

1 *Introduction*

Хорошо быть медведем, ура!
 Хорошо быть медведем, ура!
 Побежу . . .
 (Нет, победу!)
 Победу я и жару и мороз,
 Лишь бы мёдом был вымазан нос!
 Победу . . .
 (Нет, побегду!)
 Побегду, я люблю беду,
 Лишь бы были все лапки в меду! . . .¹
 Винни-Пух и все-все-все
 (Milne 1965: 75)

1.1 The problem of missing word-forms

A frequently cited truism about inflection is that it is highly productive. Inflectional morphology exists primarily to fulfill the needs of the syntax, which suggests that any given lexeme should have a form for every syntactic context in which it can appear, and this expectation is usually met. While speakers do not always have the same intuitions about what the form of a word should be, both children and adults readily produce required grammatical forms for both existing and novel lexemes (e.g., Albright and Hayes 2002; Aronoff 1980; Bauer 2001; Berko 1958; Bybee and Slobin 1982). Correspondingly, when a new lexeme comes into a language, it tends to be quickly integrated into the inflectional system, sprouting a full array of inflected forms. To take a single example, the verb *GOOGLE* famously entered English in the late 1990s and developed the forms *googles*, *googled*, *googling*, *will google*, etc.² Examples of this kind, in

¹ It is good to be a bear, hurrah! / It is good to be a bear, hurrah! / I will conquer (*pobežu*) . . . / (No, I will conquer (*pobedju*)!) / I will conquer (*pobedju*) both heat and frost, / As long as my nose is smeared with honey! / I will conquer (*pobedju*) . . . / (No, I will conquer (*pobeždu*)!) / I will conquer (*pobeždu*), I love trouble, / As long as all my paws are in honey! . . .

² The Google search engine was launched in 1998; the verb lexeme is formed by conversion from the proper noun.

2 Introduction

Table 1.1 An example of a defective Russian verb

SPROSIT' 'to ask'			POBEDIT' 'to defeat'		
	SINGULAR	PLURAL		SINGULAR	PLURAL
1st person	sprošu	sprosim	1st person	–	pobedim
2nd person	sprošiš'	sprosite	2nd person	pobediš'	pobedite
3rd person	sprosit	sprosjat	3rd person	pobedit	pobedjat

which inflectional morphology applies seemingly automatically to new (and existing) lexemes, are numerous.

Against this backdrop, it seems surprising that in many if not all inflecting languages, we find some lexemes whose inflectional paradigms have Swiss-cheese holes in them – places where syntactically required forms simply fail to exist. *Inflectional defectiveness* can be defined intuitively as a situation in which one or more inflected forms of a lexeme is missing. It is the lack of *any* word-form when we expect to find *some* word-form. (A more formal definition is offered in §2.3.) Instances of inflectional defectiveness are *paradigmatic gaps*, or simply *gaps* for short.

Paradigmatic gaps are generally considered to be inflectional anomalies. A classic example of defectiveness – the pattern represented by the Russian verb POBEDIT' 'conquer, defeat' – serves to illustrate the basic issues.

The vast majority of Russian verbs express six combinations of person–number values in the non-past tense, including first person singular (e.g., SPROSIT' 'ask'). However, a number of verbs lack 1SG forms, including POBEDIT'. Zaliznjak (1977: 159) says that the first person singular of this verb is “difficult,”³ and counts from the Russian National Corpus confirm that in the modern language, the expected 1SG form *pobežu* is not used to any observable degree, nor is any other conceivable 1SG form of this verb, despite the lexeme being fairly frequently used overall. The corpus contains 1,380 tokens of POBEDIT' in the other five non-past person–number combinations combined, but not a single token of *pobežu* as a verb form.⁴ Russian speakers also deem

³ 1 ед. затрудняется. This is how Zaliznjak usually notes lexemes with 1SG gaps, although he marks some as 1 ед. нет ('There is no 1SG.'). There is no clear, empirical difference between lexemes that have one kind of notation versus the other.

