG.PRS.IND *vegno*, 3SG.PRS.SBJV *vegna*, GER *vegnendo*. This paradigmatic distribution is exceptional from a synchronic morphological point of view, but it is attributable to a regular historical phonological change operating specifically in these verbs. These gerunds then served as an analogical model whereby many other verbs, albeit only occasionally and variably, extended their present subjunctive alternant into the gerund (Table 1.14).

Yet there is one class of verbs that never participates in this analogy,⁴⁷ namely those with velar~palatal alternants (Table 1.15). The lack of velar alternants in the gerund seems impossible to explain in purely morphological terms: other verbs with similarly distributed alternants readily

⁴⁷ We base the claim that such forms *never* occur on the observation that the *Opera del Vocabolario Italiano* (OVI) online database of thirteenth- and fourteenth-century Italo-Romance texts does not yield any such gerunds for verbs with root-final velars, whereas gerunds of the other kinds are amply attested there. See also Vanelli (2010: 1467f.).

Table 1.15. *Absence of extension of L-pattern velar alternants into the old Italian gerund*

INF	1SG.PRS.IND	3SG.PRS.SBJV	GERUND
<i>pian</i> [dʒ]ere ‘weep’	<i>pian</i> [g]o	<i>pian</i> [g]a	<i>pian</i> [dʒ]endo never ** <i>pian</i> [g]endo
<i>dire</i> ‘say’	<i>di</i> [k]o	<i>di</i> [k]a	<i>di</i> [tʃ]endo never ** <i>di</i> [k]endo
<i>cre</i> [ʃ]ere ‘grow’	<i>cre</i> [sk]o	<i>cre</i> [sk]a	<i>cre</i> [ʃ]endo never ** <i>cre</i> [sk]endo

participate in the analogy. Nor can a purely phonological explanation hold water, since, as we have seen, there is no constraint against velars before front vowels in Italian. Rather, we seem to be in the presence of something more nuanced and complex: a tentative, and ultimately ephemeral, morphological innovation seems to have been effectively ‘deterred’ by the fact that the result would have been the replacement of a palatal alternant by a velar alternant in a phonological environment in which the palatal seems more ‘natural’ (as witness the very fact that in Romance, and cross-linguistically, velars tend to palatalize before front vowels). And there are perhaps echoes here of an observation that we made earlier concerning the notable failure of otherwise general PVFV to occur in Romance present subjunctives, except that in that case it is almost as if a tentative incipient sound change was deflected by morphological pressure within the inflexional system of the first conjugation.

There have been periods in the history of linguistic theory when morphology has effectively been evicted from it altogether. A great deal of structure in morphology, and especially the type of paradigmatic alternation discussed above, originates in sound change, and the analysis of such phenomena has often been gravely distorted by what may be termed a ‘phonologizing bias’, the belief that they should be treated as phonologically conditioned even in synchrony, especially where the original conditioning environment survived at least vestigially. Many linguists, in reaction to this tendency,⁴⁸ have devoted much energy to asserting the truly morphological nature of allegedly phonological phenomena, and one example of this might be Maiden (2009b). Maiden there argues against attempts by Burzio (2004) to force a synchronic phonological analysis of the modern Italian facts including velar~palatal alternations, sometimes by what was, as Maiden avers, the illegitimate resurrection of long dead phonological conditioning environments. However, data of the kind cited above from old Italian gerunds suggests that a properly comparative and historical perspective of the facts from the velar~palatal alternations make the case for a more nuanced and less polarized theoretical stance, in which some role can

⁴⁸ For a discussion of many of the issues see, e.g., Bybee (2001).

be accorded to phonological conditioning while also recognizing that alternations may be fundamentally morphologized. For a further review of these issues, see Maiden (2013: 31–38).

1.3.4 When Does the Morphologization of a Sound Change ‘Start’? Comparative Romance Evidence

The ‘overlapping’ nature of the transition from phonological causation to morphologization of alternation is apparent from other comparative Romance data. These suggest that speakers begin to associate the alternations created by sound change with the marking of morphosyntactic distinctions at a surprisingly early stage. Clear illustrations of this claim from the history of the Romance languages are provided by two phenomena which, while not directly a matter of palatalization, are in different ways related to it. We saw in Table 1.4 the typical pattern of allomorphy produced by PVFV, such that a velar alternant survives in the first person singular and third person plural of the present, alternating with a palatalized root-final consonant in the remainder of the paradigm. However, many of the relevant verbs, such as *a linge* ‘lick’, have past participles, supines, preterites, and pluperfects in root-final -s (Table 1.16).

The past participle, supine, preterite, and plurperfect in -s is a characteristic of a great many second and third conjugation verbs, not merely those with velar-palatal alternations. Thus standard Romanian *a tunde* ‘shear’ (Table 1.17).

Table 1.16. Root-final -s in verbs which also have velar~palatal alternations

infinitive		present	subjunctive	
<i>lin</i> [ɖʒ]e	1SG	<i>lin</i> [g]	<i>lin</i> [g]	
	2SG	<i>lin</i> [ɖʒ]i	<i>lin</i> [ɖʒ]i	
	3SG	<i>lin</i> [ɖʒ]e	<i>lin</i> [g]ă	
	3PL	<i>lin</i> [g]	<i>lin</i> [g]ă	
past participle/supine			preterite	pluperfect
<i>lin</i> [s]		3SG	<i>lin</i> [s]e	<i>lin</i> [s]ese

Table 1.17. Root-final -s in verbs which also have roots ending in dentals

infinitive		present	subjunctive	
<i>tunde</i>	1SG	<i>tund</i>	<i>tund</i>	
	2SG	<i>tunzi</i>	<i>tunzi</i>	
	3SG	<i>tunde</i>	<i>tundă</i>	
	3PL	<i>tund</i>	<i>tundă</i>	
past participle/supine			preterite	pluperfect
<i>tun</i> [s]		3SG	<i>tun</i> [s]e	<i>tun</i> [s]ese

Table 1.18. *Introduction of root-final velars into verbs in root-final [d] (here [de] > [je]) in Oltenian*

	present	subjunctive
1SG	tūŋg	tūŋg
2SG	tūnz	tūnz
3SG	tūnje	tūŋgə
3PL	tūŋg	tūŋgə
past participle (etc.)		tūns

In some dialects, however, the verb *tunde* (among others) has been subject to an analogical change such that the root-final [d] has been analogically replaced by [g] in certain parts of the paradigm. The reasons for this kind of replacement are considered in detail in Maiden (2011b) but, very briefly, the fact that both verbs like *a tunde* and those like *a linge* share past participles, preterites, and pluperfects in [s] seems to have facilitated an analogical change such that *a tunde* follows *a linge* in acquiring root-final [g] in the gerund, in the first person singular and third person plural present, and in the third person subjunctive. An example from the Oltenian dialect of Godinești (NALROltenia point 942) is in Table 1.18.

What is of particular theoretical interest here, however, is the fact that the original nasal [n] has been replaced by velar [ŋ] wherever [g] has replaced root-final [d] in *a tunde*. At first glance this looks utterly uninteresting, indeed trivial. It is, after all, a wholly natural phonetic process of assimilation of a nasal to the position of articulation of an immediately following consonant; it is a banal phonetic phenomenon and one repeatedly observed across the world’s languages. Yet a comparative perspective on the data provided by the Oltenian linguistic atlas for this verb holds a major surprise. As described in more detail in Maiden (2009a; 2011b), where the analogical [g] was introduced there has, in many localities, subsequently been a counter-reaction such that the etymologically ‘correct’ dental has been restored and has ousted the analogically introduced [g]. Yet the Oltenian linguistic atlas shows that there are localities in north-eastern Oltenia (e.g., point 935 Dobrița) in which, in the third person subjunctive of *a tunde*, the replacement of [g] by [d] has not entrained the replacement of the allophone [ŋ] by the allophone [n], so that what we find is not expected **tūndə, but actually tūŋdə, with a highly marked sequence of a velar nasal followed by dental consonant. This development is so unlikely, given our normal assumptions about the phonetic behaviour of preconsonantal nasals, that one might suspect a mishearing on the part of the investigator, or even a banal misprint in the linguistic atlas where the data appear. It is important to emphasize, therefore, that this is not observed just once, in one locality: the atlas presents the morphology of this verb for multiple localities in Oltenia, and in no less than *nine separate places*, from different informants, the exact same unexpected result was obtained. The fact that the evidence is repeated

independently gives us confidence in saying that we are in the presence of something that, however unusual, is real.

