Periphrasis and inflexion: lessons from Romance

Adam Ledgeway and Nigel Vincent

1.1 Introduction

It is no accident that evidence from the Romance languages has played such a central role in general linguistic debates, particularly when, as in many chapters in the present volume, the concern is the relation between structure and change. It is hard to imagine a better test bed: a large family of languages showing various degrees of structural relatedness and rich historical records, in some cases covering more than a millennium, together with a well-documented source language going back another thousand or so years provides an ideal context for the investigation of patterns of change and the way they interact with existing structures. At the same time, the socio-historical circumstances of these developments mean that there are frequent opportunities to investigate the role of external factors, including contact between more or less closely related members of the family, contact with a variety of non-Romance languages, and even contact between the members of the family and their historical ancestor, Latin. And yet no body of historical data, however rich and varied, is ever complete. There are always gaps, inconsistencies, and puzzles, and it is at this point that general linguistic theory can step in and help to complete the picture.

It also comes as no surprise that, in this context, one of the issues that has attracted most attention has been the connection between periphrasis and inflexion, in turn part of the broader range of topics that constitute the traditional and somewhat vexed question of the relation between analysis and synthesis. Latin had rich systems of verbal and nominal inflexion that interacted in various ways with existing periphrastic patterns and with structures that incorporated the ingredients of incipient Romance

periphrases. These structures are of interest not only to those whose principal focus lies in exploring and explaining the development of the Romance languages but also to those whose concern is with the broader theoretical implications for our understanding of language structures and the ways they can change. The Romance languages thus provide a locus where history and theory can interact to mutual benefit.

Our aim in this introductory chapter is to set out the issues, both theoretical — what models are most appropriate to describe and explain the Romance data? — and empirical — what sources of data are available and what issues arise in exploiting them? We begin with the criteria that serve to distinguish analytic patterns in general before going on to examine the specific properties of periphrastic constructions. It is a natural step from there to consider the interface between periphrasis and inflexion and the different ways they are modelled in current theoretical work. From synchrony we move to diachrony and the role of grammaticalization in the genesis of new periphrases and inflexions, and interaction on the one hand between internally and externally motivated change and on the other between reconstructed and attested patterns. We conclude with some general reflections on the relation between synchrony and diachrony before opening the door to the individual case studies that constitute the core of this volume.

1.2 Theoretical issues

1.2.1 Analysis vs periphrasis

Although, superficially at least, there is unmistakable evidence in certain areas of the grammar for the greater analyticity of Romance in contrast to the typical synthetic tendencies of Latin (Meillet 1936:116-121; Bourciez 1956:23; Harris 1978:15f.;

Schwegler 1990; Posner 1996:156f.), this traditional contrast between Latin and Romance in terms of a synthetic-analytic split proves problematic on several grounds, as does the fundamental typological distinction on which it rests (cf. Schwegler 1990:4f.; Bauer 1995:10f., 138, 166; Vincent 1997a:99f., 105; Ledgeway 2012:ch. 2; 2017a). Indeed, Humboldt (1836:288f.) observed that, although individual inflexional forms were lost in the passage from Latin to Romance, inflexion as a whole was not ('Es sanken Formen, nicht aber die Form'; cf. also Vincent 2016a:40). This highlights how the Latin-Romance transition cannot be characterized *tout court* in terms of a shift from synthetic to analytic structures since Romance continues to display extensive syntheticity in both the nominal and verbal domains (see also the discussion in Dragomirescu, Nicolae, and Zafiu, this volume).[1] Consequently, the diachronic study of Romance offers the linguist a privileged data set in which to explore and model the changing relationships between inherited inflexional forms and a series of competing analytic exponents.

Yet, while it is commonplace in treatments of Romance to concentrate on the typological shift from the predominantly synthetic structures of Latin towards the characteristically analytic structures of Romance, considerably less attention has been

[1] Coseriu (1987) reinterprets the traditional synthetic-analytic dichotomy as a distinction between 'internal' and 'external' structures, respectively (cf. Coseriu [1971] 1988:211, 224). Accordingly, the Romance languages are not characterized by a tendency towards analyticity, but, rather, by a tendency to distinguish between external and internal determination and between relational and non-relational functions. Within this approach, what sets Latin apart from Romance is its failure to make any formal distinction between relational and non-relational functions, while Romance follows a new formative principle of mapping relational and non-relational functions respectively onto externally-determined (viz. analytic) and internally-determined (viz. synthetic) structures (see Ledgeway 2012:26-28).

devoted to those cases where the two intersect.[2] Thus, although at a certain level of superficial grammatical analysis it might prove descriptively adequate to place all Romance multi-word grammatical structures under the general heading of 'analytic', it is instructive to separate out those cases where the emergence of analytic exponents serves not so much a substitutive as an additive and complementary function, entering into a direct formal and functional contrast with synthetic exponents. By way of illustration, consider the examples in (1)-(2).

(1) a. statim Roma-m profectus est. (Lat., Cic., *Att.* 12.18.1)
at.once Rome-NOM set.out.PTCP.NOM.MSG be.PRS.IND.3SG
'at once he left for Rome.'

b. A plecat la Roma. (Ro.)
have.PRS.IND.3SG leave.PTCP to Rome.
'S/he has left for Rome.'

(2) a. epistulam Malli-o dedi (Lat, Cic., *Att*. 1.16.16)
letter.ACC.FSG Malius-DAT.MSG give.PFV.IND.1SGR
'I have given the letter to Mallius'

b. Le- am dat scrisoarea la trei
them.DAT= have.PRS.1SG give.PTCP letter.DEF.FSG.NOM-ACC to three
prieteni. (Ro.)
friends.M.NOM-ACC
'I gave the letter to three friends.'

[2] Notable exceptions include Vincent (1987; 2011; 2016a:§4.2.3), Börjars, Vincent, and Chapman (1997), Taylor (2011).

Ostensibly, the pairs in (1)-(2) both involve a contrast between a Latin inflexional case form – accusative marking the directional complement in (1a) and dative the RECIPIENT in (2a) – and a corresponding Romance analytic expression involving the use of the preposition *la* 'to' in the Romanian examples (1b, 2b). Given these parallels, examples such as (1b) and (2b) are standardly lumped together as canonical cases of analyticity. Yet, such a broad-brush approach misses a fundamental difference between the two cases, as becomes clear when examples such as (2c) are introduced.

(2) c. Le-am dat scrisoarea
them.DAT=have.PRS.1SG give.PTCP letter.DEF.FSG.NOM-ACC
prieteni-lor. (Ro.)
friends.M-DEF.GEN-DAT.PL
'I gave the letter to the friends.'

Now the RECIPIENT is specific and correspondingly appears in the definite form which in Romanian is realized by means of an inflexional suffix, viz. *-lor* < ILLORUM 'that' (cf. Ledgeway 2017b), which simultaneously encodes a binary case opposition (cf. *prieteni-i* 'friends.M-DEF.NOM-ACC.PL').[3] This example therefore highlights how, unlike in examples such as (1b) where a prepositional phrase constitutes the only option in conjunction with directional complements, the use of the same preposition in marking RECIPIENTS represents but one of a set of available analytic (2b) and

[3] For further detailed discussion of the relationship between periphrastic and synthetic marking within the Romanian nominal group, see Pană Dindelegan and Uță Bărbulescu, this volume.

inflexional (2c) markers of 'dative'. In the standard language at least (Maiden 2016a:101), these markers form part of a complementary system whose distribution is determined, among other things, by the availability or otherwise of an initial element (e.g., definite noun or quantifier, but not cardinal numerals like *trei* 'three') within the nominal group able to bear inflexional case marking (Iorga Mihail 2013:147; Stan 2013a:264f.).[4] While we continue to refer to examples such as (1b) under the generic label of 'analysis', we shall refer to cases such as (2b) where analyticity and inflexion intersect as 'periphrasis' (cf. also Anderson's 2011:99 distinction between 'analytical expression' and 'grammatical periphrasis').[5]

Irrespective of the theoretical approach that one adopts, an inescapable conclusion from contrasts such as (2b-c) is that a proper understanding of periphrasis must make reference to both morphology and syntax (cf. Vincent 1987:251). On the one hand, the distribution of the preposition *la* as a marker of RECIPIENTS falls out as a consequence of the idiosyncratic limits placed on the morphological marking of dative, insofar as specific nominal categories such as cardinals simply lack the relevant inflexional morphology, i.e., there are no such forms as ***trei-or* 'three-GEN-DAT.PL' (though for examples in earlier stages of the language, see the discussion in §10.2.4.1 in Pană Dindelegan and Uță Bărublescu, this volume). Periphrastic marking of RECIPIENTS can therefore be understood as the output of a Paninian Principle introduced as a last-resort mechanism whenever inflexional morphology is blocked. On the other hand, there is no getting away from the fact that the preposition *la* in (2b) combines with the relevant nominal to form a prepositional phrase (though see the

[4] Cf. also the discussion of definiteness marking via inflexions and free-standing articles in Danish in Spencer and Popova (2015:200f.).

[5] *Pace* Spencer and Popova (2015:197) who take 'analytic' to be a synonym of 'periphrasis'.

discussion in relation to examples (5)-(7) in §1.2.1.2 below), an operation which is clearly part of the syntax which builds constituent structure and determines, for example, the head-initial linearization of *la* and its complement and ensures that the latter is case-marked accusative under subcategorization by its governing prepositional head. We thus see that periphrasis is firmly situated at the interface between morphology and syntax, a characterization which, as noted by Brown *et al.* (2012:272), 'is often problematic for linguistic models since they have to accommodate not only straightforwardly morphological and straightforwardly syntactic phenomena but also phenomena that seem to share properties of both'. Where theories differ is how they draw the line between the two, if indeed they draw one, and the concomitant balance they strike between morphology and syntax in modelling periphrasis. Following Stump's (2001:15; 2018a:56f.) taxonomy, we distinguish here between lexical and inferential approaches to inflexional morphology.

1.2.1.1 Lexical theories

Broadly speaking lexical theories, whether incremental like Lieber's (1992) or realizational such as Distributed Morphology (DM; cf. Halle and Marantz 1993; 1994), assume that the association of particular word forms and their morphosyntactic properties arises in the syntax. As such there is no line to be drawn between morphology and syntax since morphology is simply syntax. This is encapsulated in the leading claim of DM that morphology is 'syntax all the way down' (Marantz 1997), with the atomic units of syntax taken to be morphemes which are subject to the same generative principles that manipulate words into larger phrasal hierarchical constituents (cf. also §2.2.1 in Smith, this volume). On this view, periphrasis ought to

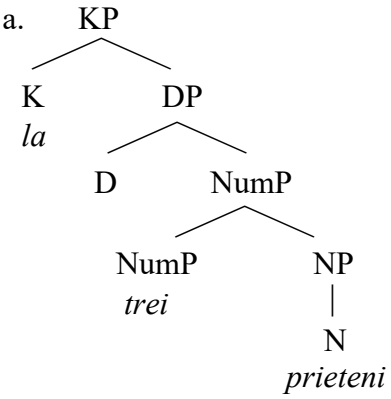
be ‘grist for the mill of Distributed Morphology’ (Kiparsky 2005:113) since it appears to offer a natural way of blending morphological and syntactic behaviours within a unified framework. Indeed, Embick (2007:3) goes so far as to claim that ‘alternations between “words” and larger syntactic objects (phrases) seem to be directly compatible with a syntactic approach to morphology’. For example, the difference between (2b-c) would presumably be modelled in terms of the application or otherwise of a post-syntactic PF process of Fusion (Halle and Marantz 1993:116) on a common syntactic derivation prior to Vocabulary Insertion when the various feature bundles (viz. abstract morphemes dubbed ‘listemes’) are given phonological exponence.[6]

On this view, the contrast witnessed in (2b-c) highlights how morphosyntactic variation in dative-marking through the formal alternation between periphrastic *la* ‘to’ and inflexional *-lor* crucially depends on whether K(ase) is realized in a scattered or

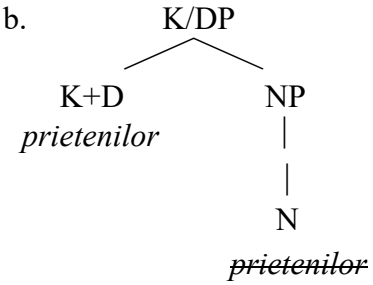
[6] In other cases the amalgamation of individual functional heads and their associated features may arise in the narrow syntax through such operations as head movement (cf. Baker’s 1985 *Mirror Principle*). For instance, the contrast between the French synthetic and periphrastic future expressions *(il) parler-a* ‘(he) speak.INF-FUT.3SG’ vs *(il) va parler* ‘(he) go.PRS.IND.3SG speak.INF’ can be understood as the respective outputs of the move vs merge operations. In the former case the lexical root – instantiated, at least historically, by the infinitive – undergoes V-movement to left-adjoin to the T-related functional head realizing affixal future features (together with person/number Agr(eement)), thereby yielding a complex head which is readily interpreted as a synthetic expression at PF, namely [$_{T*}$ [$_{V}$ *parler-*] [$_{T}$ *-a*]] (cf. Roberts and Roussou 2002; 2003:48-58; Roberts 2009). In the latter case, by contrast, the features of the T-related ‘future’ head are lexicalized by direct merger of a free-standing auxiliary such that the lexical infinitive remains *in situ* (viz. [$_{T'}$ [$_{T}$ *va*]…[$_{v\text{-}VP}$ *parler*]]) and, in the absence of post-syntactic PF rules, a periphrastic expression obtains (cf. also discussion in Andriani, Groothuis, and Silvestri 2020:§2).

syncretic fashion (cf. Giorgi and Pianesi 1997). As illustrated in (3a),[7] whenever lexical material such as the cardinal *trei* intervenes between the K° and N° nodes, as happens when the nominal is not definite and hence remains *in situ*, then these two positions are independently projected and the two heads are realized in a scattered fashion with the K° head lexicalized by *la* 'to'. When, however, the two heads are adjacent, as happens whenever the nominal is definite and lexicalizes the D° position under N-movement, then a syncretic, complex K/D head obtains under Fusion in which both case and definiteness are now inextricably bound together and phonologically spelt out in the inflexional suffix *-lor* realized on the nominal head (3b).

(3) a.



b.



The theory therefore predicts that unfused structures are realized as periphrastic datives, whereas the resulting fused structures are spelt out as synthetic datives.[8] In some contexts the surface output is variable, as in the case of quantifiers which may

[7] For ease of exposition, in (3a-b) we informally represent the features of terminal nodes with their surface PF lexicalizations.

[8] A similar line of reasoning in terms of post-syntactic 'merger' operations is proposed in Embick (2000) and Embick and Noyer (2001), and in Embick (2007) to account for synthetic and periphrastic alternations in the Latin perfect and the English comparative/superlative, respectively.

optionally undergo Fusion with K (4a), an option favoured in higher registers of the language, to produce synthetic forms such as *puțin-or* 'few-GEN-DAT.PL' alongside periphrastic *la puțini* 'to few.NOM-ACC.MPL'. This variability in the application of the post-syntactic PF process also explains sub-standard periphrastic structures such as *la cei* 'to the.NOM-ACC.MPL' in (4b) where, in contrast to the standard synthetic variant *celor* 'the.GEN-DAT.PL', Fusion fails to apply to the free-standing definite article *cei* merged under D^{o}, itself a periphrastic form of the article employed just in those structures where an intervening cardinal blocks N-to-D raising.