⁴ <http://ruscorpora.ru>; accessed August 11, 2011. There are five attestations of a historically alternative 1SG form *pobeždu*, at least four of which pre-date the lexeme becoming defective in the late nineteenth and early twentieth centuries (Baerman 2008). There were no instances of a logical third possibility, *pobedju*, which would represent a leveling of the stem shape within the

1.1 The problem of missing word-forms 3

any sentence containing the first person singular of this verb to be ungrammatical. Compare (1a), which is an internet newspaper headline,⁵ to (1b), which speakers categorically reject, despite the fact that the sentence is semantically and syntactically unproblematic.⁶

(1)

a. *Kandidaty ot opozicii pobedjat vo vseh pjati problemnyx okrugax.*
 candidates from opposition win.PFV.NPST.3PL in all five problematic regions
 'Opposition candidates will win in all five problematic regions.'

b. **Ja pobežu vo vseh pjati okrugax.*
 I win.PFV.NPST.1SG in all five regions
 'I will win in all five regions.'

The bottom line is thus that while a 1SG form of POBEDIT' is syntactically expected, there is simply no acceptable way to express the intended meaning, except by using a near synonym or other circumlocution – e.g. *oderžu pobedu* 'I will score a victory', in which *pobedu* is the accusative singular of the noun POBEDA 'victory'. POBEDIT' has a paradigmatic gap in the 1SG non-past. If inflectional morphology is really so productive, and new lexemes acquire a full set of inflected forms automatically, then the failure of the first person singular is perplexing.

Moreover, the fact that only a small and lexically idiosyncratic group of verbs is affected makes the pattern of defectiveness all the more anomalous. Halle (1973: 7) states without elaboration that in Russian, "about 100 verbs" are defective in the same way that POBEDIT' is. Alley et al. (2006) identify ninety-six lexemes, some based on the same root, that are listed as having 1SG gaps in at least one of eight reference works consulted. These include the verbs in (2), which are cited as defective in at least five of the sources. While a few of these are relatively unlikely to be used in the first person singular (e.g., *šELESTET'*, which describes the noise made by dry leaves), in other cases the lack of a 1SG form is equally as surprising as in POBEDIT'.

paradigm. Some examples of these forms can be found on the Internet, but almost all of these appear to be discussions of the form itself and thus are instances of mention, not use. In many of the remaining examples, it is clear from context that the use is sarcastic or mocking and is probably an intentional violation of norms of usage.

⁵ http://newsradio.com.ua/rus/2013_05_16/Kandidati-ot-oppozicii-pobedjat-vo-vseh-pjati-problemnyh-okrugah-Sobolev/. Accessed September 15, 2013.

⁶ This is based on my own informal survey of a few dozen educated Russian speakers over the years. I have yet to find anyone who considers any sentence containing the 1SG of POBEDIT' to be anything less than terrible sounding.

4 Introduction

(2)	<i>bdet'</i>	'keep watch'	<i>buzit'</i>	'create scandal, disorder'
	<i>galdet'</i>	'make a lot of noise'	<i>derzit'</i>	'be impertinent'
	<i>dudet'</i>	'play the fife'	<i>erundit'</i>	'do or say s.t. stupid'
	<i>zatmit'</i>	'darken; overshadow'	<i>kudesit'</i>	'do magic'
	<i>očutit'sja</i>	'find o.s.; come to be'	<i>rysit'</i>	'trot'
	<i>sosedit'</i>	'be a neighbor'	<i>ubedit'</i>	'convince'
	<i>umiloserdit'</i>	'take pity on'	<i>čudesit'</i>	'do magic'
	<i>čudit'</i>	'behave eccentrically'	<i>šelestet'</i>	'rustle'
	<i>škodit'</i>	'misbehave'		

The defective lexemes all belong to the same morphological subclass, but they represent only a small percentage of all verbs in this class. Thus, although there are some regularities to the distributional pattern, it is impossible to identify a morpho(phono)logically defined class that includes all and only the defective lexemes.⁷ Similarly idiosyncratic examples of defectiveness have been documented in a variety of languages' inflectional systems. Such gaps are often noted as rare but troublesome exceptions to the productive nature of inflection.