The explanation of what we observe is in one respect quite obvious: the continued presence of the velar nasal clearly reflects the earlier presence of [g]. The distribution of that velar was clearly morphologized, in that it was associated with a morphosyntactically specified set of paradigm cells. Yet the purely allophonic velar nasal survives in the relevant paradigm cells even after its phonological conditioner (the velar [g]) has been replaced. That this ‘allophone’ [ŋ] survives in the phonologically quite unnatural environment of a following dental is clear proof that it, too, had become ‘morphologized’, and had been associated directly with the feature specification ‘third person subjunctive’. Morphology has, so to speak, been ‘snatched from the jaws of phonology’, and a distribution which seemed exquisitely and naturally dependent on phonological environment turns out to be quite ‘unnatural’ and arbitrary, having become associated with a morphosyntactic context. The relevant dialects contain many other cases where the [n]~[ŋ] alternation remains correlated, in the verb and elsewhere, with the difference between non-velar and velar environments, e.g., Dobrița third person subjunctive [ˈplɨŋgə] ‘weep’ vs present [ˈplɨndʒe] ‘he weeps’. On the evidence of the [ˈtũŋdə] type in the same dialect, we are bound to suspect that the [ŋ] of subjunctive [ˈplɨŋgə] is already a morphologized marker of subjunctive there too. In short, the detailed comparative-historical perspective provided by, for example, a linguistic atlas can offer clear evidence that even ‘low level’, allophonic and apparently purely phonetically-conditioned alternations may become morphologized.

The case of the Oltenian morphologization of nasal assimilation before velars is not the only one in which comparative Romance dialectology has offered insights into the relation between phonological conditioning of alternation and the morphologization of alternants. ‘Metaphony’ is a phenomenon attested in the history of many Romance languages, and extensively attested in Italo-Romance, whereby an unstressed (and usually word-final) high vowel ([i] or [u]) triggers assimilatory raising of a preceding stressed vowel. In Italo-Romance varieties the potential input to metaphony typically includes stressed high mid vowels ([e] and [o]), stressed low mid vowels ([ɛ] and [ɔ]), and stressed [a]. The input to metaphony is regionally variable: whereas in most dialects the high and low mid vowels undergo the assimilation, metaphony of stressed [a] is much rarer. The output of metaphony is also variable. The ‘teleology’ of a phonological assimilation is naturally the minimization of the phonological distance between input and conditioning environment, so there is a sense in which the most ‘natural’ outcome of metaphony would be the raising of *all* stressed vowels to the height position of the conditioning vowels unstressed [i] and [u]. This never happens: stressed [e] and [u] are indeed raised, respectively, to [i] and [u]. But the low mid vowels [ɛ] and [ɔ] show two kinds of metaphonic

outcome, either as the diphthongs [je] and [wo] or as raising to [e] and [o],⁴⁹ while metaphony of [a], when it occurs, usually yields raising (and fronting) to [ɛ] or [e]. In short, metaphony of non-high vowels yields only a *partial* adjustment towards the height position of the conditioning vowels.

The foregoing facts are amply illustrated in manuals of Italo-Romance historical linguistics (e.g., Rohlfs 1966). Yet there is an intriguing type of exception to them found, principally,⁵⁰ in central and southern Italian dialects of Lazio, Marche, Abruzzo, Molise, and Puglia. This is a phenomenon which, until recently, had gone all but unnoticed, and whose significance only becomes apparent through a properly comparative survey of the linguistic atlases and by systematic scrutiny of descriptive studies of individual dialects. These exceptions are constituted by the fact that, sporadically but over a wide range of dialects, *in the verb and only in the verb, the degree of raising produced by metaphonic raising, and the range of input to metaphonic raising, is greater than in any other part of the grammar*. Before we address the structural details, it is worth pointing out how scattered and elusive the relevant data are. So much so that, for example, they escape mention even by Rohlfs in his magisterial surveys of Italo-Romance phonology and morphology (Rohlfs 1966; 1968). The first allusion to them of which we are aware occurs however in Merlo (1909: 77f.), presented again in Merlo (1922: 13, 15, 19, 20), who mentions what he regards as a ‘minor disturbance due to analogy’ in some Lazio dialects such as that of La Cervera where we find that in the verb, and only in the verb, the metaphonic output of low mid vowels is not the otherwise regular [e] and [o] (e.g., SG [ˈpede] ‘foot’ – PL [ˈpedi], SG [ˈente] ‘tooth’ – PL [ˈenti], FSG [ˈbbɔna] ‘good’ – MPL [ˈbboni], FSG [ˈmɔrta] ‘dead’ – MPL [ˈmorti]), but the output expected for higher input vowels (i.e., for the inputs [e] and [o]), namely [i] and [u]). Thus the present indicative forms of some verbs in stressed low mid vowels (Table 1.19).

Merlo (1909: 77f.) believes that such forms have been created on the analogy of verbs that contain high vowels and therefore have regular metaphonic alternants [i] and [u], such as 1SG [ˈmeno] ‘beat’ ~ 2SG [ˈmini], or 1SG [ˈsposo] ‘marry’ ~ 2SG [ˈspusi]. Such an analysis is much more problematic than Merlo appears to think, but to understand why appeal purely to morphological analogy is inadequate, a much broader comparative perspective is instructive. That perspective is difficult to achieve, because there is no dialect in which *hypermetaphony* (the label which Maiden 1991 applied to this phenomenon) applies without exceptions, so

⁴⁹ The historical relation between these two outcomes is controversial (see, e.g., Maiden 2016b), but not immediately relevant here.

⁵⁰ As Maiden (1991: 180) points out, there are similar developments in some northern Italo-Romance dialects, but we will concentrate here on central-southern Italo-Romance varieties. We will focus on the phenomenon as it is manifest in lexical roots of verbs but, as Maiden (1991) shows, the thematic vowel [a] can also be affected and, where it is affected, this is often in the absence of any other metaphony of [a].

Table 1.19. *Exceptional effects of metaphony on mid vowels in the present tense of the verb in La Cervara*

1SG	'lɛo 'lift'	'pɛrdo 'lose'	'sento 'feel'	'troo 'find'	'koʃo 'cook'	'moro 'die'
2SG	'lɪi not **lei	'pirdi not **perdi	'sinti not **senti	'trui not **troi	'kuʃi not **kotʃi	'muri not **mori
3SG	'lɛa	'pɛrde	'sente	'troa	'koʃe	'more
3PL	'leanu	'pirdu not **perdu	'sintu not **sentu	'troanu	'kuʃu not **kotʃu	'muru not **muru

the evidence is mainly scattered and fragmentary. This means that careful sifting through the available sources is needed in order to obtain historical insights. For an overview of the relevant descriptive sources (including the two major linguistic atlases, the *AIS* and the *ALI*, and some 20 independent descriptions of different dialects of central Italy), and exemplification of how they can be assembled in order to gain a deeper historical insight into the emergence of hypermetaphony, see for example Maiden (1991: 181–84). The linguistic atlases are the obvious first port of call in trying to establish the extent and nature of hypermetaphony, but even then the evidence is hard to spot. For example, and limiting ourselves here (for reasons of space) just to cases of metaphony of the low mid back vowel [ɔ], we may compare examples of regular metaphony of this vowel in non-verb forms (e.g., *AIS* maps 76 MPL *morti* 'dead'; 184 FSG *grossa*, MPL *grossi* 'large'; 710 FSG *buona*, MPL *buoni* 'good'),⁵¹ with verb forms that have the potential to display hypermetaphony: thus, *AIS* map 649/51 2SG.IMP vs 2SG.PRS.IND of *dormire* 'sleep'), 1683 (3SG.PRS.IND vs 2SG.PRS.IND of *trovare* 'find'), 1694 (3SG.PRS.IND vs 2SG.PRS.IND of *potere* 'be able'), 1696 (3SG.PRS.IND vs 2SG.PRS.IND of *morire* 'die'). The masculine plural forms in Table 1.20(a),⁵² and the second person singular forms in Table 1.20(b), ended historically -i, and their vowels were therefore subject to metaphony.

The most important thing that the various dialectological sources, *taken together*, bring to light is that in the verb we can have not only a different kind of metaphonic output from that otherwise expected, but even the presence of metaphony where none would be expected at all. At issue here is principally metaphony of stressed [a]. In Agnone (Ziccardi 1910) metaphony of [a] is historically triggered only by reflexes of historically underlying unstressed [i], never by historically underlying unstressed [u]. Moreover, the

⁵¹ In many Romance varieties, such as Italian, the adjective meaning 'dead' is also the past participle of the verb 'die'. Hypermetaphony is never observed, however, in past participles, whose morphological behaviour is generally like that of adjectives.

⁵² The same pattern of alternation illustrated in Table 1.20(a) for adjectives is found in nouns.

