

## Periphrasis as collocation

Olivier Bonami

Prefinal version of December 2014. To appear in 2015 in *Morphology*

**Abstract** This paper provides a formal theory of inflectional periphrasis, the phenomenon where a multi-word expression plays the grammatical role normally played by a single word filling a cell in an inflectional paradigm. Expanding on the literature, I first identify and illustrate six key properties that a satisfactory theory of periphrasis should account for: (i) the phenomenon of periphrasis is found in the inflection of all major parts of speech; (ii) the logic of the opposition between periphrasis and synthesis is the logic of inflection; (iii) auxiliaries as used in periphrases are morphosyntactic hybrids; (iv) some periphrases are morphosyntactically non-compositional; (v) periphrasis is independent of phrase structure, but (vi) the parts of a periphrase are linked by a grammatical function. The rest of the paper presents a lexicalist theory of periphrasis, relying on a version of HPSG (Pollard & Sag, 1994) for syntax combined with a version of Paradigm Function Morphology (Stump, 2001) for inflection. The leading idea is that periphrases are similar to syntactically flexible idioms; the theory of periphrasis is thus embedded within a more general theory of collocation. Periphrasis is accounted for in a strictly lexicalist fashion by recognizing that exponence may take the form of the addition of collocational requirements. I show how the theory accounts for all key properties identified in the first section, deploying partial analyses for periphrastic constructions in English, French, Czech, and Persian.

**Keywords** Periphrasis · Inflection · Lexicalism · Realizational morphology

### 1 Introduction

Inflectional periphrasis is the phenomenon where a multi-word expression plays the grammatical role normally played by a single word filling a cell in an inflectional paradigm. Probably the most discussed case of inflectional periphrasis is found in Latin conjugation. As shown in Table 1, while ordinary active verbs possess synthetic forms expressing the perfect, for passive and deponent verbs this role is played by the combination of a form of the copula carrying appropriate tense and mood inflection and a passive past participle.

The fact that the same morphosyntactic content can be expressed by synthetic or analytic means motivates the idea that the periphrase is part of the inflectional paradigm. Here and throughout, periphrastic forms are highlighted in boldface.<sup>1</sup>

		MONEŌ active 'advise'	MONEŌ passive 'be advised'	VEREOR deponent 'revere'
[−PERFECT]	PRS	moneō	noneor	vereor
	PST	monēbam	monēbar	verēbar
	FUT	monēbō	monēbor	verēbor
[+PERFECT]	PRS	monuī	<b>monitus sum</b>	<b>verītus sum</b>
	PST	monueram	<b>monitus eram</b>	<b>verītus eram</b>
	FUT	monuerō	<b>monitus erō</b>	<b>verītus erō</b>

**Table 1** Selected 1SG forms of Latin second conjugation verbs

Periphrasis clearly lies at the morphology–syntax interface. As Matthews (1991, 219–220) puts it, “[the form of the Latin Perfect Passive] is clearly two words, which obey separate syntactic rules (for example, of agreement); Nevertheless they are taken together as a term in what are otherwise morphological oppositions.” In a field where morphology and syntax tend to be examined by different specialists, the dual nature of periphrasis is often overlooked, if not denied. Within lexicalist approaches to grammar, syntactic studies of periphrases usually attempt to treat them as ordinary syntax, thereby ignoring how the expressions interact with the rest of the inflectional paradigm; morphological studies, on the other hand, are typically content with generating a sequence of two words, thereby ignoring the nature of the syntactic relation these two words entertain.

Starting with the seminal studies of Vincent & Börjars (1996) and Ackerman & Webelhuth (1998), the past fifteen years have witnessed a number of attempts to do justice to the dual nature of inflectional periphrasis, including Sadler & Spencer (2001); Spencer (2003); Ackerman & Stump (2004); Bonami & Samvelian (2009); Bonami & Webelhuth (2013) and Blevins (forthcoming). In parallel, a number of detailed empirical investigations (including Chumakina 2013, Nikolaeva 2013, Stump 2013, Popova & Spencer 2013, and Bonami & Samvelian in press) and typological studies (including Anderson 2006, Brown & Evans 2012, Corbett 2013 and Spencer 2013b) have broadened our understanding of the diversity of the phenomenon. Once these empirical studies are taken into account, it becomes clear that all previous theoretical proposals for the analysis of periphrasis are either too vague to be fully evaluated or too constrained to account for the known data.

The goal of the present paper is to present, justify and illustrate a novel approach to the morphology and syntax of periphrasis. The central intuition behind the approach is that a periphrase is the inflectional analogue of a flexible idiom: just like idiom parts stand in a partially flexible syntactic relation but jointly express semantic content in a non-compositional fashion, parts of a periphrase stand in a partially flexible syntactic relation and jointly express morphosyntactic content that is not necessarily deducible from the synthetic morphology on the parts. On the basis of that intuition, I present a formal theory of periphrasis that combines analytic tools of phrase-structural syntax, realizational morphology, and lexicalist theories of idioms and other collocations.

<sup>1</sup> Notice that I use *periphrasis*, a mass term, to designate the general phenomenon, and *periphrase*, a count term, to refer to particular instances of the phenomenon (that is, particular periphrastic constructions).

The structure of the paper is as follows. Drawing on a wide empirical base, section 2 presents six key properties of inflectional periphrases that any adequate theory should be able to account for. The section ends by examining formal models of periphrasis proposed in the literature, and concludes that none accounts for the full set of key properties. Sections 3 to 5 present a new theory of periphrasis. The formal proposal relies on a combination of Head-driven Phrase Structure Grammar (Pollard & Sag, 1994) and Paradigm Function Morphology (Stump, 2001). In section 3, the analogy between idioms and periphrasis is presented. I then show how a lexicalist view of flexible idioms as involving two mutually selecting lexical items can straightforwardly be adapted to account for the syntactic flexibility of inflectional periphrasis. In section 4, I present an extended view of inflection, where exponence of some morphosyntactic properties may take the form of a collocational requirement rather than that of a modification of the word's phonology. This idea is executed formally by extending the notion of a paradigm function. Finally, section 5 puts together the syntactic and inflectional aspects of the analysis, showing how they jointly account for the typological diversity of periphrases.

A difficult issue that any study of periphrasis must face is that of drawing the border between periphrasis and ordinary syntax. By definition, inflectional periphrases are multi-word expressions that realize the same kind of content as inflectional morphology. This does not mean, however, that all ways of paraphrasing inflection using syntactic constructions should be considered instances of inflectional periphrasis. To take an extreme example, from the fact that the translation of the Turkish example in (1) involves modification by an adverb, one would not conclude that indirect evidentiality is in English an inflectional category realized by combination of the verb with the adverb *apparently*.

- (1) Ali bahçe-sin-e                      bir    meşe ağac-ı    dik-miş  
 Ali garden-3SG.POSS-DAT INDF oak    tree-NC plant-PST.EVID  
 'Ali apparently planted an oak tree in his garden'      (Göskel & Kerlake, 2005, 309)

In this paper I will devote very little attention to this issue, and refer the reader to the relevant literature (notably Haspelmath 2000; Spencer 2003; Ackerman & Stump 2004; Brown, Chumakina, Corbett, Popova, & Spencer 2012). For practical purposes I adopt the view that any situation where a morphosyntactic feature value is expressed by multiple words rather than synthetic morphology on a single word qualifies as periphrasis.<sup>2</sup> The adoption of such a permissive definition is motivated by the fact that the theory presented here is not intended to be restrictive; rather, it attempts to capture the already explored part of a still largely uncharted territory. Thus there is no downside for it to be able to capture constructions whose status as periphrases rather than simple syntactic constructions is disputable.

Another difficult issue that needs to be addressed is one of vocabulary. By definition, an inflectional periphrase involves more than one word. One of the words constituting the periphrase always is a form of the lexeme that is realized: I call this word the MAIN ELEMENT of the periphrase, and I call the lexeme whose paradigm it belongs to the MAIN LEXEME. In the extant literature, the other element (or other elements) of the periphrase is usually called the AUXILIARY. This has led to much terminological confusion, due to the fact that

<sup>2</sup> Note that I use *morphosyntactic feature* where Brown et al. (2012) say *grammatical feature*. Also note that this definition makes the identification of periphrasis dependent on the definition of morphosyntactic features, and limits the attention to inflection. I thus leave aside the issue of periphrastic lexeme formation, on which see (Ackerman & Webelhuth, 1998, 219–340), Ackerman et al. (2011), and Lee & Ackerman (submitted).

many grammatical traditions isolate a class of AUXILIARY VERBS characterized by a common set of syntactic properties, rather than their use in periphrastic inflection. Thus when discussing English, it is customary to define auxiliary verbs as those verbs that verify the so-called NICE properties (Negation, Inversion, Contraction, and Ellipsis; see Huddleston & Pullum, 2002, 92–115 for a contemporary discussion). As it happens, all English verbs that can be argued to participate in the periphrastic expression of morphosyntactic features belong to this class (*be* in the progressive, *have* in the perfect, possibly *be* in the passive and *will* in the future), but for other verbs in the class (e.g. modals) a periphrastic analysis is unwarranted, and some of these verbs have other uses as main lexemes (e.g. copular *be*). To avoid some of the confusion, I will call words participating in a periphrase other than the main element ANCILLARY ELEMENTS, and by extension, I will call the corresponding lexemes ANCILLARY LEXEMES. Thus the perfect in English is expressed by a combination of an ancillary element that is a present form of the auxiliary verb *have* and a main element realized as a past participle. A further advantage of this terminological choice is that it does not tie the descriptive vocabulary to a particular part of speech: the ancillary element in a periphrase may be a verb or belong to some other part of speech.

## 2 Key properties of periphrasis

In this section I present six key properties of inflectional periphrasis that any theory should be able to account for.

### 2.1 Periphrasis is independent of part of speech

Most examples of periphrasis discussed in the literature are found in verbal inflection; this is to be expected, since verbs tend to have larger paradigms than nouns or adjectives. It is important to remember however that the possibility of periphrastic inflection is agnostic to part of speech (Haspelmath, 2000), in light of recent claims by John M. Anderson (2011, chap. 6) that periphrasis is essentially verbal. This categorial neutrality becomes evident by relying upon the criterion which Ackerman & Stump (2004) call FEATURE INTERSECTIVITY: if a multi-word expression is used to realize the combination of two feature values that are otherwise expressed synthetically, then this expression is an inflectional periphrase. Such a criterion is usually taken as a sufficient condition for establishing periphrastic status (Hockett (1958, 212), Mel'čuk (1993, 355), Haspelmath (2000, 655–660), Ackerman & Stump (2004, 126–131), Brown et al. (2012, 250–252)).

Tundra Nenets nouns provide a clear example of periphrasis in the nominal domain (Nikolaeva, 2013). Nouns inflect for three numbers (singular, dual, and plural) and seven cases: the three grammatical cases nominative, accusative, and genitive, and the four local cases dative, locative, ablative, and prosecutive.<sup>3</sup> Local postpositions also inflect for local case and take a genitive complement (2b).

- (2) a. Xidya-n<sup>o</sup>h ya-m pudabta<sup>o</sup>.  
 cup-DAT.SG flour-ACC.SG strew[AOR.3SG]  
 'He strewed flour into the cup.' (Salminen, 1997, 141)

<sup>3</sup> Tundra Nenets nouns also inflect for pronominal possessors. Only absolute subparadigms are displayed in the interest of brevity. On Tundra Nenets inflection see also Salminen (1997); Ackerman & Nikolaeva (2014); Nikolaeva (2014).

- b. Xusa-m      lata-h      nyi-h      me<sup>o</sup>-da.  
 dough-ACC.SG table-GEN.SG on-DAT take-OBJ.SG.3SG  
 ‘He put the dough onto the table.’ (Salminen, 1997, 141)

Table 2 shows the absolute paradigm of a sample noun. Although inflection is mostly synthetic, it is periphrastic in the dual for local cases: the main element is in the genitive dual, and occurs as the complement of the postposition *nya* ‘at’ inflected for the appropriate case. Notice that the distribution is featurally intersective: the dual is synthetic for nonlocal cases, and local cases have synthetic singular and plural forms.

	SG	DU	PL
NOM	ti	tex <sup>o</sup> h	tiq
ACC	tím	tex <sup>o</sup> h	tí
GEN	tih	tex <sup>o</sup> h	tíq
DAT	ten <sup>o</sup> h	<b>tex<sup>o</sup>h nyah</b>	tex <sup>o</sup> q
LOC	tex <sup>o</sup> na	<b>tex<sup>o</sup>h nyana</b>	tex <sup>o</sup> qna
ABL	tex <sup>o</sup> d <sup>o</sup>	<b>tex<sup>o</sup>h nyad<sup>o</sup></b>	tex <sup>o</sup> t <sup>o</sup>
PROS	tew <sup>o</sup> na	<b>tex<sup>o</sup>h nyamna</b>	teqm <sup>o</sup> na

**Table 2** Absolute subparadigm of the Tundra Nenets noun T1 ‘male reindeer’ (Salminen 1997)

A particularly clear case of periphrasis in adjectival inflection is provided by Ingush (Nichols, 2011). Ingush adjectives systematically inflect for case (nominative vs. oblique) and comparative grade.<sup>4</sup> Predicative adjectives form their comparative by adding the suffix *-gh* to the positive form (3a). Attributive adjectives in the positive grade modifying a non-nominative noun take the oblique suffix *-acha* (3b). For attributive adjectives in the comparative grade, however, a periphrastic strategy is used rather than the expected combination of two suffixes: the adjective carrying comparative morphology is realized as the predicative complement of the present participle of the copula, which realizes case marking (3c). The situation is summarized in Table 3. Here too the distribution is featurally intersective: exponence of comparative grade is synthetic in predicative use, as is expression of case in the positive grade.

- (3) a. Xii    benziinal    d-az-a-gh      d-y  
 water gasoline.CSN D-heavy-NOM-CMP D-be.PRS  
 ‘Water is heavier than gasoline.’ (Nichols, 2011, 516)
- b. Xaatta    hwei    q’ean-acha    Gichiiga    [...]  
 ask.IMP 2S.MIR old-OBL    Gichi.ABL  
 ‘Ask old Gichi [...].’ (Nichols, 2011, 759)
- c. So    **d-oaqq-a-gh d-olcha**    zhwaliena    bwarjga+j-eira  
 me[J] D-big.CMP    D-be.PPL.OBL [D]dog.DAT eye+J-see.WP  
 ‘The bigger dog saw me.’ (Nichols, 2011, 222)

These two examples clearly show that periphrasis as a grammatical strategy is available across parts of speech.

<sup>4</sup> Some adjectives also inflect for number and/or gender. Gender agreement is marked by the prefixes *d-*, *b-*, *j-* and *v-*. The superlative is formed by combining a comparative adjective with the preposed word *eggara* ‘(the) most’. Nouns distinguish 4 case values, but all non-nominative cases are syncretic on adjectives. This syncretic case value is noted as OBL in the glosses.

	PREDICATIVE		ATTRIBUTIVE	
	NOMINATIVE	NOMINATIVE	NOMINATIVE	OBLIQUE
POSITIVE	doaqqa	doaqqa	doaqqa	doaqqacha
COMPARATIVE	doaqqagh	<b>doaqqagh dola</b>	<b>doaqqagh dolcha</b>	

**Table 3** Partial paradigm of the Ingush adjective ‘big’ (Nichols, 2011)

## 2.2 The logic of the synthesis/periphrasis opposition is the logic of inflection

The bread and butter of inflectional analysis is the determination of the distribution of synthetic exponents. In this section we show that the distribution of synthetic and periphrastic inflection follows the same logic as the distribution of synthetic exponents. We first discuss the various situations found within the paradigm of a lexeme, and then the situations arising across paradigms.

### 2.2.1 Synthesis and periphrasis within paradigms

Alternative strategies of exponence may combine in various ways, giving rise to what Corbett (in press) calls LEXICAL SPLITS. When two exponents are in complementary distribution in the paradigm of a lexeme, three situations may occur, as exemplified in Table 4 for prototypical person/number paradigms. The paradigm may exhibit a BALANCED SPLIT, whereby a binary feature value conditions the choice of exponent. It may exhibit a PĀṆINIAN SPLIT if one exponent corresponds to the general case, but is preempted by another exponent in a more specific, coherent class of cells. Or it may exhibit a MORPHOMIC SPLIT if neither exponent corresponds to a natural class of paradigm cells.<sup>5</sup>

(a) Balanced split			(b) Pāṇinian split			(c) Morphomic split		
	SG	PL		SG	PL		SG	PL
1	<i>A</i>	<i>B</i>	1	<i>B</i>	<i>A</i>	1	<i>A</i>	<i>B</i>
2	<i>A</i>	<i>B</i>	2	<i>A</i>	<i>A</i>	2	<i>B</i>	<i>B</i>
3	<i>A</i>	<i>B</i>	3	<i>A</i>	<i>A</i>	3	<i>B</i>	<i>A</i>

**Table 4** Types of splits in the distribution of complementary exponents in a paradigm

There is a strong sense that balanced splits constitute the most simple, and somewhat uninteresting, situation. The high prevalence of Pāṇinian splits motivates the use of a specificity ordering on rules of exponence, variously implemented under the name of the elsewhere condition (Kiparsky, 1973), the subset principle (Halle & Marantz, 1993), or Pāṇini’s principle (Stump, 2001). While morphomic splits have played an important role in morphological theory since Maiden (1992) and Aronoff (1994), they are usually taken to be the exceptional rather than the typical situation.

It is notable that all three kinds of splits just discussed are also found in the arbitration between a periphrastic and a synthetic strategy for the realization of some feature. Balanced splits are very common; a prime example is the expression of the perfect in familiar

<sup>5</sup> If more than two exponents are in complementary distributions, the three situations may combine in various ways.