It is important to note that paradigmatic gaps do not necessarily result in significant disruption to communication and speakers need not be consciously aware of the existence of a gap. For instance, as a native speaker of English, I did not realize that BEWARE has a defective paradigm until it was pointed out to me (Fodor 1972). The fact that speakers are capable of navigating around paradigmatic gaps without being (consciously) aware of their existence is noteworthy. In this respect, paradigmatic gaps are a little like suppletion, heteroclisis, depency, and so on, which have the potential to be stumbling blocks, but which speakers navigate successfully. (It is mostly from the perspective of the linguist that these kinds of form–meaning mismatches seem remarkable.)

At least sometimes, however, gaps are indeed disruptive, creating impediments to communication. In everyday conversation, speakers start sentences and then cannot finish them because a required word-form does not exist. Speakers may also try out one or more possible forms before rejecting all of them (Albright 2003; Orgun and Sprouse 1999), as in the following example from the Internet, in which the writer runs into the past tense of FORGO, which for many English speakers is defective.⁸

⁷ Readers who are familiar with Russian will know that some relevant facts are being withheld, most notably that in the affected inflection class, the first person singular exhibits a stem-final alternation (see *SPROSR'* in Table 1.1). There has been debate about the role of this alternation in causing the gaps. Ultimately, I argue that the relationship is primarily historical, but the issues are complicated. See §7.2 for discussion.

⁸ <http://modelingmadness.com/scott/allies/us/rafp47.htm>; accessed August 15, 2011.

1.1 The problem of missing word-forms 5

- (3) Because of the fussiness of the decals, I forgoed . . . forewent . . . err . . . I didn't put on any of the small stencil decals.

In such circumstances, there is no choice but to start the sentence over with a different verb or find some other work around.

Humor offers a less straightforward but interesting illustration of the disruption that gaps can cause. A charming example is found in *Vinni-Pux i vse-vse-vse* (Milne 1965), a Russian translation of the A. A. Milne story *The world of Pooh*, about Winnie-the-Pooh, a lovable but not very bright bear, and his friends in the Hundred Acre Wood. In one part, quoted at the beginning of this chapter and repeated below, Winnie-the-Pooh sings about his love for honey and how he will conquer heat and frost in order to get it. However, he keeps stumbling over the 1SG form of the verb POBEDIT' 'conquer'.

- | | | |
|-----|--------------------------------------|--|
| (4) | Xorošo byt' medvedem, ura! | It is good to be a bear, hurrah! |
| | Xorošo byt' medvedem, ura! | It is good to be a bear, hurrah! |
| | Pobežu . . . | I will conquer (<i>pobežu</i>) . . . |
| | (Net, pobedju!) | (No, I will conquer (<i>pobedju</i>)!) |
| | Pobedju ja i žaru i moroz, | I will conquer (<i>pobedju</i>) both heat and frost, |
| | Liš' by mēdom byl vymazan nos! | As long as my nose is smeared with honey! |
| | Pobedju . . . | I will conquer (<i>pobedju</i>) . . . |
| | (Net, pobeždu!) | (No, I will conquer (<i>pobeždu</i>)!) |
| | Pobeždu, ja ljublju bedu, | I will conquer (<i>pobeždu</i>), I love trouble, |
| | Liš' by byli vse lapki v medu! . . . | As long as all my paws are in honey! |
- (Milne 1965: 75)

The song plays with the fact that Russian verbs are expected to express six combinations of person–number values, yet POBEDIT' lacks a 1SG form. Winnie-the-Pooh tries out three different possible forms – *pobežu*, *pobedju*, and finally, *pobeždu* – but he simply cannot hit on an acceptable one. The humor thus lies in the fact that the structure of the Russian language makes it impossible for the sometimes-hapless Winnie-the-Pooh to conquer (!) the verb; his attempt to navigate the language is ill-fated from the start. While this playful use of language is obviously intentional, it illustrates that even from the perspective of speakers, a form is expected but missing.⁹