Table 1.20. *Ordinary metaphony of low mid vowels vs hypermetaphony: examples from the AIS*

(a) Regular metaphony ([ɔ] > [o]) in adjectives				
	FSG	MPL	FSG	MPL
Capestrano	'bona	'bɔni	'grossa	gross
Trasacco			'rossə	'rossə
Sonnino			'rossa	rosso
S.Giovanni Rotondo	'bonə	'bɔnə	'grossə	'grossə
(b) Hypermetaphony ([ɔ] > [u]) in some verbs				
	3SG	2SG	3SG	2SG
Capestrano			po	pu
Trasacco				
Sonnino	'trova	'trove	po	po
S.Giovanni Rotondo	'trova	'truve	po	pu
			'more	'mure
			'more	'mure
			'dormə	'durme
			'dormə	'durme
			'dormə	'durme
			'dormə	'durme

Table 1.21. *Metaphony of [a] in Agnone*

SG	PL
'a.sənə (< *'a.se.nu) 'donkey'	'ɛ.sənə (< *'a.se.ni)
'kwan.də (< *'kwan.tu) 'how much'	'bɛf.fə (< *'baf.fi) 'moustache'
ju.'kɛa.tə (< *'jo.'ka.tu) 'played'	sul.'dʒe.tə (< *'sol.'dat.i) 'soldiers'

regular output of metaphony of [a] is differentiated according to phonological environment: in closed syllables and in proparoxytones it yields [ɛ] while in open syllables it yields the diphthong [je] (Table 1.21).

However, in the verb, and only in the verb, we find that the output of metaphony of [a] in closed syllables is systematically that expected in open, rather than closed, syllables, namely [je]. It should be added that [je] is also the regular output of metaphony of the low mid vowel [ɛ] (e.g., 2SG.PRS 'sjendə < *'senti 'feel; hear'). The fact that in this dialect metaphony of [a] is never triggered by unstressed [u] would also lead us to expect that metaphony of [a] would never occur in verbs whose third person plural present ended historically in *-u(no); but in Agnone metaphony does occur in just such verb forms (Table 1.22).⁵³

The situation in the dialect of Arpino (Lazio; Parodi 1892: 300) is even more dramatic. There is no metaphony of [a] at all here, *except in the verb*: e.g., 1SG.PRS.IND 'mannə 'send' (< *'mando) ~ 2SG.PRS.IND 'mjennə

⁵³ Given the broadly expository nature of this presentation, we cannot here enter into finer details of analysis. Suffice it to say that there are grounds for scepticism about Ziccardi's suggestion (1910: 431) that such forms in the third person plural are purely analogical (and therefore independent of phonological conditioning). Rather, there is comparative evidence to suggest that third person plural metaphony of [a] occurs only where there was an original final unstressed [u].

Table 1.22. *Metaphony of [a] in the verb in Agnone*

1SG	'partə 'leave'	< *parto	'kandə 'sing'	< *kanto
2SG	'pjertə (not **pɛrtə)	< *parti	'kjendə	< *kanti
3SG	'partə	< *parte	'kandə	< *kanta
3PL	'pjertənə (not **partənə)	< *partuno	'kandənə	< *kantano

(< *'mandi) vs SG 'annə 'year' (< *'annu) ~ PL 'annə (< *'anni).⁵⁴ The same situation, with metaphony of stressed [a] found only in the verb, is repeated in the dialect of Martinsicuro (Mastrangelo Latini 1976) on the Abruzzo/Marche border. In some dialects (e.g., of the Gargano peninsula, or Veroli in Lazio) a similar situation is found for stressed [ɔ]: this vowel is either not subject to metaphony at all, or not subject to metaphony by *u in closed syllables, but in the verb, and only there, this vowel undergoes hypermetaphony in metaphonizing environments (see Maiden 1991: 185).

The implication of these facts is remarkable. What would we say about hypermetaphony if, instead of being rigidly confined to the verb, it occurred more or less sporadically across the lexicon, without morphologically defined restrictions? Surely we would say that it is an unsurprising and wholly *natural* phonetic phenomenon?⁵⁵ And that is exactly what it is. It is natural that an assimilatory process should *assimilate*, that it should yield outputs maximally close, phonologically, to the conditioning environment. It is equally natural that the assimilation should apply to the maximum possible range of phonological inputs, that it should bring all more open vowels closer to the assimilating vowels. Hypermetaphony is a paradox: a phonetically natural phenomenon somehow confined to a purely morphological environment.

Some linguists, for example already Merlo (1909: 77f.), might react to hypermetaphony by trying to argue that it is, in fact, essentially a matter of morphological analogy. Merlo thinks that the hypermetaphonic type 1SG ['perdo] 'lose' ~ 2SG ['pirdi], 1SG ['mɔro] 'die' ~ 2SG ['muri], is analogically modelled on regular metaphonic alternation in verbs with high mid vowels, such as 1SG ['meno] 'beat' ~ 2SG ['mini], 1SG ['sposo] 'marry' ~ 2SG ['spusi]. But such an analysis is problematic because ['meno] ~ ['mini], ['sposo] ~ ['spusi] involve a *different alternation* from [ɛ] ~ [i], [ɔ] ~ [u], and there is simply no precedent in the language for the novel alternation [ɛ] ~ [i], [ɔ] ~ [u]. One might respond that there is a precedent insofar as [i] or [u] alternate in the inflexional paradigm with other stressed vowels. But in this case one would predict that either of these two metaphonic vowels could come to alternate

⁵⁴ Parodi (1892) and Mastrangelo Latini (1976) view this situation as a *remnant* of once more widespread metaphony of [a]. But there is no good reason to take this view: see Maiden (1991: 184) and, more generally, Maiden (1987).

⁵⁵ Perhaps one that was diffusing gradually through the lexicon.

with any other vowel: why not ['parto] 'leave' ~ **['purti] or ['dormo] 'sleep' ~ **['dirmi]? And why should only the metaphonic alternants [i] and [u], and not also [e] and [o], alternate in this way? Why not ['parto] 'leave' ~ **['porti] or ['dormo] 'sleep' ~ **['dermi]? If Merlo's hypothesis does not seem at first glance implausible, it is because of an assumption that he does not explicitly state, namely that the supposed analogy applies to the two input vowels, [e] and [o], which are those *phonetically most similar* to [e] and [o] and for which [i] and [u] would actually be quite 'natural' phonological outputs of metaphony.⁵⁶ Merlo's analysis tacitly acknowledges the 'phoneticity' of the phenomenon. If we then take all the comparative evidence for hypermetaphony into account, and particularly the propensity of stressed [a] to undergo metaphony in the verb but not outside the verb, appeals to morphological analogy seem wholly inadequate. Rather, the maximal comparative-historical perspective suggests, quite simply, that the phonetic process of metaphony has a particular propensity to occur within the inflexional paradigm of the verb. Hypermetaphony is simply a manifestation of the phonetic process of metaphony which is also exquisitely sensitive to a morphologically specified context, that of the verb. How this is to be explained remains moot: the point is that the facts taken together show us that something needs explaining. For example, Maiden (1991: 232–45) surmises that the inflexional paradigm of finite verb forms is generally 'vulnerable' to incipient sound change in the Romance languages,⁵⁷ precisely because the inflexional paradigms of Romance verbs comprise multiple word forms which may be differentiated from each other by patterns of sometimes extreme allomorphy which may involve considerable phonological distance between alternants, whereas nouns and adjectives characteristically show rather little root allomorphy (usually only for number and/or gender). Maiden speculates that this characteristic 'looseness' of the verb in some way licensed speakers to 'unleash' latent phonological tendencies that were liable to produce more extreme forms of allomorphy.⁵⁸ Maiden's

⁵⁶ An alternative analysis of the data presented in Table 1.20 might say that hypermetaphony is purely phonological and that the local output of metaphony of low mid vowels [e] and [o], namely [e] and [o], simply 'feeds' metaphony of high mid vowels, in that it undergoes further raising to [i] and [u]. Such an analysis could not escape that hard morphological fact that hypermetaphony is restricted to the verb and therefore irreducibly morphological. Nor would it be able to explain those cases where vowels that do not otherwise undergo metaphony actually do so just in the verb, or the existence of hypermetaphony in dialects where the output of metaphony of low mid vowels is not [e] and [o] but rather a diphthong (on metaphonic diphthongs see Maiden 2016b).

⁵⁷ Past participles, although part of the inflexional paradigm of the Romance verb, have inflexional properties of adjectives and, like other adjectives, never show hypermetaphony. Other non-finite forms of the verb (gerund, infinitive) do not show generally the conditioning environment for metaphony.

⁵⁸ Contrast, for example, modern Italian, where the inflexional paradigms of nouns and adjectives contain at most four different word forms and where allomorphy in the root is rare (e.g., MSG *alto* 'high', FSG *alta*, MPL *alti*, FPL *alte*), and verbs—which may display around 40 different word-forms, sometimes with idiosyncratic kinds of root-alternation (for example, 3SG.PRS.IND *vale* 'is worth', 3SG.PRS.SBJV *valga*, 3SG.FUT *varrà*, 3SG.PST.PFV *valse* or 3SG.PRS.IND *ha* 'has', 3SG.PRS.SBJV *abbia*, 3SG.FUT *avrà*, 3SG.PST.PFV *ebbe*). It is also noticeable that suppletion (see Chapter 12, this

idea is a surmise whose plausibility could probably only be tested by typological comparison: the question one would need to ask is whether it is the case cross-linguistically, in language whose word-classes show very different degrees of morphological variation, whether the more morphologically varied classes also show systematic susceptibility to phonological innovation.⁵⁹ The crucial point here is that careful scrutiny of geolinguistic variation in the Italo-Romance data suggest that morphologization and the processes of phonological change are not mutually exclusive but may overlap and interact from the very inception of a sound change.⁶⁰ And this in turn should lead us to ask why.