(4) a. Dau scrisoarea la puțini / puținor
give.PRS.1SG letter.DEF.FSG.NOM-ACC to few.NOM-ACC.MPL few.GEN-DAT.PL
prieteni. (Ro.)
friends.M
'I'll give the letter to a few friends.'

b. Dau scrisoarea ?la cei / celor
give.PRS.1SG letter.DEF.FSG.NOM-ACC to the.NOM-ACC.MPL the.GEN-DAT.PL
doi prieteni. (Ro.)
two.M friends.M
'I'll give the letter to the two friends.'

To sum up, *la* represents the scattered spell-out (cf. Giorgi and Pianesi 1997) of a single [Kase] feature in contrast to its syncretic spell-out in the feature bundle [Kase, Definite] (⇒ *-lor*, *celor*) or [Kase, Q] (⇒ *puținor*). However, it is not immediately clear what, if any, independent syntactic motivation there is for positing post-syntactic operations such as Fusion and other PF readjustment rules other than to get the

morphology to fit the syntax by manipulating and rearranging the hierarchical structure generated by the narrow syntax (cf. Vincent 1987:251; Stewart and Stump 2007:384f.).[9] Apart from being unconstrained in this way, processes such as Fusion also face an additional timing problem in that the premise of their application rests on a 'look-ahead' principle (cf. Siddiqi 2018:155f.): fused forms obtain just in those cases when the Vocabulary contains an item such as *prietenilor* which requires two heads to fuse, but this information is not available at the time of fusion which follows syntax but precedes Vocabulary Insertion. By the same token, variation between fused and unfused variants such as *puținor/celor* and *la puțini/? la cei* proves at best hit-and-miss, with the theory apparently hedging its bets between which of the two available Vocabulary Items might be inserted.

More worrying, however, for our immediate concerns are the consequences that such approaches have for our understanding of periphrasis. To put it bluntly, periphrasis has no place in a theory like DM which simply fails to recognize it as a distinct category at all. Rather, it is forced to treat cases like (2b) under the all-embracing general heading of 'analysis', indiscriminately lumping it together with examples such as (1b), for they both represent cases in which Fusion has not applied and, crucially, in which there is no *a priori* expectation for Fusion to have failed to apply in one or other of the two cases. In short, the system simply outputs objects, irrespective of whether the morphophonological packaging of the relevant set of terminal nodes surfaces in a one- or multi-word form. On this view, periphrasis turns out to be rather unspectacular, just another undifferentiated case of analysis. Yet, we have seen in §1.2.1 that there are very good reasons to keep examples like (2b)

[9] For an overview of such post-syntactic processes and DM in general, see Bobaljik (2017) and Siddiqi (2018).

separate from examples such as (1b) and failure to do so misses some non-trivial empirical and theoretical generalizations.

1.2.1.2 Inferential theories

In contrast to lexical approaches where word forms and their associated morphosyntactic properties are mediated entirely within the syntax, inferential theories, whether incremental (e.g., Steele 1995) or realizational such as Paradigm Function Morphology (PFM; Stump 1991; 2001; 2016; 2018; Stewart and Stump 2007), formalize such associations within an autonomous system of morphology. Above we saw in relation to contrasts such as (2b-c) that our understanding of periphrasis crucially relies on the assumption that the domain of inflexion is not limited to the conventional building blocks and units of traditional morphological analysis, but should be interpreted in such a way as to be able to also accommodate multi-word sequences whose forms and distribution are generally taken to fall within the realm of syntax. In our view, such a state of affairs is only compatible with inferential-realizational models of the Word-and-Paradigm type since these are able to integrate both single synthetic forms and multi-word expressions with the morphosyntactic realization of individual paradigm cells (cf. Blevins, Ackerman, and Malouf 2018:273f.). Key therefore to our definition of periphrasis is the role of the inflexional paradigm and the relations that hold between the interdependent exponents – whether synthetic or analytic – of individual cells of a given lexeme.[10] Indeed, the

[10] Of course, paradigmaticity plays no role in configuration-based lexical theories such as DM which do not recognize the paradigm in their analyses of such phenomena as synthetic/analytic alternations (Embick 2000; 2007) and suppletion (Pomino and Remberger 2019). In short, to recognize the paradigm within DM would involve a special case of the problem of transderivationality which has

notion of paradigm figures prominently in most definitions of periphrasis, witness Haspelmath's (2000:655) now classic formulation as 'a situation in which a multi-word expression is used in place of a single word in an inflectional paradigm' (cf. also Vincent 1987:240; Stump 2001:43; Stewart and Stump 2007:408; Anderson 2011:143).

On this view, the inflexional and periphrastic forms in (2b-c) can now be laid out as part of a structured inflexional paradigm along the lines of Table 1.1.

Table 1.1: Inflexional paradigm of Romanian PRIETEN 'friend'

	Indefinite		Definite	
	Singular	Plural	Singular	Plural
Nom-Acc	prieten	prieten-i	prieten-ul	prieteni-i
Gen-Dat	a/la prieten	a/la prieteni	prieten-ului	prieteni-lor

With very few exceptions, the paradigmatic relations of form and function illustrated in Table 1.1 are representative of all masculine nouns where the distribution of individual forms can be stated in terms of a network of intersecting features relating to (in)definiteness, number (singular vs plural), and case (nominative-accusative vs genitive-dative) (Maiden 2016a:100f.). In most of the cells the exponents of the available feature combinations are synthetic, the exception being those cells realizing the pairing genitive-dative and indefinite where genitive and dative are marked, respectively, by *a* and *la* in combination with the unmarked nominative-accusative indefinite forms.[11] Within their paradigmatic representation, these periphrastic

always bedevilled derivational models since it is hard to see how the derivation of one structure can 'look across' to see if it is compatible with the derivation of another structure.

[11] Although genitive and dative are syncretic when realized synthetically (e.g., *-lui*, *-lor*), they are separated out in their periphrastic realizations, namely *a* vs *la*, respectively. Given our focus here on

exponents can therefore be interpreted as a 'repair strategy' for language-specific deficiencies in the morphological realization of particular feature combinations which find no formal expression in the synthetic morphology of the language (cf., by contrast, the well-formedness of Lat. (in)definite AMIC-I/-ORUM 'frend.-GEN.MSG/-GEN.PL').

Confirmation that the periphrastic realizations in Table 1.1 are effectively part and parcel of the inflexional paradigm is evidenced by a number of considerations. First, the cells involved do not represent accidental gaps in the paradigm of particular defective lexemes which are otherwise regularly realized by synthetic exponents in the inflexional paradigms of other lexemes belonging to the same category. Rather, as the examples in (5)-(6) illustrate, the paradigmatic pattern characterizes all masculine nouns of the same inflexional class without exception.

(5) a. Îi vorbește popor-/comitet-/grup-ului. (Ro.)
DAT.MSG= speak.PRS.IND.3SG people/committee/group-DEF.GEN-DAT.MSG
'S/he speaks to the people/committee/group.'

b. Îi vorbește la tot popor-/comitet-/grup-ul. (Ro.)
DAT.MSG=speak.PRS.IND.3SG to all.MSG people/committee/group-DEF.NOM-ACC.MSG
'S/he speaks to all the people/committee/group.'

(6) a. Le vorbește studenți-/muncitori-/musafiri-lor. (Ro.)
DAT.PL= speak.PRS.IND.3SG students.M/workers.M/guests.M-DEF.GEN-DAT.PL

---

dative, we ignore in what follows the periphrastic genitive marker *a,* though it should be noted that its distribution is conditioned by exactly the same factors as those operating on *la*.

'S/he speaks to the students/workers/guests.'

b. Le vorbește la niște studenți/muncitori/musafiri. (Ro.)
DAT.PL= speak.PRS.IND.3SG to some students/workers/guests.NOM-ACC.MPL
'S/he speaks to some students/workers/guests.'

Second, just like the synthetic members of the inflexional paradigm (5a, 6a), the periphrastic exponents can also be doubled by a dative clitic (5b, 6b), although prepositional objects introduced by *la* do not otherwise allow clitic-doubling (cf. Cornilescu 1995:13), witness the ungrammaticality of (7).

7 (**Le) recurge la niște studenți/muncitori/musafiri. (Ro.)
DAT.3PL= resort.PRS.IND.3SG to some students/workers/guests.NOM-ACC.MPL
'S/he has recourse to some students/workers/guests.'

Finally, it is standardly assumed that one of the constraints operating on coordination is the requirement that both conjuncts be of equal status and, in the case of DP coordination, that both DP conjuncts bear the same case form. Consequently, coordinate structures such as (8a) where a single nominal functions as the complement of two coordinated verbs, the first subcategorizing for an accusative DP and the second for an oblique PP headed by *la* 'to', prove ill-formed since the complement cannot simultaneously meet the conflicting categorial and case requirements of both verbs. Against this backdrop, the grammaticality of examples such as (8b), in which inflexional and periphrastic datives are felicitously coordinated, highlights that we are indeed dealing with the same type of grammatical beast,

irrespective of its surface synthetic or analytic form to which the grammar is seemingly blind.

(8) a.**Citeam și mă uiteam (la) un
read.PST.IPFV.1SG and me.ACC= look.PST.IPFV.1SG to a.NOM-ACC.MSG
sms. (Ro.)
sms.NOM-ACC.MSG
'I was reading and looking at a text message.'

b. Le vorbeam studenților și la niște
DAT.3PL= speak.PST.IPFV.1SG students.DEF.GEN-DAT.PL and to some
prieteni. (Ro.)
friends.NOM-ACC.MPL
'I was speaking to the students and some friends.'

In conclusion, these brief considerations highlight how periphrastic forms are fully integrated into otherwise predominantly synthetically-based inflexional paradigms. Despite their superficial analytic appearance, they have been shown to pattern in many respects (e.g., clitic doubling, coordination), not with canonical phrasal prepositional constituents formed in the syntax, but, rather, with the monolectal forms that constitute the atomic objects of inflexional morphology with which they share key properties. This highlights how all that is analytic is not necessarily syntactic, but can equally be morphological at the appropriate level of abstraction.

At the same time, it is not at all clear how lexical-based models such as DM, which do not make reference to paradigms, would be able to handle and make sense of the paradigmatically predictable synthetic-periphrastic alternations in (5a,6a) and (5b,6b)

and the concomitant categorial and case distinctions highlighted in the distribution of clitic-doubling and coordination. In particular, it is hard to see how the relevant synthetic-periphrastic alternations can be accounted for other than by stipulating a post-syntactic process of Fusion precisely and only in those cases where synthetic datives are not available. Similar problems face the clitic doubling and coordination contrasts in (5)-(7) and (8a-b). In the former case one might have to appeal to an abstract distinction, the motivation for which is not particularly strong in relation to these clitic doubling facts, between, say, structural and inherent (or lexical) Cases, on the assumption that the dative clitics *îi/le* are a spell-out of structural, not inherent Case. Following this same line of reasoning, it could be argued that when marking structural dative, *la* heads a KP (cf. 3a-b), in contrast to its uses in inherent/lexical Case contexts such as (7, 8a) where it heads a full-fledged PP, and hence is amenable to coordination in (8b) with a synthetic dative since they both instantiate (K/)DP constituents.

1.2.2 Formal issues in modelling periphrases

Having reviewed some basic theoretical assumptions about the nature of periphrasis and its formal and functional relationship to inflexion and analyticity more generally, we now turn to explore a number of key issues that have figured prominently in the theoretical literature in defining and modelling inflexional periphrasis. In particular, our aim here will be to investigate some of the central ideas and claims that have emerged in the formal definition and identification of periphrasis as a distinct linguistic category by considering a selection of Romance examples which have not previously figured in the relevant literature. The Romance evidence, it will be shown, brings to light some significant theoretical questions and challenges which force us to

reassess and refine in a number of ways our understanding of periphrasis and the interfaces between morphology, syntax, and semantics.

1.2.2.1 Intersectivity

As briefly noted above, a fundamental criterion in the definition of inflexional periphrasis in paradigmatic approaches is intersectivity (cf. also §2.4.2 in Smith, this volume), a notion which allows us to distinguish between 'pure' cases of analyticity, which fall entirely within the realm of syntax, and those incidences of 'superficial' analyticity, namely periphrasis, which can only be appropriately defined in relation to the syntax-morphology interface (Ackerman and Stump 2004; Vincent 2011:425; Brown *et al.* 2012:246; Spencer and Popova 2015:204). Essentially, intersectivity arises whenever within a given paradigm the realization of permissible combinations of different grammatical features results in the integration of cells which, for the most part, are realized by monolectal synthetic expressions with one or more cells whose exponence is exceptionally realized by multi-word periphrastic expressions, producing what Haspelmath (2000:656) calls 'paradigm symmetry'. As observed in the Romanian examples above (cf. Table 1.1), within the nominal domain intersectivity is commonly determined by the convergence of core nominal morphosyntactic features such as (in)definiteness, number, gender, and case, but arguably not by purely morphological features such as declensional class.[12] Within the

[12] There is a not insignificant number of native (e.g., *ianuarie* 'January', *luni* 'Monday', *b* 'letter b', *Ion*) and non-native (e.g., *dingo* 'dingo', *euro*, *miss* 'beauty contestant', *yoga*) nouns in Romanian which are invariant and hence inflexionless (cf. Ledgeway 2017b:§2.2). Unlike most nouns (cf. Table 1.1), their definite genitive-dative forms are in most (but not all) cases therefore supplied by the unbound prenominal determiner *lui*, e.g., *începetul lui ianuarie* 'the beginning of January', a potential case of periphrasis. However, even if one were to treat such invariant nouns as constituting a distinct

verbal domain, by contrast, the relevant intersecting parameters that yield structured synthetic-periphrastic alternations relate to person, number and, in particular, tense, aspect, mood, and voice. Person and number, whether individually or together, are less strong determinants of intersectivity as witnessed by the paucity of relevant Romance examples. However, one plausible candidate is the Italian present and past subjunctive paradigms set out in Table 1.2.

Table 1.2: Italian present and past subjunctive paradigms

	Present			Past		
	cant- 'sing'	vend- 'sell'	serv- 'serve'	cant- 'sing'	vend- 'sell'	serv- 'serve'
1SG	cant-i	vend-a	serv-a	cantass-i	vendess-i	serviss-i
2SG	**(tu) cant-i	**(tu) vend-a	**(tu) serv-a	**(tu) cantass-i	**(tu) vendess-i	**(tu) serviss-i
3SG	cant-i	vend-a	serv-a	cantass-e	vendess-e	serviss-e
1PL	cant-iamo	vend-iamo	serv-iamo	cantass-imo	vendess-imo	serviss-imo
2PL	cant-iate	vend-iate	serv-iate	canta-ste	vende-ste	servi-ste
3PL	cant-ino	vend-ano	serv-ano	cantass-ero	vendess-ero	serviss-ero

Although Italian arguably represents an exemplary case of a pro-drop language, it exceptionally requires the use of the overt second-person singular subject pronoun *tu* 'you' in the present and past subjunctive (9a-b) where, in contrast to non-subjunctive contexts, the use of an overt pronoun does not license the typical focus or topic-shift interpretations (Benincà 1993:258-262; Cardinaletti 1997:51f.; Cordin 2001:540; Ledgeway 2016b:216).

(9) a. Non sapevano che pro$_{1/**2/3SG}$ cant-i / -assi. (It.)
neg=know.PST.IPFV.3PL that pro$_{1/**2/3SG}$ sing-PRS.SBJV.SG -PST.SBJV.1-2SG

b. Non sapevano che tu cant-i /-assi. (It.)

declensional class from those to which the majority of regular inflecting nouns belong, the relevant forms would still not involve any notion of intersectivity because, as inflexionless nouns, they do not belong to a paradigm. Rather, they represent a genuine case of inflexional defectivity.

neg=know.PST.IPFV.3PL that you.SG sing-PRS.SBJV.SG -PST.SBJV.1-2SG

'They didn't know that I/you/(s)he sing(s)/sang.'