⁹ A similar example is found in the animated cartoon *Kak griby s goroxom voevali* (How the mushrooms fought with the peas) (Aksenčuk 1977). In it, subjects in the land of the mushrooms vie for the hand of the mushroom princess. When the kingdom is threatened by the king of the peas, each suitor declares how he will defend it. The “smartest” mushroom suitor declares: “Ja ego intellektom pobedju . . . pobežu . . . pobež . . . du . . . Zab'ju mozgami, koroče govorja.” (‘With (my) intellect I will defeat [*pobedju*] . . . defeat [*pobežu*] . . . defeat [*pobež . . . du*] . . . him. I will

6 Introduction

POBEDIT' thus serves to highlight the fundamental issue surrounding inflectional defectiveness: Inasmuch as gaps are disruptive to a language's morphosyntactic system, we should wonder why they arise and why speakers do not quickly and automatically generate word-forms to fill the empty paradigm cells, especially since gaps occur in only a small and idiosyncratic group of verbs. We might expect paradigmatic gaps to be fleeting phenomena – rare to occur and quickly smoothed out when they do surface. Yet they arise occasionally, and when they do, they can persist indefinitely. Many of the Russian 1SG gaps have persisted for over a hundred years (Baerman 2008), and they continue to be passed on to and learned by new generations of speakers. Paradigmatic gaps are thus fascinating primarily for the way that they seem to contradict the fundamental nature of inflectional morphology and fly in the face of speakers' tendency to generalize. If for no other reason, this makes inflectional defectiveness an important topic for morphological study.

In this book, I explore what inflectional defectiveness can reveal about the structure of inflectional systems. Specifically, I am interested in the following questions:

- (1) What mechanisms can account for the fact that inflectional morphology applies robustly to new lexemes, and also the fact that paradigmatic gaps arise and persist? Are these contradictory outcomes and how are they balanced against each other?
- (2) What conditions lead to defectiveness, and why do speakers sometimes choose to leave a paradigm cell empty, rather than applying some kind of repair strategy?
- (3) Once gaps arise, how are they learned by new generations? In particular, why do gaps often persist in a language, even in the absence of the conditions that originally gave rise to them?
- (4) How should paradigmatic gaps be represented within a theory of inflection? Are paradigmatic gaps inflectional anomalies? Should defectiveness be accounted for outside of the normal functioning of inflectional structure or in a way that is integrated with it?

I argue that the answers to these questions lead to deeper insights about inflectional structure. One of the best ways to understand how linguistic systems work is by looking at how they break.

score with (my) brains, in other words.') The cartoon is mocking the pomposity of the elite – the "smartest" suitor has so little smarts that he cannot even speak Russian without being tripped up.

1.2 Anomalies, epiphenomena, or morphological objects? 7

1.2 Random anomalies, epiphenomena, or (almost) normal morphological objects?

A useful place to start is by looking at how previous research has answered this last question. Should defectiveness be accounted for outside of the normal functioning of inflectional structure, or in a way that is integrated with it? In essence, this asks how much we should expect to learn about inflectional structure from studying defectiveness. Three broad positions – what I call the “gaps-as-anomalies” position, the “gaps-as-epiphenomena” position, and the “gaps-as-morphological-object” position – are identified and illustrated in this section. Note that the analyses discussed here should not necessarily be taken as definitive; the data under investigation have all been subject to alternative analysis. But each reflects a general philosophical perspective on the relationship between defectiveness and inflectional structure that has some degree of currency in linguistics.

The tension between the tendency of lexemes to have complete paradigms on the one hand and the existence of paradigmatic gaps on the other has long been recognized, but defectiveness has not until recently been a topic of much theoretical investigation. Prior to the early 2000s, if paradigmatic gaps were discussed at all, they tended to be treated as random anomalies. In an early account that looked at the Russian 1SG gaps, Halle (1973) posited that the verb form corresponding to the 1SG cell is generated according to productive inflectional rules, but in defective lexemes, a lexically specified feature [–Lexical Insertion] acts as a filter, preventing the generated form from being inserted into syntactic structure. The implication was that the Russian gaps are accidental exceptions to the normal functioning of the inflectional system. He labeled the Russian gaps as “arbitrary.”