1.3.5 ‘Standard Language Bias’ in Historical Linguistic Analysis

Another kind of deviation from the historically predicted morphological effects of palatalization occurs in old Romanian. In some respects it may seem quite minor, indeed it is rarely commented on at all, but a recent attempt to account for it happens to exemplify a distorting perspective on the analysis of language change which, in fact, can be observed in a number of other respects in Romance linguists (and no doubt beyond). There are times when Romance linguistics can provide models of what *not* to do in historical linguistics!

Latin verbs of the fourth conjugation (with a few other verbs of the third conjugation) had third person plural present indicative forms that ended in *-IUNT*. The *-i-* of such forms would have become [j] in early Romance and this [j] should, in turn, automatically have triggered palatalization of a preceding consonant. A similar development would have been expected to occur in the first person singular present indicative and throughout the present subjunctive (the relevant changes are discussed and illustrated in Section 1.3.1). Yet comparison of Romanian with Italian yields a surprise: both languages historically duly display the expected effects of palatalization in the first person singular present indicative and in the present subjunctive, but Romanian systematically deviates from Italian in

volume) in Romance languages is found almost exclusively in verbs, and scarcely ever in nouns and adjectives.

See also Maiden (2018: 296–300) for the radically different morphological nature of Romance nouns and adjectives as opposed to verbs.

⁵⁹ Readers may sense an apparent contradiction with another claim by Maiden, mentioned in Section 1.3.2, that the verb may have been *resistant* to an incipient sound change, namely the palatalization of velars. In fact, the two claims are mutually reinforcing, although we do not have space here to go into details (for which see Maiden 2018: 277–83). Briefly, the verb forms at issue are first conjugation present subjunctives, and Maiden argues that first conjugation verbs are distinguished from other conjugation classes precisely by their relative lack of root allomorphy (and especially consonantal allomorphy), rather as nouns and adjectives are distinguished overall from verbs by lack of root allomorphy. So the fundamental claim is that sound changes may be sensitive to contrasting degrees of characteristic allomorphy.

⁶⁰ The issue continues to preoccupy synchronic linguists working from an experimental perspective, as well. See, for example, Strycharczuk and Scobbie (2016).

Table 1.23. *Presence in old Italian vs absence in old Romanian of expected palatalization in the third person plural present indicative*

	Latin	old Italian	palatalized?	old Romanian	palatalized?
1SG.PRS.IND	UENIO	<i>vegno</i> ‘come’	yes	<i>viu</i> ‘come’	yes
3SG.PRS.IND	UENIT	<i>viene</i>	no	<i>vine</i>	no
3PL.PRS.IND	UENIUNT	<i>vegnono</i>	yes	<i>vin</i>	no
3SG.PRS.SBJV	UENIAT	<i>vegna</i>	yes	<i>vie</i>	yes
	Latin	old Italian	palatalized?	old Romanian	palatalized?
1SG.PRS.IND	SALIO	<i>saglio</i> ‘go up’	yes	<i>saiu</i> ‘jump’	yes
3SG.PRS.IND	SALIT	<i>sale</i>	no	<i>sare</i>	no
3PL.PRS.IND	SALIUNT	<i>sagliono</i>	yes	<i>sar</i>	no
3SG.PRS.SBJV	SALIAT	<i>saglia</i>	yes	<i>saie</i>	yes

the third person plural present indicative by not showing the expected effect of palatalization. As Table 1.23 shows, unlike old Italian,⁶¹ old Romanian third person plural present indicatives pattern with (for example) the unpalatalized third person singular:

Our concern here is not to identify the particular reason for the unexpected behaviour observed in Romanian,⁶² but to underscore the perils for reliable linguistic reconstruction of a perspective which ignores the full range of comparative-historical data. So long as we limit ourselves to comparing these two standard Romance languages (Italian and Romanian), Italian seems ‘well-behaved’, taking into account etymology and the expected regularity of sound change, while Romanian appears to be a deviant requiring an explanation specific to it. It seems evident, then, that this explanation should be sought in some circumstance peculiar to Romanian, and an obvious candidate is language contact. Indeed, in the Slavonic variety with which Romanian is known to have been in contact during the Middle Ages (‘middle Bulgarian’) a similar pattern of alternation can be shown to have existed,⁶³ with an alternant (caused by a historically underlying yod) present in the first person singular present but absent for regular historical phonological reasons in the third person plural of the present tense. Details of the relevant middle Bulgarian patterns and their history will be found in Elson (2017: 879). Since Italian, in contrast, has never had any significant contact with Slavonic, the apparently deviant behaviour of Romanian seems plausibly explicable by a particular historical fact, contact with Slavonic, which is absent from the history of Italian and distinctive of Romanian alone.

⁶¹ The relevant effects are best illustrated from older forms of these standard languages. As mentioned in Section 1.3.1, they have been complicated or effaced by analogical changes in the more recent history.
⁶² For an attempt to do this, at least in part, see Maiden (2020).
⁶³ But not, overall, an identical one given that, for example, Slavonic has no present subjunctive.

Such an explanation has indeed been proposed,⁶⁴ and at first sight it may appear an attractively elegant one. There is, after all, abundant evidence of Slavonic influence on Romanian morphology – although, it must be said, in the realm of derivational, not inflexional, morphology (see Maiden 2021). But it is an explanation which begins to deflate just as soon as one begins to consider the facts from an adequately comparative perspective. The first question any historical linguist should ask of any proposed explanation for a particular language, is ‘How do the same historically underlying structures play out in all the other, cognate, varieties to which we have access?’ The plausibility of any proposed historical explanation stands or falls on the comparative evidence. In the case at hand a wealth of relevant dialectal data from non-standard varieties is available, and we are bound to ask what they tell us. The information regarding the fate of the expected palatalization in third person plurals from the dialects of Romania and from Romanian’s sister Daco-Romance varieties does indeed support (or at least does not contradict) what we see in (old) Romanian. The problem is comparison of Romanian, or more broadly Daco-Romance, with just ‘Italian’. ‘Italian’ may today be the national standard language of Italy, but it is essentially the continuant of the medieval dialect of Florence – it is merely one of thousands of different ‘Italo-Romance’ dialects, and we cannot legitimately take it as a proxy for the vastly varied Italo-Romance domain. We need to ask how the other Italo-Romance dialects behave with respect to the third person plural morphology at issue, if we are to make any valid comparison with Romanian.

Even a glance at the Italian linguistic atlases begins to make us see both ‘Italian’ and Romanian in a different light. The *Sprach- und Sachatlas Italiens und der Südschweiz* (AIS) has several maps or plates providing information on the relevant third person plural forms (e.g., AIS 1661, 1691, 1695, 1696, 1699), and they contain a surprise. In most parts of Italy for which an answer is discernible,⁶⁵ notably the centre and south, the relevant third person plural present indicatives turn out to behave *exactly as they do in Romanian*: there is no palatalization, even though the effects of the expected palatalization (or allomorphs reflecting the presence of original palatal allomorphs, see, e.g., Maiden 2011a: 236–40) are firmly present in the first person singular present indicative. Table 1.24 gives the example of the verb COME from localities in central and southern Italy, and compares it with Romanian.

This picture, showing (original) palatalization in the first person singular but not in the third person plural, is faithfully confirmed by other linguistic

⁶⁴ See Elson (2017: 889f.).

⁶⁵ In some areas, especially the north, phonological or morphological changes have simply obscured the historically underlying developments.

Table 1.24. *Non-palatalized reflexes of Latin 3PL UENIUNT ‘come’ vs reflexes of (original) palatalization in 1SG UENIO ‘come’ in some central and southern Italian dialects (AIS)*

	1SG.PRS	3PL.PRS
Scanno, Abruzzo	ˈvjɛŋə	ˈvjɛ:nəɐ̯
Trevico, Campania	vɛŋk	ˈvjennə
Vernole, Puglia	ˈɛŋu	ˈɛ:nunɐ
San Chirico Raparo, Basilicata	ˈvɛŋgu	ˈvjɛ:ninu
Mistretta, Sicily	viˈɛŋu	viˈɛ:nu
old Romanian	viu	vinu

atlases (e.g., *ALI* map 1695) and by an abundance of monographic and journal studies of individual Italo-Romance dialects (see Maiden 2020; 2021 for more details). In fact, with the exception of Italian and some other dialects of central Italy, in the overwhelming majority not just of Italo-Romance, but of Romance languages for which it is possible to have an answer, the position turns out to be that found in Romanian: the expected palatalization does not occur in the third person plural present indicative. This means that while the lack of palatalization in the third person plural is indeed anomalous from a historical phonological point of view, from a comparative and morphological point of view it turns out to be Italian that is ‘anomalous’ while Romanian is simply a ‘normal Romance language’. This fact is invisible just so long as one focuses too closely on the standard language, or treats it as a proxy for a dialect area of which it is a mere member, but once it becomes visible it simply nullifies appeal to Slavonic influence as the historical explanation of the Romanian facts.