Romance and Germanic languages, where [−PERFECT] is synthetic and [+PERFECT] is periphrastic.

Pāṇinian splits are also common. Most of the time, synthesis is the general situation, and periphrasis the specific case. In Tundra Nenets nominal declension, as witnessed in Table 2, periphrastic inflection applies in the specific combination of dual number and local case, while synthetic inflection is used in all other situations. A similar situation is found for Ingush adjectives: as witnessed in Table 3, the general strategy is synthetic inflection, periphrasis being used only for attributive comparatives. Likewise, the Latin future infinitive is periphrastic, whereas (in active, non-deponent verbs) the rest of the conjugation is synthetic, including nonfuture infinitives and non-infinitive futures. The opposite kind of Pāṇinian split, with periphrasis as the general case and synthesis as the specific case, is found in Persian (Bonami & Samvelian, in press). At first sight, there seems to be a balanced split between periphrastic perfect forms and synthetic non-perfect forms, as shown in Table 5. Closer examination shows, however, that the present perfect has morphologized into a synthetic form.<sup>6</sup>

		BOUNDED	UNBOUNDED	PERFECT
PRESENT		—	mi-xar-ad	xarid-e-ast
PAST	DIR.	xarid	mi-xarid	<b>xarid-e bud</b>
	IND.	xarid-e-ast	mi-xarid-e-ast	<b>xarid-e bud-e-ast</b>
SUBJUNCTIVE			be-xar-ad	<b>xarid-e bāš-ad</b>

**Table 5** Distribution of the Persian perfect periphrase (Bonami & Samvelian, 2009)

Good examples of a morphomic split between periphrastic and synthetic realization are not common.<sup>7</sup> One convincing example is provided by Archi. Focusing on the expression of tense and aspect in the present and past, and simplifying somewhat, the situation can be depicted in Table 6, based on Chumakina (2013). In this subparadigm, verbs possess two synthetic forms which in isolation realize the generic/habitual present and the bounded past respectively. All other paradigm cells are filled by a periphrastic construction, where the main element is a converb realizing [±PERFECT] and the ancillary element is a form of the copula realizing tense. Clearly, neither periphrasis nor synthesis covers a natural class of cells in the paradigm: on the one hand, the same periphrastic construction used systematically in the unbounded past is used only in the progressive for the present; on the other hand, the past does use a synthetic form for bounded aspect.<sup>8</sup>

<sup>6</sup> The present perfect is syncretic with the bounded past indirect evidential, and historically based on the morphologization of the clitic form of the copula. However the present perfect forms exhibit a cohesiveness that is not found in copular constructions based on the clitic copula, and present idiosyncratic morphophonological fusion in colloquial usage; hence whatever their historical source they are clearly synchronically synthetic words. See Bonami & Samvelian (in press) for details.

<sup>7</sup> Corbett (2013, 172) cites the suppletive paradigm of the verb ‘go’ in the Romanian dialect of Fundătura documented by Maiden (2004, 242), which exhibits a morphomic split along the N-pattern between inherent reflexive forms such as 1SG *mă duk* and nonreflexive forms such as 1PL *merem*. However this is not a clearcut case, since Romanian reflexive markers have been argued to be affixal (Monachesi, 2005).

<sup>8</sup> This table is unfaithful to Chumakina’s description in two respects. First, Chumakina points out that a periphrastic realization is also possible for the present habitual, although it is dispreferred. Second, Chumakina labels ‘past perfective’ the forms labeled here as ‘past perfect’, and takes the present/past distinction to be irrelevant for the form labeled here ‘present perfect’; in addition the form labeled here ‘bounded past’ is labeled ‘simple past’, and my ‘unbounded’ forms are ‘imperfective’. This slightly different view of the

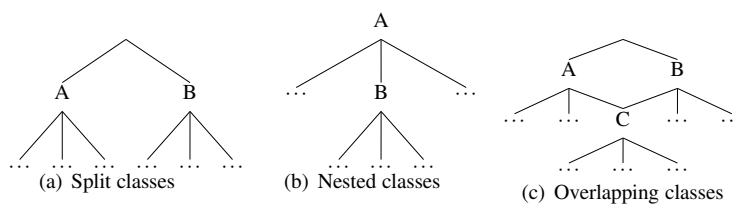
	BOUNDED	UNBOUNDED		PERFECT
		PROGRESSIVE	HABITUAL	
PRESENT	—	<b>ko-r-ši i</b>	ko-r	<b>ko-li i</b>
PAST	ko	<b>ko-r-ši edi</b>	<b>ko-r-ši edi</b>	<b>ko-li edi</b>

**Table 6** Partial paradigm of the Archi verb ‘hear’

### 2.2.2 Synthesis and periphrasis across paradigms

Up to now I have discussed competition between exponence strategies within the paradigm of a single lexeme. But alternative exponence strategies are also found ACROSS lexemes: different exponents may be used for the realization of the same properties, depending on the lexeme. As in the previous section, I first review the types of situations found with synthetic exponence, and to then show that the same situations are found when comparing synthetic and periphrastic realizations.

Figure 1 reviews common distributions for alternative synthetic exponence strategies, in the form of schematic hierarchies of classes of lexemes. Sometimes two strategies apply to numerically comparable sets of lexemes; this usually leads to the assumption that the set is split in two inflection classes. Sometimes one strategy is much more common than the other; in such situations, it is customary to assume nested inflection classes, and to invoke specificity again to account for the distribution of exponents: exponents of the smaller class preempt the use of exponents from the larger class. A third, less common situation, however, is to have overlapping classes: there are two well-defined inflection classes, but some lexemes share exponence properties found in both classes.



**Fig. 1** Types of distribution of exponence strategies across lexemes

Where inflection classes overlap, two situations may arise, which I will illustrate using the Czech nouns shown in Table 7.<sup>9</sup> The more common situation is HETEROCLISIS: items in the overlapping class pattern with one or the other superclass in different paradigm cells. This is illustrated by the neuter noun *kuře* ‘chicken’: it uses the same exponents as soft neuter nouns like *moře* ‘sea’ in the singular, but uses instead those of hard neuter nouns like *město* ‘town’ in the plural.<sup>10</sup> Another possibility however is to resolve the conflict through OVER-

data does not affect the point that neither the synthetic nor the periphrastic strategy covers a natural class of paradigm cells: quite on the contrary, by more strongly collapsing the opposition between tense and aspect, Chumakina’s view of the data is even more strikingly morphomic. I am indebted to Marina Chumakina for extended discussion of the issue and for kindly providing the unpublished partial paradigm in Table 6.

<sup>9</sup> Czech nouns inflect for 2 numbers and 7 cases. In the interest of space, Table 7 only shows 4 case values.

<sup>10</sup> In the Czech grammatical tradition, the class of *kuře* is treated as a separate inflection class, under the assumption of a segmentation where the stem is constant and the *-et-*, *-at-* arguments are part of inflection. A



ABUNDANCE (Thornton, 2012): both strategies are equally grammatical for members of the overlapping class. This is illustrated by the masculine inanimate noun *pramen* ‘spring’:<sup>11</sup> in the plural, *pramen* inflects like a hard declension masculine inanimate noun such as *most* ‘bridge’. In the singular, it alternates between exponents typical of the hard and soft declension.<sup>12</sup>

		MASCULINE			NEUTER		
		hard	mixed	soft	hard	mixed	soft
SG	NOM	most	pramen	pokoj	měst-o	kuř-e	moř-e
	GEN	most-u	pramen-u~pramen-e	pokoj-e	měst-a	kuřet-e	moř-e
	DAT	most-u	pramen-u~pramen-i	pokoj-i	měst-u	kuřet-i	moř-i
	ACC	most	pramen	pokoj	měst-o	kuř-e	moř-e
PL	NOM	most-y	pramen-y	pokoj-e	měst-a	kuřat-a	moř-e
	GEN	most-ů	pramen-ů	pokoj-ů	měst	kuřat	moř-í
	DAT	most-ům	pramen-ům	pokoj-ům	měst-ům	kuřat-ům	moř-ím
	ACC	most-y	pramen-y	pokoj-e	měst-a	kuřat-a	moř-e
		‘bridge’	‘spring’	‘room’	‘town’	‘chicken’	‘sea’

Table 7 Partial paradigm of six Czech nouns

Turning to arbitration between periphrasis and synthesis, one finds again all the same situations. The Czech future tense exhibits a nice combination of split and nested classes (Short, 1993, 481–491). The future is always synthetic with perfective verbs, and it is generally periphrastic with imperfectives. There are however a few imperfective verbs with a synthetic future: the copula *být* and manner of motion verbs such as *jít* ‘go on foot’. Table 8 provides relevant examples.

	PERFECTIVE	IMPERFECTIVE		
1SG	vstanu	<b>budu čekat</b>	budu	půjdu
2SG	vstaneš	<b>budeš čekat</b>	budeš	půjdeš
3SG	vstane	<b>bude čekat</b>	bude	půjde
1PL	vstaneme	<b>budeme čekat</b>	budeme	půjdeme
2PL	vstanete	<b>budete čekat</b>	budete	půjdete
3PL	vstanou	<b>budou čekat</b>	budou	půjdou
	‘get up’	‘wait’	‘be’	‘go on foot’

Table 8 Future tense of a sample of Czech verbs

Czech imperfective verbs are an example where periphrasis is the default strategy, and synthesis the more specific strategy, used only with a dozen verbs—this is what Haspelmath (2000, 659) calls ‘anti-periphrasis’. Czech adjectives provide an example of the opposite situation (Short, 1993, 478). As shown in Table 9, an overwhelming majority of Czech adjec-

more satisfactory analysis, which is hinted at in Cvrček et al. (2010, 189), is to assume that these augments are segmented separately, either as portion of stem alternants or as separate suffixes.

<sup>11</sup> See Cvrček et al. 2010, 167. I am indebted to Jana Strnadová for extended discussion of the Czech data.

<sup>12</sup> The symbol ‘~’ notes free variation. In the nominative and accusative, there is no contrast between the masculine inanimate hard and soft declensions, and hence no variation in form.

tives inflect synthetically for comparative and superlative grade. A smaller class, including undeclinable adjectives such as *blond* ‘blond’, resort to a periphrastic strategy.

	‘old’	‘small’	‘late’	‘blond’
POS	starý	malý	pozdní	blond
CMP	starší	menší	pozdnější	<b>víc blond</b>
SUPER	nejstarší	nejmenší	nejpozdnější	<b>nejvíc blond</b>

**Table 9** Nominative masculine singular form of a sample of Czech adjectives

Turning now to overlapping classes, one finds instances of both situations documented above for arbitration between synthetic exponence strategies. Latin semi-deponent verbs are a case of heteroclisis: as exemplified in Table 10, while ordinary active verbs such as *monéo* ‘advise’ have synthetic forms in the *perfectum*, deponent verbs such as *vereor* ‘rever’ (like passive verbs) use analytic forms. Perfect semi-deponent verbs such as *audeo* ‘dare’ pattern like deponents in using periphrasis in the *perfectum*, while they otherwise pattern like ordinary active verbs; the imperfect semi-deponent *revertor* on the other hand patterns like a deponent verb in the *inflectum* but arbitrates in favor of synthesis, using exponents characteristic of normal active verbs, in the *perfectum*.<sup>13</sup>

		‘advise’	‘dare’	‘come back’	‘revere’
[−PERFECT]	PRS	monēō	audeō	revertor	vereor
	PST	monēbam	audēbam	revertēbar	verēbar
	FUT	monēbō	audēbo	revertēbor	verēbor
[+PERFECT]	PRS	monūī	<b>ausus sum</b>	revertī	<b>verītus sum</b>
	PST	monueram	<b>ausus eram</b>	reverteram	<b>verītus eram</b>
	FUT	monuerō	<b>ausus erō</b>	reverterō	<b>verītus erō</b>

**Table 10** Selected 1SG forms of 4 Latin second conjugation verbs

English adjectives, on the other hand, provide a convincing example of an overlapping class leading to overabundance. Although a few adjectives categorically use a synthetic or periphrastic strategy for comparative and superlative grade, the vast majority is compatible with both (Boyd, 2007; Aronoff & Lindsay, in press), and have been for centuries (Gonzalez-Diaz, 2008).

	easy	odd	friendly	dangerous
CMP	easier	odder ~ <b>more odd</b>	friendlier ~ <b>more friendly</b>	<b>more dangerous</b>
SUPER	easiest	oddest~ <b>most odd</b>	friendliest~ <b>most friendly</b>	<b>most dangerous</b>

**Table 11** Paradigm of four English adjectives

<sup>13</sup> This observation is independent of the debate as to whether deponents themselves should be seen as exhibiting a morphosyntactic mismatch (Stump, 2002; Hippisley, 2007) or as heteroclitics (Kiparsky, 2005; Walther, 2013).

To conclude this discussion, languages use the same strategies to arbitrate competition between alternative synthetic exponence and competition between synthesis and periphrasis, both within paradigms and across paradigms. This systematic similarity provides a strong argument for the view that periphrasis forms part of the inflection system. It also provides evidence against the view that periphrastic expression is generally preempted by synthetic expression: while that is sometimes the case, it is by no means necessary. Hence the distribution of periphrases can't be accounted for by assuming a general priority for synthesis, *contra* Poser (1992); Bresnan (2001b); Kiparsky (2005).

### 2.3 Ancillary lexemes are autonomous lexical items

Since ancillary lexemes typically are homophonous with ordinary lexemes, it is worth asking whether the ancillary use in a periphrastic construction and the main use of a lexeme can be equated. There are two main reasons for resisting the impulse to take them to be a single unit. First, ancillary lexemes typically only share *some* of the properties of the lexemes they evolved from. Second, ancillary lexemes typically have defective paradigms, as a consequence of the fact that they are used to express morphosyntactic features rather than convey semantic content.

#### 2.3.1 Partial overlap between ancillary lexemes and their diachronic sources

Since periphrastic inflection typically arises from the grammaticalization of syntactic constructions, ancillary lexemes tend to exhibit some similarity with the lexemes they emerged from. However that similarity is partial.

From the point of view of morphology, the paradigm of an ancillary lexeme may be more or less abnormal. Catalan has a periphrastic past tense based on the combination of present forms of *anar* 'go' with the infinitive of the main lexeme. However whereas the main verb *anar* exhibits a suppletive stem alternation in the present tense 1PL and 2PL, this alternation has been leveled out in the ancillary lexeme, as shown in Table 12 (Wheeler, 1979, 68-69).

	1 SG	2 SG	3 SG	1 PL	2 PL	3 PL
PRESENT	vaig	vas	va	anem	aneu	van
periphrastic PAST	<b>vaig anar</b>	<b>vas anar</b>	<b>va anar</b>	<b>vam anar</b>	<b>vau anar</b>	<b>van anar</b>

**Table 12** Partial indicative subparadigm of Catalan *anar* 'go'

Similarly, the Persian future is formed from a combination of a present form of *xâstan* 'want' followed by the short infinitive of the main lexeme (4a). The form of *xâstan* however is abnormal in not being marked with the imperfective prefix *mi-*. The absence of *mi-* reflects an older irregular conjugation of *xâstan* which has been abandoned for the full lexeme (4b), but retained for the ancillary lexeme.

- (4) a. Maryam in tâblo=râ      **xâh-ad**      **foruxt.**  
 Maryam this painting=DDO want.PRS-3SG sell[SINF]  
 'Maryam will sell this painting'      (Bonami & Samvelian, in press)

- b. Maryam mi-xâh-ad            in    tâblo=râ            be-foruř-ad.  
 Maryam UNBD-want.PRS-3SG this painting=DDO SBJV-sell.PRS-3SG  
 ‘Maryam wants to sell this painting.’

Although I have not found a completely convincing example of this, it is conceivable that the main verb serving as the diachronic source of an ancillary lexeme might lose all of its uses as a main verb.

Turning to syntactic behavior, it is customary to observe that ancillary lexemes have a different, often more restricted, distribution than the corresponding main verbs. Once again examples abound. In many varieties of English, the perfect auxiliary *have* is an auxiliary verb, but the main verb *have* is not (cf. *He hasn't come* vs. *%He hasn't any money*). In Persian, as (4) illustrates, the future auxiliary *xâstan* takes a nonfinite complement and must be adjacent to the main element, while the main verb *xâstan* takes a finite subjunctive complement and need not be adjacent to the embedded verb. Turning to a different type of case, Czech expresses past through a periphrase in the first and second person. The Czech periphrastic past combines forms of the copula used as an ancillary element with what is historically an *l*-participle (5). However there are two important differences between the copula and the ancillary element. First, whereas the true copula is a full word, the ancillary element is a second position clitic (see e.g. Franks & Holloway King, 2000, 91–97): witness the fact that the copula can start a clause (6), whereas the ancillary element rigidly occurs after one major constituent (7). Second, the copula has a third person form (8a), and the use of that form is obligatory (8b). However, this form cannot be used in the periphrastic expression of the past: rather, the third person past is synthetic, consisting of a bare *l*-participle unaccompanied by any ancillary element (9).