More recently, Orgun and Sprouse (1999) put forward a substantially similar analysis of inflectional gaps in Turkish and Tene that they frame within Optimality Theory. They argue for a constraint evaluation component, CONTROL, which applies after EVAL. Within EVAL, candidate forms are evaluated according to the principles of classical Optimality Theory: there is competition between candidates to identify the one that best satisfies ranked constraints and constraints are violable if superseded by a higher-ranked constraint. Within CONTROL, however, only the winning candidate from the EVAL component is evaluated and constraints are never violable, so a candidate form which violates any CONTROL constraint results in a gap.

Orgun and Sprouse’s formalization takes advantage of a general shift in phonological theory towards evaluation of surface well-formedness, and in

8 Introduction

this respect their analysis has a more general theoretical basis than Halle's does. However, like the feature [–Lexical Insertion], the CONTROL component serves a single purpose – to prevent productively generated forms from being inserted into syntactic structure. And like Halle, Orgun and Sprouse locate the formal account of defectiveness outside of the productive grammar. Fanselow and Féry (2002) expand on arguments for the CONTROL component and make an even clearer claim that defectiveness should be accounted for outside of the functioning of “core” grammatical mechanisms.

There is very little about this kind of approach to suggest that gaps interact in meaningful and interesting ways with other aspects of grammar, or that they even have the potential to do so. Since linguists are generally interested in productive grammatical phenomena, it would be easy to assume that inflectional defectiveness is not a rich topic for investigation and that we can simply relegate paradigmatic gaps to the status of idiosyncratic exceptions and move on to investigating more productive areas of the grammar. It is unsurprising that with the exception of Hetzron (1975), Halle's paper inspired few theoretical studies of inflectional defectiveness.

However, more recent research has hinted that paradigmatic gaps do, in fact, interact with inflectional structure in interesting ways. One line of thought has argued that defectiveness is an epiphenomenal result of morphophonological or morphosyntactic competition. For instance, Hudson (2000) explores the English negative 1SG present gap in the copula that is represented by **amn't*. He argues that the lack of a form for the negative 1SG cell is rooted in the organization of inflectional structure in terms of a multiple default inheritance hierarchy. In an inheritance hierarchy, general information that is shared by different elements is specified higher in the tree and then inherited by default by lower nodes, capturing the shared quality. The information structure thus becomes more specific as you move down the tree. Essentially, Hudson argues that the morphosyntactic feature specification of the negative 1SG present subsumes that of both *aren't* (BE, {present, negative}) and *am* (BE, {present, 1st, singular}), and so it should inherit its form from both. However, since the stem shapes *are-* and *am-* are in conflict, multiple non-orthogonal inheritance causes an unresolvable conflict, and thus, a gap. The important issue here is that the *amn't* gap is not lexically specified. In fact, there is no direct correspondent to the gap in Hudson's account. It is an epiphenomenal by-product of the conflicting morphosyntactic feature specifications.¹⁰

¹⁰ The **amn't* gap has been a popular topic for investigation, with a wide range of accounts. See Broadbent (2009) for an argument about the causes of **amn't* that is rooted in patterns of phonological reduction and the history of dialectal variation in English. Other treatments include

1.2 Anomalies, epiphenomena, or morphological objects? 9

Similarly in some respects, but operating more directly at the level of form, Albright (2003) derives two patterns of defectiveness among Spanish verbs from morphophonological rule competition. Both patterns of gaps are connected to stem alternations; for instance, defectiveness occurs in exactly the inflection classes and present tense paradigm cells that are (potentially) affected by diphthongization or raising of the stem vowel under stress – i.e., in all present tense singular forms and third person plural, but not first or second persons plural. (This “N-pattern allomorphy” [Maiden 2004] is observed in many Romance languages and it is tied to defectiveness in several.) Based on the principles of the Minimal Generalization Learner (Albright and Hayes 2002), Albright proposes that low lexeme frequency and low reliability of the alternation combine to produce gradient uncertainty within the inflectional system, the extreme result of which is a paradigmatic gap. In essence, he argues that the grammar is indeterminate because the affected lexemes fall into an area of the lexicon in which there is no island of reliability (high confidence implicative relation) from which to derive the form. This derives the distribution of paradigmatic gaps from the core functioning of the inflectional system. Crucially, however, it also treats them as epiphenomena and denies them any direct grammatical status.

