

# 1

## 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,<sup>1</sup> 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

<sup>1</sup> 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, Gleßgen, 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; Paná 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:<sup>2</sup> 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).<sup>3</sup> 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

<sup>2</sup> 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).

<sup>3</sup> 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	∅	eum	∅	∅	coxisse	∅	panem.
Fr.	Je dis/veux	<b>qu'</b>	il	<b>a/ait</b>	<b>fait</b>	cuire	<b>du</b>	pain.
It.	Dico/Voglio	<b>che</b>		<b>ha/abbia</b>	∅	cotto	<b>(del)</b>	pane.
Ro.	Spun/Vreau	<b>că/să</b>		<b>a/fi</b>	∅	copt	∅	pâine.
	I.say/want	that <sub>(REALIS/IRREALIS)</sub>	him/he	has <sub>IND/(be)<sub>SBJV</sub></sub>	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).<sup>4</sup> 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).<sup>5</sup>

<sup>4</sup> For examples and discussion of a phonological parameter, see Section 5.7.

<sup>5</sup> See Munaro (1998), Ambar et al. (2001), Munaro, Poletto, and Pollock (2001), Munaro and Poletto (2002), Benincà and Poletto (2005), Kato and Mioto (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 **ei**ne a cumpărat cărțile? (Ro.)  
 b. ¿**Quién** cree John que **quié**n 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**) **quié**n ha comprado **qué**? (Sp.)  
 c. **Cine** ce **ei**ne a cumpărat ee? (Ro.)  
 who what bought what  
 ‘Who has bought what?’

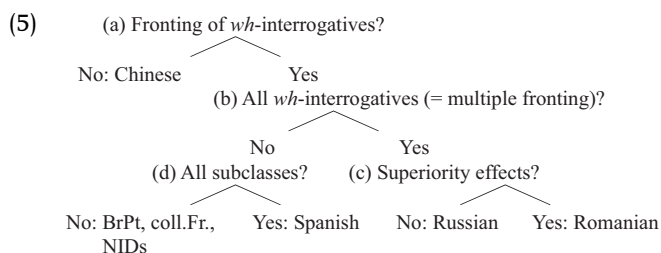
In this respect the behaviour of Romanian appears to parallel that of Slavonic, hardly a surprising result given the widespread borrowing, not just of lexical, but also of functional features across languages of the so-called Balkan Sprachbund. Nonetheless, a closer look at the Romanian facts reveals 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 **ei**ne a cumpărat ee? (Ro.)  
 what who bought  
 ‘Who bought what?’



- c. (\*\*Când) Cine când ~~ei~~e a cumărat-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 *who*]]?

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:<sup>6</sup>

- (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

<sup>6</sup> 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’.