- (5) **Čeka-l**            **jsem**            na Jard-u.  
 wait-PST.PTCP[M.SG] be.PRS.1SG on Jarda-ACC.SG  
 ‘I was waiting for Jarda.’
- (6) Jsem            rád.  
 be.PRS.1SG happy[NOM.M.SG]  
 ‘I’m happy.’
- (7) a. Na Jard-u            **jsem**            **čeka-l**.  
       on Jarda-ACC.SG be.PRS.1SG wait-PST.PTCP[M.SG]  
       ‘I was waiting for Jarda.’  
 b. \* **Jsem čekal** na Jardu.<sup>14</sup>  
 c. \* Na Jardu **čekal jsem**.
- (8) a. Jard-a            je            rád.  
       Jarda-NOM.SG be.PRS.3SG happy[NOM.M.SG]  
       ‘Jarda is happy.’  
 b. \* Jard rád.
- (9) a. Čeka-l            na Jard-u.  
       wait-PST.PTCP[M.SG] on Jarda-ACC.SG  
       ‘He was waiting for Jarda.’  
 b. \* **Čekal je** na Jard-u.

<sup>14</sup> Initial position of the auxiliary is marginally possible in colloquial spoken Czech (Franks & Holloway King, 2000, 114), but definitely not in standard written Czech (Karlík et al., 1995, 648).

Finally, although there is evidently a diachronic connection between periphrases and their historical sources, there is generally no way of deriving synchronically the morphosyntactic content expressed by an ancillary element from the semantics of the corresponding main verb. Of course this point is a lot harder to argue convincingly without entering the details of the grammar of a particular language; however there is telling circumstantial evidence. As a case in point, cognate periphrases based on the present of *go* combined with an infinitive complement express the past in Catalan (see Table 12 above) and the near future in French (see Table 13 below): it is unclear how one could take both the expression of past and future to be variants of the expression of the same semantic content.

From this discussion I conclude that ancillary lexemes may be distinct from the corresponding main lexemes in all linguistic dimensions, and hence must be given separate lexical entries.

### 2.3.2 Partial defectivity of ancillary lexemes

Let us now turn to an examination of the shape of the paradigm of ancillary lexemes. The generalization that emerges is that ancillary lexemes typically exhibit inflection expected in some independent part of speech, although their paradigm may be defective to various degrees. At one end of the spectrum, some ancillary lexemes have the full paradigm appropriate for their category. For instance French perfect auxiliaries *avoir* and *être* have the same set of synthetic forms as all other nondefective verbs in the language. Sometimes the ancillary lexemes exhibit MOTIVATED DEFECTIVITY: the distribution of the ancillary lexeme is simply limited by the subpart of the paradigm of the main lexeme it serves to inflect. One clear example of this is the Czech future auxiliary exemplified in Table 8. Since the auxiliary realizes future tense, it does not have cells corresponding to past or present. To take another example, in Tundra Nenets, the postposition *nya* used to realize local cases in the dual (see Table 2) does not exhibit inflection for pronominal complements in that use, because that is incompatible with its function as an ancillary element.

In other cases, the synthetic paradigm of the ancillary lexeme contains more or less arbitrary gaps. One telling example is that of the French prospective (or ‘near future’) periphrase, based on the combination of the verb *aller* ‘go’ with an infinitive main verb. As Table 13 shows, the ancillary lexeme is found only in the present and past imperfective. However, there is no obvious motivation for the presence of gaps. Witness the fact that the idiomatic expression *être sur le point de*, literally ‘to be on the verge of’, which is a near paraphrase of the prospective periphrase, is available for all paradigm cells.

	periphrase	ordinary syntax
PRESENT	Il va se réveiller.	Il est sur le point de se réveiller.
FUTURE	*Il ira se réveiller.	Il sera sur le point de se réveiller.
SIMPLE PAST	*Il alla se réveiller.	Il fut sur le point de se réveiller.
PAST IMPERFECTIVE	Il allait se réveiller	Il était sur le point de se réveiller.
SUBJUNCTIVE	*qu’il aille se réveiller	qu’il soit sur le point de se réveiller
CONDITIONAL	*Il irait se réveiller	Il serait sur le point de se réveiller.
INFINITIVE	*aller se réveiller	être sur le point de se réveiller.

**Table 13** Comparison of periphrastic and ordinary constructions expressing the near future in French

At the other end of the spectrum, one finds cases where an ancillary lexeme has a single form. A case in point is the Bulgarian negative future periphrase (Popova & Spencer, 2013),

which is based on a combination of the negative 3SG present form of *imam* ‘have’ with a finite clause containing the main verb agreeing with the subject (10a). This is despite the existence of a full paradigm of negative present forms for *imam*, including crucially 1SG *njamam* (10b).

- (10) a. **Njama**            neprekāsno az    da    **xodja**            za xlab.  
 NEG.have[3SG] incessantly 1.SG COMP go[PRS].1SG for bread  
 ‘I will not be the one to go and buy the bread all the time.’
- b. \* **Njama-m**        neprekāsno az    da    **xodja**            za xlab.  
 NEG.have-1SG incessantly 1.SG COMP go[PRS].1SG for bread  
 (Popova & Spencer, 2013, 201)

### 2.3.3 Taking stock

In this section I have shown that ancillary lexemes are autonomous lexical items, distinct from the full lexemes that are their historical sources. First, they have their own lexical identity, with morphological, syntactic and semantic properties distinguishing them from the full lexemes they derive from. Second, as a class they differ from ordinary lexemes in being typically defective, the shape of their paradigm being conditioned by the expression of morphosyntactic content. An adequate theory of periphrasis should be flexible enough to accommodate such a multidimensional gradient of possibilities.

## 2.4 Periphrases need not be morphosyntactically noncompositional

A basic property of synthetic inflection is that inflectional exponents realize morphosyntactic properties of phrasal relevance. As a typical example, consider the Czech example in (11). The instrumental case suffix *-ou* on the noun *kniha* reflects the fact that the whole NP is instrumental, and that a semantic predicate roughly corresponding to the meaning of the preposition ‘with’ in English applies to the NP’s denotation. In this example, morphosyntactic information flows between the head and the phrase. But this is not the only possibility. Another common situation is for information to flow between the phrase and one of its edges (Miller, 1992; Halpern, 1995), as in English genitive marking (12): here the suffix *-s* realizes a property of the phrase *a friend of Mary’s* on a word embedded in its non-head daughter, but that crucially is the last word of the phrase.<sup>15</sup>

- (11) Praštil-a    ho            [ tou            strašně těžk-ou            knih-ou].  
 hit.PST-F.SG ACC.M.SG DEM.F.SG.INS horribly heavy-F.SG.INS book[F]-SG.INS  
 ‘She hit him with that incredibly heavy book.’

- (12) a friend of Mary’s hat

There is thus a kind of compositionality in the flow of morphosyntactic information in phrases: for any morphosyntactic feature expressed in inflection, the value carried by a

<sup>15</sup> The HEAD and EDGE situations are far from exhausting the typology of loci of realization of morphosyntactic properties. A spectacular example in Persian is discussed by Samvelian & Tseng (2010), where pronominal complements of a verb are realized on the edge of its least oblique remaining complement, irrespective of the position of that complement in the clause.

phrase is a function of the values carried by its parts, the rule used to combine them, and the identity of the particular feature under consideration.

Periphrases may disrupt this normal flow of information, leading to a kind of morphosyntactic non-compositionality (Ackerman & Stump, 2004). Familiar examples of perfect periphrases in Romance and Germanic illustrate: in *John has left the room*, the whole sentence is in the present perfect, but synthetic exponence on the auxiliary realizes a present non-perfect, and that on the main verb realizes a nonfinite form, thus a non-present.

In the general case, one finds various situations of information flow between parts of a periphrase and the construction as a whole.<sup>16</sup> The Czech periphrastic future, exemplified in Table 8, illustrates a case of canonical information flow. The auxiliary carries inflection that is appropriate for a future form—indeed, it is indistinguishable from the synthetic future of the copula—and no morphosyntactic content is shared between the ancillary and the main element (negation is expressed synthetically on the auxiliary and can't be realized on the main verb). When such an ideal situation does not obtain, one finds situations of DISTRIBUTED EXPONENCE (Ackerman & Stump, 2004): a morphosyntactic property of the whole may be realized by synthetic exponents on the ancillary element, on the main element, on both, or on neither. Consider again the Bulgarian negative future discussed above in (10). Negation is realized on the ancillary element only; subject agreement is realized on the main verb only; future is realized on neither: both verbs are present forms, and thus future is expressed by the use of the periphrastic construction itself, rather than by morphology on one of the elements it combines.

The Persian progressive provides a clear example of a situation of redundant synthetic exponence on the main and ancillary element (Bonami & Samvelian, in press)—what one might call periphrastic multiple exponence. In this periphrase, the ancillary and the main verbs realize redundantly tense, evidentiality, and agreement (unbounded aspect is overt only on the main element, due to a morphological peculiarity of the verb *dâštan*).

- (13) a. Maryam **dâr-ad**      tâblo=râ      **mi-foruš-ad.**  
 Maryam have.PRS-3SG painting=DDO UNBD-sell.PRS-3SG  
 'Maryam is selling the painting.'
- b. Maryam **dâšt**      tâblo=râ      **mi-foruxt.**  
 Maryam have.PST[3SG] painting=DDO UNBD-sell.PST[3SG]  
 'Maryam was selling the painting.'
- c. Maryam **dâšte-ast**      tâblo=râ      **mi-foruxte-ast.**  
 Maryam have.IND-3SG painting=DDO UNBD-sell.IND-3SG  
 'Reportedly, Maryam was selling the painting.'

Thus one may conclude that periphrasis involves different kinds of departures from the expected morphosyntactic information flow between heads and phrases: the synthetic morphology on the head of the phrase may correspond only in part with the morphosyntactic properties of the phrase it heads.

<sup>16</sup> This discussion presupposes that the determination of the head of the relevant construction can be made independently of the morphological expression of phrasal properties, by the sole examination of the syntactic distribution of the ancillary and main elements. Of course in many instances the relevant evidence is partial or lacking. Note the contrast between the our and Gregory Anderson's (2006) use of 'head': Anderson calls a periphrase AUX-headed (resp. LEX-headed) if and only if all phrasal properties that are realized inflectionally are realized on the ancillary (resp. main) element.

## 2.5 Periphrasis is independent of phrase structure

Bonami & Webelhuth (2013) emphasize another crucial empirical property that any satisfactory theory of the phenomenon of periphrasis needs to capture: the syntactic parts of a periphrase can stand in diverse phrase-structural configurations both across languages and within one language. The paper presents syntactic evidence that the perfect periphrases of German, English and French, despite all being based on a combination of a finite form of *have* with a past participle, have contrasting phrase structures: the two verbs combine in a verb cluster (= VC) in German (14a), as sister verbs in a flat VP in French (14b), and as the respective heads of two nested VPs in English (14c). Moreover, the components of periphrastic predicates can even be separated by clause boundaries in Persian (14d) and Bulgarian (14e), as argued in Bonami & Samvelian (in press) and Popova & Spencer (2013), respectively: in both cases, the main element is a finite verb, and shows no sign of being in a tighter syntactic relationship with the ancillary element than the head of a finite complement clause is with its governing verb.

- (14) a. *dass das Buch jemand [vc gekauft hat]*  
 COMP the book somebody buy.PST.PCPL have.PRS.3SG  
 ‘that somebody bought the book’ (German)
- b. *Paul [vp a lu ce livre ]*.  
 Paul have.PRS[3SG] read.PST.PCPL that book  
 ‘Paul read that book.’ (French)
- c. *Paul [vp has [vp read that book ]]*. (English)
- d. *[s Maryam dâšt [s madrase mi-raft ]]*.  
 Maryam have.PST[3SG] school IPFV-go.PST[3SG]  
 ‘Maryam was going to school.’ (Persian)
- e. *[s Njama neprekâsnato az [cp da xodja za xlab]]*.  
 have.NEG.3SG incessantly 1.SG COMP go[PRS].1SG for bread  
 ‘I will not be the one to go and buy the bread all the time.’ (Bulgarian)

In addition, in French, the perfect periphrase contrasts with the ‘near future’ periphrase discussed in section 2.3.2, where the main verb heads a VP. One piece of evidence for this difference is the distribution of pronominal affixes, which must be realized as prefixes to the ancillary element in the perfect, as shown in (15), but as prefixes to the main verb in the ‘near future’, as shown in (16).<sup>17</sup>

- (15) a. *Paul [ lui a parlé de toi ]*.  
 Paul 3SG.DAT have.PRS.3SG speak.PST.PTCP of 2SG  
 ‘Paul spoke to him about you.’
- b. \* *Paul [ a lui parlé de toi ]*.  
 Paul have.PRS.3SG 3SG.DAT speak.PST.PTCP of 2SG
- (16) a. *Paul [ va [ lui parler de toi ]]*.  
 Paul go.PRS.3SG 3SG.DAT-speak.INF of 2SG  
 ‘Paul will to him about you.’

<sup>17</sup> See Abeillé & Godard (2002) for detailed discussion of the contrasts between these two structures. Importantly, neither structure is reserved for periphrastic inflection: the flat structure is also characteristic of causative constructions, the nested structure is also characteristic of modal verbs.





Since Gazdar (1982) this has been taken as evidence that the two sentences have analogous structure, and that the same syntactic relation holds between *has* and *may* and the nonfinite verbs they combine with, both in the inverted sentences in (18) and in their non-inverted counterparts (19). Indeed, under the analysis of Ginzburg & Sag (2000), in both cases the nonfinite verb heads a VP which has the function of complement of the matrix verb. In fact, the matrix verbs in both pairs of sentences are realizations of the same schematic lexical entry:

$$(20) \left[ \begin{array}{l} \text{HEAD} \\ \text{VAL} \end{array} \left[ \begin{array}{l} \left[ \begin{array}{l} \textit{verb} \\ \text{AUX} \quad \pm \end{array} \right] \\ \left[ \begin{array}{l} \text{SUBJ} \quad \langle \text{NP} \rangle \\ \text{COMPS} \quad \langle \text{VP} \rangle \end{array} \right] \end{array} \right]$$

In inverted sentences, the matrix verb is realized as [AUX +] and in uninverted sentences as [AUX –]. This differential marking correlates with the phrasal constructions in which the verb can occur: uninverted verbs combine with their VP-complement in a *head-complement phrase* and form another VP they head. This VP then combines with the subject in a *head-subject phrase*. The inverted verb, in contrast, heads a *subject-auxiliary phrase*, in which it simultaneously combines with its subject and its VP-complement in a ternary configuration. Crucially for our present purposes, the relationship between the matrix verb and the embedded VP in (18)-(19) remains constant in terms of grammatical function. Thus, as long as the inflectional component specifies a head-complement relationship between an auxiliary verb and the VP headed by the main verb (as in (20)), the auxiliary and the VP can appear in any syntactic configuration that maintains that relationship.

This same point can be illustrated with a long-distance dependency construction. In the sentences below, the embedded VP is preposed. As before, it does not matter whether the matrix verb and the embedded verb jointly express an inflectional periphrase or not.

- (21) a. [VP Left the room ] [S I believe [S she has – ] ].  
 b. [VP Leave the room ] [S I believe [S she may -- ] ].

Here, despite its position, the preposed VP is as much a complement of the finite verb contained in the subordinate clause of (21) as it is in the in-situ construction in (19). In fact, the verbs *has* and *may* in (21) both instantiate the uninverted versions of the lexical entry (20) that was already used in the canonical sentences (19). The sentences differ from each other in that (21) realizes the nonfinite VP complement as a gap which is filled by the preposed VP. But in both (18a) and (21a), the auxiliary selects a VP-complement which is a projection of the main verb that forms an inflectional periphrase together with the auxiliary.

I thus conclude that the link between the main and ancillary elements in a periphrase are established by a grammatical functional relation rather than by constraints on constituent structure. The point just made about English clearly generalizes to languages where the integration of periphrasis in the inflectional paradigm is tighter; for instance, Bonami & Samvelian (in press) show that in the Persian perfect periphrase, whose paradigmatic distribution was discussed in section 2.2, the main element can be topicalized.

- (22) **Foruxte** [ fekr ne-mi-kon-am [ \_\_\_ **bâš-ad** in  
 sell.PRF.PTCP thought NEG-UNBND-do.PRS.1SG be.SBJV-3SG this  
 tâblo=râ]].  
 painting=DDO

‘I don’t think she has sold this painting.’ (Bonami & Samvelian, 2009, 33)

Saying that parts of a periphrase are linked by grammatical functional relation does not entail, of course, that the identification of the relevant grammatical function is always trivial. Where the periphrase is syntactically analogous to a non-periphrastic syntactic combination, as in the examples just discussed, the determination of the relevant function is easy. In other cases, the exact identity of the function may be more disputable. Consider the expression of polarity in Tundra Nenets (Wagner-Nagy 2011, Nikolaeva 2014, 213–219, 272–282.<sup>20</sup> Negation is expressed by inflecting the main element as a special nonfinite form, the connegative, and combining it with a negative auxiliary inflected for tense, mood, and agreement.

- (23) a. Wera ti-m tær’i xada-da-s’<sup>o</sup>  
 Wera reindeer-ACC DP kill-3SG>SG.OBJ-PST  
 ‘Wera killed the reindeer needlessly.’ (Nikolaeva, 2014, 216)
- b. Pet’a n’ī-da ti-m xada-q  
 Petya NEG-3SG>SG.OBJ reindeer-ACC kill-CONNeg  
 ‘Petya didn’t kill the reindeer.’ (Nikolaeva, 2014, 218)

In this case, the finiteness contrast between the two forms clearly favors an analysis where the ancillary element is the head and the main element is its complement. However, the word order pattern exhibited by the construction provides conflicting evidence: Tundra Nenets is predominantly verb final, and thus the position of the ancillary element is unexpected (Wagner-Nagy, 2011, 94). Example (24) shows this to be the case for combinations of a finite head and a nonfinite complement.