As Table 1.2 illustrates, there is an obvious functionalist explanation for this distribution, in that in the present there is syncretism in the synthetic exponents of the paradigm between all three persons in the singular and between the first and second persons singular in the past. Obligatory use of the pronoun therefore serves to (partially) disambiguate reference between the singular persons and thereby reduce the computational burden placed on the hearer.

Even if such functionalist explanations might lie behind the emergence of this usage, synchronically they are untenable, especially since there are many other Romance varieties with similar subjunctive paradigms where such surface inflexional ambiguity is readily tolerated (Manasantivongs 2005). For instance, the overt pronoun is still required even if no such ambiguity arises, as happens in cases like (10a) where the subjunctive verb is inherently reflexive and second-person singular reference is already unambiguously marked by the coreferential reflexive clitic. Similarly, the pronoun cannot be omitted in (10b) where the matrix verb is first person singular and the subjunctive verb is past, even though the latter can only be interpreted as second person singular on account of the obviation effect.

(10) a. Vuole/Voleva che **(tu) ti penta /

want.PRS.IND/PST.IPFV.IND.3SG that you.SG yourself.SG=repent.PRS.SBJV.SG

pentissi. (It.)

repent.PST.SBJV.1-2SG

'(S)he wants/wanted you to repent'

b. Volevo che **(tu) ti pentissi. (It.)
want.PST.IPFV.IND.1SG that you.SG yourself.SG= repent.PST.SBJV.1-2SG
'I wanted you to repent.'

Rather, the distribution of the overt pronoun is clearly driven by syntax rather than by functionalist considerations, although we see once again that we must also make reference to morphology since the relevant two-word form is clearly integrated into the inflexional paradigm of the subjunctive where it spells out the intersection of the features second person singular and subjunctive. Indeed, the morphological nature of this periphrastic expression is supported by the observation that its presence within the inflexional paradigm of the present subjunctive gives rise to a potentially morphomic distribution of forms whereby the second person singular (e.g., *tu canti*) is formally contrasted with the first/third person singular (e.g., *canti*). In this way, distinctive marking of the second person singular is responsible for placing together first/third singular in a single syncretic category which otherwise lacks any syntactic or semantic motivation, rather than say first/second person singular (the discourse participants) vs third person singular, as happens in the inflexional morphology of the past subjunctive (viz. *-i* vs *-e*).

In other respects too, there are reasons to treat the combination *tu*+V_{SBJV} increasingly as a single morphological unit, rather than as the output of the syntax. First, in northern and central varieties of Italian, where linguistic prestige is highest, the second person singular nominative pronoun is not *tu* as in the official standard language and in southern varieties, but has been replaced by the erstwhile tonic oblique pronoun *te* (11a), which also occurs in all varieties in 'strong' positions such as exclamative vocatives and under coordination (11b). As a result, the only context

in which northern and central speakers use *tu* (on a par with southern speakers) is the subjunctive (although here too they use *te* if the subject is given any particular prominence), a distribution which eliminates *tu* from the paradigm of pronouns and relegates it to a greatly weakened preverbal second person singular subjunctive marker more akin to an inflexional element than an autonomous piece of phrasal syntax.

(11) a. Io rimango, ma tu/te che fai? ((S)It / N-CIt.)
I remain.PRS.IND.1SG but you.NOM/OBL what do.PRS.IND.2SG
'I'm staying, but what are going to do'

b. Povero te /**tu! // Io e te /**tu… (It.)
poor.MSG you.OBL NOM I and you.OBL NOM
'Poor you! / I and you…'

Finally, a similar conclusion comes from examples of complementizer deletion in subjunctive contexts (Wanner 1981; Giorgi and Pianesi 1997; 2004; Poletto 2001; Scorretti 1994; Giorgi 2010:43-47). For most speakers, complementizer deletion also entails deletion of the preverbal subject position, such that canonical preverbal subjects such as *lui* 'he' and *Ida* in (12a) are obligatorily realized in postverbal position under complementizer deletion in (12b). The exception, however, is *tu* which uniquely survives under complementizer deletion, a distribution which reinforces our conclusion that it must be treated as part of the inflexional structure of the verb, rather than as a tonic pronoun since there is ostensibly no available position for preverbal subjects in contexts such as (12b).

(12) a. Penso che tu / lui / Ida sia in grado di
think.PRS.IND.1SG that you he Ida be.PRS.SBJV.SG in position of
aiutarlo. (It.)
help.INF=him

b. Penso tu /**lui /**Ida sia lui / Ida in grado di
think.PRS.IND.1SG you he Ida be.PRS.SBJV.SG he Ida in position of
aiutarlo. (It.)
help.INF=him
'I think (that) you/he/Ida is able to help him.'

Our example of the Italian subjunctive also highlights another useful distinction in relation to the morphological space across which intersectivity may range. In this case, the relevant intersection of features occurs at the level of the 'local' paradigm, in the sense that the alternation between synthetic and periphrastic forms arises specifically within the individual sub-paradigms of the present and past subjunctive without any need for us to make reference to the wider verb system. In other cases, by contrast, the distribution of synthetic and periphrastic expressions can only be stated in terms of the 'extended' paradigm (cf. also reference to the 'extended paradigmatic set' in Vincent 2011:424). Classic examples are the Romance periphrastic perfects (Vincent 1987:238f.; 2011:430-432; Taylor 2011:405-407; Brown *et al.* 2012:271 n.11) such as in northern regional Italian where the relevant intersection between imperfective (13a) and perfective (13b) aspect cuts across distinctions of person/number (13c), tense (13d), mood (13e), and finiteness (13f) to hold across all individual sub-paradigms of the extended active verb system.

(13) a. Parlo ‘speak.PRS.IPFV.IND.1SG’, parlavo ‘speak.PST.IPFV.IND.1SG’, … (NIt.)

b. Ho parlato ‘have.PRS.IPFV.IND.1SG speak.PTCP’, avevo parlato ‘have.PST.IPFV.IND.1SG speak.PTCP’, … (NIt.)

c. Parlano/parlavano ‘speak.PRS/PST.IPFV.IND.3PL’ vs Hanno/avevano parlato ‘have.PRS/PST.IPFV.IND.3PL speak.PTCP’,… (NIt.)

d. Parlerò ‘speak.FUT.IPFV.1SG’ vs Avrò parlato ‘have.FUT.IPFV.1SG speak.PTCP’,… (NIt.)

e. Parli/parlassi ‘speak.PRS/PST.IPFV.SBJV.1SG’ vs Abbia/avessi parlato ‘have.PRS/PST.IPFV.SBJV.1SG speak.PTCP’ (NIt.)

f. Parlare ‘speak.INF.IPFV’ vs Aver parlato ‘have.INF.IPFV speak.PTCP’,… (NIt.)

As noted by Blevins (2015:88), typically the members of a nominal paradigm (viz. declensional system) are organized into a single network of forms differentiated by specific featural combinations based on number, gender, case, and (in)definiteness (cf. Table 1.1), whereas verbal paradigms (viz. conjugations) introduce a further level of complexity in that individual verbal lexemes are organized not just in terms of person and number distinctions, but are also generally structured into multiple sub-paradigms based on such distinctions as tense, aspect, mood, and voice. It follows that the relevant intersection that produces synthetic-periphrastic splits typically applies locally within the global paradigm in the case of the nominal domain (cf. Table 1.1), but across different (sets of) sub-paradigms belonging to the extended paradigm in the case of the verb (cf. 13a-f). While this is undoubtedly the usual state of affairs, our example of the Italian subjunctive in Table 1.2 is, however, revealing since it highlights that in the verbal domain the relevant intersection can also occur ‘locally’ within individual sub-paradigms without reference to the extended verbal paradigm.

Finally, we briefly consider cases of so-called anti-periphrasis (Haspelmath 2000:658f.) where the canonical distributional balance between synthetic (the majority) and periphrastic (the minority) expressions within the paradigm is reversed. By way of example, consider the partial Romanian paradigms in Table 1.3:

Table 1.3: Romanian (3SG) imperfective and perfective paradigms SERV- 'serve'

	Present	Past	Future	Conditional	Infinitive	Subjunctive
IPFV	servește	servea	va servi	ar servi	a servit	să servească
PFV	a servit	servise	va fi servit	ar fi servit	a fi servit	să fi servit

In contrast to the forms of the imperfective which are either synthetic (present, past) or periphrastic (cf. future/conditional auxiliaries *va/ar*, infinitival/subjunctive markers *a/să*), the forms of the perfective are, on a par with the (northern) Italian forms reviewed above, periphrastic consisting of auxiliary (*a/fi* 'have/be') and perfect participle (*servit*), with the notable exception of the past perfective form (the so-called pluperfect) which, in the standard language at least, is synthetic (cf. *servise*; see also Dragomirescu, Nicolae, and Zafiu, this volume). We thus see that the synthetic past perfective sub-paradigm enters into a formal opposition with all the other available perfective sub-paradigms which are periphrastic, leading us to question whether the paucity of available monolectal forms warrants defining all the remaining perfective forms as examples of intersective periphrasis (cf. also discussion in §2.4.2 in Smith, this volume). Indeed, Spencer and Popova (2015:207) prefer in such cases to talk of 'overdifferentiation' for morphosyntactic properties in a particular sub-paradigm which are otherwise realized syntactically elsewhere in the extended paradigm. It is doubtful, however, that such a characterization is correct. In the prototypical case, overdifferentiation is taken to arise where a particular grammatical distinction fails to extend beyond a restricted number of lexemes (cf. Brown 2007), whereas the

distribution of the Romanian synthetic pluperfect knows no such restrictions but, rather, is a fully productive category extended to all verbal lexemes with a perfective stem (e.g., *mâncasem / putusem / mersesem / plânsesem / servisem…* 'I had eaten/believed/gone/cried/served…'). Furthermore, it should be remembered that we are dealing here with two intersective relationships. The first operates at the level of the extended paradigm between imperfective sub-paradigms, realized both synthetically and periphrastically, and perfective sub-paradigms predominantly realized periphrastically. It is across this latter local network of sub-paradigms that the second and most relevant intersection operates where the synthetic nature of the pluperfect represents an aberration from the otherwise predominantly periphrastic realization of the perfect. We thus see that the intersection of the features perfective and past exceptionally sets off this sub-paradigm of the perfect from all other perfective sub-paradigms. The fact, however, remains, that, irrespective of the number of synthetic sub-paradigms and cells that make these up, there is a formal split within the extended paradigmatic set of the perfective that places both synthetic and periphrastic expressions alongside each other as part of a single (sub)system. We conclude therefore that it seems more appropriate within this perspective to treat the Romanian pluperfect as an anti-periphrastic perfect, albeit one which enters into an intersective relationship with the periphrastic perfective forms.

1.2.2.2 Non-compositionality

The latter example involving the Romanian pluperfect and its relationship to other perfective sub-paradigms also exemplifies an important semantic aspect of inflexional periphrasis which has frequently been considered definitional, namely non-compositionality (cf. Ackerman and Stump 2004:142; see also the discussion in

§2.4.6 in Smith, this volume). Arguably, at a certain level of analysis the periphrastic forms of the perfect exemplified in Table 1.3 can said to be compositional, in that the overall meaning of each of them transparently corresponds to the sum of its individual parts; for instance a present form of the auxiliary *a* in conjunction with the past participle *servit* straightforwardly gives rise to a present perfect, in that same way the future form of the auxiliary *va fi*, itself periphrastic, yields a future perfect in conjunction with the participle, and so forth. The synthetic pluperfect *servise*, by contrast, does not lend itself to such a transparent compositional analysis between form and meaning, but is characterized by a degree of morphological opacity. At most, one can inferentially recognize a perfective stem *servì-* (cf. preterite *servì* '(s)he served', participle *servìt* 'served'),[13] whereas it proves more difficult, at least from a purely synchronic perspective, to transparently associate *-s-* with past tense.[14] There thus emerges an asymmetry across the perfective sub-paradigms between the transparency of the periphrastic forms and the opacity of the synthetic forms, highlighting how non-compositionality represents a frequent, though not a necessary or definitional, characteristic of inflexional morphology.

In view of such tendencies, it is natural to assume, if periphrasis is to be defined in terms of its membership of a structured inflexional paradigm, that it too might develop and show varying degrees of non-compositionality. This view is particularly strongly espoused within PFM where, within a theory of autonomous morphology, the meaning of a periphrasis is not derived from the contentive information associated

[13] We indicate here word stress by the use of the grave accent, although in standard orthography word stress is not marked.

[14] Historically, the relevant forms of the Romanian synthetic pluperfect continue the Latin pluperfect subjunctive, cf. SERUISSET (cf. Dragomirescu, Nicolae, and Zafiu, this volume).

with the sum of its individual parts but, rather, is specified morpholexically (Ackerman and Stump 2004:142). This contrasts with frameworks such as DM which struggles to accommodate non-compositionality since it predicts a strong one-to-one mapping between form and meaning. And indeed non-compositionality is what we find in many cases of Romance periphrasis. Returning to the Romanian forms of the perfect, we have already noted how under its present perfect reading *a servit* is transparently compositional, but what we have not yet considered is its preterite uses in all but formal written registers, the outcome of a process of aoristic drift (Schaden 2012; Bertinetto and Squartini 2016:944f.; Bossong 2016:70; cf. also §1.3.2), where the relevant past punctual value cannot be transparently inferred from the sum of its parts (cf. also Spencer 2001:283; Taylor 2011:§18.5.1; Vincent 2011:430). Rather, what we have is a textbook case of a non-compositional reading.

Interestingly, in south-eastern varieties of Romanian spoken in Oltenia the synthetic preterite (cf. *servi*) survives alongside the periphrastic perfect (cf. *a servit*), but, curiously, the respective past punctual and present perfect values associated with these paradigms in other Romance varieties are reversed (Haase 1995:141; Zafiu 2013:59; Bertinetto and Squartini 2016:942; Maiden 2016a:109). Consequently, as illustrated in (14), the periphrastic forms mark pre-hodiernal values, whereas the synthetic forms convey hodiernal readings.

(14) jeri̯ dimiˈne̯atsə m am skuˈlat m am
yesterday morning me=have.PRS.1SG raise.PTCP me= have.PRS.1SG
ɨmbrəˈkat m am spəˈla ʃɨ jo. […] tot aˈʃa
dress.PTCP me=have.PRS.1SG wash.PTCP and I also thus
fəˈkui̯ ʃɨ̯ ˈazi̯ dimiˈne̯atsə. mə skuˈlai̯ […] ˈdupə kum

do.PFV.1SG and today morning me= raise.PFV.1SG after as

spuˈseɨ̯ […] mulˈseɨ̯ ˈvaka […] fəˈkuɨ̯ miŋˈkare. (Câlnic)

say.PFV.1SG milk.PFV.1SG cow.DEF do.PFV.1SG food

'Yesterday morning I too got up, got dressed, washed. […] I've done the same this morning too. I've got up […] as I've said […] I've milked the cow […] I've prepared some food.'