What matters for our present purposes is not the specific explanation of the facts, but the need to utilize to the maximum the perspective of dialectal variation when setting about any kind of historical reconstruction. The case of the Romanian third person plural is not isolated. There is a sense, indeed, in which ‘familiarity’ (the fact of being characteristic of some well-known standard language) sometimes seem to breed, if not exactly ‘contempt’, at least a misleading disregard for historical reality in Romance linguistics. As another example, in which, this time, it is a rather too narrow focus on standard Romanian which is to blame, we may look at the suppletive morphology of the Romanian verb *BE*. Most parts of the paradigm of this verb continue forms of Latin *ESSE* ‘be’, but the subjunctive (together with the infinitive and gerund) is derived from a different Latin verb *FERI* ‘happen, become, be made’ (Table 1.25).

Analysis of the historical morphology of this verb has been distorted by viewing it through the ‘lens’ of modern standard Romanian. Because in the modern language the reflexes of *FERI* are (among the finite verb forms) uniquely associated with the subjunctive, many historical linguists

Table 1.25. Reflexes of Latin *FERI* 'happen, become, be made' continued in the Romanian verb *BE*

	PRS	SBJV
1SG	<i>sunt</i>	<i>fiu</i>
2SG	<i>ești</i>	<i>fii</i>
3SG	<i>este</i>	<i>fie</i>
1PL	<i>suntem</i>	<i>fim</i>
2PL	<i>sunteți</i>	<i>fiți</i>
3PL	<i>sunt</i>	<i>fie</i>

Table 1.26. Reflexes of Latin *FERI* (> *hi-*) in the Aromanian verb *BE*

	PRS.IND	SBJV
1SG	'esku/h'iu	h'iu
2SG	'eŋt' / h'ii	h'ii
3SG	'easte	'hie/h'ibə
1PL	h'im	h'im
2PL	h'its'	h'its'
3PL	sən/sun	'hie/h'ibə

(Philippide 2011: 479; Capidan 1925: 174; Bourciez 1956: 541; Rothe 1957: 114; Rosetti 1986: 148) have simply assumed that these forms are derived directly from the Latin present subjunctive of *FERI*, namely 1SG *FIAM*, 2SG *FIAS*, 3SG *FIAT*, 1PL *FIAMUS*, 2PL *FIATIS*, 3PL *FIANT*. While it is very likely that the third person singular and plural form *fie* derives from subjunctive *FIAT* and *FIANT*, derivation of the other forms from the subjunctive of *FERI* is surprisingly insouciant of historical phonology: the present subjunctives *FIAM*, *FIAMUS*, *FIATIS*, at least,⁶⁶ simply *cannot* yield *fiu*, *fim*, *fiți*; the only phonologically possible outcomes would be **fie*, **fiem* or **fiam*, **fiați*. Rather, *fiu*, *fii*, *fim*, *fiți* are, instead, all perfectly regular outcomes of the Latin present *indicatives* *FIO*, *FIS*, *FIMUS*, *FITIS*, a set of forms which happen not to be continued as present indicatives in Romanian. The moment the wider dialectal picture is explored, it becomes even more surprising that the Romanian forms should ever be analysed as original subjunctives. For in at least two (Aromanian and Megleno-Romanian) of the other three major branches of Daco-Romance,⁶⁷ reflexes of indicative *FIMUS*, *FITIS* provide the sole first and second person present indicative forms of the verb *BE*, while in some varieties of these two branches *FIO*, *FIS* provide the first and second person singular forms as well (Table 1.26).

⁶⁶ The status of 2SG *fii* is more ambiguous, but it is not a direct phonological reflex of *FIAS*.

⁶⁷ There is also some, more problematic, evidence for the presence of reflexes of *FERI* in the present indicative of the verb *BE* in the fourth branch, Istro-Romanian (see Maiden and Uță Bărbulescu in progress).

Now the usual reaction of linguists to these facts has arguably been ill-judged: for example, Streller (1904: 5, 11), Capidan (1925: 173; 1932: 488), and Rosetti (1986: 148) have all asserted that these present indicative forms in Aromanian or Megleno-Romanian are analogically imported *from the subjunctive*, despite the very obvious fact that not only formally, but also functionally, they correspond perfectly to the Latin present indicatives. The only conceivable motivation (as overtly acknowledged by Streller 1904: 11) for their making this assumption is the fact that such forms are only found in the subjunctive but not in the indicative *in standard Romanian*. We have here what might be termed ‘standard language bias’. Rather than allowing the comparative dialectal facts to guide the analysis of the standard language itself, linguists have allowed the modern standard language to guide the analysis of the dialects and thereby the entire morphological history of the verb *BE*, even in the standard language. The approach taken even ignores the ‘internal’ comparative evidence of each of Aromanian, Megleno-Romanian, and Daco-Romanian, for in each the first and second person subjunctive forms of *all* verbs are demonstrably derived, historically, from the present indicative. Neglect of comparative data has led to a serious methodological failure, by obfuscating the obvious implication that proto-Daco-Romance must have had reflexes of *FIERI* in the present indicative. This conclusion is, by the way, again supported by comparative Romance data: medieval northern Italian dialects (Rohlf 1968: 272; 1969: 129f.; Michaelis 1998; Cennamo 2003) shows present-tense reflexes of *FIERI*, where this verb was used both as the passive auxiliary *BE*, and also survived in its original meaning of ‘become, happen’. For example, from the early fourteenth century Milanese *Volgarizzamento in antico milanese dell’Elucidario* (78):

- (78) Doncha nuy **fimo** crucificadi con Criste al mondo [...] per la vivanda del So corpo nuy **fimo** una medexima cossa con Criste

‘So we **are** crucified with Christ in the world [...] by the food of His body we **become** one with Christ.’

Sardinian also preserves reflexes of what were originally present indicative forms of *FIERI* (see Wolf 2014).

This is by no means the only example in Romance linguistics in which the historical analysis of an aspect of a standard language has been compromised by too narrow a focus with the standard language itself, lacking a more balanced comparative perspective. Loporcaro (1997a) points out the distorting effects of taking analyses devised to explain the structure of *modern Italian* and projecting them backwards into diachrony to explain phonological history. At issue is Italian *raddoppiamento fonosintattico* (see Sections 1.2.2.4 and 6.4.3.2), that phenomenon whereby the initial consonant of a word is lengthened principally if it is immediately preceded by a word-final stressed vowel. For modern standard Italian, the phenomenon lends itself

to explanation as an effect of prosodic constraints on syllable structure. However, Loporcaro (1997a: 40) writes ‘any hypothesis on the origin of R[addoppiamento] F[onosintattico] in Italian, formulated on the basis of internal data, must on the one hand be evaluated by sifting through the available direct documentation, following forward through time the diachronic path that led from Latin to Tuscan-based standard Italian, in its written attestations. And, on the other hand, that hypothesis will need to be confirmed by the reconstructive arguments which can be developed on the strength of evidence provided by other Romance varieties’. This is exactly what Loporcaro does, and the path of evolution that he reconstructs via careful comparison with ‘other Romance varieties’ (especially other Italo-Romance varieties) offers a different and far more plausible account of the development of the phenomenon than could be achieved by projection backwards in time from modern Italian, showing on comparative grounds that *raddoppiamento* originates in Italian, as elsewhere in Italo-Romance, as an effect of consonant assimilation at word boundaries, which later undergoes locally differentiated types of extension and reanalysis over time.

A further possible example of the benefits of a comparative as opposed to a ‘standard-language-oriented’ perspective involves the controversial question of the origins of diphthongization of proto-Romance low mid vowels in open syllables in Italian,⁶⁸ e.g., *‘tɛ.ne > *tiene* ‘s/he holds’, *‘kɔ.re > *cuore* ‘heart’.⁶⁹ Maiden (2016b: 32) detects a historic tendency, particularly among Italian scholars, and perhaps most of all in the work of Arrigo Castellani,⁷⁰ to treat Italian (and the Tuscan dialect from which it emerged) as a ‘special case’ which has somehow developed in an independent way from other (Italo-)Romance varieties.⁷¹ But is Italian so special? Taking a comparative-historical perspective, however, Maiden emphasizes how the diphthongization found in Tuscan resembles diphthongization in other (particularly central and southern) Italian dialects where it is the product of the assimilatory process of metaphony (see Section 1.3.4). Maiden (2016b: 19–23), pointing out that an alleged significant difference between metaphonic diphthongization of low mid vowels and Tuscan diphthongization of low mid vowels, namely that in Tuscan the phenomenon is limited to open syllables, is actually compatible with what we also know about medieval forms of metaphony in central Italy, where there is evidence that metaphony, too, was restricted to open syllables. He also shows that the exceptions to diphthongization in Italian and modern Tuscan dialects nearly always involve words that would not historically have met the phonological environment for metaphony, or would not have alternated

⁶⁸ See also Section 7.3.2.2.