- (24) Pet’a ti-m xada-wa-n<sup>o</sup>h xərwa<sup>o</sup>  
 Petya reindeer-ACC kill-IMPf-AN-DAT want.3SG  
 ‘Petya wants to kill a reindeer.’ (Ackerman & Nikolaeva, 2014, 125)

This example clearly shows that periphrases need not participate in a syntactic pattern that is otherwise attested in the language, a point also made by Lee & Ackerman (submitted). This however does not invalidate the observation that parts of periphrases are linked by a grammatical functional relation; rather, it shows that the familiar situation where the correct syntactic analysis for some construction is underdetermined by the empirical data extends to periphrastic constructions.

## 2.7 The challenge

The upshot of the discussion so far is that inflectional periphrases lead a double life. With one foot they firmly stand within the inflection system, in particular within the paradigms of lexemes. But with their other foot, they equally firmly stand within syntax, their parts being linked by a grammatical functional relation that may or may not permit them to be separated from each other within the sentence. Clearly, a theory of periphrasis needs to capture this double life if it is to count as satisfactory.

At this point we encounter the conceptual and technical hurdle that, within a lexicalist view of grammar, the optimal theories of syntax and (inflectional) morphology each make

<sup>20</sup> I am indebted to Farrell Ackerman for pointing out this dataset and its relevance. Periphrastic expression of the future in Persian presents similar problems.

use of different designs. Thus, I will follow what we take to be the majority of working morphologists at this point in assuming that inflection systems are best described in word-and-paradigm approaches (see among many others Robins (1959); Hockett (1967); Matthews (1972); Anderson (1992); Zwicky (1992); Aronoff (1994); Stump (2001); Blevins (2006)). In contrast, I assume with many syntacticians that syntactic systems are best described in phrase-structural terms, as incrementally built combinations of signs (see among many others Harman (1963); Bresnan (1978); Gazdar et al. (1985); Pollard & Sag (1987); Steedman (1996)). To put it in the terms of Stump (2001): inflection is inferential-realizational, syntax is lexical-incremental. The hurdle mentioned above now consists in adjusting the desirable theories of syntax and inflection in such a manner that the double life of periphrasis in both grammatical domains is captured by the overall framework.

A number of proposals have been made in pursuit of such a framework, but none of them is completely satisfactory. Probably the earliest attempt within formal grammar was contained in Ackerman & Webelhuth (1998). The authors worked out the rudiments of a theory of periphrasis that permitted auxiliary-main verb and verb-particle combinations to be treated as lexical representations whose parts could be mapped into phrase structure separately. However, the syntactic theory in that work was too inflexible, as Müller (2002, 392–401) argues at length. In particular, it was unable to handle extraction of pieces of periphrastic predicates; and, by treating auxiliatation as a specific grammatical function, was unable to account both for the syntactic parallels between periphrases and non-periphrastic constructions and for the fact that different periphrases may have different syntactic properties within the same language. The theories of Spencer (2003), Booij (2010) and Blevins (forthcoming), where periphrases are generated directly by phrase structure schema, suffer from similar problems, at least in the absence of an explicit interface with a theory of extraction and variable word order. Sadler & Spencer (2001), Ackerman & Stump (2004) and Ackerman et al. (2011) display the opposite problem. None of these theories constrain the syntactic relations that may obtain between the two pieces realizing a periphrastic predicate and thus all leave many details of the analysis of periphrasis open. Bonami & Samvelian (2009, in press) present an analysis of inflectional periphrasis in Persian. While it does not present the same problems mentioned earlier, it violates two desirable design properties: its morphological component fails to be completely realizational, as noted by Stump & Hippisley (2011), and it entails a view of the lexicon that does not adhere to the Principle of Lexical Modification (Ackerman et al., 2011, 326), whereby lexical properties such as argument structure cannot be altered by inflection. Bonami & Webelhuth (2013) address these concerns, and deal with the phrase structural diversity of the realization of periphrasis illustrated above by contrasting a general *synthetic inflection* construction with multiple periphrastic constructions. That same approach is applied to Sanskrit by Stump (2013). However, Bonami and Webelhuth's theory has at least two unsatisfactory aspects. First, Pāṇini's Principle cannot arbitrate between synthetic and periphrastic realization. And second, the theory cannot deal with periphrases that rest on the modifier-head relation.

The present work constitutes an attempt to develop earlier theories with the goal of overcoming the shortcomings just mentioned. Towards that end, I will draw on similarities between periphrastic predicates and certain collocations.

### 3 The syntactic status of periphrases

#### 3.1 The analogy between periphrasis and flexible idioms

From a syntactic point of view, periphrases have three key properties that were highlighted in sections 2.4, 2.5 and 2.6:

- (25) a. Periphrases can be morphosyntactically non-compositional: the morphosyntactic features conveyed by a periphrase may be different both from the morphosyntactic features conveyed by the main element and by those conveyed by the ancillary element.<sup>21</sup>
- b. Periphrases are syntactically flexible: the two parts of a periphrase may stand in more than one phrase-structural relation.
- c. Periphrases are syntactically linked: the two parts of a periphrase are tied by a grammatical functional relation, such as the head-complement or the head-adjunct relation.

These three properties are highly reminiscent of the literature on idioms and other phraseological expressions, as already noted by Spencer (2003) and Booij (2010). Property (25a) is clearly the morphosyntactic analogue of the defining property of idioms as multi-word expressions with non-compositional meaning. Properties (25b) and (25c) also have analogues in the more specific class of syntactically flexible idioms.

Since the seminal work of Wasow et al. (1984), Fillmore et al. (1988) and Nunberg et al. (1994), it is well established that the class of idioms encompasses various subclasses. Here I adopt the classification and descriptive vocabulary of Weibelhuth et al. (forthcoming). First, a basic distinction must be drawn between SYNTACTICALLY FROZEN and SYNTACTICALLY FLEXIBLE idioms. Syntactically frozen idioms such as *kick the bucket* do not allow for any syntactic variation, although they allow for morphological variation: the verb *kick* may take any of its inflected forms (26), but the phrase-structural relationship between *kick*, *the* and *bucket* cannot be disrupted by any syntactic operation (27).<sup>22</sup>

- (26) a. He just kicked the bucket.  
 b. He may kick the bucket at any time.  
 c. His kicking the bucket caused great concern.
- (27) a. \* The bucket was finally kicked.  
 b. \* It was the bucket that he kicked.  
 c. \* The bucket he kicked was spectacular.

Syntactically flexible idioms, on the other hand, do authorize various degrees of freedom in the syntactic relation between idiom parts. Thus in the idiom *spill the beans*, the verb may be passivized (28a), but the NP may not be extracted (28b,c). The idiom *pull strings* on the other hand does not obey this restriction (29b,c), and even allows for the occurrence of the idiomatic NP *strings* in a syntactic context where *pull* is not present, as long as the idiomatic meaning has been established in the previous discourse (29d).

<sup>21</sup> The Persian progressive, discussed in section 2.4, is an example of the limiting case where the main element on its own is compatible with the content expressed by the periphrase. In that case the sole role of the periphrastic combination is the *exclusion* of other readings—specifically, while the synthetic form on its own is compatible with either a progressive or a habitual reading, the periphrase only conveys the progressive.

<sup>22</sup> This is well known to be a slight idealization, since *kick the bucket* allows for limited, metalinguistic internal modification in expressions such as *kick the proverbial bucket*.

- (28) a. The beans appeared to be spilled when he opened his mouth. (Horn, 2003, 262)  
 b. \* The beans that Joe spilled caused us a lot of trouble. (Horn, 2003, 262)  
 c. \* Which beans did Harry spill? (Horn, 2003, 264)
- (29) a. Strings seem to be pulled every time he applies for a promotion. (Horn, 2003, 261)  
 b. We were surprised at the strings that were pulled to get Joe's promotion. (Horn, 2003, 261)  
 c. How many strings did he pull to get the promotion? (Horn, 2003, 261)  
 d. Pat and Chris graduated from law school together with roughly equal records. Pat's uncle is a state senator, and he pulled strings to get Pat a clerkship with a state supreme court justice. Chris, in contrast, didn't have access to any strings, and ended up hanging out a shingle. (Wasow et al., 1984, 93)

Going back to inflectional periphrases, the licit syntactic relationships between the ancillary and the main element appear to fall somewhere between those observed for *spill the beans* and *pull strings*: as with *pull strings*, there does not need to be a local phrase structural relation between the two parts of a periphrase, as they can be set apart by extraction. On the other hand, as with *spill the beans*, there must be a single, invariable grammatical function linking the two parts: the locality of the relation can be disrupted only by syntactic operations such as extraction or coordination, which do not affect grammatical functions; and parts of a periphrase never license each other across an intervening predicate.

It may thus be concluded that the combinatory relation between parts of a periphrase closely resembles that between parts of a syntactically flexible idiom. Of course the grammatical status of the two types of multi-word expressions is otherwise very different: while idioms carry semantic content, and are thus multi-word equivalents of lexemes, periphrases carry morphosyntactic content, and are thus multi-word equivalents of inflected words. However, the similarity is close enough that analytic techniques innovated for the treatment of idioms can be redeployed to account for periphrases.

### 3.2 Reverse selection

As Nunberg et al. (1994) argue forcefully, the syntactic flexibility of idioms such as *spill the beans* or *pull strings* makes it necessary for idiom parts to be given autonomous lexical entries separate from those of the corresponding non-idiomatic lexemes. In such a context, one main challenge of a theory of idioms is to avoid overgeneration: if idiomatic *spill* and idiomatic *beans* have their own lexical entries, how does one make sure that the two sentences in (30) are ungrammatical? Although there is more than one possible answer to this question, one fruitful possibility is to assume that the two words stand in a relation of mutual selection: idiomatic *spill* selects for the lexical identity of (the head of) its theme argument, while idiomatic *beans* selects for the lexical identity of the verb that takes it as an argument.

- (30) a. \* John spilled the secret.  
 b. \* John revealed the beans.

Pushing the analogy between flexible idioms and periphrases one step further, the same question (how does one ensure that the two elements in an inflectional periphrase occur together?) may be given the same answer: the ancillary element and the main element stand

in a relation of mutual selection. The problem then is to embed the analysis of periphrastic constructions within a framework that allows for such relations of mutual selection to be stated. Over the last decade, Manfred Sailer and colleagues (Sailer, 2000; Richter & Sailer, 2003; Soehn & Sailer, 2003; Soehn, 2006; Richter & Sailer, 2010) have developed a general HPSG theory of collocation that allows individual lexical entries to place collocational requirements on words or phrases in their environment. Although the theory is by no means limited to the treatment of VP idioms (see e.g. Richter & Sailer (2003) on cranberry words, Richter & Soehn (2006) on negative polarity items, Richter & Sailer (2010) on phraseological clauses), it allows among other things for the lexical entry of idiomatic *beans* to specify a requirement that it head an NP selected as the theme argument by the idiomatic verb *spill*. Here I build on this line of work to propose a specific implementation of mutual selection relations in inflectional periphrases. The proposal adopted here is clearly too restricted to account for the full range of collocational requirements within grammar as a whole, but is sufficient to deal with the issue at hand while being formally simpler than the full theory of Sailer and colleagues.

The central analytic device is the notion of REVERSE-SELECTION. Intuitively, reverse selection is the situation where some lexical item places a selectional requirement on another item in a way that goes in the opposite direction from ordinary selection: a complement selects properties of the head, rather than the head selecting properties of the complement; a head selects properties of an adjunct, rather than the other way around; and so on. Specifically, reverse selection is defined as parasitic on normal selection (31). The analysis of inflectional periphrasis then relies on the two assumptions in (32).

(31) *Reverse selection* (informal definition)<sup>23</sup>

A reverse selection requirement  $\varphi$  carried by a word  $w_1$  is satisfied if and only if  $w_1$  is syntactically selected by a word  $w_2$  and  $w_2$  verifies property  $\varphi$ .

- (32) a. Ancillary elements select morphosyntactic properties of the main element through normal syntactic selection, as specified in the ancillary lexeme's lexical entry.  
 b. Main elements realize morphosyntactic content partly by synthetic inflection, partly by placing reverse selection requirements ensuring the presence of an appropriate ancillary element.

Let us review the consequences of these assumptions on a concrete example. In the interest of readability I discuss the English perfect periphrase, although parallel analyses hold for the other periphrastic constructions discussed in section 2, modulo differences in phrase structure. In accordance with (32), and as shown in Figure 2, I assume that the main verb reverse-selects the auxiliary *have*, which in turn selects for a past participial complement.

More precisely, the construction is licensed by a rule of periphrastic exponence which can be informally stated as follows:

- (33) To realize the perfect of lexeme  $l$ , use a word whose form is that of the past participle of  $l$ , and which carries a reverse selection requirement for the tense, finiteness and appropriately agreeing form of the auxiliary *have*.

In the present perfect, the morphosyntactic content to be realized consists of present tense, perfect aspect, and third person singular agreement. (33) specifies that this is done by realizing a main element whose morphological form is that of a past participle, and which

<sup>23</sup> A more precise definition of reverse selection in the vocabulary of HPSG can be found in the appendix.

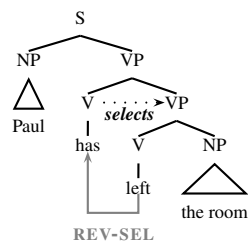


Fig. 2 A simple example of the English perfect periphrase

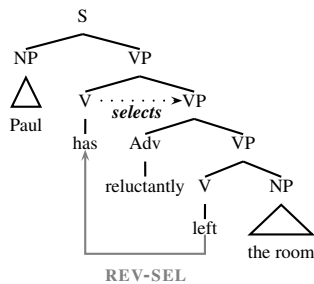


Fig. 3 Adverb intervening in an English perfect periphrase

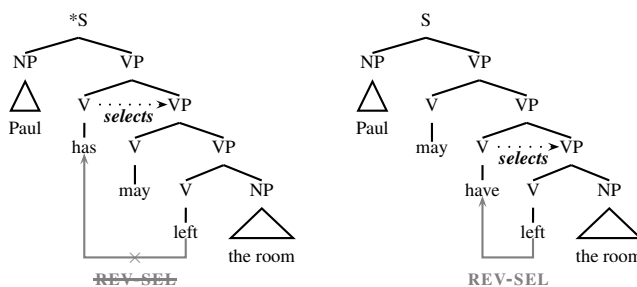


Fig. 4 Interaction between periphrasis and other complement structures

carries a reverse selection requirement for the auxiliary *have* in the present third singular. This requirement is obviously verified in the syntactic configuration in Figure 2.

The definition of reverse selection is flexible enough to authorize an appropriate but limited amount of syntactic flexibility, depending on the precise definition of the notion of selection. Thus adverbs occurring between the two verbs are licensed: in Figure 3, despite the presence of the adverb, the main verb *left* is still the head of the embedded VP, and thus selected by the auxiliary. On the other hand, the left hand side configuration in Figure 4 is ungrammatical, because there is no direct selection relation between *has* and *left*: here *left* is selected by *may* and *may* is selected by *has*, but *left* is not selected by *have*. This is in contrast with the right hand side configuration, where a local selection relationship obtains between the bare infinitive form of the auxiliary (*have*) and the participle (*left*), which jointly form the perfect bare infinitive of *leave*, the appropriate form for the complement of *may*.

Likewise, coordination of participles is predicted to be grammatical: in Figure 5, *close* and *left* are respectively partial realizations of the present perfect of the lexemes *close* and *leave*, and both carry a reverse selection requirement for the auxiliary *has*. This requirement



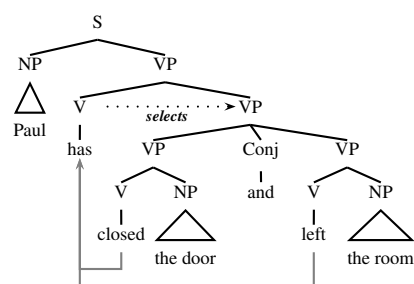


Fig. 5 Syntactic analysis of a periphrase with coordinated main elements

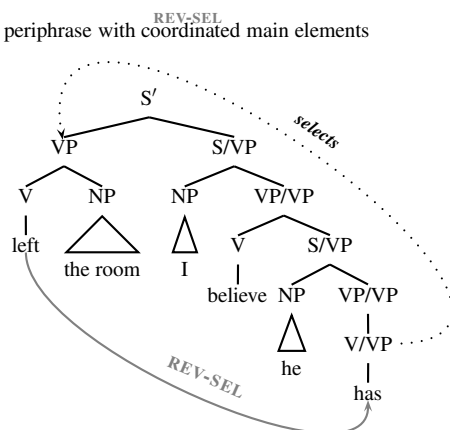


Fig. 6 Syntactic analysis of a periphrase with an extracted main element

is satisfied by the presence of the auxiliary as a local selector of the coordination of two phrases headed by the two main words.

Finally, in combination with standard HPSG assumptions on extraction structures, the current approach correctly captures the possibility of extracting the main element in a periphrase. In the HPSG approach to extraction, fillers are licensed through the percolation of the SLASH feature, which ensures that the distant filler is interpreted as satisfying the selectional requirements imposed by the head at the extraction site. Thus in a sentence such as the one in Figure 6, *left the room* counts as being selected by the auxiliary *has* exactly in the same sense as it does in Figure 2—technically, in both cases it is the head of a phrase whose LOCAL value occurs on the auxiliary's argument structure list.<sup>24</sup>

To sum up, the use of reverse selection requirements correctly captures the syntactic distribution of the parts of a periphrase: the main element is required to be in the direct syntactic dependence of the ancillary element, and this dependence is defined in terms of grammatical functions, rather than in terms of phrase-structure configurations. This directly captures the two key syntactic properties of inflectional periphrases discussed in sections 2.5 and 2.6. The theory of periphrasis outlined in the previous section also is clearly agnostic concerning part of speech: the only requirement for the theory to be applicable is for the main element to be syntactically selected by the ancillary element. Thus the analysis extends directly and appropriately to the nominal and adjectival domains. In Tundra Nenets periphrastic nouns (see Table 2), the main element is a noun selected as a complement by the postposition *nya*.