Evidence like this supports the view that the mapping between function and form, be it synthetic or periphrastic (*pace* Corbett 2007), is often only weakly correlated and at times even entirely opaque as in the Oltenian case. Yet, at the same time it would be disingenuous not to recognize that periphrases often do retain something, if not all, of their original compositional semantics, even in later stages of advanced grammaticalization (Vincent 2011; 2015). It is for this reason that, in contrast to Ackerman and Stump (2004:142), Haspelmath (2000:661), Brown *et al.* (2012:§5.3), and Spencer and Popova (2015:211f.) reject the idea that non-compositionality is a sufficient, if not a necessary, condition in defining periphrasis. Rather, we must recognize that morphosyntactic non-compositionality is a possible, though not necessary, by-product of periphrasis which often emerges with the shifting balance between syntactic and morphological properties that accompanies the increasing morphologization and concomitant opacity of multi-word expressions as they are integrated into the inflexional paradigm (cf. discussion of gradience in §1.2.2.5 below).

Finally, we return again to the Romanian periphrastic perfective forms reviewed above. Even under its present perfect reading, *a servit* and the other periphrastic paradigms of the perfect in Table 1.3 are arguably not strictly compositional. In our

discussion so far, we have deliberately overlooked the role of the participle in the various periphrastic formations: despite what our compositional analysis would suggest, there is nothing inherently active about the participle. Rather, the Romance participle is essentially neutral with respect to voice (for extensive diachronic discussion, see Vincent 2011:§19.3.2; cf. also Taylor 2011:406f.), as evidenced by its use in the formation of the passive across Romance, including in Romanian (15).

(15) Ion a / este servit. (Ro.)
Ion have.PRS.IND.3SG be.PRS.IND.3SG serve.PTCP.MSG
'Ion has/is served.'

Given such active/passive contrasts in the interpretation of the participle, Vincent (1982; 2011:429f.) convincingly draws on the origins of the two constructions to argue that the voice value of the participle can be straightforwardly derived from the choice of auxiliary, while still retaining the compositionality of the two readings. In particular, selection of HAVE specifically licenses the active value which, as the output of a Panini Principle, allows the participle to retain its original default passive value in all other contexts, whether employed in isolation (e.g., *un client odată servit…* 'a customer once (having been) served*,…*') or in conjunction with BE or one of a series of other possible passive auxiliaries (see Ledgeway 2021:§3). While this one-to-one transparent mapping between the choice of auxiliary and the voice value of the participle is met in most Romance varieties, the Romanian evidence warns us against assuming such isomorphism as a given. With the exception of the present perfect (cf. 15), the other periphrastic forms of the perfect in Table 1.3 all employ the auxiliary BE, viz. *fi*, in accordance with a distribution of the two auxiliaries arranged along

parameters of finiteness (Ledgeway 2014). This immediately raises a problem for Vincent's account, or at the very least forces us to recognize the emergence of non-compositionality as a result of some quite spectacular readjustments in the distribution of the auxiliaries (Ledgeway 2019:348-354), since the value of the participle cannot be inferred from the choice of auxiliary (cf. further discussion §1.2.3). Indeed, the perfective forms with BE in Table 1.3 are genuinely ambiguous in that, alongside their active readings (16a), they also readily license transparently passive readings, at least when the subject is masculine singular (16b).[15]

(16) Ion va fi / ar fi servit // Înainte de a fi servit… (Ro.)
Ion FUT.3SG be COND.3SG be serve.PTCP.MSG before of to be serve.PTCP
a. 'Ion will/would have served // Before having served…'
b. 'Ion will/would be served // Before being served…'

The result is an asymmetrical distribution of non-compositional and compositional readings of the active and passive, respectively where, not accidentally, the non-

[15] The participle displays the default masculine singular form in the active, but not in the passive where it agrees with the surface subject for number and gender. Furthermore, there is no ambiguity in the subjunctive where, in contrast to the invariant form of auxiliary *fi* 'be' employed in the active, the passive requires an inflected form of auxiliary FI 'be'. Both of these facts are illustrated in (i.a-b).

(i) a. Vor să fi servit Ioana. (Ro.)
want.PRS.IND.3PL that.IRR be serve.PTCP.MSG Ioana
'They want Ioana to have served.'
b. Vor să fie servită Ioana. (Ro.)
want.PRS.IND.3PL that.IRR be.PRS.SBJV.3SG serve.PTCP.FSG Ioana
'They want Ioana to be served.'

compositional reading arises precisely in the case of the active where we know that there have been some quite exceptional changes in the distribution of the Romanian perfective auxiliaries (Ledgeway 2014; 2019), unlike in the passive whose morphosyntax has essentially remained unchanged. This underlines how in large part the emergence of non-compositionality and concomitant opacity is intimately connected to questions of diachrony (cf. Vincent 2015) and, in particular, to the extent to which the relevant periphrasis has been integrated into the inflexional paradigm (cf. §1.3.2). While DM might be able to conveniently circumvent the problem of the contradictory values of the participle, presumably in terms of underspecification for the feature [voice] with the 'passive' value treated as the elsewhere form (cf. the treatment of aspect in the Latin participle in -URUS in Embick and Halle 2005), there is no natural way to account for the observed correlations between diachrony and non-compositionality.

1.2.2.3 Distributed exponence[16]

Further to intersectivity and non-compositionality, the third defining property for periphrasis proposed by Ackerman and Stump (2004:144f.) is distributed exponence, which arises in cases where different elements of the component parts of the periphrasis are individually responsible for realizing different parts of its content (cf. also §2.4.4). A particularly revealing example comes from the Ariellese pluperfect (D'Alessandro and Ledgeway 2010), an Italian dialect of eastern Abruzzo, where alongside the expected periphrastic formation consisting of a past form of the

[16] Of course, the use of 'distributed' here in the sense of a non-continuous morphological realization is not to be confused or equated with the use of the term in the name of the framework 'Distributed Morphology'.

auxiliary in conjunction with the participle (17a), we also find a compound periphrasis (see §1.2.2.4) involving the use of two finite auxiliaries in conjunction with the past participle (17b).

(17) a. Avive parlate. (Arielli, EAbr.)
have.PST.IPFV.2SG speak.PTCP

b. Si' 'vé parlate. (Arielli, EAabr.)
be.PRS.2SG have.PST.IPFV speak.PTCP
'You had spoken'

In (17b) the higher auxiliary references the features of the subject, also in part through a classic person-based split identical to that found in the present perfect (viz. 1/2 ⇒ BE vs 3 ⇒ HAVE; cf. D'Alessandro and Roberts 2010; see also the chapters by Bach and Štichauer and by Loporcaro,), whereas the lower auxiliary surfaces as a past tense form of HAVE, namely *'vé*, a reduced and undifferentiated syncretic form of the singular persons (< *avé/avive/avè* < HABEBA-M/-S/-T; cf. Verratti 1998:71, 73), which constitutes the sole exponent of imperfective past tense in the construction (arguably together with singular number). Consequently, Ariellese presents an interesting case of feature-spreading in the discontinuous expression of agreement and tense features across two distinct auxiliaries, in contrast to its non-compound variant in (17a) where both are realized syncretically on a single auxiliary. In this way, the contrast between (17a-b) highlights what probably lies at the heart of Ackerman and Stump's original observation, namely that the features which are distributed in this way should typically involve those which are otherwise expected (in a given language) to be realized in a monolectal expression (Spencer and Popova 2015:214). This leads us to

conclude following Vincent (2011:426) that distributed exponence cannot be considered 'even a mild signal of a possible periphrasis, leave alone a sufficient condition' (Brown *et al.* 2012:263), but simply reflects deviation from a language's syntactic norms (Spencer and Popova 2015:230). Indeed, from a realizational perspective the occurrence of distributed exponence would *a priori* seem to favour a more transparent and discrete mapping between meaning and forms, whereas we have seen in §1.2.2.2 that a strong hallmark of periphrasis is growing non-compositionality.

Irrespective of the status of distributed exponence, it strikes us that this might be an area where frameworks such as DM are particularly well placed to model the relevant facts since it has a powerful post-lexical machinery which can readily manipulate the distribution and (multiple) exponence of the relevant features. In particular, while examples such as (17a) arguably involve the PF spell-out of bundled features brought together in the narrow syntax through head-movement of the auxiliary through Tense and Agr heads (cf. footnote 6 above), (17b) is to be interpreted as the output of a post-syntactic process of 'Fission' (Bobaljik 2017:§3.4; Siddiqi 2019:148) which splits a single node and its features into two or more nodes in the morphological representation (for an analysis along these lines, see D'Alessandro and Ledgeway 2010). In principle, distributed exponence is not incompatible with WP approaches either which can otherwise readily handle morphological examples of distributed and/or extended exponence such as Pt. *pãozinho* 'roll' (< *pão* 'bread' + DIM *-zinho*), which in the plural is marked not only through the desinence of the diminutive, but also in the nasal vowel alternation on the nominal stem, namely, *pã**e**-zinho-**s*** 'bread\PL-DIM.M-PL'. By allowing one or more morphosyntactic properties of a given cell to be expressed by realization rules that may apply in more than one block (Stump 2018a:65; 2018b:289), a framework such as PFM is able to handle both

distributed and extended exponence and, by implication, similar cases involving periphrastic cells of the paradigm.

1.2.2.4 Compound periphrases

A common, though relatively understudied, aspect of periphrasis is compound periphrasis.[17] By way of example, in (18) the auxiliary of a given verbal periphrasis, the passive in (18a) and the progressive in (18b), intersects, in turn, with the grammatical feature perfect, which itself is realized periphrastically, namely *era* (> *era estat) aidat* and *so* (> *so istatu*) *travallande*.

(18) a. Qu' era estat aidat per un
that be.PST.IPFV.IND.3SG be.PTCP.MSG help.PTCP.SG by a
amic. (Gsc., Darrgrand 1974:117)
friend
'He had been helped by a friend.'

b. So istatu travallande. (Srd., Jones 1993:141)
be.PRS.IND.1SG be.PTCP work.PTCP
'I've been working.'

The embedding of one periphrasis under another in this way presents challenges for standard definitions of periphrasis and, in particular, for the question of how the boundary between syntax and morphology is to be drawn. On the one hand, recursion of the type that we see here is canonically taken to be a hallmark of syntax, hence

[17] See, however, the extensive discussions of the so-called *temps surcomposés* ('double compound tenses') in Vincent (2011:430-432; 2014:12f.; 2015:111f.).

offering us a simple way-out if we go down the avenue of a DM-style analysis where morphology is ultimately syntax all the way down. On the other, we have adduced incontrovertible evidence above for the view that periphrasis should be understood in terms of a paradigmatic convergence between syntax and morphology. This therefore requires us to accommodate compound periphrases within a theory of inflexional morphology, ultimately to be interpreted as the result of the cross-categorization of morphosyntactic and morphosemantic features (Brown *et al.* 2012:269f.).

Essentially, this amounts to treating compound periphrasis, not as simplex periphrasis embedded within syntax, but, rather, as complex multi-word expressions which realize two or more intersective featural relationships within a given cell of the paradigm. This leads us to expect compound periphrasis to unite in greater or lesser measure aspects of syntactic and morphological properties so as to display exactly the same characteristics (e.g., intersectivity, non-compositionality, distributed exponence) that we have identified above for cases of simplex periphrasis. *A priori* there is no reason to believe that this is not the case. For example, we have already observed how the examples in (18a-b) combine a primary intersection of features, e.g., passive and third person singular or progressive and first person singular, with a subsequent intersective relation with the feature perfect. At the same time, both examples display distributed and extended exponence with, for example, the person features of the subject encoded on the finite auxiliary and its gender specification on the participial auxiliary and, in the case of (18b), also on the lexical participle. Furthermore, although the result in (18a-b) is arguably compositional, there are many examples of compound periphrasis where this is not the case, witness the Romanian contrast in (19a-b).

(19) Ion a fost servit. (Ro.)

Ion have.PRS.IND.1SG be.PTCP.MSG serve.PTCP.MSG

a. 'Ion has been served.'

b. ?'Ion had served.'

Demonstrably, the structure in (19) transparently delivers a present perfective passive interpretation, as indicated by its translation in (19a). However, for many speakers in a large dialectal area (including North Moldavia, Maramureș, Crișana, Transylvania, Banat, and Muntenia), as well as in earlier stages of the written language, the structure in (19) is also compatible with the substandard active pluperfect reading given in (19b).[18] While the passive variant in (19a) licenses a one-to-one form-function compositional reading, the same cannot be said of the variant in (19b) where the same structural template with the exact same lexicalizations yields an active non-compositional interpretation which cannot be directly read off the sum of its component parts, exactly along the lines of what we observed above in (16a-b).

1.2.2.5 Gradience

If periphrasis is to be modelled in terms of structured paradigms, as maintained here, then a reasonable expectation is that periphrastic expressions should share other typical properties of inflexional morphology, as well as syntax. Indeed, in the canonical case periphrasis 'is canonical syntax but within the paradigmatic organization of canonical morphology', such that 'some properties of periphrastic

[18] For further details, see Densusianu ([1938] 1961:144), Nevaci and Todi (2009142), Dragomirescu (2012:208f.), Pană Dindelegan (2012:568), Zafiu (2013:34f.), Ledgeway (2014:20f.), and Dragomirescu, Nicolae, and Zafiu, this volume.

constructions follow from their syntactic nature and some from their morphological nature' (cf. Brown *et al.* 2012:245). Rarely if ever, however, are actual cases so canonical, with individual examples exhibiting varying degrees of morphological and syntactic interdependence, the balance of which is often subject to change over time (cf. Vincent 2011:424), especially as original multi-word expressions are increasingly integrated within the inflexional paradigm as their grammatical ingredients progressively move down the word > clitic > affix cline (Hopper and Traugott 2003:7), a typical effect of grammaticalization (see §1.3.1). By way of example, consider the contrasts in (20)-(21).

(20) a. Son benguts / An barradas las
be.PRS.IND.3PL come.PTCP.MPL have.PRS.IND.3PL close.PTCP.FPL the.FPL
fenestras. (Périgord Occ., Miremont 1976:54f.)
windows.F

b. Han venido / Han cerrado las
have.PRS.IND.3PL come.PTCP.MSG have.PRS.IND.3PL close.PTCP.MSG the.FPL
ventanas. (Sp.)
windows.F
'They have come / They have closed the windows.'

(21) a. A trôp travalhat. (Périgord Occ., Miremont 1976:147)
have.PRS.IND.3SG too.much work.PTCP.MSG

b. Ha (**demasiado) trabajado demasiado. (Sp.)
have.PRS.IND.3SG too.much work.PTCP.MSG too.much
'S/he has been working too much.'

All Romance varieties show, with varying degrees of grammaticalization and different shades of meaning, a perfective periphrasis built on the auxiliary BE and/or HAVE in conjunction with the participle, a construction which ultimately has its roots in Latin (for in-depth discussion and bibliography, see Ledgeway 2012:130-134). As the example in (20a) illustrates, Occitan, together with French and many Italo-Romance varieties including Italian, shows a rather conservative behaviour, in that the periphrasis is still transparently organized around an original active-stative alignment in which the syntax is sensitive to the semantic nature of the available arguments (Ledgeway 2012:319-326). Thus, the auxiliary alternates in accordance with the semantic nature of the surface subject, namely, $A/S_A \Rightarrow$ HAVE vs $S_O \Rightarrow$ BE (Vincent 1982; Loporcaro 2016:802f., 812f.; Ledgeway 2019:364-366), and the participle continues to agree with (various subclasses of) stative arguments, viz. O or S_O (Smith 1991; 1996; 1999; Loporcaro 1998:64-78; 2016:§49.2.3; Manzini and Savoia 2005, II:§5.1; Ledgeway 2012:3326f.). The Spanish example (20b), by contrast, which is representative of Ibero-Romance as a whole as well as modern Romanian, displays a more innovative behaviour in line with a core nominative-accusative alignment, such that auxiliary alternation has given way to the generalization of HAVE and participle agreement has disappeared entirely. We thus observe in the contrast (20a-b) a progressive desyntacticization of the perfective construction in a process of increasing morphologization (Joseph and Janda 1988; Joseph 2008). The latter strips away the former alternations of the syntax proper in favour of non-alternating forms in line with the growing cohesion of the composite parts of the periphrasis, as it moves along a cline from periphrasis to inflexion and the borders between the two become increasingly fuzzy. This process of univerbation manifested in the increased

integration between the individual elements of the periphrasis is further visible in the contrast between (21a) and (21b), where the adjacency of auxiliary and participle may still be interrupted in Occitan by the interpolation of elements such as adverbs, but not in Spanish where the effects of desyntacticization have been greater (cf. also the discussion in §2.4.5 in Smith, this volume). Ultimately, in the latter case the fusion between an albeit still fully inflected auxiliary and participle gives rise to a distribution and behaviour of the periphrasis more akin in many respects to monolectal expressions.