⁶⁹ For a description of the history of the issues, see Sánchez Miret (1998).

⁷⁰ E.g., Castellani (1980a–1980e).

⁷¹ See also Section 5.5.1.

morphologically with forms historically in inflexional [-u] or [-i] subject to metaphony, such as the invariant *b[ɛ]ne* ‘well’, *n[ɔ]ve* ‘nine’, or also, for example, dialectal *m[ɛ]le* ‘honey’, effectively a *singulare tantum*, lacking a plural (and therefore not alternating with a form in [-i]). Those interested may judge for themselves which account they find more convincing, but there is no doubt that the issues cannot be adequately assessed without viewing Italian in a properly comparative perspective, and this is valid for the Romance languages more generally. Standard Romance languages are obviously ‘special’ in that, for different reasons, they have risen to prominence as prestigious varieties. But from a strictly linguistic perspective, it can be a serious mistake to consider them as any more ‘special’ than their sister varieties.

1.3.6 What Is a Romance Language? Could There Be an Answer in Morphology?

Perhaps surprisingly, it will be suggested here that the morphological repercussions of Romance palatalization might even provide one kind of answer to the fundamental question ‘what is a Romance language?’. This is a question to which there is, perhaps equally surprisingly, no obviously ‘right’ answer. The standard response, as given for example by the *Encyclopaedia Britannica*,⁷² is that Romance languages are those historically ‘descended’ or ‘derived’ from Latin. This is a purely historical definition, which carries the implication of continuous intergenerational transmission. It does not explain where the *difference* with Latin lies, although the Romance languages are conventionally labelled in a way that distinguishes them from their ancestor.⁷³ There is a lively controversy about the historical and cultural circumstances that might have induced Romance speakers to perceive themselves as speaking something other than Latin, which we will not address here (see, e.g., Wright 1982; 2016; Varvaro 2013). The question we want to explore, rather, is whether there is any possible, purely linguistic, definition of what a Romance language is. What, if anything, do all Romance languages have linguistically in common that would distinguish them from other languages, including Latin?

One might seek some unique, and shared, trait in phonology or in morphosyntax, but it is not clear that any characteristic exists that strictly

⁷² See www.britannica.com/EBchecked/topic/508379/Romance-languages (consulted 10 March 2021). The *Encyclopaedia* actually says that the Romance languages are ‘derived’ from ‘Vulgar Latin’, defined as a ‘spoken form of non-Classical Latin’.

⁷³ However, see the term ‘neo-Latin’ in the title of Tagliavini (1969), *Le lingue neolatine*. Note also that some Romance varieties, *Ladino* (a name for Judaeo-Spanish), or *Ladin* (the Romance variety spoken in the region of the Dolomites) preserve popular forms of the word ‘Latin’ (see Müller 1963). Terms such as *rumantsch/romansh* or the Romanian term *român* ‘Romanian’ also show continuity with Latin *ROMANICUS* or *ROMANUS* ‘Roman’. For more discussion of the naming of Romance languages see, e.g., Wright (2013; 2016).

meets these criteria. A great deal of what the Romance languages have in common in those domains recurs repeatedly across the world's languages. For example, probably all Romance languages have the five 'cardinal' vowels [i, e, a, o, u], or distinguish singular from plural, or have at least two grammatical genders, or have a basic subject-verb-object word order,⁷⁴ but these are properties observable across numerous languages of the world. Chapter 9 of this volume explores a number of typologically unusual features in Romance phonology, for example, but none of these seems to meet the criterion of being unique to, yet shared across, the Romance languages.

Clearly a uniquely defining linguistic characteristic of Romance languages will need to be maximally *idiosyncratic*, and thereby maximally unlikely to be shared by other languages. Could we perhaps find it in the *lexicon*, where the principle (cf. Saussure 1968: 100f.) of arbitrariness in the relation between form and meaning is at its most prominent? There are several reasons why this, too, would be unsatisfactory. It is certainly true that there is a common core lexicon inherited by all Romance languages (see for example Dworkin 2016: 580), comprising semantically basic items such as DIE, BE BORN, SKY, TOOTH, HAND, FINGER, SON, DAUGHTER, as well as a range of function words such as personal pronouns or prepositions, but this etymologically shared lexicon still fails to unite the Romance languages. This is because of the disruptive effects of sound change: all Romance languages, for example, retain reflexes of Lat. DIGITUS for 'finger', but Sp. *dedo*, Cat. *dit*, Fr. *doigt* ([dwa]), Srs. *det*, Ro. *deget* have little more than the initial voiced dental in common, and even that is not true for all Romance varieties. There are southern Italo-Romance varieties where, by regular sound change, this word begins with some other consonant or with none: e.g., Nap. ['ritə], Isc. [lit], Sora (province of Frosinone) ['itə] (see Rohlfs 1966: 204f.).⁷⁵

Another problem is that lexical criteria could lead us to count as 'Romance' some languages which are generally regarded as deserving a different classification. Thus, some 'creole' languages have predominantly Romance lexicons, but that does not automatically make them Romance languages. The 'Romance' creoles (see Bollée and Maurer 2016: 447–67) are widely considered not to be 'Romance languages' precisely because, although they have a largely Romance-derived lexicon, they have not inherited the distinctive phonological, morphological, or syntactic systems

⁷⁴ The first two criteria do not in fact distinguish Romance from Latin, either. For further discussion of word order see, e.g., Salvi (2016).

⁷⁵ It needs to be said, also, that lexical items can be borrowed from one language into another, and far more easily than grammatical or phonological structures. For example, on one estimate (Finkstaedt and Wolff 1973), Latin- or Romance-based vocabulary constitutes nearly 60 per cent of the lexicon of modern English, but that does not make English a 60 per cent Romance language.

of the relevant Romance languages.⁷⁶ The final difficulty is that probably all of the ‘shared’ vocabulary is shared because it is inherited from Latin, so that the lexical criterion would not differentiate Romance from Latin.

‘Arbitrariness’ is of course also encountered in morphology. Could morphology be the domain in which linguistic ‘Romanceness’ could be located? The answer is probably yes, but in a very specific respect. Derivational affixes or inflexional desinences often present similar arbitrariness in the relationship between form and meaning as we find in the lexicon, but we run into exactly the same problems as we have seen for the lexicon. What is common to all Romance languages in their derivational or inflexional inventory is likely also to be shared with Latin yet to have become phonologically differentiated beyond recognition across Romance. For example, virtually all Romance languages share a second person plural marker derived from Latin *-TIS*,⁷⁷ but *-TIS* is not uniquely Romance precisely because it is shared with Latin.⁷⁸ It cannot constitute a modern pan-Romance second person plural desinence because of the multiplicity of its modern phonological reflexes: e.g., *CANTATIS* ‘you sing’ > Pt. *cantais*, Cat. *canteu*, Fr. *chantez* [ʃɑ̃te], Log. [kanˈtaːðes], Srs. *canteis*, It. *cantate*, Ro. *cântați* [kinˈtatɕ]. However, it is suggested in what follows that it is in another aspect of inflexional morphology that we might encounter phenomena that meet our definitional criteria. They are ‘arbitrary’ to a very high degree yet, if the following analysis is correct, they could be part of all Romance speakers’ abstract and active knowledge of the organization of their morphological system, in such a way, indeed, as to guide and condition diachronic change.

Our example is the pattern of allomorphy historically caused by the palatalizations. As we saw in Section 1.3.1, an effect of the palatalizations in early Romance was to confer on all Romance varieties a pattern of alternation such that presence of a root allomorph in the present subjunctive cells, or in the first person singular present indicative cell, implies the presence of that same alternant in all the other cells. In central

⁷⁶ The view that Romance-based creoles are not strictly Romance languages does not, of course, mean that they are one bit less worthy of study by linguists, but it does mean that they may be more effectively understood from a ‘creole’ rather than a ‘Romance’ perspective. See, for example, the rich literature on the theoretical significance of the emergence of creoles (e.g., Holm 1988; Manessy 1995; Hazaël-Massieux 1996; Mufwene 2002). However, for an intriguing example of a characteristic and idiosyncratic feature of Romance verb morphology, namely conjugation-marking vowels, being continued in a creole language see Luís (2011).