<sup>24</sup> In Figure 6 I conservatively assume that bare verbs head non-branching VPs. This is by no means a necessary assumption, see e.g. Abeillé & Godard (2000); Müller (2013).

In Ingush periphrastic attributive adjectives (see Table 3), the main element is a predicative adjective selected as a (predicative) complement by the present participle of the copula; in more familiar Czech or English periphrastic adjectives (see Tables 9 and 11), the main element is a positive adjective selected through the modifier-head relation by the appropriate degree adverb.

#### 4 The inflectional status of periphrases

In this section we address the distinctly inflectional property of paradigm integration discussed in section 2.2.

##### 4.1 Reverse selection as exponence

Since Sadler & Spencer (2001), theoretical linguists working on inflectional periphrasis have attempted to capture the intuition that periphrasis amounts to syntactic exponence of morphosyntactic features: just as the suffix *-s* in English is the exponent of third singular subject agreement in the present, the exponent of perfect aspect is the combination of a past participle with a form of the auxiliary *have*. The main appeal of this intuition is that it allows one to see periphrasis as part of the inflection system, and thus to account for the paradigmatic organization of the arbitration between synthesis and periphrasis.

While this is a powerful intuition, its concrete execution in a realistic grammatical framework has proven elusive (Ackerman & Webelhuth, 1998; Ackerman & Stump, 2004; Ackerman et al., 2011; Blevins, forthcoming; Bonami & Samvelian, 2009; Bonami & Webelhuth, 2013; Bonami & Samvelian, in press). Arguably, the difficulty stems from the fact that periphrasis seems to go against standard assumptions concerning the morphology–syntax interface. First, within lexicalist frameworks, it is assumed that inflectional morphology outputs syntactic atoms. On the face of it, periphrases are not atoms for syntax: the interface must thus be revised. One obvious revision is to allow for morphology to output either phrases or sequences. However, neither will do: as Bonami & Webelhuth (2013) highlight, whether the two parts of a periphrase form a phrase (or indeed are adjacent) is a parochial syntactic matter. In some languages they do, in others they don't; the possibility of periphrasis should not be sensitive to that parameter. Second, there is a consensus among morphologists that the organization of inflection systems rests in part on paradigmatic relations, in the form of comparisons of alternative inflection strategies (appealed to in various guises under names such as the Elsewhere condition (Kiparsky, 1973), disjunctive rule ordering (Anderson, 1992), the subset principle (Halle & Marantz, 1993), or Pāṇini's principle (Stump, 2001)). This is felt as both empirically unavoidable and computationally innocuous, since the search space for inflectional alternatives is finite and small. On the other hand, most formally explicit surface-oriented syntactic frameworks avoid relying on the online comparison of alternative syntactic strategies,<sup>25</sup> with three motivations: paradigmatic aspects of organization are much more limited in syntax than in morphology, analytic techniques that avoid direct comparison of alternatives are readily available (Malouf, 2003), and comparison of alternatives is computationally untractable over syntactic domains (Johnson & Lappin, 1999; Kuhn, 2003). This set of observations presumably explains at least part of the reluctance on the part of formal grammarians to take inflectional periphrasis at face value,

<sup>25</sup> A notable exception is OT-LFG as developed in Bresnan (2001a) and Kuhn (2003).

and the continuing insistence on attempting to reduce periphrasis to ordinary syntax by ignoring its paradigmatic aspects (see among many others Abeillé & Godard (2002), Müller (2002, 2010)).

The view of periphrasis proposed here provides a novel solution to this problem. Under the current proposal, periphrasis is not literally syntactic exponence: rather, periphrasis amounts to exponence of a morphosyntactic feature by a reverse selection requirement. For instance, rule (33) explicitly states that the perfect in English is realized by a reverse selection requirement carried by the participle. Thus the statement of the dependency involved in periphrasis is local to the main element, just as the statement of synthetic inflection rules is local to the word carrying their synthetic exponent. This has three conceptual advantages.

First, a reverse selection requirement is a kind of collocational requirement, and collocational requirements are an unavoidable feature of natural language grammars, that is needed independently of periphrasis to account for various kinds of lexical dependencies, as discussed in section 3.1. Thus the postulation of reverse selection requirements in periphrases does not in any way extend the descriptive power of a realistic grammatical framework: the only extension is to allow for different inflected forms of the same lexeme to have different collocational requirements, while usually collocational requirements are thought of as being attached to the lexeme itself. Second, periphrastic inflection rules can be stated as constraints on words. Thus the treatment of periphrasis does not entail any deep modification of the morphology–syntax interface: the role of inflection still is to output syntactic atoms. Third and finally, periphrastic inflection rules play the same general role as synthetic inflection rules, as partially constraining the relation between the morphosyntactic features expressed by words and the forms occurring in sentences. The only difference is that synthetic inflection expresses features locally on the word, whereas periphrastic inflection expresses them as conditions on the context of occurrence. Given that the two kinds of inflection strategies are of the same formal nature, and divide up a finite domain of feature value combinations, paradigmatic arbitration between synthesis and periphrasis can be accounted for in just the same way as arbitration between inflection strategies, without running the risk of seeing competition between alternatives spill over into phrasal syntax.

#### 4.2 A paradigm function approach to reverse selection as exponence

Now that I have conceptually motivated the view that periphrasis amounts to the use of a collocational requirement for the purposes of exponence, it remains to be seen how this idea can be implemented in a formal theory of inflection. I do so by sketching a version of Paradigm Function Morphology (Stump, 2001) where the paradigm function outputs collocational requirements in addition to phonological shapes.

##### 4.2.1 *Paradigm Function Morphology*

I briefly illustrate the workings of Paradigm Function Morphology on the basis of a slightly simplified version of Bonami & Samvelian's (in press) analysis of Persian conjugation, applied to examples from Table 5. A subset of the REALIZATION RULES for Persian are shown in (34). Rules (34a) and (34b) are rules of stem selection, taking care of the fact that there is an arbitrary relation between stem allomorphs in Persian, so that each lexeme contains a specification for two stems in its lexical entry. The rest of the rules are rules of EXPONENCE, introducing individual affixes in the context of some morphosyntactic property set to be re-

alized. For instance, (34c) states that for all verbs ( $V$ ), unbounded aspect in the indicative mood ( $\{\text{MOOD } ind, \text{ASP } unbd\}$ ) may be realized by prefixation of *mi-* on the input string  $X$ .

- (34) a. I,  $X_V, \sigma : \{\text{VFORM } fin\} \rightarrow X$ 's basic stem  
 b. I,  $X_V, \sigma : \{\text{VFORM } fin, \text{TNS } pst\} \rightarrow X$ 's second stem  
 c. II,  $X_V, \sigma : \{\text{MOOD } ind, \text{ASP } unbd\} \rightarrow miX$   
 d. III,  $X_V, \sigma : \{\text{POL } -\} \rightarrow naX$   
 e. III,  $X_V, \sigma : \{\text{MOOD } ind, \text{ASP } unbd, \text{POL } -\} \rightarrow neX$   
 f. V,  $X_V, \sigma : \{\text{EVID } ind\} \rightarrow Xe$   
 g. V,  $X_V, \sigma : \{\text{PER } 1, \text{NB } sg\} \rightarrow Xam$   
 h. V,  $X_V, \sigma : \{\text{PER } 2, \text{NB } sg\} \rightarrow Xi$   
 i. V,  $X_V, \sigma : \{\text{TNS } prs, \text{PER } 3, \text{NB } sg\} \rightarrow Xad$   
 j. V,  $X_V, \sigma : \{\text{TNS } pst, \text{EVID } ind, \text{PER } 3, \text{NB } sg\} \rightarrow Xast$

A crucial design feature of Paradigm Function Morphology is that realization rules are segregated into RULE BLOCKS, which serve to specify which rules stand in paradigmatic opposition. Rules (34d) and (34e) belong to the same block III, and hence are mutually exclusive. On the other hand, rules (34c) and (34e) are not in the same block, and thus may jointly participate in the exponence of unbounded aspect for negative indicative paradigm cells. The choice of the appropriate rule within a block is arbitrated by Pāṇini's principle: the rule with the more specific scope is used. In the case at hand, when inflecting for an unbounded indicative form, both rules (34d) and (34e) are applicable, but since (34c) has a more specific scope, by Pāṇini's principle, only that rule can be used.<sup>26</sup> When the grammar of a language provides no rule within a block for the realization of some morphosyntactic property set, a universal IDENTITY FUNCTION DEFAULT (IFD) rule ensures that the output of a block is identical to its input. As a consequence, the absence of any rule in block V for inflecting past verbs in the 3SG indicates zero exponence of person and number in the past, rather than defectivity.

Example (35) illustrates the notation  $[B : \langle X, \sigma \rangle]$ , which denotes the output of the most specific rule in some rule block B for realizing the morphosyntactic property set  $\sigma$  on the input string  $X$ . In the case at hand, we are taking the input form *xarid* and looking for the output of the most specific rule in block III realizing an indicative third singular form with unbounded aspect. The most specific rule for this property set is (34e), and hence the output is *nemixarid*.

- (35)  $[III : \langle mixarid, \{\text{MOOD } ind, \text{ASP } unbd, \text{POL } -, \text{PER } 3, \text{NUM } sg\} \rangle]$   
 $= \langle nemixarid, \{\text{MOOD } ind, \text{ASP } unbd, \text{POL } -, \text{PER } 3, \text{NUM } sg\} \rangle$

The final crucial ingredient of a PFM analysis is the specification of the PARADIGM FUNCTION. By definition, the paradigm function for a language is the function that associates each lexeme and appropriate complete morphosyntactic property set with the form filling the corresponding cell in the lexeme's paradigm. Although there are many ways this function can be specified, in PFM this is usually done through clauses such as the one in (36). This indicates the default inflection strategy for verbs, which consists of successively applying the narrowest rules in blocks I to V.

<sup>26</sup> I follow Bonami & Stump (forthcoming) in assuming that situations where two rules are unordered by specificity while being more specific than all other rules in the block lead to overabundance: there is more than one optimal exponence strategy.

- (36) If  $l$  is a verb and  $s$  is  $l$ 's basic stem, then for any morphosyntactic property set  $\sigma$  corresponding to a cell in verbal paradigms,

$$\text{PF}(l, \sigma) = [\text{v} : [\text{IV} : [\text{III} : [\text{II} : [\text{I} : \langle s, \sigma \rangle ]]]]]].$$

Example (37) summarizes the derivation of an inflected Persian verb using (36) and the rules in (34).

- (37) For lexeme *xaridan* with basic stem  $s = \text{xar}$ , second stem *xarid*, and property set  $\sigma = \{\text{MOOD } ind, \text{ASP } unbd, \text{POL } -, \text{PER } 3, \text{NUM } sg\}$ :

$$\begin{aligned} [\text{v} : [\text{IV} : [\text{III} : [\text{II} : [\text{I} : \langle s, \sigma \rangle ]]]]] &= && \text{by rule (34b)} \\ [\text{v} : [\text{IV} : [\text{III} : [\text{II} : \langle \text{xarid}, \sigma \rangle ]]]] &= && \text{by rule (34c)} \\ [\text{v} : [\text{IV} : [\text{III} : \langle \text{mixarid}, \sigma \rangle ]]] &= && \text{by rule (34e)} \\ [\text{v} : [\text{IV} : \langle \text{nemixarid}, \sigma \rangle ]]] &= && \text{by IFD} \\ [\text{v} : \langle \text{nemixarid}, \sigma \rangle ] &= && \text{by IFD} \\ \langle \text{nemixarid}, \sigma \rangle &= && \end{aligned}$$

The paradigm function is often specified through a disjunction of multiple clauses with the same format as (36). In that situation, the choice of the appropriate clause to choose for a given lexeme and property set is decided on the basis of Pāṇini's principle (see e.g. Stump 2006). I will illustrate the situation by adding a second clause (38), which is a statement of the directional syncretism noted in Table 5, and refers the realization of any form of the present perfect to the corresponding form of the evidential bounded past. The notation ' $\sigma! \tau$ ' denotes the superset of  $\tau$  that also contains all feature value pairs of  $\sigma$  that are not incompatible with  $\tau$ .<sup>27</sup>

- (38) If  $l$  is a verb, then for any property set  $\sigma \supseteq \{\text{MOOD } ind, \text{TNS } prs, \text{PRF } +\}$ ,

$$\text{PF}(l, \sigma) = \text{PF}(l, \sigma! \{\text{MOOD } ind, \text{TNS } pst, \text{ASP } bnd, \text{EVID } indir, \text{PRF } -\}).$$

Although clauses in the definition of the paradigm function are presented in prose, these systematically mention a class of lexemes and a description of morphosyntactic property sets to be realized. These jointly define the scope of the clause. Pāṇini's principle then applies in the usual way: clause  $\alpha$  is more specific than clause  $\beta$  if either the class of lexemes  $\alpha$  mentions is a subset of the class  $\beta$  mentions or  $\alpha$  realizes a superset of the set of morphosyntactic properties realized by  $\beta$ . In the case at hand, clause (38) mentions a specific morphosyntactic property set whereas (36) doesn't, and both clauses apply to all verbs. Hence Pāṇini's principle will arbitrate in favor of (38) wherever is it applicable. Thus the negative present perfect 3SG form of *xaridan* will be referred to its negative past imperfective indirect 3SG form, which by application of (36) will be derived as *nemixarideast*.

#### 4.2.2 Adding periphrasis to PFM

With these preliminaries out of the way, we can turn to the issue of supplementing PFM with rules of periphrastic exponence. I argue that these rules should be stated at the level of the

<sup>27</sup> Notice that Bonami & Samvelian (2009) obtain the effect of (38) using a block-internal rule of referral rather than an implicative statement in the definition of the paradigm function. See Bonami & Stump (forthcoming) for a discussion of the differences between these two approaches to directional syncretism.

paradigm function, rather than at the level of rule blocks. First, periphrasis is an alternative to synthetic exponence as a whole, rather than the use of a specific affix within a rule block. Second, rules of periphrastic exponence systematically entail the use of a specific shape for the main element, through referral to a distinct paradigm cell. This shape is then licensed by ordinary synthetic inflection.

I thus propose that rules of periphrastic exponence be implemented as clauses in the definition of the paradigm function. In general then, a paradigm function does not output a paradigm cell, but a pair of a paradigm cell and a set of reverse selection requirements.<sup>28</sup> The format is exemplified by the rule for the Persian perfect given in (39). This states that inflection of a verb for the perfect is done by, on the one hand, selecting as a phonological form the shape  $\varphi$  of the corresponding perfect participle, and, on the other hand, reverse selecting for the appropriate bounded aspect nonperfect positive form of the auxiliary *budan*. The reverse selection requirement itself is written as a pair of a lexeme identifier and a morphosyntactic property set, whose syntactic interpretation will be discussed in the next section.

(39) If  $l$  is a verb, then for any property set  $\sigma \supseteq \{\text{PRF } +\}$ ,

$$\text{PF}(l, \sigma) = \langle \varphi, \sigma, \{ \langle \textit{budan-prf}, \sigma! \{ \text{ASP } \textit{bnd}, \text{PRF } -, \text{POL } + \} \} \} \rangle,$$

where  $\text{PF}(l, \sigma! \{ \text{FORM } \textit{part} \}) = \langle \langle \varphi, \sigma! \{ \text{FORM } \textit{part} \} \rangle, \emptyset \rangle$ .

Since rules of periphrasis are clauses in the definition of a paradigm function, they participate in Pāṇinian competition on a par with other clauses in that definition. The case at hand illustrates the interesting situation where a rule of periphrasis is both more specific than the general rule for synthetic inflection (36), and less specific than the implicative rule for present perfects (38). This captures correctly the place of periphrasis in the system, as specific to the perfect but excluded in the present. Other situations discussed in section 2.2 are also readily captured by the current proposal. As representative examples, I outline the contrasting analyses of periphrastic expression of comparative grade in Czech, French and English. As we saw in Table 9, in Czech the expression of comparative grade is synthetic in general and periphrastic for a few subclasses, including the class of undeclinables. This is captured by the two statements in (40). The rule of periphrasis in (40b) states that undeclinable adjectives have comparative forms whose shapes are identical to those of the corresponding positive grade forms, but which reverse-select for the adverb *víc* in the comparative grade. The scope in (40b) is narrower than that of (40a), both in terms of lexeme classes (restricted to undeclinables) and morphosyntax (restricted to comparative grade). Hence, by Pāṇini's principle, periphrasis is preferred to synthesis where both are available.<sup>29</sup>

(40) *Paradigm function statement for Czech adjectives*

a. For any adjective  $l$  and property set  $\sigma$ , if  $\varphi$  is  $l$ 's basic stem,

$$\text{PF}(l, \sigma) = [\text{III} : [\text{II} : [\text{I} : \langle \varphi, \sigma \rangle ]]]$$

b. For any undeclinable adjective  $l$  and property set  $\sigma \supseteq \{ \text{GRADE } \textit{comp} \}$ ,

<sup>28</sup> In the case of statements about synthetic inflection, the set of reverse selection requirements is empty. In that case, in the interest of backwards compatibility, I take the notational liberty to write the output of the paradigm function as a paradigm cell: I write ' $\langle \varphi, \sigma \rangle$ ' for what should really be ' $\langle \langle \varphi, \sigma \rangle, \emptyset \rangle$ '.