Similar developments in the shifting balance between syntactic and morphological properties, arguably reaching a more advanced stage of desyntacticization and morphologization, are also observable in the progressive periphrasis of many dialects of Apulia. In these varieties the progressive originally involves a pseudo-coordinate structure in which a reflex of STARE 'stand' is coordinated by means of the conjunction AC 'and' (itself often subject to surface deletion, though still signalled by initial consonantal lengthening of the following word) with a finite lexical predicate, as illustrated in Table 1.4.

Table 1.4: Aspectual pseudo-coordinate periphrases with STARE 'stand' in Apulia

	(a) Putignano 'do'	(b) Martina Franca 'call'	(c) Lecce 'lose'
1SG	stok a fˈfattsə	stɔ cˈcɛmə	sta pˈpɛrdu
2SG	ste fˈfaʃə	stɛ cˈcɛmə	sta pˈpɛrdi
3SG	ste fˈfaʃə	stɛ cˈcɛmə	sta pˈpɛrdɛ
1PL	sta ffaˈʃeimə	stɛ ccaˈmɛmə	sta ppɛrˈdimu
2PL	sta ffaˈʃeitə	stɛ ccaˈmɛtə	sta pˈpɛrˈditi
3PL	ston a fˈfaʃənə	stɔnə (a) cˈcamənə	sta pˈpɛrdɛnu

As shown in Ledgeway (2016a) and Andriani (2017:ch. 5), the forms of STARE exhibit varying degrees of inflexional attrition (cf. also Manzini and Savoia 2005, I:688-701). In particular, it is possible to identify three patterns which display increasing degrees

of inflexional attrition as we move southwards across the region from Pugliese to Salentino dialects. For instance, in the Pugliese dialect of Putignano (province of Bari) in (a) STARE displays a near full inflexional paradigm (albeit with syncretisms in the second/third person singular and first/second person plural), whereas in the more southerly Pugliese dialect of Martina Franca (province of Taranto) in (b) the paradigm shows reduced agreement with distinctive forms, based on the stem stɔ(-), in just the first person singular and the third person plural according to a morphomic U-pattern widespread across Romance (cf. Maiden 2011a; 2016b:§43.2.3). More drastic are Salentino dialects such as Leccese in (c) where STARE has lost all agreement and is now reduced to an invariable aspect marker occurring as a bare stem, a clear counterexample to Spencer and Popova's (2015:208) claim that periphrasis is not expected to produce uninflectability (cf. also discussion in §1.2.3 below).[19]

These synchronic cross-dialectal patterns of inflexional attrition may also be read vertically as evidence to reconstruct different diachronic stages in the progressive desyntacticization of the periphrasis. In particular, this morphological reflex highlights an important change in the nature of STARE, inasmuch as a full inflexional paradigm constitutes the hallmark of what it means to belong to the lexical category of verb in Romance. Consequently, any attrition in the inflexional paradigms of STARE can be taken to represent a weakening in its defining verbal characteristics and, at the same time, to signal a concomitant change in its category from auxiliary verb to functional prefixal marker. Indeed, the prefixal nature of STARE in the final stages of this development is clearly revealed by the fact that, while in some dialects object

[19] Another example of Romance uninflectability already seen is the Romanian invariant BE stem *fi* used for all persons in the active perfect subjunctive (cf. Table 1.3), in contrast to the inflected forms found in the present passive (see footnote 15 above for examples).

clitic pronouns may still surface between invariant *sta* and the finite lexical verb (e.g., Scorrano *sta* ***te*** *visciu* '*sta*.PROG you= I.see (= I can see you)'), in more progressive dialects clitics must precede *sta* which cannot be separated from the finite verb by any intervening material (e.g., Matino ***ci*** *sta penzu* 'to.it= *sta*.PROG I.think (= I'm thinking about it)').

Examples like these, together with those in (20)-(21), serve to remind us that periphrasis is not an immutable, absolute category. Rather, it must be conceived of as a fluid label that can be equally applied to a range of multi-word expressions showing varying degrees of greater or lesser core syntactic and morphological properties, a view championed in approaches such as Canonical Typology (Brown, Chumakina, and Corbett 2013). Given this gradient interpretation of periphrasis (cf. similar remarks about analyticity in Vincent 1997a:100; Ledgeway 2013:§2.2.2; 2017a:§3.2), we can readily accommodate both synchronic variation across a range of unrelated examples of periphrasis, as well as diachronic variation in related examples of periphrasis within the same language (family). Ultimately, changes in the balance between syntactic and morphological properties in the compositional make-up of individual periphrases, whereby we note over time a tendency towards desyntacticization and concomitant growing morphologization, are to be understood as a direct effect of paradigm membership. As members of a paradigm in which they realize one or more cells, the relevant periphrases enter into direct opposition with synthetic inflexional forms with which they come to be inextricably bound both formally and functionally. *A priori* growing morphologization is therefore expected as a natural consequence of this paradigmatic interdependence which aligns originally syntactic elements alongside synthetic inflexional elements within the domain of morphology. By contrast, in configuration-based lexical approaches to morphology

such as DM the relationship between periphrases and their related inflexional forms is not formalized by way of a paradigm, such that any such relationships are purely accidental, as are observed diachronic tendencies towards desyntacticization and concomitant morphologization which can only be stipulated, but not captured naturally.

1.2.3 Paradigmaticity

Much of the Romance data that we have discussed so far in relation to periphrasis unquestionably point to the need to make reference to a set of paradigmatically related forms, a conclusion which is out of the reach of lexical-realization theories like DM. A natural consequence of placing periphrasis firmly within the realm of inflexional morphology in this way is that it leads us to expect periphrasis to display many of the recurrent features characteristic of morphology and, in particular, inflexional paradigms. *Pace* Spencer and Popova (2015:208), this prediction, we contend, is in large part confirmed by the Romance data (cf. also Corbett 2013:180-186), a brief survey of which follows below.

We begin with Spencer and Popova's (2015:208) claim that it is difficult to imagine cases of periphrastic defectivity and uninflectability on a par with the equivalents of defective and uninflectable lexemes, since the exponence of periphrasis is through syntax which, unlike morphology, is held to be less prone to such irregularities (cf. also Kiparsky 2005:§2.2). However, we have already seen in relation to the Salentino progressive, an erstwhile pseudo-coordination structure (cf. Table 1.4), how grammaticalization of the construction over time leads to differing degrees of inflexional attrition and erosion of the auxiliary STARE 'stand', ultimately

resulting in its uninflectability in varieties such as Leccese where it has been reduced to an invariable preverbal affixal marker of aspect (viz. *sta*).[20]

These same varieties simultaneously exemplify defectivity – or 'underexhaustivity' in the words of Spencer (2006:291; cf. also Vincent 1987:241; Haspelmath 2000:661; Baerman 2015:§7.4) – since, as shown in detail by Andriani (2017:ch.5), many dialects in the Pugliese-Salentino continuum show defective paratactic paradigms like those exemplified in Table 1.5 where we observe arbitrary feature restrictions on the distribution of the paratactic pseudo-coordinate structure. More specifically, the shaded cells indicate the distribution of the paratactic construction and the unshaded cells the paradigmatic gaps, which are realized instead by a hypotactic construction consisting of fully inflected STARE + *a* 'to' + infinitive.

Table 1.5: Defective pseudo-coordinate periphrases with STARE 'stand' in Pugliese (Andriani 2017:ch.5)

	(a) Gravina 'play'	(b) Rutigliano	(c) Conversano
1SG	stóuc'a ssuné	stɔŋg a ʃʃɔ:uk	stéc'a ssonə
2SG	sté ssunə	ste ʃʃuk	st'a sunə
3SG	sté ssóunə	ste ʃʃɔ:uk	st'a sónə
1PL	stəm'a ssuné	stəm a ʃə'kwe	stém'a ssənə́
2PL	stət'a ssuné	stətə a ʃə'kwe	stét'a ssənə́
3PL	stònn'a ssuné	stɔnn a ʃə'kwe	stànn'a ssònənə

[20] Also relevant here are aspectual periphrases of the TORNARE type discussed by Parry, this volume, which in some Romance varieties have also been reduced to invariable aspectual verbal markers (cf. Cos. *torna* 'again', Sal. *ntorna* 'again'). In the case of Salentino, we can even find sequences such as (i) in which both invariant preverbal markers may co-occur.

(i) Ntorna sta trona! (Matino, Fiorentino 1998:53)
again PROG thunder.PRS.3SG
'It's thundering again!'

In addition to illustrating periphrastic defectivity, the paradigms in Table 1.5 also highlight three further typical properties of inflexional morphology within a periphrastic structure. The first is suppletion (cf. Vincent 1987:242; Anderson 2011:§3.2), in that the three paradigms in Table 1.5 variously show gaps in the distribution of the paratactic construction in specific persons which are filled suppletively by the hypotactic structure.[21] This blended distribution between paratactic and hypotactic periphrastic forms follows arbitrary restrictions on feature realization characteristic of morphological paradigms but not, significantly, of syntactic constructions. The second property evidenced by the paradigms in Table 1.5 regards the extension of the paratactic paradigm according to a morphomic pattern (see also the chapters by Cruschina and by Vincent and Wheeler, this volume). In particular, we see that in its earliest stages (cf. Gravina) it involves just the second/third persons singular, but can subsequently extend to the first person singular (cf. Rutigliano) to yield a singular/plural contrast within the paradigm – itself a frequent split in inflexional paradigms –, from where it can then be further extended to include also the third person singular (cf. Conversano), a classic example of Maiden's N-pattern (Maiden 2016b:712-716), before finally extending to the whole paradigm (cf. Table 1.4). The third and final significant property of inflexion evinced by the paradigms in Table 1.5 is heteroclisis. Given traditional definitions of heteroclisis (cf. Stump 2006; 2015:§6.4.2.3) such as 'the property of a lexeme whose inflectional paradigm contains forms built upon stems belonging to two or more

[21] For another interesting case of periphrastic suppletion, see the discussion of the distribution of the Italo-Romance passive auxiliaries ESSERE 'be' and VENIRE 'come' in Vincent (1987:248f.) and Ledgeway (2021:§3.3).

distinct inflection classes' (Stump 2006:279), it is tempting to view the blending of paratactic and hypotactic forms above as part of a heteroclitic distribution.

Another frequent property of inflexional paradigms and, in particular, declensional classes, is syncretism (cf. Baerman 2015:§7.2). If it can therefore be shown to also occur in periphrases, this will provide additional proof of the paradigmatic nature of periphrasis (Vincent 1987:241f.). Above we have already seen several examples, but perhaps the most spectacular involves examples such as (19a-b), repeated here in (22a-b).

(22) Ion a fost servit. (Ro.)
Ion have.PRS.IND.1SG be.PTCP.MSG serve.PTCP.MSG
a. 'Ion has been served.'
b. ?'Ion had served.'

It is striking to note that this Romanian example comes very close to the other frequent example of periphrastic syncretism discussed in the literature (cf. 23), namely the active-passive syncretism found in the progressive of some (obsolete) varieties of English (cf. Anderson 2011:116).

(23) John's house was building.
a. 'John's house was built.'
b. 'John was building his house.'

We turn now to the opposite of syncretism, overabundance (Thornton 2011; 2016; 2019b; Cappellaro 2018; see also Sornciola, this volume). In contrast to Ackerman

and Stump (2004:126f.), Brown *et al.* (2012:245) maintain that we must allow for the possibility that periphrasis and synthesis may compete for the realization of a given cell (cf. also Haspelmath 2000:659; Taylor 2011:405). They cite the example of definiteness marking in Scandinavian where a suffixed form of the noun competes with combinations of a free-standing article + noun, and conditions imposed by the syntax determine which is chosen. Remarkably, this mirrors again what we have already seen for Romance, in this case the alternation between suffixal and free-standing prenominal forms of the definite article in Romanian (§§1.2.1-1.2.1.2) in such alternations as *prieteni-i* 'friends.M-DEF.NOM-ACC.PL' vs *cei trei prieteni* 'the.NOM-ACC.MPL three friends.M-DEF.NOM-ACC.PL'. Unlike in the Scandinavian case, however, in some contexts the alternation is not determined exclusively by conditions of the syntax, but also by distinctions of style and register (cf. also Spencer and Popova 2015:224), witness the alternation between standard *Ane-i* 'Ana.GEN-DAT-DEF.GEN-DAT.FSG' and substandard *?lui Ana* 'the.GEN-DAT.SG Ana.DEF.NOM-ACC.FSG' (for detailed discussion, see also Pană Dindelegan and Uță Bărbulescu, this volume). We can also draw a further distinction here between cases like those just considered (e.g., *-i* vs *cei*), where there is competition between syncretic and periphrastic expressions in the realization of a given paradigmatic cell, and cases where there arises an alternation between two (or more) periphrastic forms competing for the realization of a given paradigmatic cell. An obvious example here is the Catalan periphrastic GO-preterite (cf. Paoli and Wolfe, this volume) where the function word in the periphrasis, viz. auxiliary ANAR 'go', shows multiple competing forms in accordance with complex stylistic and diatopic variation (cf. Wheeler, Yates, and Dols 1999:301f., 307, and the discussion in §3.3 in Vincent and Wheeler, this

volume) including, among others, *vaig/ˀvàreig menjar* ‘GO.AUX.IND.1SG eat.INF (= I ate)’ and *vam/ˀvàrem/ˀvem menjar* ‘GO.AUX.IND.1PL eat.INF (= we ate)’.

Closely related to overabundance is the phenomenon of overexhaustivity which obtains whenever ‘a periphrastic expression gives rise to more combinations than might be expected given the ordinary rules of syntax in a language’ (Spencer 2006:291). A possible example here involves once again the Catalan periphrastic GO-preterite which in the subjunctive enters into competition with the synthetic forms of the past subjunctive, e.g., *vagi menjar* ‘GO.AUX.SBJV.1SG eat.INF’ vs *menjés* ‘eat.PST.SBJV.1SG’. While in the indicative alternation between periphrastic and synthetic preterite forms is purely a question of stylistic and diatopic variation (Wheeler, Yates, and Dols 1999:343), inasmuch as their semantic values and functions overlap exactly (24a-b), in the subjunctive there arises a partial aspectual distinction (25a-b; Wheeler, Yates, and Dols 1999:380f.; and discussion in §3.3 in Vincent and Wheeler, this volume).

(24) a. S’ acostà a la porta i va
self= draw.PFV.3SG to the door and GO.AUX.IND.3SG
trocar-hi dos cops. (Cat.)
knock.INF=there two blows

b. Es va acostar a la porta i hi
self= GO.AUX.IND.3SG draw.INF to the door and there=
trucà dos cops. (Cat.)
knock.PFV.3SG two blows
‘He went up to the door and knocked twice.’