Hudson argues that the **amn’t* gap – and by implication, all paradigmatic gaps – must fall out from conflicting grammar principles in this way because otherwise they would not be learnable.¹¹ Albright is neutral on the question of whether *all* gaps must fall out epiphenomenally from the grammar. However, he argues that lexical specification of the type proposed by Halle predicts that gaps will be randomly distributed in the lexicon. The implication is that gaps of the Spanish (and Russian 1SG) type cannot be adequately treated as lexically specified because their distributional facts indicate that they are not random accidents. He thus seems to make a sharp distinction between gaps that are grammatically motivated (and appropriately treated as epiphenomena) and gaps that are random accidents (and appropriately treated via lexical specification).

In short, previous analyses have often either treated gaps as the by-product of morphological rule competition or dismissed gaps as random accidents to be accounted for through ad hoc lexical specification. Either gaps are regulated by productive grammatical principles to the point of not requiring (and perhaps

Gazdar et al. (1982), Bresnan (2001), and Frampton (2001). Note also that not all dialects have a gap. There are dialects in which *amn’t* is used, as well as dialects in which *ain’t* is used regularly in the first person singular negative. This is not to deny the existence of a gap for at least some speakers (including me), but it does speak to how complicated the issues are.

¹¹ Questions of learnability are discussed in Chapter 7.

10 Introduction

not even allowing!) direct generalization, or they are anomalous to that structure and lie entirely outside of it. These represent opposite approaches in many respects. However, in reducing gaps either to epiphenomena or to anomalies, proposals of the type discussed above *share* the assumption that defectiveness has a status that is fundamentally different from other kinds of morphological patterns.

This assumption has been challenged, however. For instance, Anderson (2010) investigates the Surmiran Romansch (Indo-European > Romance, spoken in Switzerland) verb DUEIR ‘should’, several forms of which are missing. As in Spanish, the key generalization relates to the N-pattern distribution of stems. Some classes of Surmiran Romansch verbs have different stem shapes when stress falls on the stem (in present tense, all singular forms and 3PL) than when it falls on the desinence (1PL and 2PL). Defectiveness in DUEIR follows the same distribution: gaps exist exactly and only in those paradigm cells where we would expect the stressed stem allomorph. Moreover, interestingly, the verbs that are close phonological neighbors to DUEIR (e.g., STUEIR ‘must, should’) are suppletive. The core of Anderson’s argument is that DUEIR is defective precisely because one of its required stems is not lexicalized and the resources of the language do not offer an obvious model on which to form the stressed stem allomorph, or at least, not one that speakers prefer to simply using a different verb.

In some respects, Anderson’s argument has features in common with Albright’s analysis, even beyond the fact that they both investigate variants of the Romance verb stem alternation pattern. In particular, Anderson’s claim that there is no valid model on which a stressed stem of DUEIR could be formed requires that such generalizations are made on a very local scale. This has some overlap with Albright’s notion of an island of reliability – both argue that gaps arise out of localized failures in the generalizability (productivity) of inflectional form. However, a crucial part of Anderson’s analysis is the claim that the alternation in the verb stems is morphologized and that the stem alternants are therefore morphomic in the sense of Aronoff (1994).¹² The structure of the stem space is thus an element in the pattern of defectiveness, independently of the issue of grammar competition/indeterminacy.

Boyé and Cabredo Hofherr (2010) argue even more directly that defectiveness in French and Spanish verbs (including some of the same Spanish facts considered by Albright) is sensitive to the organization of stem space, meaning,

¹² A morphome is a purely morphological structure or function, one that mediates between, but is independent of, phonology and syntax. It is distinct from the concept of a morpheme as a lexical bundle of form and meaning. The morphomic level captures principles of autonomous morphological organization.