<sup>29</sup> Implicit in (40a) is an analysis of synthetic inflection relying on 3 rule blocks for superlative prefixes, comparative suffixes, and case/number suffixes. The slightly simplified analysis sketched here does not capture the common properties of periphrastic comparatives and superlatives. Doing so entails treating comparatives and superlatives as sharing a common morphosyntactic property.

$$\text{PF}(l, \sigma) = \langle \varphi, \sigma, \{ \langle \text{v}i\text{c}, \{ \text{GRADE } \textit{comp} \} \} \rangle \rangle,$$

$$\text{where } \text{PF}(l, \sigma! \{ \text{GRADE } \textit{pos} \}) = \langle \langle \varphi, \sigma! \{ \text{GRADE } \textit{pos} \} \rangle, \emptyset \rangle.$$

French presents a situation that is almost the mirror image of that of Czech: with the vast majority of adjectives, comparative grade is expressed periphrastically through modification by the adverb *plus* ‘more’, but a handful of adjectives, including *bon* ‘good’, possess synthetic forms. (41) captures this situation: inflection is synthetic in general (since adjectives always inflect synthetically for number and gender agreement), but periphrastic for comparative grade (41b). Exceptions to this second rule are stated as lexeme-particular clauses, such as the one in (41c), which introduce a suppletive basic stem that must go through the regular synthetic inflection rule blocks. Since we are dealing with a handful of cases and suppletive stems must be introduced anyway, these highly specialized clauses do not result in any unwarranted redundancy in the description.

(41) *Paradigm function statement for French adjectives*

- a. For any adjective  $l$  and property set  $\sigma$ , if  $\varphi$  is  $l$ 's basic stem,

$$\text{PF}(l, \sigma) = [\text{II} : [\text{I} : \langle \varphi, \sigma \rangle]]$$

- b. For any adjective  $l$  and property set  $\sigma \supseteq \{ \text{GRADE } \textit{comp} \}$ ,

$$\text{PF}(l, \sigma) = \langle \varphi, \sigma, \{ \langle \textit{plus}, \emptyset \rangle \} \rangle,$$

where  $\text{PF}(l, \sigma! \{ \text{GRADE } \textit{pos} \}) = \langle \langle \varphi, \sigma! \{ \text{GRADE } \textit{pos} \} \rangle, \emptyset \rangle.$

- c. For any  $\sigma \supseteq \{ \text{GRADE } \textit{comp} \}$ ,  $\text{PF}(\text{BON}, \sigma) = [\text{II} : [\text{I} : \langle \textit{meilleur}, \sigma \rangle]]$ .

In English the situation is still different: as was noted in section 2.2, some lexemes exhibit overabundance between a synthetic and a periphrastic strategy for the expression of comparative grade. To capture this, I define two overlapping classes of adjectives, simply named here  $A$  and  $B$ :  $A$  contains all adjectives inflecting synthetically,  $B$  all adjectives inflecting periphrastically, and hence their intersection  $A \cap B$  contains overabundant lexemes. Given this setup of the inflection class system, neither of the two rules (42a) and (42b) is more specific than the other. Thus when inflecting a lexeme from  $A \cap B$ , such as *odd* or *friendly*, neither clause is ruled out by Pāṇini's principle, and hence both inflection strategies are licensed.<sup>30</sup> Note that rules are slightly simpler than for Czech and French, because grade is the only dimension of inflectional variation for English adjectives.<sup>31</sup>

(42) *Paradigm function statement for English adjectives*

- a. For any adjective  $l$  from class  $A$  and property set  $\sigma$ , if  $\varphi$  is  $l$ 's basic stem,

$$\text{PF}(l, \sigma) = [\text{I} : \langle \varphi, \sigma \rangle]$$

- b. For any adjective  $l$  from class  $B$ ,

$$\text{PF}(l, \{ \text{GRADE } \textit{comp} \}) = \langle \varphi, \sigma, \{ \langle \textit{more}, \{ \text{GRADE } \textit{comp} \} \} \} \rangle,$$

where  $\text{PF}(l, \{ \text{GRADE } \textit{pos} \}) = \langle \langle \varphi, \{ \text{GRADE } \textit{pos} \} \rangle, \emptyset \rangle.$

<sup>30</sup> Of course this is only the first step of an adequate account of overabundance. As Boyd (2007) shows, the relative frequency of the synthetic and periphrastic strategies varies considerably from adjective to adjective, with various phonological, morphological, syntactic or lexical factors acting as partial predictors of the observed distribution. Modeling such effects can only be approached within a probabilistic view of grammar.

<sup>31</sup> Technically, the treatment of overabundance proposed in Bonami & Stump (forthcoming) and adopted here entails that paradigm functions output sets of cells. Hence the usual notation involving identity statements of the form ‘ $\text{PF}(l, \sigma) = X$ ’ is misleading, and would more cogently be replaced by a notation of the form ‘ $X \in \text{PF}(l, \sigma)$ ’. In the interest of readability and backwards compatibility though, I have refrained from introducing a new notation. However, in section 5, where I will need to refer to the output of the paradigm function rather than individual clauses in its definition, set membership will explicitly be used.

A final virtue of the present analysis is that it allows for an intuitive account of the phenomenon of auxiliary selection. Many languages concurrently use two auxiliary verbs in perfect periphrases, typically cognates of *have* and *be*. As is well-known, auxiliary selection tends to correlate with lexical semantics, but has to be recognized as partially arbitrary (see e.g. Sorace (2000); Abeillé & Godard (2002); Legendre (2007)). This is reminiscent of the status of inflection classes: similar lexemes tend to cluster in the same classes, but there are exceptions. In the present approach, auxiliary selection is literally a matter of inflection class: just as different classes of lexemes may trigger the use of distinct rules of synthetic exponence for the expression of the same feature, they may likewise trigger the use of distinct rules of periphrastic exponence. To take a concrete example, let us consider the situation in French. Perfect forms (which are also used for the expression of the bounded past) are normally inflected using *avoir* ‘have’ as their ancillary element (43a). There are two types of exceptions. First, a few dozen intransitive verbs use *être* ‘be’ instead (43b). Second, any verb qualifying as reflexive uses *être*. To this class belong (i) verbs with a reflexive accusative or dative argument realized as a pronominal affix (43c), and (ii) so-called ‘intrinsic’ reflexives, which carry an affix of the same class without that affix corresponding to an argument of the verb (43d).

- (43) a. Le soleil **a disparu** à 8 heures.  
 DEF.M.SG sun[m] have.PRS.3SG disappear.PST-PTCP at 8 hour.PL  
 ‘The sun disappeared at 8.’
- b. Le soleil **est apparu** à 8 heures.  
 DEF.M.SG sun[m] be.PRS.3SG appear.PST-PTCP at 8 hour.PL  
 ‘The sun appeared at 8.’
- c. Paul s’ **est tué**.  
 Paul REFL.3 be.PRS.3SG kill.PST-PTCP  
 ‘Paul killed himself.’
- d. Paul s’ **est suicidé**.  
 Paul REFL.3 be.PRS.3SG commit\_suicide.PST-PTCP  
 ‘Paul committed suicide.’

This situation may be captured by positing the three rules of periphrastic exponence in (44). The default rule (44a), licensing *avoir* as the ancillary lexeme, is overridden either when the verb belongs to a restricted lexical class (44b), or when it is reflexive (44c). The rule of referral in (44d) accounts for the use of the present perfect for the expression of the simple bounded past.<sup>32</sup>

(44) *Paradigm function statements for the French perfect periphrase*

- a. If  $l$  is a verb, then for any property set  $\sigma \supseteq \{\text{PRF} +\}$ ,
- $$\text{PF}(l, \sigma) = \langle \varphi, \sigma, \{\langle \text{avoir-prf}, \sigma! \{\text{PRF} -\} \rangle\} \rangle,$$
- where  $\text{PF}(l, \sigma! \{\text{VFORM } \textit{pst-ptcp}\}) = \langle \langle \varphi, \sigma! \{\text{VFORM } \textit{pst-ptcp}\} \rangle, \emptyset \rangle$ .
- b. If  $l$  is a verb of class *être-auxiliary*, then for any property set  $\sigma \supseteq \{\text{PRF} +\}$ ,
- $$\text{PF}(l, \sigma) = \langle \varphi, \sigma, \{\langle \text{être-prf}, \sigma! \{\text{PRF} -\} \rangle\} \rangle,$$
- where  $\text{PF}(l, \sigma! \{\text{VFORM } \textit{pst-ptcp}\}) = \langle \langle \varphi, \sigma! \{\text{VFORM } \textit{pst-ptcp}\} \rangle, \emptyset \rangle$ .

<sup>32</sup> The analysis presented here is simplified in various ways, in the interest of space, notably by not taking into account so-called ‘overcompound’ (‘surcomposé’) forms, or an explicit account of pronominal affix realization on the auxiliary.



- c. If  $l$  is a verb, then for any property set  $\sigma \supseteq \{\text{PRF } +, \text{REFL } +\}$ ,

$$\text{PF}(l, \sigma) = \langle \varphi, \sigma, \{\langle \text{etre-prf}, \sigma! \{\text{PRF } -\} \rangle\} \rangle,$$

where  $\text{PF}(l, \sigma! \{\text{VFORM } \textit{pst-ptcp}\}) = \langle \langle \varphi, \sigma! \{\text{VFORM } \textit{pst-ptcp}\} \rangle, \emptyset \rangle$ .

- d. If  $l$  is a verb, then for any  $\sigma \supseteq \{\text{VFORM } \textit{pst}, \text{ASP } \textit{bnd}\}$ ,

$$\text{PF}(l, \sigma) = \text{PF}(l, \sigma! \{\text{VFORM } \textit{prs}, \text{PRF } +\} \rangle).$$

To sum up, I have shown how the notion of a paradigm function can be extended to generate reverse selection requirements in addition to phonological shapes. This then ensures that the various analytic techniques usually deployed within Paradigm Function Morphology to deal with various situations of arbitration between synthetic exponence strategies can be applied to deal with analogous situations of arbitration between synthetic and periphrastic exponence.

## 5 Periphrasis at the morphology–syntax interface

In the two previous sections I have presented a novel approach to periphrasis from two standpoints. From the point of view of syntax, periphrases are morphosyntactic analogues of idioms. A treatment of periphrasis in terms of collocational requirements on the main element allows one to state the appropriate constraints on the relationship between parts of a periphrase. From the point of view of inflection, reverse selection requirements can be generated by paradigm functions, accounting for the paradigmatic properties of periphrases. In this section I make explicit the interface between the morphological and the syntactic part of the analysis. For concreteness I do this in the context of an HPSG approach to syntax.

### 5.1 The hybrid status of words in a periphrase

In periphrasis, both the main and the ancillary element lead a double morphosyntactic life. To understand why, let us focus again on the English perfect as illustrated in Figure 2, and assume that English paradigms are partially described using a feature  $\text{VFORM}$  whose values include  $\textit{prs}$  (present),  $\textit{pst}$  (past),  $\textit{inf}$  (bare infinitive),  $\textit{prs-ptcp}$  (present participle) and  $\textit{pst-ptcp}$  (past participle) and two binary features  $\pm\text{PRF}$  (perfect) and  $\pm\text{PROG}$  (progressive). In the example at hand, the main VP should clearly be thought of as both  $[\text{VFORM } \textit{prs}]$  and  $[\text{PRF } +]$ , for the purposes of possible selection by an embedding predicate and semantic interpretation. The embedded VP should clearly be a  $[\text{VFORM } \textit{pst-ptcp}]$ , as this constituent has the syntactic properties of a nonfinite VP (it can combine with constituent negation, be fronted in topicalization, be elided under VP ellipsis, etc.). At the level of words, things are not so clear. By our hypothesis that periphrases are really inflected forms of the main element, the word *left* should be  $[\text{VFORM } \textit{prs}, \text{PRF } +]$ . But the form of *left* is that of a past participle, and, as we saw, the syntactic properties of the phrase it projects are that of a nonfinite form. Turning to the auxiliary, the situation is still different. The shape of the auxiliary is that of a simple present, specifically the present of *have*; but the phrase it heads is  $[\text{VFORM } \textit{prs}, \text{PRF } +]$ , which, by usual principles of feature percolation, leads us to expect that the auxiliary itself carries these features.

The problems raised by this dual nature of ancillary and main elements is part of the motivation that led Sadler & Spencer (2001) to propose a dual encoding of features: for

one and the same dimension of inflectional variation, Sadler and Spencer distinguish a morphological and a syntactic feature, whose values do not always match. Here I propose a variant of that idea, and posit that words distinguish two parallel structures for the representation of features relevant to inflection: an INFL structure is added to the sign, collecting features which serve as the input to rules of inflection. Every feature within INFL corresponds to some feature already present somewhere in the syntactic representation.<sup>33</sup> For ordinary synthetic words, syntactic features and corresponding INFL features have matching values. In periphrases they typically differ, both on the ancillary and the main element. Figure 7 provides an appropriate representation for an English present perfect. In addition to the VFORM and PRF feature, the representation exhibits the LID feature, whose purpose is to allow for the identification of lexemes by selectors, constructions, or morphological processes (Sag, 2012; Spencer, 2013a). Token-identities between feature values, indicated by boxed integers, indicate the flow of morphosyntactic information that needs to be established. The goal is to inflect a main verb that is a present (1) perfect (2) form of the lexeme *leave* (3). This is done by projecting into syntax a past participle (4) and reverse-selecting for a form of the auxiliary *have* which is present (1) but non-perfect. The lexical entry for the auxiliary needs to ensure that its syntactic VFORM matches its inflectional VFORM, but that its syntactic values for PRF and LID are taken over from the embedded VP. As a consequence, the relevant syntactic features of the embedding VP match the inflectional features of the main verb. This captures the intuition that periphrasis is a roundabout way of getting the same effect as synthetic inflection: projection at phrase-level of inflectional features.

## 5.2 Integrating periphrases in the grammar

The licensing of the representation in Figure 7 rests on three ingredients. First, one must make explicit the interface between paradigm functions and syntactic representations. I assume without discussion that a bijection can be established between PFM complete morphosyntactic property sets and lexeme indices on the one hand, and the typed feature structures constituting INFL values in an HPSG grammar on the other hand.<sup>34</sup> The morphology–syntax interface can then be stated as in (45), where **pf** is a function associating INFL values with a set of realizations in the form of a pair of a phonological representation and a set of reverse selection requirements. For normal synthetic words, REV-SEL will be empty.

### (45) Morphology–Syntax Interface

$$word \rightarrow \begin{bmatrix} \text{PHON} & \boxed{1} \\ \text{INFL} & \boxed{2} \\ \text{REV-SEL} & \boxed{3} \end{bmatrix} \wedge \langle \boxed{1}, \boxed{3} \rangle \in \mathbf{pf}(\boxed{2})$$

Second, lexical entries for ancillary lexemes make explicit both the relationship between their syntactic and INFL features, and constraints on the syntactic features of the main element. Specifically, in the case of the English perfect, the syntactic LID is identified with that

<sup>33</sup> The locus of these syntactic features within the sign is variable. All features discussed in exemplification in this paper are HEAD features, that is, features whose value is projected from head to phrase. However it is known that some features realized by inflection need to be considered MARKING (Tseng, 2002a) or EDGE (Tseng, 2002b) features.

<sup>34</sup> An obvious alternative is to implement within HPSG an approach to morphology sharing the main design features of PFM. See Crysmann & Bonami (2012); Bonami & Crysmann (2013) for explicit proposals of this kind. The interface between INFL and morphosyntactic property sets could also be seen as the locus of the distinction between content and form paradigms advocated by Stump (2006). I leave aside these issues which are orthogonal to the concerns of this paper.

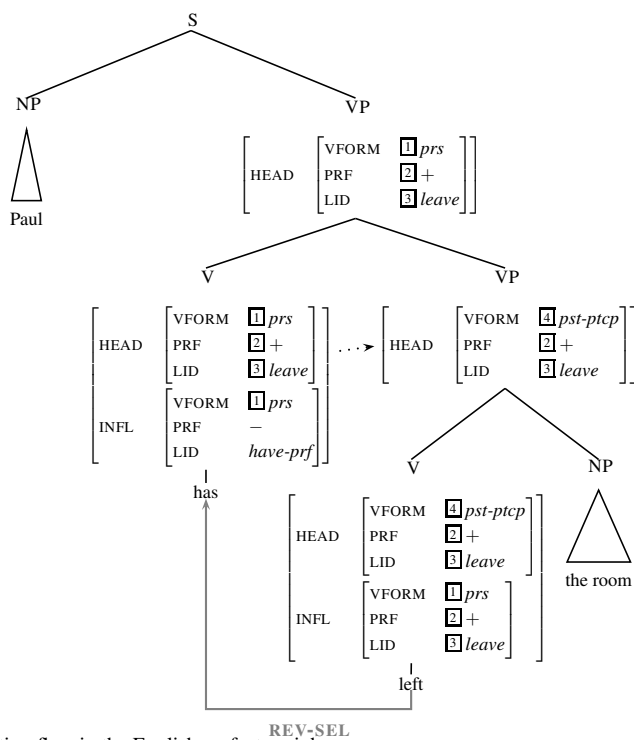


Fig. 7 Information flow in the English perfect periphrase

of the main element through a selection feature, here the feature ARG-ST. The syntactic PRF feature is specified as +, which does not match the value of the INFL feature: this is crucial to ensure that the rule of periphrastic inflection for [PRF +] will not apply to *have*. The syntactic VFORM value, on the other hand, is identical to the INFL value: this captures appropriately the fact that the VFORM of the periphrase is congruent with the synthetic inflection on the auxiliary. Finally, *have* appropriately constrains its complement to be a nonfinite form.