(25) a. És impossible que ho escrivís ella. (Cat.)
be.PRS.IND.3SG impossible that it= write.PST.SBJV.3SG she
'It's impossible that she was writing/wrote it.'

b. És impossible que ho vagi escreure ella. (Cat.)
be.PRS.IND.3SG impossible that it= GO.AUX.SBJV.3SG write.INF she
'It's impossible that she wrote it.'

The synthetic past subjunctive form in (25a) is aspectually ambiguous between the imperfective and perfective readings in accordance with the typical Romance pattern, whereas the periphrastic formation in (25b) introduces a novel perfective distinction which is otherwise neutralized in other Romance varieties. Consequently, from a comparative Romance perspective, though not necessarily from a Catalan-internal one, the periphrastic preterite gives rise to more aspectual combinations than would ordinarily be expected given canonical Romance restrictions on the formal realization of aspect in subjunctive paradigms.

We conclude this section with a brief remark about the possibility of pure periphrastic morphology. Vincent (2011:434) asks the question whether there could be anything such as 'pure periphrastic morphology, that is to say situations where a non-finite form occurs in a range of periphrases which are not linked by common features and indeed may even have contradictory feature values'. He hypothesizes that one possible candidate is the Romance perfect participle which licenses an active reading in the perfect but passive (or perhaps better, stative) reading elsewhere, but he ultimately dismisses this possibility on the assumption that their respective values can be derived from the effects of the different auxiliaries they combine with (cf. also discussion in §1.2.2.2 above). While such arguments hold for many standard

Romance varieties where there is a robust correlation between auxiliary choice and voice value (cf. 20-21), we have already seen above that, with the exception of the present perfect, the sub-paradigms of the Romanian perfective (cf. Table 1.3) and the substandard pluperfect (cf. 16a-b/19a-b) are genuinely ambiguous, at least when the subject is masculine singular, between active and passive readings. Similarly, the correlation between auxiliary choice and the active/passive interpretation of the participle fails to show a one-to-one mapping in many central-southern dialects of Italy (see the chapters by Bach and Štichauer and by Loporcaro, this volume) where, among other things, auxiliary choice in the active can also be determined by mood (±realis), tense (±past), person (typically, though not exclusively, 1/2 vs 3), and/or a combination of these (for discussion, see Ledgeway 2019). By the same token, in some Pugliese and south-eastern Basilicatese dialects the auxiliary of the passive can even be HAVE (26), a rather extraordinary development from a cross-linguistic perspective (Loporcaro 1988:291-299; 2012:179-180; Ledgeway 2021:§3.4).

(26) ləlɔčč e ğğəsɛppə non ávənə paɣeːtə. (Irsina, Loporcaro 2012:179)
Raffaele and Giuseppe not have.PRS.3PL pay.PTCP
'Raffaele and Giuseppe are not (being) paid.' (Loporcaro, 2012, p. 179)

Given these observations, there is then a *a priori* case for taking Vincent's original supposition seriously to make some allowance for some, albeit limited, cases of pure periphrastic morphology.

1.3 Diachrony

Taken together, the case studies in the present volume demonstrate the essential complementarity of synchronic and diachronic perspectives on the description and explanation of linguistic phenomena. And already in this chapter we have seen diachronic and synchronic factors interacting in various ways. In this section, we address the historical side of the question in more detail, focusing on four themes which are especially pertinent to understanding the way periphrasis and inflexion develop and interact over time both in the specific case of Romance and more generally in cross-linguistic terms.

1.3.1 Grammaticalization and directionality

The obvious place to start is with the phenomenon of grammaticalization both because of its centrality in debates over the last four decades and because many of the classic examples are drawn from Romance (see also the discussion in §2.4.9 in Smith, this volume): the emergence of definite articles from distal deictics, future periphrases incorporating a GO verb, progressives built out of the STAND verb plus a gerund, locative prepositions assuming a complementizing function, and the like. At the same time, Romance is also home to developments that appear to go against the standard predictions, as with the Cat. ANAR 'go' + infinitive (see chapters by Vincent and Wheeler and by Paoli and Wolfe, this volume) as the expression of simple past exemplified in (24) and (25) or the Lombard and Piedmontese TENER 'hold, keep' + perfective participle expressing continuative or iterative meaning as in (27) compared to the use of the non-past *-ing* form in its English translation (see Vincent 2014:14f. for discussion and further references):

(27) Ten spendasciaa. (OMil., G.P. Lomazzo 1589)

KEEP.PRS.3SG waste.money.PTCP

'He keeps on wasting money.'

The paradox is that while grammaticalization is a key factor in the development of periphrases such as these and others investigated in this volume, the diachronic dimension does not figure greatly in the literature on this class of constructions. It is, for example, only mentioned in passing in the survey by Spencer and Popova (2015) and not at all in the essays collected in Chumakina and Corbett (2012). We hope that the contributions here will go some way towards building the necessary bridge between the history and structure of these items.

To begin with, it is important to be clear about what we mean by the term 'grammaticalization' since, beside the extensive literature on the topic, there is a recurrent strand of criticism and even denial of its very existence. Thus, Joseph (2011; 2020), while agreeing that there certainly exists a phenomenon that deserves the name grammaticalization, objects to labelling it a process or a mechanism and prefers instead to call it simply the result of independently motivated mechanisms of change. In other words, his definition is static and synchronic, and hence his suggestion that structures of the kind we have been discussing here should be labelled 'more grammatical' and not 'more grammaticalized' (2011:196) when compared with their historical source constructions. We will instead suggest that the Romance data strongly confirm the view that there is a mechanism of grammaticalization with its own distinctive properties, and one for which such a family of closely related languages with its rich attestation over a long time span, provides an ideal research environment. We do, on the other hand, agree with Joseph and others that grammaticalization is not a theory in the usual sense of the term but rather a

phenomenon in need of appropriate theoretical modelling (Vincent and Börjars 2010). Other relevant aspects of the questions Joseph raises, in particular those concerning the role of language contact and reconstruction, will be addressed in §1.3.3 and §1.3.4 below.[22]

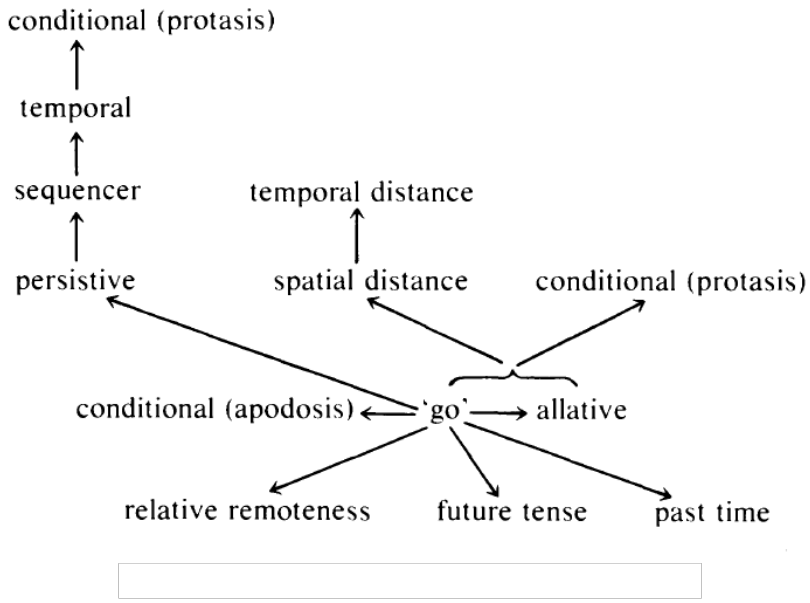
What then are the properties of the grammaticalization process and how do the Romance data help to confirm its validity as a concept? We will start with the issue of the directionality of change, which in turn breaks down into two sub-questions: a) are the directions unidimensional? And b) is there evidence of a reverse type of change, so-called degrammaticalization? For both, the history of the Romance languages provides valuable evidence. To deal with the latter first, what is striking is the absence of any convincing instances within Romance. As Willis (2017) shows, there are inherent reasons why a reverse change from grammatical to lexical, if possible at all, is likely to be a rare event but even so, given the richness of the Romance empirical base, it is somewhat surprising that few if any examples to put beside Bulgarian *nešto* 'something' > 'thing' (pronoun > noun), Pennsylvania German *wotte* 'will' > 'want' (auxiliary > lexical verb) and the like have been highlighted in any of the Romance languages or dialects to date. Negative evidence of this kind cannot be conclusive, but it does tend to confirm the view that degrammaticalization is intrinsically a rare and unlikely diachronic development, and certainly not one which undermines the existence of grammaticalization as a phenomenon in its own right.

[22] One aspect of Joseph's critique that we will not discuss is his concern that too much attention has been devoted to the topic by comparison with other aspects of language change. Fair or otherwise, this is a charge against the scholarly community and the sociology of the discipline but does not impinge on the nature of the phenomenon in and of itself.

Let us return to changes which follow Meillet's classic route from lexical to grammatical or, in Kuryłowicz's alternative formulation, from lexical to less grammatical to more grammatical. Most of the relevant literature concentrates on recurrent single trajectories defined in semantic terms. Thus, the much cited typological study by Bybee *et al.* (1994) contrasts futures derived from agent-oriented modalities such as desire and obligation with those derived from movement verbs and notes that both are well attested in languages and language families all around the world. Perhaps not surprisingly, then, Romance is home to examples of all these sub-types. Romance GO futures have been much discussed and further treatment is to be found in the chapters in this volume by Vincent and Wheeler and by Paoli and Wolfe. A future periphrasis built on a verb meaning 'desire' can be seen in Ro. *voi pleca* 'I will leave' with the auxiliary *voi* < Lat. UOLO 'I want', and one built on an obligation verb occurs in Srd. *des àer cantau* 'you.SG will have sung' with *des* < Lat. DEBES 'you.SG owe, must'. Bybee *et al.* (1994:269) further note sub-patterns such as the tendency for futures derived from movement verbs to express immediate or incipient futurity, as with the English imminent *It's going to rain in a minute* beside the more generally predictive *It will rain.* A similar contrast has been noted for the French future periphrasis with ALLER 'go' + infinitive in contrast to the synthetic future (see Mooney 2020 and references there).

An alternative approach starts not from a grammatical endpoint like futurity but from a particular lexical concept such as verbs meaning GO or COME and tracks the range of possible developments, which can be quite varied. For instance, in Figure 1.1 we reproduce the semantic map compiled by Lichtenberk (1991:499) on the basis of a survey of developments of the GO verb in Oceanic:

Figure 1.1: Grammaticalization of GO forms



In discussing trajectories like this Lichtenberk (1991:476) uses the term 'heterosemy', in his words, 'to refer to cases (within a single language) where two or more meanings or functions that are historically related, in the sense of deriving from the same ultimate source, are borne by reflexes of the common source element that belong in different morphosyntactic categories'. In fact, however, he goes well beyond a single language and demonstrates how the historically connected forms can be found within a whole family of related languages. In similar vein, Craig (1991) uses the term 'polygrammaticalization' to characterize the developments of the GO verb in the Nicaraguan language Rama.

It is instructive to compare the patterns attested in Figure 1.1 with what we find in Romance. In addition to the GO pasts and futures already alluded to, there is the Italian progressive or continuative in (28a) which is close in meaning to the Houaïlou example in (28b) (= Lichtenberk's (34)), which he labels as 'persistive':

(28) a. Gli andavo dicendo che … (It.)
MSG.DAT= GO.IPFV.1SG say.GER COMP

‘I was saying to him that …’

b. Na cuɛ vi. (Houaïlou)

he sit GO

‘He goes on sitting (there).’

The difference is that the Italian (28a) refers simply to an ongoing activity while (28b) also implies the beginning and the continuing of the activity, as indeed does the English GO ON construction. That this distinction is also available in Romance can be seen in the Spanish minimal pair in (29), due originally to Ignacio Bosque and discussed by Bertinetto (2000:577f.):

(29) a. Juan estuvo colocando libros de 3 a 5. (Sp.)

Juan BE.PFV.3SG place.GER book.PL from 3 to 5

‘Juan was placing books (on the shelf) from 3 until 5.’

b. Juan fue colocando libros de 3 a 5. (Sp.)

Juan GO.PFV.3SG place.GER book.PL from 3 to 5

‘Juan kept putting books back (on the shelf) from 3 until 5.’

The story is complicated here by the fact that *fue* in (29b) is the suppletive past of IR ‘go’ as well as being the past of SER ‘be’, but in Bertinetto’s words ‘while (a) simply depicts a durative situation, (b) adds to it the idea that the event be conceived of as a sequence of identical gestures […] To put it differently, (a) describes a static scenario while (b) presents a dynamic one’.

Romance not only confirms the patterns identified by Bybee, Craig, Lichtenberk, and others but also adds to them, as with the Italian passives built on GO + past participle in (30):

(30) a. Questo libro va letto. (It.)
this book GO.PRS.3SG read.PTCP
'This book must be read.'

b. La casa è andata distrutta durante il terremoto. (It.)
DEF house BE.PRS.3SG GO.PTCP destroy.PTCP during DEF earthquake
'The house was destroyed during the earthquake.'

Here too there are sub-developments. In (30a) not only does ANDARE 'go' serve as the passive auxiliary but a modal component of obligation is introduced. In (30b), on the other hand, ANDARE + past participle expresses change of state but with a semantic restriction to negative or undesirable outcomes and hence the ungrammaticality of ***la casa è andata costruita* 'lit. the house has gone built'. The modal component in (30a) is associated with imperfective meaning but is not present in combination with completives as in (30b). Moreover, as Mocciaro (2014) shows, the beginning of these patterns is to be seen in texts from the thirteenth and fourteenth centuries and provides evidence for two different diachronic trajectories. She further shows that, contrary to what is claimed by Giacalone Ramat (2000), these trajectories develop after the GO verb had developed as a passive auxiliary.

Examples like (30a) with a modal interpretation are not attested before the sixteenth century, while the plain passive use is older as in (31):

(31) a. andatagli la elezione confirmata dal
GO.PTCP.FSG=MSG.DAT DEF.FSG election.FSG confirm.PTCP.FSG by.DEF
papa (OIt., Sacchetti *Trecentonovelle*, 14th c.)
pope
'his election having been confirmed by the Pope'

b. se il detto notaio […] andasse electo
if DEF.MSG said.MSG lawyer.MSG GO.PST.SBJV.3SG elect.PTCP.MSG
o vero nominato (OIt., Statutes, Siena, 13th c.)
or indeed nominate.PTCP.MSG
'if the said lawyer was elected or indeed nominated'

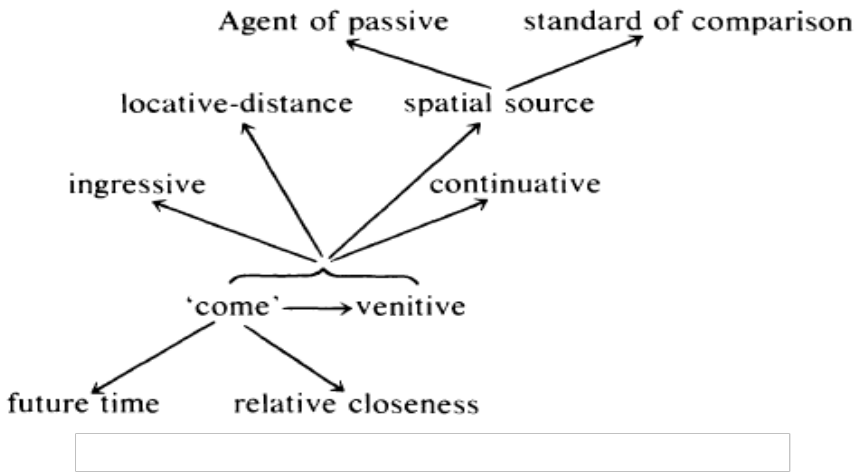
c. va più velata che nulla scienza (OIt., Dante,
GO.PRS.3SG more veil.PTCP.FSG COMP no science
Convivio 1304-7)
'it is more veiled than any other science'

We see here the co-occurrence of ANDARE in a passive with a wider range of lexical items than is found in the modern language. Mocciaro (2014:64) interprets the (31a) type as 'a focus shift' from 'result reached through a path' to 'movement towards a result which is not yet reached and, for metonymical contiguity, should be reached', with the modal component now an integral part of the construction in the modern language. The (31b) development, by contrast, is part of a wider cross-linguistic pattern whereby the GO verb participates in idiom patterns with seemingly arbitrary semantic restrictions. Compare English where one can say *I told her the news and she went wild/nuts/quiet/silent*, but not ***she went noisy/angry/happy,* etc. Not only do examples like these show how Romance can contribute to the wider typological

picture, they also demonstrate how such changes can be tracked over time in a way that is not possible for languages lacking a rich historical documentary tradition.