⁷⁷ The reservation is prompted by the possibility that the Italian second person plural endings originate in the Latin *imperative* ending *-TE* (see Maiden 2007: 159–61). Even then, the formative [t] would be underlyingly common (in the historical sense) to all Romance languages.

⁷⁸ It is also non-distinctive by being part of the common Indo-European inheritance: compare, for example, Lat. *UIDETIS* ‘you.2PL see’, It. *vedete*, but also Croatian (Slavonic) *videte*. On this basis an unwary observer might conclude that these three languages belong to the same family, while French (*voyez* [vwajɛ]) was unrelated to any of them!

Italo-Romance and for a subset of verbs in Romanian (the reasons are ultimately a matter of regular historical phonology: see, e.g., Maiden 2020). Maiden (e.g., 2018: 84) labels this paradigmatic distributional pattern the ‘L-pattern’; the variant that includes the third person plural present indicative is labelled the ‘U-pattern’. These two labels are strictly arbitrary; indeed, they are necessarily arbitrary since the relevant patterns are not describable in terms of any phonological or functional labels. However, for ease of reference here we will use the label ‘L-pattern’ to cover both of them (after all, the U-pattern subsumes the L-pattern). Two assumptions are crucial: that the original phonological conditioning of the L-pattern is defunct, and that the set of paradigm cells defined by the effects of the original sound changes is fundamentally ‘incoherent’, in that it is not plausibly reducible to any semantic or functional common denominator. What, after all, does the class of present subjunctive and first person singular present indicative cells have significantly and uniquely in common? Now these assumptions are far from uncontroversial, but our purpose here is expository: see especially Maiden (2018: 161–66) for references to, and discussion of, the relevant debate. Maiden’s view is that the L-pattern belongs to that class of entities which Aronoff (1994) describes as ‘morphemes’, autonomously morphological entities conceived as functions lacking any inherent connexion with a specific form or a specific meaning yet serving systematically to relate form and meaning in the morphological system (see further Sections 11.3–4). In our case, and put informally,⁷⁹ the palatalizations gave rise to a pattern of allomorphy which has persisted long after the original phonological motivation has become defunct. Its paradigmatic distribution becomes synchronically arbitrary, lacking any common set of distinguishing morphosyntactic or phonological features. Yet, crucially, the L-pattern, historically found across the Romance languages, commonly displays diachronic ‘coherence’, in that real or potential exceptions to it tend to be removed or resisted, and changes affecting any of the specified cells in the L-pattern equally affect all the others ‘in lockstep’. Moreover, the L-pattern often acts as a ‘template’ for the distribution of innovatory kinds of alternation which have no connexion whatever with the sound changes which originally created the L-pattern.

It is not possible here to describe anything like the full range of coherence and innovations respecting the L-pattern over the history of the Romance languages (for detailed exemplification see Maiden 2018: 93–122), and all we can do here is to give some brief indicative examples. Table 1.27 illustrates L-pattern effects produced by phonetically regular palatalizations in Galician-Portuguese and in Italian. Galician-Portuguese is represented here

⁷⁹ For more detailed theoretical discussions of the issues, see Maiden (2018: 236–41).

Table 1.27. 'L-pattern' effects of sound change in Portuguese and old Italian

(a) Portuguese						
	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	<i>tenho</i> 'have'	<i>tens</i>	<i>tem</i>	<i>temos</i>	<i>tendes</i>	<i>têm</i>
PRS.SBJV	<i>tenha</i>	<i>tenhas</i>	<i>tenha</i>	<i>tenhamos</i>	<i>tenhais</i>	<i>tenham</i>
PRS.IND	<i>veja</i> 'see'	<i>vês</i>	<i>vê</i>	<i>vemos</i>	<i>vedes</i>	<i>vêem</i>
PRS.SBJV	<i>veja</i>	<i>vejas</i>	<i>veja</i>	<i>vejamos</i>	<i>vejaís</i>	<i>vejam</i>
PRS.IND	<i>faço</i> 'do'	<i>fazes</i>	<i>faz</i>	<i>fazemos</i>	<i>fazeis</i>	<i>fazem</i>
PRS.SBJV	<i>faça</i>	<i>faças</i>	<i>faça</i>	<i> façamos</i>	<i>façaís</i>	<i>façam</i>
PRS.IND	<i>venho</i> 'come'	<i>vens</i>	<i>vem</i>	<i>vimos</i>	<i>vindes</i>	<i>vêm</i>
PRS.SBJV	<i>venha</i>	<i>venhas</i>	<i>venha</i>	<i>venhamos</i>	<i>venhais</i>	<i>venham</i>
PRS.IND	<i>digo</i> 'say'	<i>dizes</i>	<i>diz</i>	<i>dizemos</i>	<i>dizeis</i>	<i>dizem</i>
PRS.SBJV	<i>diga</i>	<i>digas</i>	<i>diga</i>	<i>digamos</i>	<i>digais</i>	<i>digam</i>
(b) Old Italian						
	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	<i>soglio</i> 'am wont'	<i>suoli</i>	<i>suole</i>	<i>solemo</i>	<i>solete</i>	<i>sogliono</i>
PRS.SBJV	<i>soglia</i>	<i>sogli</i>	<i>soglia</i>	<i>sogliamo</i>	<i>sogliate</i>	<i>sogliano</i>
PRS.IND	<i>veggio</i> 'see'	<i>vedi</i>	<i>vede</i>	<i>vedemo</i>	<i>vedete</i>	<i>veggiono</i>
PRS.SBJV	<i>veggia</i>	<i>veggi</i>	<i>veggia</i>	<i>veggiamo</i>	<i>veggiate</i>	<i>veggiano</i>
PRS.IND	<i>vegno</i> 'come'	<i>vieni</i>	<i>viene</i>	<i>venimo</i>	<i>venite</i>	<i>vegnono</i>
PRS.SBJV	<i>vegna</i>	<i>vegni</i>	<i>vegna</i>	<i>vegnamo</i>	<i>vegnate</i>	<i>vegnano</i>
PRS.IND	<i>dico</i> 'say'	<i>dici</i>	<i>dice</i>	<i>dicemo</i>	<i>dite</i>	<i>dicono</i>
PRS.SBJV	<i>dica</i>	<i>dichi</i>	<i>dica</i>	<i>dicamo</i>	<i>dicate</i>	<i>dicano</i>

by modern Portuguese, and Italian by the medieval form of the language (which better exemplifies the effects of regular palatalization than the modern language).

The diachronic 'coherence' of these alternation patterns lies in the fact that they either survive intact, or are lost completely. The alternations are often subject to analogical levelling, but when that happens it affects all the relevant cells of the paradigm at once. That is to say that there are no 'halfway houses', with the alternant removed in some cells of the paradigm but not in others, and therefore the L-pattern is never violated. Sometimes, the alternants may undergo novel adjustments in their form, but these adjustments operate in all cells in which those alternants occur, apparently at once. Thus, in modern Italian, the historical alternation between *veggi*-[veddʒ]- and *ved*- has been removed in favour of *ved*-, but it has been removed *completely* (not, say, just in the subjunctive, or just in the first person singular and third person plural present indicative). The alternant *vegn*-[veɲɲ]-, has been analogically replaced in the modern language by

Table 1.28. 'Coherent' levelling and innovatory adjustments of the Italian variant of the L-pattern in modern Italian

	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	<i>vedo</i> 'see'	<i>vedi</i>	<i>vede</i>	<i>vediamo</i>	<i>vedete</i>	<i>vedono</i>
PRS.SBJV	<i>veda</i>	<i>veda</i>	<i>veda</i>	<i>vediamo</i>	<i>vediate</i>	<i>vedano</i>
PRS.IND	<i>vengo</i> 'come'	<i>vieni</i>	<i>viene</i>	<i>veniamo</i>	<i>venite</i>	<i>vengono</i>
PRS.SBJV	<i>venga</i>	<i>venga</i>	<i>venga</i>	<i>veniamo</i>	<i>veniate</i>	<i>vengano</i>

Table 1.29. Innovatory L-pattern alternation in Portuguese *perder* 'lose'

	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	<i>perco</i> 'lose'	<i>perdes</i>	<i>perde</i>	<i>perdemos</i>	<i>perdeis</i>	<i>perdem</i>
PRS.SBJV	<i>perca</i>	<i>percas</i>	<i>perca</i>	<i>percamos</i>	<i>percais</i>	<i>percam</i>

Table 1.30. Creation of L-pattern suppletion in Galician dialects

	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	<i>'koʎo</i> 'fit'	<i>'kaβes</i>	<i>'kaβe</i>	<i>ka'βemos</i>	<i>ka'βeðes</i>	<i>'kaβeŋ</i>
PRS.SBJV	<i>'koʎa</i>	<i>'koʎas</i>	<i>'koʎa</i>	<i>ko'ʎamos</i>	<i>ko'ʎaðes</i>	<i>'koʎaŋ</i>

veng- [vɛŋg]-,⁸⁰ but this replacement, again, has happened equally in the present subjunctive⁸¹ and in the two cells of the present indicative (Table 1.28).