(46) Partial lexical entry for the ancillary lexeme *have*

HEAD	<table style="border-collapse: collapse;"> <tr><td style="padding: 2px 5px;">VFORM</td><td style="padding: 2px 5px;">[1]</td></tr> <tr><td style="padding: 2px 5px;">PRF</td><td style="padding: 2px 5px;">[2] +</td></tr> <tr><td style="padding: 2px 5px;">LID</td><td style="padding: 2px 5px;">[3]</td></tr> </table>	VFORM	[1]	PRF	[2] +	LID	[3]			
VFORM	[1]									
PRF	[2] +									
LID	[3]									
INFL	<table style="border-collapse: collapse;"> <tr><td style="padding: 2px 5px;">VFORM</td><td style="padding: 2px 5px;">[1]</td></tr> <tr><td style="padding: 2px 5px;">PRF</td><td style="padding: 2px 5px;">-</td></tr> <tr><td style="padding: 2px 5px;">LID</td><td style="padding: 2px 5px;"><i>have-prf</i></td></tr> </table>	VFORM	[1]	PRF	-	LID	<i>have-prf</i>			
VFORM	[1]									
PRF	-									
LID	<i>have-prf</i>									
ARG-ST	<table style="border-collapse: collapse;"> <tr> <td style="padding: 2px 5px;"> <table style="border-collapse: collapse;"> <tr><td style="padding: 2px 5px;">VFORM</td><td style="padding: 2px 5px;"><i>pst-ptcp</i></td></tr> <tr><td style="padding: 2px 5px;">PRF</td><td style="padding: 2px 5px;">[2]</td></tr> <tr><td style="padding: 2px 5px;">LID</td><td style="padding: 2px 5px;">[3]</td></tr> </table> </td> <td style="padding: 2px 5px;"> <table style="border-collapse: collapse;"> <tr><td style="padding: 2px 5px;">[ ]</td></tr> </table> </td> </tr> </table>	<table style="border-collapse: collapse;"> <tr><td style="padding: 2px 5px;">VFORM</td><td style="padding: 2px 5px;"><i>pst-ptcp</i></td></tr> <tr><td style="padding: 2px 5px;">PRF</td><td style="padding: 2px 5px;">[2]</td></tr> <tr><td style="padding: 2px 5px;">LID</td><td style="padding: 2px 5px;">[3]</td></tr> </table>	VFORM	<i>pst-ptcp</i>	PRF	[2]	LID	[3]	<table style="border-collapse: collapse;"> <tr><td style="padding: 2px 5px;">[ ]</td></tr> </table>	[ ]
<table style="border-collapse: collapse;"> <tr><td style="padding: 2px 5px;">VFORM</td><td style="padding: 2px 5px;"><i>pst-ptcp</i></td></tr> <tr><td style="padding: 2px 5px;">PRF</td><td style="padding: 2px 5px;">[2]</td></tr> <tr><td style="padding: 2px 5px;">LID</td><td style="padding: 2px 5px;">[3]</td></tr> </table>	VFORM	<i>pst-ptcp</i>	PRF	[2]	LID	[3]	<table style="border-collapse: collapse;"> <tr><td style="padding: 2px 5px;">[ ]</td></tr> </table>	[ ]		
VFORM	<i>pst-ptcp</i>									
PRF	[2]									
LID	[3]									
[ ]										

As the third and final ingredient in the licensing of Figure 7, constraints must be stated to the effect that the main element determines the lexical identity of a periphrase. Thus I assume the general constraint in (47) on main elements, that the lexical identity realized

in their morphology matches their lexical identity as visible to syntax. Since LID is a head feature, it projects further from head to phrase, where it can be selected for by the ancillary element, as specified in the lexical entry for auxiliary *have*.

$$(47) \text{ main-word} \rightarrow \left[ \begin{array}{l} \text{HEAD} \left[ \begin{array}{l} \text{LID} \quad \boxed{1} \end{array} \right] \\ \text{INFL} \left[ \begin{array}{l} \text{LID} \quad \boxed{1} \end{array} \right] \end{array} \right]$$

The appropriate rule of periphrasis (33) is stated again in (48) in the format defined in section 4.2.

(48) If  $l$  is a verb, then for any property set  $\sigma \supseteq \{\text{PRF } +\}$ ,

$$\text{PF}(l, \sigma) = \langle \varphi, \sigma, \{\langle \text{have-prf}, \sigma! \{\text{PRF } -\} \rangle\} \rangle,$$

$$\text{where } \text{PF}(l, \sigma! \{\text{VFORM } \textit{pst-ptcp}\}) = \langle \langle \varphi, \sigma! \{\text{VFORM } \textit{pst-ptcp}\} \rangle, \emptyset \rangle.$$

Together with the principle of reverse selection, (46) and (48) ensure that the two parts of the periphrase stand in the appropriate mutually constraining relation: the main element constrains the auxiliary to express in its inflection the VFORM value that needs to be realized. The VFORM value is passed up the head path to the matrix VP by the auxiliary's lexical entry. On the other hand, the auxiliary constrains the VFORM value of the embedded VP to be that of a past participle, a specification that is congruent with the phonological shape of the main element. A notable feature of the analysis is that for the main element, the relationship between syntactic features and INFL features is not stated directly, but only indirectly through reverse selection of an ancillary element that selects for particular features of the element it combines with.

### 5.3 Variations

In the preceding paragraphs we have seen how the syntactic and inflectional aspects of the present proposal combine at the morphology–syntax interface to provide a full analysis of a typical inflectional periphrase. Here I show how the proposal scales up to address the diversity of periphrastic strategies documented in the languages under investigation.

#### 5.3.1 Stacked periphrases

In languages with a rich system of periphrastic inflection, it is often the case that the realization of some paradigm cells involves the combination of two periphrases. In the current approach, such STACKED PERIPHRASES can be dealt with in one of two ways: either the combination of two separate rules of periphrasis, or a single rule introducing simultaneously two reverse selection requirements. The first solution is appropriate wherever the periphrases appear to make separate contributions. For instance, it provides an elegant analysis of the English perfect progressive. I assume this is dealt with by the two rules of periphrasis in (49). The tree in Figure 8 outlines the analysis. The main element to be inflected is present, perfect and progressive. By rule (49a), this is realized by the shape of a present participle and selection of the auxiliary *be-prog* in the present perfect nonprogressive. This word in turn is inflected according to (49b), through the shape of a past participle and a reverse selection requirement for a present, nonperfect, nonprogressive form of the auxiliary *have-perf*. Note that (49b) is a minimal variant of (48) restricting its application to [PROG –] paradigm cells. This is sufficient to ensure that *\*Paul was having slept* is ungrammatical.

	PRESENT	FUTURE
PRF –	mislja	šte mislja
PRF +	săm mislila	šte săm mislila
	*băda mislila	šte băda mislila

**Table 14** 1SG forms of the Bulgarian verb *mislja* ‘think’ (Popova & Spencer, 2013, 206)

(49) a. If  $l$  is a verb, then for any property set  $\sigma \supseteq \{\text{PROG} +\}$ ,

$$\text{PF}(l, \sigma) = \langle \varphi, \sigma, \{\langle \text{be-prog}, \sigma! \{\text{PROG} -\} \rangle\} \rangle,$$

where  $\text{PF}(l, \sigma! \{\text{VFORM } \textit{prs-ptcp}\}) = \langle \langle \varphi, \sigma! \{\text{VFORM } \textit{prs-ptcp}\} \rangle, \emptyset \rangle$ .

b. If  $l$  is a verb, then for any property set  $\sigma \supseteq \{\text{PRF} +, \text{PROG} -\}$ ,

$$\text{PF}(l, \sigma) = \langle \varphi, \sigma, \{\langle \text{have-prf}, \sigma! \{\text{PRF} -\} \rangle\} \rangle,$$

where  $\text{PF}(l, \sigma! \{\text{VFORM } \textit{pst-ptcp}\}) = \langle \langle \varphi, \sigma! \{\text{VFORM } \textit{pst-ptcp}\} \rangle, \emptyset \rangle$ .

The second solution, which could be called multiple periphrastic exponence, is appropriate in situations where the description of a periphrase with two ancillary elements cannot be reduced to the combination of two simpler periphrases. One such case is discussed by Popova & Spencer (2013, 206–208). As shown in Table 14, Bulgarian possesses a future periphrase based on the ancillary element *šte* combined with a present tense main verb, and a perfect periphrase based on the copula—in the 1SG, *săm*—combined with a participle. In the future perfect, the two may be combined. An alternative, however, is to use a combination of *šte* and *băda*, historically also a present form of the copula. This form cannot be used in the present perfect. In a sense, then, *băda* is a cumulative exponent of future and perfect, and the exponence of future is distributed over two ancillary elements. A rule of multiple periphrastic exponence can easily capture this situation and appropriately restrict the use of *băda* to future perfects.

### 5.3.2 Variable compositionality in periphrases

As was noted in section 2.3, the flow of morphosyntactic information in an inflectional periphrase is highly variable: the exponents carried by the main and ancillary elements may correspond more or less faithfully to the morphosyntactic properties expressed by the periphrase. The theory of periphrasis proposed here is flexible enough to allow for this variation. To show that this is the case, let us quickly address two extreme examples. Remember that in the Persian progressive, the auxiliary verb and the main verb both realize tense, aspect, evidentiality, and subject agreement. This is readily accounted for by assuming the rule of periphrasis in (50), where the morphology of the main and ancillary elements are essentially identified: progressive is the only morphosyntactic property that is overwritten, both for the determination of the shape of the main element, and in the reverse selection requirement.

(50) If  $l$  is a verb, then for any property set  $\sigma \supseteq \{\text{PROG} +\}$ ,

$$\text{PF}(l, \sigma) = \langle \varphi, \sigma, \{\langle \textit{dăştan-prog}, \sigma! \{\text{PROG} -\} \rangle\} \rangle,$$

where  $\text{PF}(l, \sigma! \{\text{PROG} -\}) = \langle \langle \varphi, \sigma! \{\text{PROG} -\} \rangle, \emptyset \rangle$ .



where  $\text{PF}(l, \sigma! \{ \text{TNS } prs, \text{POL } pos \}) = \langle \langle \varphi, \sigma! \{ \text{TNS } prs, \text{POL } pos \} \rangle, \emptyset \rangle$ .

Intermediate situations, such as that presented by the Czech future (see Table 8), rest on an asymmetry between the main and ancillary element: here the shape of the main element is that of a positive infinitive form, but the morphosyntactic property set of the ancillary element coincides with that of the periphrase as a whole.

(52) If  $l$  is a verb, then for any property set  $\sigma \supseteq \{ \text{VFORM } fut \}$ ,

$$\text{PF}(l, \sigma) = \langle \varphi, \sigma, \{ \langle \text{být-fut}, \sigma \rangle \} \rangle,$$

where  $\text{PF}(l, \{ \text{VFORM } inf, \text{POL } pos \}) = \langle \langle \varphi, \{ \text{VFORM } inf, \text{POL } pos \} \rangle, \emptyset \rangle$ .

## 6 Conclusion

In section 2 of this paper I presented six key properties of periphrasis that any adequate theory should be able to account for. In sections 3 to 5 I have developed a theory of periphrasis at the morphology–syntax interface crucially relying on the notion of collocation: in periphrasis, the exponence of some morphosyntactic property set takes the form of a collocational requirement rather than the selection of a specific bit of synthetic morphology. As a result, the main and ancillary elements stand in a relation of mutual selection not unlike that found in lexically flexible idioms.

I will now review how the present theory accounts for the six key properties. First, periphrasis is independent of part of speech. Under the current view this falls out naturally from the fact that periphrasis is just a variety of inflection: as all parts of speech may be subject to inflection, all parts of speech may inflect periphrastically. Second, arbitration between periphrasis and synthesis follows the logic of inflection, with different kinds of splits both within lexemes and across lexemes. Under the present analysis this follows directly from the fact that rules of periphrastic exponence are integrated in the definition of a language’s paradigm function; thus any kind of split that can be found within synthetic inflection is expected to be found between synthetic and periphrastic inflection. Third, ancillary lexemes are morphosyntactic hybrids. I have accounted for this property by taking them to be lexemes in their own right, distinct from the full lexemes that constitute their historical source. Fourth, periphrases need not be morphosyntactically compositional. The present theory accounts for this property by defining a bidimensional representation of morphosyntactic information between syntactic and INFL features. While ordinary synthetic words have matching representations in the two dimensions, parts of a periphrase give rise to mismatches. Fifth and sixth, parts of a periphrase are linked by grammatical functions rather than phrase-structural relations. In the present theory this is accounted for by defining collocational requirements as reverse selection requirements, which in turn are defined in terms of grammatical functions: in essence, the main element in a periphrase selects the ancillary element by checking that this element selects for it. This ensures that the two elements may stand in any phrase-structural relation allowed by the language for elements linked by that particular grammatical function.

At the beginning of this paper I quoted Matthews’s cogent (1991) characterization of nature of periphrases, which are “clearly two words” but are “taken together as a term in what are otherwise morphological oppositions”. This characterization makes periphrases paradoxical from a lexicalist perspective, where syntactic atoms are usually assumed to constitute the interface between morphology and syntax. This is what Ackerman et al. (2011) refer to as the Principle of Unary Expression (53).

- (53) In syntax, a lexeme is uniformly expressed as a single morphophonologically integrated and syntactically atomic word form. (Ackerman et al., 2011, 326)

Ackerman et al. (2011) argue the adoption of this principle creates a paradox for lexicalist theories when they confront morphological periphrases. In this paper I have proposed to solve the paradox without renouncing unary expression, by relying on the theory of collocation. In the view advocated in this paper, the main element in a periphrase *is* the single syntactic atom expressing the lexeme together with its morphosyntactic content. However, as all syntactic atoms, the main element is a multidimensional sign, which, among its characteristics, may place collocational conditions on its environment of occurrence. From the point of view of morphology, this collocational condition constitutes the exponence of some morphosyntactic property set; in that sense and in that sense only, the main and ancillary element function “together as a term”. From the point of view of syntax, each of the two words constitutes a cohesive syntactic atom realizing a distinct lexeme. A surprising result of the present study is thus that in the end, periphrasis presents no threat to strong lexicalism.

**Acknowledgements** This paper started out as an elaboration of Bonami and Webelhuth (2013), designed jointly with Gert Webelhuth and presented at the eighth Mediterranean Morphology Meeting (Cagliari, september 2011) and at the first Frankfurt HPSG workshop (May 2012). I thank the audience at these events for their comments and suggestions. Although Gert Webelhuth could not participate in the writing of this paper and insisted on not co-signing it, his influence on its content is considerable. For providing expert help, and often data, on various languages, I thank Farrell Ackerman (Tundra Nenets), Marina Chumakina (Archi), Johanna Nichols (Ingush), and Jana Strnadová (Czech). Discussions on idioms with Sascha Bargmann were also very helpful. Farrell Ackerman, Berthold Crysman, Philip Miller, Andrew Spencer and Gert Webelhuth provided important feedback on pre-final versions, leading to many improvements. Any remaining errors are of course my own.

This work was written as the author was a member of the Institut Universitaire de France. It was partially supported by a public grant overseen by the French National Research Agency (ANR) as part of the “Investissements d’Avenir” program (reference: ANR-10-LABX-0083).

## References

- Abeillé, A. & Godard, D. (2000). French word order and lexical weight. In R. D. Borsley (Ed.), *The Nature and Function of Syntactic categories*, Syntax and semantics (pp. 325–360). New York: Academic Press, Inc.
- Abeillé, A. & Godard, D. (2002). The syntactic structure of French auxiliaries. *Language*, 78, 404–452.
- Ackerman, F. & Nikolaeva, I. (2014). *Descriptive Typology and Linguistic Theory. A study in the morphosyntax of relative clauses*. Stanford: CSLI Publications.
- Ackerman, F. & Stump, G. T. (2004). Paradigms and periphrastic expression. In L. Sadler & A. Spencer (Eds.), *Projecting Morphology* (pp. 111–157). Stanford: CSLI Publications.
- Ackerman, F., Stump, G. T., & Webelhuth, G. (2011). Lexicalism, periphrasis and implicative morphology. In R. D. Borsley & K. Börjars (Eds.), *Non-Transformational Syntax: Formal and Explicit Models of Grammar* (pp. 325–358). Wiley-Blackwell.
- Ackerman, F. & Webelhuth, G. (1998). *A Theory of Predicates*. Stanford: CSLI Publications.
- Anderson, G. (2006). *Auxiliary verb constructions*. Oxford: Oxford University Press.
- Anderson, J. M. (2011). *The Substance of Language Volume II: Morphology, Paradigms, and Periphrases*. Oxford: Oxford University Press.
- Anderson, S. R. (1992). *A-Morphous Morphology*. Cambridge: Cambridge University Press.