Lichtenberk offers a similar though less diversified profile for COME verbs, as illustrated in Figure 1.2:

Figure 1.2: Grammaticalizations of COME forms



And here once again Romance both duplicates and complements the patterns he describes. In particular, we find the COME verb evolving as a passive auxiliary in both Italian (32a) (Giacalone Ramat and Sansò 2014) and Romanian (34b) (Dragomirescu and Nicolae 2014):

(32) a. La questione veniva discussa ieri. (It.)
DEF.FSG matter.FSG COME.IMPF.3SG discuss.PTCP.FSG yesterday
'The matter was being discussed yesterday.'

b. Casa aceea vine aşezată aici. (Ro.)
house.DEF that.FSG COME.PRS.3SG place.PTCP.FSG here
'The house is/will be/should be placed here.'

In Italian, the VENIRE ‘come’ + PTCP construction expresses an event and not a state, in contrast to the BE passive in many languages, including Italian, which typically has both values. However, unlike the BE passive, it is restricted in its distribution to the inflexional but not the periphrastic forms hence the contrast between (33a) and (33b):

(33) a. La questione è stata discussa ieri. (It.)
DEF.FSG matter.FSG BE.PRS.3SG BE.PTCP.FSG discuss.PTCP.FSG yesterday
‘The matter was discussed yesterday.’

b.**La questione è venuta discussa
DEF.FSG matter.FSG BE.PRS.3SG COME.PTCP.FSG discuss.PTCP.FSG
ieri. (It.)
yesterday

The sequence *è venuta* by itself is perfectly grammatical with the meaning ‘She has come’ so the ungrammaticality of (33b) involves not the status of the component parts but rather their combinability.

A different development of the COME verb, corresponding to Lichtenberk’s ‘relative closeness’ and defined as ‘an aspectual auxiliary of immediate anteriority’ by Bres and Labeau (2015), is seen in a French example like (34) (= Bres and Labeau’s (1a)):

(33) Mon copain vient de rompre … (Fr.)
my boyfriend COME.PRS.3SG DE break.INF
‘My boyfriend has just broken up with me.’

In this case too the combination with the periphrastic perfect fails and hence the ungrammaticality of (34b) in contrast to (34a):

(34) a. Il vient de pleuvoir. (Fr.)
it COME.PRS.3SG DE rain.INF
'It just rained.'

b.**Il est venu de pleuvoir. (Fr.)
it BE.PRS.3SG COME.PTCP DE rain.INF

c.**Il vint de pleuvoir.
it COME.PFV.3SG DE rain.INF

The difference here is that the simple past as in (34c) is also not acceptable, and the explanation for the ungrammaticality of both (34b) and (34c), as Bres and Labeau (2015) demonstrate, does not in this instance concern the combination of forms but rather the combination of meanings. Simple past defines a moment in time rather than an ongoing state and therefore does not permit the aspectual sub-domain defined by VENIR DE to intrude. And once again, thanks to the richness of the historical documentation, they can show how this aspectual meaning has developed by stages from the motion sense, which was the only one available prior to the twelfth century, with the aspectual use not becoming predominant until the sixteenth century.

Romanian is different again. We have already seen in (32b) that VENI 'come' can occur as a passive marker, but the multiple glosses there demonstrate that an additional modal meaning component is also a possibility. The details of the various available interpretations are too complex to enter into here, but Dragorimescu and

Nicolae (2014:93) provide a neat summary: '[…] in contrast with the regular 'be'-passive, which is static, the *veni*-passive is dynamic; in contrast with the reflexive passive, the *veni*-passive contributes a stronger deontic or iterative value […] in the imperfect the passive auxiliary *veni* yields a habitual, dispositional reading'. What, however, VENI shares with the BE verb is the ability to act as a copula, as in (35):

(35) Ion îmi vine cumnat. (Ro.)
Ion 1SG.DAT= COME.PRS.3SG brother-in-law
'Ion is my brother-in-law.'

If we ask, how such complexity and variety in the development of a single lexical item comes about, the answer once again is that these various uses are natural outcomes of the mechanisms of semantic change including the passage from concrete to abstract, as detectable in the changing choice over time of accompanying verbs, and the bleaching of the directional component of meaning.

To complete the picture for reflexes of Lat. UENIRE, mention needs to be made of its use with the gerund to express progressive or continuative meaning, in this respect paralleling what we have already seen with GO in examples like the Italian (28a) and Spanish (29b). Thus for the modern languages we have:

(36) a. Gianni veniva scrivendo articoli. (It.)
Gianni COME.PST.IPFV.3SG write.GER article.PL
'Gianni kept writing articles.'

b. Hace años que viene diciendo lo mismo. (Sp.)
make.PRS.3SG year.PL COMP COME.PRS.3SG say.GER DEF same

‘He’s been saying the same thing for years.’

And once again there is historical evidence to show the emerging pattern – as demonstrated for instance, for Italian in the recent compilation by Colella (2020), from which the examples in (40) are drawn:

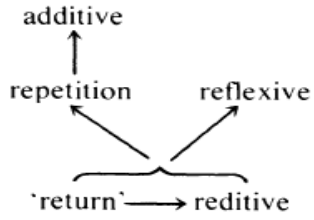
(37) a. quando il sole si viene tramontando (14th-c. OFlo., *La Metaura di Aristotile*)
when DEF.MSG sun REFL= COME.PRS.3SG set.GER
‘when the sun was setting’

b. sarà venuto augmentando lo calore a passo a passo (OIt., Restoro d’Arezzo, *Composizione,* 1282)
BE.FUT.3SG COME.PTCP increase.GER DEF heat to step to step
‘the heat will have been gradually increasing’

(37b) is particularly striking since it involves a periphrastic future perfect of VENIRE combined with a gerund, a combination which is no longer acceptable in the modern language. There are also cross-linguistic generalizations insofar as the periphrases with the COME verbs are less frequent than those with the GO verbs, co-occur with a narrower range of lexical verbs and commonly preserve a degree of deictic orientation (Bertinetto and Squartini 2016:949f.). In short, grammaticalization starts earlier and moves faster and farther with GO than with COME. Once again, then, the Romance data both support and extend Lichtenberk’s generalizations and allow them to be documented over close to a millennium of textual attestation.

Even more restricted is the third profile proposed by Lichtenberk, namely that for reflexes of the RETURN verb, as illustrated in Figure 1.3.

Figure 1.3: Grammaticalizations of RETURN forms



Here too there are Romance parallels as demonstrated in detail in the chapter by Parry in this volume, where she shows how in a number of northern Italian dialects the verb TORNARE 'return' offers synchronically and diachronically documentable evidence confirming all the dimensions and directionality of change identified by Lichtenberk.

1.3.2 Compositionality and univerbation

How, one may ask, do all the examples we have just considered relate to the theoretical questions raised above in §1.2? The general lesson is that they show how forms and meanings can separate over time. More precisely, they help us to address in more detail the question of (non-)compositionality and its relation to periphrasis (cf. also the discussion in §2.4.6 in Smith, this volume). To this end, the principle of compositionality is best taken together with a second principle which, following Langacker (1987:448), we will call 'analysability' and which predicts that complex forms can in general be expected to break down transparently into their constituent parts (Vincent 2015). The possible relations between these two dimensions can be set out as in Table 1.6.

Table 1.6: Dimensions of language structure

	+ analysable	**– analysable**
+ compositional	regular, 'transparent' morphosyntactic structures	some instances of suppletion
– compositional	some compounds and periphrases, idioms, and folk etymologies	arbitrary signs *à la* Saussure

The combination [+analysable, +compositional] serves as a baseline for defining the way words and morphemes are combined in natural languages given the needs of communication and expression over a potentially infinite range of topics and contexts. Change, however, may disrupt such form-meaning transparency in a variety of ways. At one extreme, both analysability and compositionality are lost with the consequence that a new lexical item is created. This is particularly evident in the case of adverbial particles and prepositions such as Fr. *encore* 'still, again' < Lat. *HINC HAC ORA 'hence this hour' or It. *dopo* 'after' < Lat. DE POST 'from after'. Alternatively, analysability may be retained but compositionality lost as we have seen with so-called 'aoristic drift' where perfect periphrases such as Fr. *il est parti* or (northern) It. *è partito* have acquired the simple past meaning 'he left' even though they incorporate the forms *est* and *è* 'BE.PRS.3SG'. Note here, however, that the loss of compositionality is not historically immediate but depends crucially on an intermediate stage in which there is still compositionality but involving a lexical head and a grammatical tense/aspect marker rather than between two lexical items. Most grammaticalization is of this latter kind as we have seen with our examples of GO, COME, and RETURN verbs. A more complex intervening stage is what is labelled 'Dahl's paradox' in Vincent (2015:110-112). This is the case with the *surcomposé* verb forms such as Fr. *il a eu mangé* 'he had eaten' where the relation between the auxiliary constituents *a* 'HAVE.PRS.3SG' and *eu* 'HAVE.PTCP' is non-compositional — they mean 'had' not 'has

had' — but where the relation between these two items and the lexical head *mangé* 'eat.PTCP' is compositional: 'had' + 'eaten' = 'had eaten' (see also the discussion in §2.4.2 in Smith, this volume).

An intriguingly different kind of development is one where analysability is introduced without concomitant suppletion as happens with what is traditionally called 'folk etymology' as for example It. *battisuocera* 'cornflower', where the source word Lat. BAPTISECULA has been reanalysed as *batti* 'beat' + *suocera* 'mother-in-law'. As Maiden (2008) explains, developments like this are not the random curiosities they might at first seem but evidence that analysability of form is in principle independent of compositionality of content. They are thus on a par with the diachronic emergence of morphomic patterns, that is to say structures which unite members of a verbal paradigm in analysable and diachronically reproducible classes but which are not susceptible to morphosyntactic compositionality.

That said, the boundaries are not always easy to draw and are the focus of ongoing theoretical debate. Thus, the emergent Romance future and conditional paradigms built from the lexical stem plus endings derived from Lat. HABERE yields a morphologically productive and analysable pattern, as for example with the future and conditional of CORRER 'run' in Portuguese exemplified in Table 1.7:

Table 1.7: Portuguese future and conditional forms of RUN

	Future		Conditional	
1SG	correrei	< CORRERE HABEO	correria	< CORRERE HABEBAM
2SG	correrás	< CORRERE HABES	correrias	< CORRERE HABEBAS
3SG	correrá	< CORRERE HABET	correria	< CORRERE HABEBAT
1PL	correremos	< CORRERE HABEMUS	correríamos	< CORRERE HABEBAMUS
2PL	correreis	< CORRERE HABETIS	correríeis	< CORRERE HABEBATIS
3PL	correrão	< CORRERE HABENT	correriam	< CORRERE HABEBANT

If we subtract the infinitival stem *correr-* from these forms, we have a set of endings that apply to all verbs in the language whatever their other (ir)regularities might be. And at the outset it is reasonable to suppose that this regular morphological analysability was matched by semantic compositionality: CORRERE HABEO 'I have to run' and CORRERE HABEBAM 'I had to run'.[23] Two developments, however, complicate this picture. On the one hand, the fact that main stress fell originally on the auxiliary for some verbs led to phonetically reduced forms of the infinitival stem being found in both the future and the conditional, as in Pt. *direi* 'I will say' and *diria* 'I would say' beside *dizer* 'to say' and *traremos* 'we will bring' and *traríamos* 'we would bring' beside *trazer* 'to bring'. The extent and detail of this alternation between the full infinitive and the reduced future/conditional stem varies from language to language but the principle of such an alternation holds across most of the family. Maiden (2018a:ch 8) provides rich documentation of the pattern and, following Esher (2014), adopts the label Fuèc for this special stem. Crucially, he goes on to argue that the morphological shape remains even when the semantic link between future and conditional has weakened or disappeared, and that therefore the Fuèc stem is a further instance of morphomic patterning. Alternatively, if one adopts the semantics of the conditional as a combination of [FUTURE] and [PAST] as proposed by von Fintel and Iatridou (2008), the structures both in their original periphrastic version and in their modern univerbated form are still compositional. The jury is still out on this one.

[23] For the sake of simplicity, we omit discussion here of the alternative development whereby the conditional forms derive from the past perfect rather than the imperfect of the auxiliary, as with It. *correrebbe* 's/he would run' < CORRERE HABUIT. However, the logic of the argument is the same in both cases.

A different version of the compositionality question presents itself when we consider the outcomes of the combination of the GO verb with the infinitive as in Fr. *nous allons danser* 'we will dance (lit. we GO.PRS.1PL dance.INF)' beside Cat. *vam ballar* 'we danced (lit. GO.PRS.1PL dance.INF)'. Both languages have structures which are analysable as GO + INF and both are compositional but the underlying semantics which define the composition differ.

Finally, something which Table 1.6 does not allow for is simple non-availability of forms, or defectiveness. This has always been something of a puzzle within the domain of inflexion: why, for example, is the expected past participle form ***permasto* not available for the Italian verb PERMANERE 'stay, continue' when the structurally parallel *rimasto* is in regular use as the participle of the virtually synonymous RIMANERE 'stay, remain', and when the Latin simple verb MANERE 'stay', of which both are compounds, does have the form **mansum* and indeed old Italian provides evidence for a form *permaso*? The compositional components are, as it were, ready to go, but the lack of the necessary analysable constituents creates an arbitrary gap. And as we have seen, gaps of this kind exist for periphrases as well, so that while the Spanish compound periphrasis in (38a) is acceptable its Italian counterpart (38b) is not:

(38) a. He estado pensando que … (Sp.)
HAVE.PRS.1SG BE.PTCP think.GER COMP
'I've been thinking that …'

b.**Sono stato pensando que … (It.)
BE.PRS.1SG BE.PTCP think.GER COMP

However, unlike with inflexion, what we have also seen is that in some cases the absence of the periphrastic forms is not a matter of arbitrary gaps but of semantically motivated failure of compositionality as with the case of the Fr. VENIR DE construction exemplified in (34).