Portuguese, like Italian, shows 'coherent' levellings and adjustments, but it also acquires at least one completely novel and unprecedented alternation, whose origins are frankly mysterious. The old Galician-Portuguese verb *perder* 'lose' would not have been expected to show any kind of L-pattern consonantal alternation. Yet at some point, and for some reason (Maiden 2018: 115), *perd-* began to acquire consonantal alternants including *perc-* (where *c* = [k]). This novel alternation, wherever it comes from, takes as its distributional template the L-pattern (Table 1.29).

In some Galician dialects (which show the same kind of inherited L-pattern alternation in the verb as their sister variety Portuguese), two etymologically distinct verbs meaning 'fit, be containable in', namely *coller* and *caber*, have coalesced suppletively (see also Section 12.2.5) into a single inflexional paradigm. The distributional pattern they adopt is again the L-pattern (Table 1.30).

⁸⁰ For a discussion of the motivation, see Maiden (2011a: 238f.)

⁸¹ The first and second persons plural of the present subjunctive are exceptions and develop in a different way. This is a widespread phenomenon in Italian, whose nature is explored in Maiden (2010).

These are just a few illustrative examples from two far-separated Romance languages. The broader claim, made for example by Maiden (2018) is that this (and some other) patterns, unprecedented in Latin, exist across the Romance languages and generally show the same historical characteristics of ‘coherence’ in the face of morphological innovations. This would mean that we have a genuine candidate for a defining structural characteristic of a Romance language. This characteristic is arbitrary in that it is morphomic, and the kind of diachronic evidence that we have seen suggests that this pattern is ‘psychologically real’ for all native Romance speakers. Coherence, and the fact that the L-pattern is repeatedly seen to act as a template for morphological innovation, are only explicable if Romance speakers generally have an abstract sense that the allomorph of any one of relevant set of cells must be identical to all of the others. Our claim that the pattern is a defining characteristic of Romance would of course be compromised if it turned out that an identical pattern existed outside Romance, but precisely because the Romance ‘L-pattern’ is the complex consequence of series of disparate historical sound changes acting on a disparate and idiosyncratic set of paradigm cells in the early Romance verb it is inherently unlikely that exactly this pattern could ever recur in another language family.⁸²

This claim is certainly a bold one, but it has the virtue of being empirically testable on modern speakers as well as in diachrony. Very little experimental testing of the psychological reality of such structures in modern speakers has actually been done, however. The results of the tests that have been done to date (Nevins et al. 2015) have been interpreted as showing that our claim is not in fact true for modern adult speakers of Portuguese, Spanish, and Italian, but the methods and conclusions of that research are problematic (see, e.g., Maiden 2018: 165f.; 2021: 90) and it is certain that further experimental research is needed to test speakers’ reactions to artificially engendered violations of the L-pattern, or to tasks in which speakers are asked to produce the set of present indicative and present subjunctive forms for made-up verbs with novel, invented, forms of root allomorphy. The prediction is, of course, that such violations would tend to be rejected, and that the experimentally made-up allomorphs will tend to be distributed according to the L-pattern. Such work remains to be done, but the diachronic evidence nonetheless indicates that there must at least

⁸² This observation does not imply, as has recently been claimed (Herce 2019; but Maiden 2021a), that typological comparisons of morphomic structures across language families are impossible. Rather, it is inherently unlikely (although not logically impossible) that the particular details of a morphomic structure such as the L-pattern could ever occur outside the Romance languages. Claims that allegedly unique morphomic patterns are in reality not morphomic because they are more widely attested cross-linguistically tend to rest on a failure to grasp the precise details of such patterns (cf. Andersen 2010; Nielsen 2019). It is just possible, also, that morphomic patterns might be *borrowed* from one language into another, although the evidence for this is elusive (see Maiden in progress).

have been a point in the past where the L-pattern was a defining characteristic of Romance languages. It may be added that even those who are doubtful about the validity of morphomic claims of this kind for the *modern* synchrony of Romance languages (see, e.g., Embick 2016: 304; cf. Maiden 2020: 90), tend to concede that the relevant diachronic developments attest to the existence in the past of the kind of abstract morphological knowledge which is postulated.

Whether the ‘L-pattern’ can be regarded as a defining characteristic of Romance languages will continue to be debated. It can at least stand as an example of how one might go about seeking to establish such a definition. That approach consists of looking for diachronic and comparative evidence for speakers’ knowledge of, and ability to replicate, truly arbitrary patterns of mapping between form and grammatical meaning unique to a particular language family. In fact, it recasts the question ‘what is a Romance language?’ as ‘what kind of linguistic knowledge is specifically shared by all and only native speakers of Romance languages?’

1.4 Conclusion

Although modern linguistics as a discipline developed in large part out of the philological study of individual languages and language families such as Romance, it is a striking characteristic of contemporary research in both linguistics and in Romance studies that the traditional link between the two disciplines is often not as strong as it might be. To be sure, analyses within Romance linguistics which fail to take account of enlightening ideas and principles from linguistic theory risk overlooking and/or misconstruing the relevance of all or part of the available empirical evidence they are so at pains to correctly reconstruct, evaluate, and interpret. By the same token, linguistic analyses which are blindly driven by theory-internal considerations with insufficient regard for actual data as offered by textual corpora and the numerous dialectal varieties of Romance run the risk of presenting a largely idealized and unhelpfully selective representation of the available linguistic evidence. The result may be a partial theory which is only capable of accounting for a subset of the available data and that largely ignores the imperfections and irregularities characteristic of real linguistic productions.

The discussions of the preceding sections may seem eclectic and heterogeneous, but their common theme has been that when theory and Romance evidence are considered together, the results are mutually enhancing. The discussion has also shown how careful and detailed consideration of the Romance linguistic data can open our eyes to problems, possibilities, and approaches sometimes overlooked by mainstream linguistic theory.

Selected References

Below you can find selected references for this chapter. The full references can be found online at the following page: www.cambridge.org/Romancelinguistics

- Ledgeway A. (2007a). 'La posizione dell'aggettivo nella storia del napoletano'. In Bentley, D. and Ledgeway, A. (eds), *Sui dialetti italo-romanzi. Saggi in onore di Nigel B. Vincent (The Italianist 27, Special supplement 1)*. Norfolk: Biddles, 104–25.
- Ledgeway A. (2010b). 'The clausal domain: CP structure and the left periphery'. In D'Alessandro, R., Ledgeway, A., and Roberts, I (eds), *Syntactic Variation. The Dialects of Italy*. Cambridge: Cambridge University Press, 38–51.
- Ledgeway A. (2012a). *From Latin to Romance: Morphosyntactic Typology and Change*. Oxford: Oxford University Press.
- Ledgeway A. (2015a). 'Parallels in Romance nominal and clausal microvariation', *Revue roumaine de linguistique* 60: 105–27.
- Ledgeway A. (2017a). 'Syntheticity and analyticity'. In Dufter, A. and Stark, E. (eds), *Manual of Romance Morphosyntax and Syntax (Manuals of Romance Linguistics)*. Berlin: De Gruyter, 837–84.
- Ledgeway A. (2020b). 'The north–south divide: parameters of variation in the clausal domain', *L'Italia Dialettale* 81: 29–77.
- Maiden, M. (1991). *Interactive Morphology. Metaphony in Italy*. London: Routledge.
- Maiden, M. (1992). 'Irregularity as a determinant of morphological change', *Journal of Linguistics* 28: 285–312.
- Maiden, M. (2013). 'Semi-autonomous morphology: a problem in the history of the Italian (and Romanian) verb'. In Cruschina, S., Maiden, M., and Smith, J. C. (eds), *The Boundaries of Pure Morphology*. Oxford: Oxford University Press, 24–44.
- Maiden, M. (2016a). 'Romanian, Istro-Romanian, Megleno-Romanian, and Aromanian'. In Bentley, D. and Ledgeway, A. (eds), *Sui dialetti italo-romanzi. Saggi in onore di Nigel B. Vincent (The Italianist 27, Special supplement 1)*. Norfolk: Biddles, 91–125.
- Maiden, M. (2016b). 'Italo-Romance metaphony and the Tuscan diphthongs'. *Transactions of the Philological Society* 114: 198–232.
- Maiden, M. (2018). *The Romance Verb. Morphomic Structure and Diachrony*. Oxford: Oxford University Press.
- Maiden, M. (2021a). 'The morphome'. *Annual Review of Linguistics* 7: 89–108.
- Maiden, M. (2021b). 'Establishing contact: Slavonic influence on Romanian morphology?' *Journal of Language Contact* 14: 24–52.