- Aronoff, M. (1994). *Morphology by itself*. Cambridge: MIT Press.
- Aronoff, M. & Lindsay, M. (in press). Partial organization in languages: la langue est un système où la plupart se tient. In S. Augendre, G. Couasnon-Torlois, D. Lebon, C. Michard, G. Boyé, & F. Montermini (Eds.), *Proceedings of the 8th Décembrettes* (pp. 1–14). Toulouse: CLLE-ERSS.
- Blevins, J. P. (2006). Word-based morphology. *Journal of Linguistics*, 42, 531–573.
- Blevins, J. P. (forthcoming). Periphrasis as syntactic exponence. In F. Ackerman, J. P. Blevins, & G. T. Stump (Eds.), *Paradigms and Periphrasis*. Stanford: CSLI Publications.
- Bonami, O. & Crysmann, B. (2013). Morphotactics in an information-based model of realisational morphology. In S. Müller (Ed.), *Proceedings of HPSG 2013* (pp. 27–47). Stanford: CSLI Publications.
- Bonami, O. & Samvelian, P. (2009). Inflectional periphrasis in Persian. In S. Müller (Ed.), *Proceedings of the HPSG 2009 Conference* (pp. 26–46). Stanford: CSLI Publications.
- Bonami, O. & Samvelian, P. (in press). The diversity of inflectional periphrasis in Persian. *Journal of Linguistics*, 51(2).
- Bonami, O. & Stump, G. T. (forthcoming). Paradigm Function Morphology. In A. Hippisley & G. T. Stump (Eds.), *Cambridge Handbook of Morphology*. Blackwell.
- Bonami, O. & Webelhuth, G. (2013). The phrase-structural diversity of periphrasis: a lexicalist account. In M. Chumakina & G. G. Corbett (Eds.), *Periphrasis: The role of syntax and morphology in paradigms* (pp. 141–167). Oxford: Oxford University Press.
- Booij, G. (2010). *Construction morphology*. Oxford: Oxford University Press.
- Boyd, J. (2007). *Comparatively speaking: A psycholinguistic study of optionality in grammar*. Ph.D. thesis, University of California, San Diego.
- Bresnan, J. (1978). A realistic transformational grammar. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic Theory and Psychological Reality* (pp. 1–59). Cambridge: MIT Press.
- Bresnan, J. (2001a). Explaining morphosyntactic competition. In M. Baltin & C. Collins (Eds.), *Handbook of contemporary syntactic theory* (pp. 11–44). Oxford: Basil Blackwell.
- Bresnan, J. (2001b). *Lexical-functional syntax*. Oxford: Basil Blackwell.
- Brown, D., Chumakina, M., Corbett, G. G., Popova, G., & Spencer, A. (2012). Defining ‘periphrasis’: key notions. *Morphology*, 22, 233–275.
- Brown, D. & Evans, R. (2012). Morphological complexity and unsupervised learning: validating Russian inflectional classes using high frequency data. In F. Kiefer, M. Ladányi, & P. Siptár (Eds.), *Current Issues in Morphological Theory: (Ir)regularity, analogy and frequency* (pp. 135–162). Amsterdam: John Benjamins.
- Chumakina, M. (2013). Periphrasis in Archi. In M. Chumakina & G. G. Corbett (Eds.), *Periphrasis: The role of syntax and morphology in paradigms* (pp. 27–52). Oxford: Oxford University Press.
- Corbett, G. G. (2013). Periphrasis and possible lexemes. In M. Chumakina & G. G. Corbett (Eds.), *Periphrasis: The role of syntax and morphology in paradigms* (pp. 169–189). Oxford: Oxford University Press.
- Corbett, G. G. (in press). Morphosyntactic complexity: a typology of lexical splits. *Language*.
- Crysmann, B. & Bonami, O. (2012). Establishing order in type-based realisational morphology. In S. Müller (Ed.), *Proceedings of HPSG 2012* (pp. 123–143). Stanford: CSLI publications.
- Cvrček, V., Kodýtek, V., Kopřivová, M., Kovářiková, D., Sgall, P., Šulc, M., Táborcký, J., Volín, J., & Waclawičová, M. (2010). *Mluvince sousčasné češtiny*, vol. 1. Prague: Karolinum.

- Fillmore, C. J., Kay, & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language*, 64, 501–538.
- Franks, S. & Holloway King, T. (2000). *A Handbook of Slavic Clitics*. Oxford: Oxford University Press.
- Gazdar, G. (1982). Phrase structure grammar. In P. Jacobson & G. K. Pullum (Eds.), *The Nature of Syntactic Representations*. Reidel.
- Gazdar, G., Klein, E., Pullum, G. K., & Sag, I. A. (1985). *Generalized Phrase Structure Grammar*. Harvard: Harvard University Press.
- Ginzburg, J. & Sag, I. A. (2000). *Interrogative Investigations. The Form, Meaning, and Use of English Interrogatives*. Stanford: CSLI Publications.
- Gonzalez-Diaz, V. (2008). *English Adjective Comparison: A Historical Perspective*. Amsterdam: John Benjamins.
- Göskel, A. & Kerlake, C. (2005). *Turkish: A comprehensive grammar*. Oxon: Routledge.
- Halle, M. & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale & S. J. Keyser (Eds.), *The view from building 20*, (pp. 111–176). Cambridge: MIT Press.
- Halpern, A. (1995). *On the placement and morphology of clitics*. Stanford: CSLI Publications.
- Harman, G. H. (1963). Generative grammars without transformation rules. a defense of phrase structure. *Language*, 39, 597–616.
- Haspelmath, M. (2000). Periphrasis. In G. Booij, C. Lehmann, & J. Mugdan (Eds.), *Morphology: an international handbook on inflection and word-formation*, vol. 1 (pp. 654–664). Berlin: Walter de Gruyter.
- Hippisley, A. (2007). Declarative deponency: A network morphology account of morphological mismatches. In M. Baerman, G. G. Corbett, D. Brown, & A. Hippisley (Eds.), *Deponency and Morphological Mismatches* (pp. 145–173). Oxford: Oxford University Press.
- Hockett, C. F. (1958). *A Course in Modern Linguistics*. New York: Macmillan.
- Hockett, C. F. (1967). The Yawelmani basic verb. *Language*, 43, 208–222.
- Horn, G. M. (2003). Idioms, metaphor, and syntactic mobility. *Journal of Linguistics*, 39, 245–273.
- Huddleston, R. & Pullum, G. K. (2002). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Johnson, D. & Lappin, S. (1999). *Local constraints vs. economy*. Stanford: CSLI Publications.
- Karlík, P., Nekula, M., & Rusínová, Z. (1995). *Příruční mluvnice češtiny*. Prague: Lidové Noviny.
- Kiparsky, P. (1973). 'Elsewhere' in phonology. In S. R. Anderson & P. Kiparsky (Eds.), *A Festschrift for Morris Halle* (pp. 93–106). New York: Holt, Rinehart and Winston.
- Kiparsky, P. (2005). Blocking and periphrasis in inflectional paradigms. In G. Booij & J. v. Marle (Eds.), *Yearbook of Morphology 2004* (pp. 113–135). Dordrecht: Springer.
- Kuhn, J. (2003). *Optimality-Theoretic Syntax: a Declarative Approach*. Stanford: CSLI Publications.
- Lee, L. & Ackerman, F. (submitted). Thai passives and the morphology–syntax interface. UCSD.
- Legendre, G. (2007). On the typology of auxiliary selection. *Lingua*, 117, 1522–1540.
- Maiden, M. (1992). Irregularity as a determinant of morphological change. *Journal of Linguistics*, 28, 285–312.

- Maiden, M. (2004). When lexemes become allomorphs — on the genesis of suppletion. *Folia Linguistica*, 38, 227–256.
- Malouf, R. (2003). Cooperating constructions. In E. Francis & L. A. Michaelis (Eds.), *Linguistic Mismatch: Scope and Theory* (pp. 403–424). Stanford: CSLI Publications.
- Matthews, P. H. (1972). *Inflectional Morphology. A Theoretical Study Based on Aspects of Latin Verb Conjugation*. Cambridge: Cambridge University Press.
- Matthews, P. H. (1991). *Morphology*. Cambridge: Cambridge University Press, 2nd edn.
- Mel'čuk, I. A. (1993). *Cours de morphologie générale*, vol. 1: Introduction et première partie. Montréal & Paris: Presses de l'Université de Montréal & CNRS Éditions.
- Miller, P. (1992). *Clitics and constituents in Phrase Structure Grammar*. New York: Garland.
- Monachesi, P. (2005). *The Verbal Complex in Romance*. Oxford: Oxford University Press.
- Müller, S. (2002). *Complex Predicates: Verbal Complexes, Resultative Constructions, and Particle Verbs in German*. Stanford: CSLI Publications.
- Müller, S. (2010). Persian complex predicates and the limits of inheritance-based analyses. *Journal of Linguistics*, 46, 601–655.
- Müller, S. (2013). Unifying everything: Some remarks on simpler syntax, construction grammar, minimalism, and HPSG. *Language*, 89, 920–950.
- Nichols, J. (2011). *Ingush Grammar*. Berkeley: University of California Press.
- Nikolaeva, I. (2013). Periphrasis in Tundra Nenets. In M. Chumakina & G. G. Corbett (Eds.), *Periphrasis: The role of syntax and morphology in paradigms* (pp. 77–104). London and Oxford: British Academy and Oxford University Press.
- Nikolaeva, I. (2014). *A grammar of Tundra Nenets*. Berlin: Mouton de Gruyter.
- Nunberg, G., Sag, I. A., & Wasow, T. (1994). Idioms. *Language*, 70, 491–538.
- Pollard, C. & Sag, I. A. (1987). *Information-based syntax and semantics*. Stanford: CSLI Publications.
- Pollard, C. & Sag, I. A. (1994). *Head-driven Phrase Structure Grammar*. Stanford: CSLI Publications; The University of Chicago Press.
- Popova, G. & Spencer, A. (2013). Relatedness in periphrasis: A paradigm-based perspective. In M. Chumakina & G. G. Corbett (Eds.), *Periphrasis: The role of syntax and morphology in paradigms* (pp. 191–225). Oxford: Oxford University Press.
- Poser, W. J. (1992). Blocking of phrasal constructions by lexical items. In I. A. Sag & A. Szabolcsi (Eds.), *Lexical matters* (pp. 111–130). Stanford: CSLI Publications.
- Richter, F. & Sailer, M. (2003). Cranberry words in Formal Grammar. In C. Beyssade, O. Bonami, P. Cabredo Hofherr, & F. Corblin (Eds.), *Empirical issues in formal syntax and semantics*, vol. 4 (pp. 155–171). Paris: Presses de l'Université de Paris-Sorbonne.
- Richter, F. & Sailer, M. (2010). Phraseological clauses in constructional hpsg. In *Proceedings of HPSG 2009* (pp. 297–317). Stanford: CSLI Publications.
- Richter, F. & Soehn, J.-P. (2006). Braucht niemanden zu scheren - NPI licensing in German. In S. Müller (Ed.), *Proceedings of HPSG 2006* (pp. 421–440). Stanford: CSLI Publications.
- Robins, R. H. (1959). In defense of WP. *Transactions of the Philological Society* (pp. 116–144).
- Sadler, L. & Spencer, A. (2001). Syntax as an exponent of morphological features. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 2000* (pp. 71–96). Dordrecht: Kluwer.
- Sag, I. A. (2012). Sign-based construction grammar: An informal synopsis. In H. Boas & I. A. Sag (Eds.), *Sign-Based Construction Grammar* (pp. 69–202). Stanford: CSLI Publications.

- Sailer, M. (2000). *Combinatorial semantics and idiomatic expressions in Head-driven Phrase Structure Grammar*. Ph.D. thesis, Eberhard-Karls-Universität Tübingen.
- Salminen, T. (1997). *Tundra Nenets Inflection*, vol. 227 of *Mémoires de la Société Finno-Ougrienne*. Helsinki: Suomalais-Ugrilainen Seura.
- Samvelian, P. & Tseng, J. (2010). Persian object clitics and the syntax-morphology interface. In *Proceedings of HPSG 2010* (pp. 212–232). Stanford: CSLI Publications.
- Short, D. (1993). Czech. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic Languages* (pp. 455–532). London: Routledge.
- Soehn, J.-P. (2006). *Über Bärendienste und erstaunte Bauklötze – Idiome ohne freie Lesart in der HPSG*. Frankfurt: Peter Lang.
- Soehn, J.-P. & Sailer, M. (2003). At first blush on Tenterhooks. About selectional restrictions imposed by nonheads. In G. Jäger, P. Monachesi, G. Penn, & S. Wintner (Eds.), *Proceedings of Formal Grammar 2003* (pp. 123–138).
- Sorace, A. (2000). Gradients in auxiliary selection with intransitive verbs. *Language*, 76, 859–890.
- Spencer, A. (2003). Periphrastic paradigms in Bulgarian. In U. Junghanns & L. Szucsich (Eds.), *Syntactic Structures and Morphological Information* (pp. 249–282). Berlin: Mouton de Gruyter.
- Spencer, A. (2013a). *Lexical relatedness: a paradigm-based model*. Oxford: Oxford University Press.
- Spencer, A. (2013b). Sentence negation and periphrasis. In M. Chumakina & G. G. Corbett (Eds.), *Periphrasis: The role of syntax and morphology in paradigms* (pp. 227–266). London and Oxford: British Academy and Oxford University Press.
- Steedman, M. (1996). *Surface structure and interpretation*. Cambridge: MIT Press.
- Stump, G. T. (2001). *Inflectional Morphology. A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.
- Stump, G. T. (2002). Morphological and syntactic paradigms: Arguments for a theory of paradigm linkage. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 2001*. Dordrecht: Kluwer Academic Press.
- Stump, G. T. (2006). Heterocclisis and paradigm linkage. *Language*, 82, 279–322.
- Stump, G. T. (2013). Periphrasis in the Sanskrit verb system. In M. Chumakina & G. G. Corbett (Eds.), *Periphrasis* (pp. 105–138). British Academy and Oxford University Press.
- Stump, G. T. & Hippisley, A. (2011). Valence sensitivity in Pamirian past-tense inflection: a realizational analysis. In A. Korn, G. Haig, S. Karimi, & P. Samvelian (Eds.), *Topics in Iranian linguistics* (pp. 103–116). Wiesbaden: Ludwig Riechert Verlag.
- Thornton, A. M. (2012). Reduction and maintenance of overabundance. a case study on Italian verb paradigms. *Word Structure*, 5, 183–207.
- Tseng, J. (2002a). Remarks on marking. In F. Van Eynde, L. Hellan, & D. Beermann (Eds.), *The Proceedings of the 8th International Conference on Head-Driven Phrase Structure Grammar* (pp. 267–283). Stanford: CSLI Publications.
- Tseng, J. L. (2002b). Edge features and french liaison. In J.-B. Kim & S. Wechsler (Eds.), *The Proceedings of the 9th International Conference on Head-Driven Phrase Structure Grammar* (pp. 313–333). Stanford: CSLI Publications.
- Vincent, N. & Börjars, K. (1996). Suppletion and syntactic theory. In *Proceedings of the first LFG conference*.
- Wagner-Nagy, B. (2011). *On the Typology of Negation in Ob-Ugric and Samoyedic Languages*. Helsinki: Suomalais-Ugrilainen Seura.
- Walther, G. (2013). *De la canonicité en morphologie: perspective empirique, théorique et computationnelle*. Ph.D. thesis, Université Paris Diderot.

- Wasow, T., Sag, I. A., & Nunberg, G. (1984). Idioms: an interim report. In S. Hattori & K. Inoue (Eds.), *Proceedings of the XIIIth International Congress of Linguists* (pp. 102–115). Tokyo: Nippon Toshi Center.
- Webelhuth, G., Bargmann, S., & Götze, C. (forthcoming). Idioms as evidence for the proper analysis of relative clauses. In M. Krifka, R. Ludwig, & M. Schenner (Eds.), *Reconstruction Effects in Relative Clauses*. Berlin: Akademie Verlag.
- Wheeler, M. W. (1979). *The Phonology of Catalan*. Oxford: Blackwell.
- Zwicky, A. M. (1992). Some choices in the theory of morphology. In R. D. Levine (Ed.), *Formal Grammar: Theory and Implementation* (pp. 327–371). Oxford: Oxford University Press.

### Appendix: A statement of reverse selection

The notion of reverse selection has been characterized informally in section 3.2. Here I provide a partial formalization of the notion in the context of HPSG. The following definitions assume a standard post-Pollard & Sag (1994) feature geometry, with selection of subjects, specifiers and complements via the feature ARG-ST and selection of heads by adjuncts via the feature MOD.

It will be assumed that lexical signs possess a specific representation LEX outside of *synsem* that collects various features that are only indirectly represented in syntax, including ARG-ST and the syntactic and INFL features discussed above. Reverse selection is enacted through a set-valued feature REV-SEL inside LEX. I assume that elements of REV-SEL are of type *infl*, because I have not encountered situations where more information needs to be referred to via reverse selection. However nothing crucial hinges on this choice.

Reverse selection can be stated as a principle constraining REV-SEL requirements to be met in the immediate environment:

#### (54) *Projection*

- a. Every sign is a projection of itself.
- b. A phrase is a projection of its head.
- c. A coordination is a projection of each of its daughters.

#### (55) *Selection*

- a. A sign selects all signs whose *local* value is identical to that of an element on its ARG-ST.
- b. A sign selects any sign whose *local* value is identical to that of an element on its MOD.

#### (56) *Reverse Selection Principle*

If a word  $w$  carries a reverse selection requirement  $s$  in its REV-SEL, then  $s$  must be token-identical to the INFL value of a word  $w'$  selecting for a projection of  $w$ .

As stated, the Reverse Selection Principle will ensure that the main element in a periphrase must appear in the syntactic dependency of the ancillary element. However this is not quite sufficient to avoid overgeneration: the ancillary element may still be found in combination with a sign that is not reverse-selecting for it. Although this situation could be avoided by the addition of control features, it is more elegant to rule it out as a matter of principle. This is the intent of the following principle.

- (57) All words belong to one of the three types *simple-word*, *main-word* and *ancillary-word*.
- (58) *Ancillary element licensing*  
Any word of type *ancillary-word* must be licensed by the presence in the same utterance of a *main-word* whose REV-SEL contains an element token-identical with the *ancillary-word*'s INFL value.