1.3.3 Attestation and reconstruction

At various points in our treatment so far we have underscored the need to integrate the investigation of language structure and change, and the inherent complementarity between the two. This theme emerges in a different guise when we come to consider the relation between attestation and reconstruction, a topic which has always been a matter of discussion among Romance linguists. Elcock (1960:33) famously observed that '[i]t is the special privilege of Romance philologists that they are not compelled to rely entirely upon reconstruction.' Beside this we can set the following remark from R.A. Hall at the beginning of his *Proto-Romance Phonology* (Hall 1976:1f.): 'In the Romance languages we find numerous sets of obviously related forms […] Our procedure should always be to reconstruct the Proto-Romance forms first and then to compare the reconstructions with attested Latin material.' In essence, the techniques of reconstruction apply principles derived from general theory to help open up the lost linguistic past, something which applies as much in the domain of morphosyntax as in phonology (Walkden 2014). In the case of language families with no documentary history they are the only option, but when there is ample documentation, as there clearly is in the case of Latin and in the early stages of Romance even for many of the so-called dialects, there is always the temptation to dispense with the deductive methods of general theory and work inductively on the basis of the surviving texts. An extreme expression of this latter view is Varvaro (2011:300) who asserts that to

use reconstruction when texts are available is, in his words, 'like studying the history of Napoleonic France with the methods that are normal in the domain of prehistory'.[24] In fact, however, there is a whole literature on the infectious diseases that were the scourge of some of Napoleon's armies with the evidence being derived from the analysis of excavated bodies using the same methods as deployed by archaeologists when they study prehistory (Raoult *et al.* 2006). Rather than giving priority to one or the other source of evidence it makes more sense to see them as complementary both in the domain of phonology and of morphosyntax.

As a case in point, we may consider the phenomenon of differential object marking (DOM):

(39) a. Dejad de atormentar a-l gato. (Sp.)
stop.IMP COMP tease.INF DOM-DEF cat
'Stop teasing the cat!'

b. să prinzi pre Iambru Diiac (Ro., 1600, Mardale 2015, ex 26b)
SBJV catch.2SG DOM Iambru scribe
'You should catch Iambru the Scribe'

c. ajju istu a tɔ sureḍḍa. (Crs., Manzini and
HAVE.PRS.1SG see.PTCP DOM your sister
Savoia 2005:II,513)
'I have seen your sister.'

d. si alchuno homini hochirit at alcuno atteru
if some man kill.IPFV.SBJV.3SG DOM some other

[24] For more discussion of the context in which Varvaro develops this view and further references, see Vincent (in press).

homini (14th-c. Srd., *Carta de Logu d'Arborea* 4.1)

man

'If a man killed another man'

e. ad te accareza la pultronaria (15th-c. Nap., *Brancati* 108.23)

DOM you.SG caress.PRS.3SG DEF idleness

'You're tempted by idleness'

Within Romance, DOM is widespread and displays many regional variations (see Ledgeway, in press for a rich compilation of examples, sources, and variant structures). The common factor, as the term implies, is that in certain contexts the direct object is distinctively marked by a preposition, typically *a* 'to' or whatever other form is used in the variety in question to mark an indirect object, such as reflexes of IN MEDIO (AD) 'in middle (to)' > *ma/me* 'to' in central Italy (Rohlfs 1969:15; Berizzi 2013) and DE+AB 'from' > *da* 'to' in Gallo-Italic varieties of Sicily as well as in some Umbrian dialects (Rohlfs 1969:8, 15; 1971:333f.; Manzini and Savoia 2005,II:502, 528). At first sight, this might seem relatable to the fact that in Latin for verbs such as SUBUENIRE 'help', PARCERE 'spare', INUIDERE 'envy', and SUADERE 'persuade' the object is marked with the dative rather than the more usual accusative case. And indeed, as Sornicola (1997) notes, there is a degree of correspondence between such items and many of the verbs with which DOM is found in early Italo-Romance and Ibero-Romance texts. However, the Romance phenomena differ in that the presence of DOM depends on information structure and the nature of the object as much if not more than on the governing verb, hence there may be patterns of alternation even when the verb remains the same. There is also the striking formal exception to the general pattern in Romanian where an entirely different

preposition, which does not have the function of indirect object marking has assumed this grammatical function, namely *pe* (< Lat. SUPER 'above'). What this means is that, interesting and instructive though the Latin parallels may be, our understanding of the origin of the construction type must depend on comparative reconstruction of the argument marking hierarchies (topics, animacy, pronominals, etc). Only in a second stage can the outcome of such analyses be compared to and integrated with the data provided by Latin. This is yet more true, as Mardale (2015) demonstrates, when it comes to Romanian, where, as we have seen, the formal marking differs and where the textual attestation is relatively late and varies according to register.[25] Even so, the links with particular verb classes are still there in the earliest Romanian texts

[25] Also relevant here is the earliest textual evidence provided by Sardinian which, in contrast to all other early Romance varieties, shows from as early as the eleventh century a systematic distribution of DOM (i.a-b; cf. also 39d), independently of verb class (Lombardi 2007:143f.). If medieval Sardinian is a verb-initial language, as argued by Virdis (1987) and Lombardi (2007), then the emergence of DOM in old Sardinian could be argued to be a purely endogenous development to ensure unambiguous marking of postverbal animate subjects and objects, often also involving object clitic doubling (Virdis 1987:15f.) as in (i.a).

(i) a. mi la furait Petru Tecas a nNastasia de Funtana aue domo
me= her= carry.PFV.3SG Pietro Tecas DOM Anastasia de Funtana from house
dessu thiu (OSrd., *Condaghe di San Pietro di Silki* 15)
of.the uncle
'Pietro Tecas abducted Anastasia de Funtana from her uncle's house'

b. Levait Dorgotori et issos frates a Maria et a Gavini (OSrd.,
take.PFV.3SG Dorgotori and thes brothers DOM Maria and DOM Gavini
Condaghe di San Pietro di Silki 68)
'Dorgotori and his brothers took Maria and Gavini'

(Mardale 2015:240), but the issue is complicated by the type and style of the text. Here, as elsewhere, philology and linguistics are companions not opponents.[26]

1.3.4 Language contact

In his critique of grammaticalization, Joseph (2011; 2020) deploys as one of his arguments the history of Spanish adverbs such as *claramente* 'clearly'. At first sight this seems neatly to parallel forms such as Fr. *clairement* and It. *chiaramente* and to derive via the grammaticalization of the ablative of the Latin noun MENS 'mind' in an expression such as *clara mente* 'clear.FSG.ABL mind.FSG.ABL (= with a clear mind)'. The problem, as Karlsson (1981:87) notes, is that in that case we would expect the Spanish form of the adverbial suffix to be *-miente* since in general Lat. stressed [ĕ] diphthongizes both in open and closed syllables in Spanish: *tiene* < TENET '(s)he holds > has', *hierba* < HERBAM 'grass'. And indeed in early texts forms such as *de buena miente* 'of good mind, willingly' are attested. However, according to the data collected by Karlsson, by far the most common form is in fact *mientre* as in *vera mientre* 'truly', *demesurada mientre* 'immoderately', where the presence of *-r-* is due to crossing with the Latin adverbial suffix -(I)TER as in CELERITER 'quickly'. What is not attested, however, until the fourteenth century is the suffix *-mente*, which is now the standard Castilian form. The explanation lies in the fact that Spanish *-mente* is not a direct continuer of Latin but a literary borrowing, either direct from Latin or more probably from old French or old Occitan (Karlsson 1981:106). That the original structure here was a learnèd form is also the conclusion reached by Hummel (2013;

[26] Further examples of cases where attestation and reconstruction are complementary rather than contradictory are discussed in Vincent (in press:§6).

2014). Examples such as this do not, *pace* Joseph, counter-exemplify grammaticalization but they do show that diachronic paths can be crossed or blocked as a result of contact. What is special about this case, and about Romance in general, is the way the members of the family provide evidence for these contact effects operating not only between related varieties but also, and unusually, between the daughter languages and their diachronic source.

The adverb formation mechanism seen in *claramente* is a good example of a development from analytic to synthetic without an intermediate periphrastic stage. The original Latin construction had a degree of syntactic freedom, so that examples are attested with both ADJ + N and N + ADJ orders (*mente sincera* 'with a sincere mind' Plautus *Bacchides* 509; *turbata mente* 'with a worried mind' Tacitus *Annals* 4.22), and with intervening material (*mente ferant placida* 'that they may bear with a calm mind' Ovid *Metamorphoses* 13.214). The only surviving freedom is that some varieties allow co-ordinated adjectives as with Sp. *sincera y claramente* 'sincerely (lit. sincere) and clearly' while others require the adverb suffix on both items as with It. *sinceramente e chiaramente*. At the same time, it is worth noting that the Romance patterns, regardless of the variety of attested patterns, require us to reconstruct the order ADJ + *mente* even though the unmarked order in the Romance languages is N + ADJ. Moreover, this reconstructed form wins out over the sequence ADJ + MODO 'way', which Bauer (2006) shows to have originally been the more frequently attested form in Latin.

An instance where contact interacts with periphrasis involves complex gerund forms such as Fr. *étant parti* 'BE.GER leave.PTCP (= having left)' or Pt. *tendo cantado* 'HAVE.GER sing.PTCP (= having sung)'. Although from a synchronic perspective they look simply to belong to the periphrastic part of the verbal paradigm with the choice

of auxiliary HAVE/BE varying from language to language and verb to verb, these are in fact relatively late innovations due to contact with – and more specifically translation from – Classical Latin. As Menoni (1988) demonstrates, these structures are coined at a time when classical texts were being translated into the vernacular. Thus, to take one of her examples, in Arrigo Simintendi's fourteenth-century version of Ovid's *Metamorphoses*, the sentence *amplexuque dato ait* 'embrace.ABL.SG=and give.PTCP say.PFV.3SG' (*Met* II.42) is rendered in Italian as *e abbiendolo abbracciato li disse* 'and HAVE.GER=MSG.ACC embrace.PTCP 3SG.DAT= say.PFV.3SG (= and having embraced him, he said to him)'. A further consequence of this route of development is that while, as we have seen, structures with simple gerunds are part of the core grammar of all the Romance languages, the compound gerunds are for the most part restricted to higher registers of the spoken, and in particular the written, language. They also always have the straightforward present perfective interpretation while the finite compound forms may develop differently, as for instance in Portuguese, where *tendo cantado* means 'having sung' but *tenho cantado* means not 'I have sung' (which is expressed by the synthetic *cantei* 'I sang' or 'I have sung'), but the aspectually marked 'I have been singing/have kept on singing'.

Contact via translation is also at work in some instances of the Romanian VENI 'come' passive construction discussed above. Dragorimescu and Nicolae (2014) cite examples like (40):

(40) a. Nişte corăbii neguţătoreşti […] vin
some ship.FPL commercial.FPL COME.PRS.3PL
arse (ORo., *Foletul novel*, 1693-1704)
burn.PTCP.FPL

‘Some commercial ships […] are burned.’

b. construzione […] care dă Greci vine numită
contruction REL by Greek.PL COME.PRS.3SG name.PTCP.FSG
sintaxis (ORo., Văcărescu, *Gramatica,* 1787)
syntax
‘construction […] which was/is called syntax by the Greeks.’

These examples and many similar ones derive either from translations from or imitations of Italian literary models and once again therefore are restricted to older and higher registers in contrast to the more popular usage of the structure discussed earlier.

The pattern with compound gerunds also provides an interesting counterpoint to the case of the Italian VENIRE ‘come’ passive periphrasis discussed above. There we saw that, although the structural ingredients are available to create compound forms such as ***è venuta discussa*** or ****venire discussa***, these are considered ungrammatical by native speakers. In the case of the compound gerund, the structural ingredients are in use independently in the texts from the early medieval period but attestation of the combined structure only occurs some half a millennium later and in very specific cultural circumstances. With the VENIRE construction the early texts shed a different kind of light. Although, as we have said, the modern patterns are considered ungrammatical, examples such as (41) are attested:

(41) disse esserle venuta involata […] una sua
say.PST.3SG BE=DAT.3.FSG COME.PTCP.FSG steal.PTCP.FSG a her
collana (16th-c. It., Ascanio de Mori)

necklace

‘she said a necklace of hers had been stolen’

The author of this example is from the Lombard city of Mantua, and it attests to contact between the language of literature and dialects of the region where VENIRE is commonly the unmarked, and sometimes the only, passive auxiliary (Vincent 1987:249, 2014:18f.). Once again, the combination of structural analysis, grammatical change, textual attestation, and regional variation provide a rich and stimulating cocktail for linguists of all persuasions.

We have obviously only been able to scratch the surface of the question of language contact and we have concentrated on cases where the Romance data are particularly instructive. Further dimensions are discussed in Finbow and O’Neill’s chapter in this volume where they compare the different mechanisms involved in contact between related and unrelated languages.[27]

Language adjacency does not, of course, necessarily lead to contact effects and hence the caution about the so-called ‘Balkan’ influences expressed by Maiden (in press) and other contributors to Gardani, Loporcaro, and Giudici (in press). Indeed, at the other end of the spectrum is the phenomenon we might label ‘anti-contact’. A case in point is Guardiol, the variety of Occitan originally spoken in Piedmont and which has survived in the Calabrian village of Guardia Piemontese. The cause of the dislocation from its region of origin was religious persecution and as a result since the sixteenth century the dialect has remained isolated (Kunert 1994). Although it has borrowed vocabulary from the other dialects of the region and from standard Italian, structurally it has resisted change preserving, for example, a past tense built from GO

[27] For a recent survey, see D’Alessandro (2021).

+ infinitive, which has died out in other varieties of Occitan (Jacobs and Kunert 2014, and see also the chapters by Vincent and Wheeler and by Paoli and Wolfe, this volume).

1.4 Conclusions

If there is one general conclusion to emerge from this chapter it is the essential complementarity of synchrony and diachrony. This in turn has important implications for our understanding of broader theoretical issues. Within the literature on (morpho)syntactic change over the last four decades or so, there have been two diametrically opposed stances. On the one hand the Chomskyan tradition in its various current versions (cartography, nanosyntax, universal spine, etc) argues for the primary importance of defining the properties of UG, in terms of which it is then possible to investigate and model the data of language change. Although the conceptual framework and the theoretical primitives are very different, this is essentially a continuation of the Saussurean dichotomy (or 'firewall' in the terms of Kiparsky 2015:87), with priority being given to synchrony and language change being considered, if at all, in a second phase of enquiry. The alternative view lies behind the question posed by Hopper and Traugott (2003:16):

> Do we need in our analyses to 'stop the film' and fix the grammar of a language as we investigate its structure, or do we need to view 'grammar' as a provisional way-station in our search for the more general characteristics of language as a process for organizing cognitive and communicative content?

On this view, the diachronic evidence is as important, maybe even more important, than the synchronic because it reveals the way natural languages emerge and change as the result of a series of intersecting external circumstances rather than being defined and restricted by a set of innate universal principles and mechanisms. It further implies that there is no necessity for there to be a fixed 'core grammar' in order to define what counts as a given language. And even if many will find this last conclusion too extreme, the fact remains that there is considerable variation in patterns within any one 'language/dialect', and this in turn argues for the desirability of not prioritizing either synchrony or diachrony but rather of treating both as complementary if we wish to achieve a full understanding of any given linguistic situation. Another way of expressing this same conclusion lies in Roger Lass's words from his inaugural lecture 'Language and time: a historian's view' at the University of Cape Town in 1983:

> A language never simply 'is'; it is always looking back over its shoulder, and ahead. It is simultaneously in three states: 'having been', 'being' and 'going to be'. Only 'being' is the concern of the normal speaker; all three are the linguist's concern.

In different ways, all three also figure in the chapters that follow.