

Paul O'Neill*

Defective verbs in Portuguese: a morphomic approach

<https://doi.org/10.1515/cog-2023-0034>

Received February 27, 2023; accepted January 7, 2025; published online April 2, 2025

Abstract: This article provides evidence via a statistical analysis of corpus data that defectivity in Portuguese constitutes a psychological reality for speakers. It then argues that the morpheme-based explanation for Spanish defective verbs is the most appropriate to explain defectivity in Portuguese. Morphemes are abstract distributional patterns of allomorphy based on form-form correspondences alone. The importance of these patterns has not been fully recognised within cognitive linguistics due to their lack of reference to meaning. Morphemes may not be ‘meaningful’ in the narrow sense, i.e. relating to ‘system-external information’, however, they are in the broader sense of them being ‘informative’. Morphemes are extremely informative since they exhibit systematic patterns of form that have a high predictive value within the morphological system. Within such a broader conception of meaning, morphemes have essentially the same predictive value as ‘meaning-driven’ patterns. In this article I present further evidence in favour of morphemes and I argue that Cognitive Linguistics should be open to the possibility of the importance of form-form relationships.

Keywords: morphemes; paradigms; frequency; defective verbs

1 Introduction

Inflectional morphology is usually considered to be extremely productive; neologisms and borrowings are supposedly automatically assigned a full set of inflectional forms, just as native lexemes are. Inflectional defectiveness is the phenomenon whereby certain forms of certain lexemes can be missing. Portuguese grammars identify two major types of morphological defectivity in the verbal paradigm, which coincide with patterns of syncretic allomorphy present in a number of lexemes both in Portuguese and across other Romance languages. These patterns have been defined as autonomous morphological structures or *morphemes* (Aronoff 1994). There is a wealth of diachronic evidence showing how these purely paradigmatic

*Corresponding author: Paul O'Neill, Institute for Romance Philology, Ludwig-Maximilians-Universität, Munich, Germany, E-mail: p.oneill@lmu.de

patterns and distributions condition morphological change, indicating their cognitive reality (e.g., patterns of whole word syncretism in Gallo-Romance (Hinzeln 2012) and Russian (Baerman 2004: 810), morphophonological levelling in Ibero-Romance (O'Neill 2014) and patterns of suppletion in Romance in general (Maiden 2004c, 2018)). In addition to this diachronic evidence there is also much cross-linguistic, synchronic evidence for morphemes: Herce (2023) identifies some 110 morphomic structures across a wide variety of the world's languages and a recent psycholinguistic study on Italian speakers (Cappellaro et al. 2024) provided synchronic experimental evidence for the psychological reality of a particular morphome (L/U-pattern morphome).

Despite such evidence, morphemes have largely been overlooked in Cognitive Linguistics (henceforth CL) due to their lack of reference to meaning: morphemes cannot be reduced to any coherent and convincing semantic feature, they are merely a subset of cells in the paradigm that are particularly cohesive formally and have a greater inter-predictability or diagnostic function with respect to the other cells. Meaning is often seen as central to CL since one of its basic tenets is that language is a system whose primary function is to convey meaning. Morphemes may not be 'meaningful' in the narrow sense, i.e., relating to 'system-external information', however, they are in the broader sense of them being 'informative'. Morphemes are extremely informative since they exhibit systematic patterns of form that have a high predictive value within the morphological system, a value that increases in proportion to the size and diversity of the inflected forms of the system. Within such a broader conception of meaning, morphemes have essentially the same predictive value as 'meaning-driven' patterns. Moreover, upon the assumption that children are not born with innate concepts, it is feasible to hypothesize that the acquisition of inflectional morphology can, in the first instance, be based on form-form relations only – and in this way L1 acquisition of morphology differs radically from L2 acquisition (Clahsen et al. 2010). Systematic form-form correspondences in the linguistic input with a high predictive function are accordingly extremely informative, and therefore meaningful, for the acquisition and ultimately the organisation of morphology.

CL approaches, particularly within the Usage-based tradition, place a primary focus on communication, rather than on a kind of abstract, disembodied 'meaning' and so they are well suited to model morphemes. Also, learning is a central component of/prerequisite for the use of a communication system, and any systematic correlations, whether these be system-external (ostensibly meaningful) or system-internal (morphomic), are of predictive value and thus contribute to a learner's ability to extrapolate beyond the input that they directly encounter. Moreover, according to Schmid (2016: 543), the three main premises which motivate 'the cognitive-linguistic enterprise' are (a) that language interacts with and follows the same principles as

other domains of cognition, (b) that grammatical structure derives from usage, and (c) that grammar is emergent and continuously reorganised under the influence of language use. Usage, however, has produced various purely morphological phenomena which make no reference to information outside the morphological system, the classic example being inflectional classes (for other examples and a discussion see Enger 2019). Morphemes are another instance of such purely morphological phenomena, created by recurrent patterns of form across lexemes. Note that pattern recognition is a fundamental aspect of human cognition (Hawkins and Blakeslee 2005; Koch 2004; Neisser 2014; Shepard 1987). In this article, I present further synchronic evidence in favour of morphemes, claiming that they are responsible for patterns of defective verbs in Portuguese. I also suggest that CL should be open to the possibility of the importance of form-form relationships, which are extremely meaningful in the broader sense of them being informative.

2 The Portuguese data

Portuguese grammars identify two major types of defectivity in the verbal paradigm characterised by the verb *abolir* ‘abolish’ in Table 1 and the verb *blandir* ‘brandish’ in Table 2. This latter verb supposedly does not possess any inflectional forms for the 1SG.PRS.IND, all of the PRS.SBJ and the imperative forms syncretic with the subjunctive; this type of defectivity will be called L-pattern defectivity. The other type of defectivity, as characterised by the verb *abolir*, combines this pattern but also includes the 2SG, 3SG, 3PL PRS IND and the syncretic imperative forms, this will be referred to as N&L-pattern defectivity. These names, as will be elaborated on, refer to predictive patterns of syncretism (morphemes) present in a number of lexemes across the Romance languages.

Cross-linguistically, defective forms have been attributed to semantic and/or pragmatic, or even phonological reasons (see Sims (2016: 58–63) and Bermel and Brown (2023) for a general overview of morphological defectivity). However, on the basis of comparative evidence from the extremely closely related language, Spanish (Maiden and O’Neill 2010; O’Neill 2009, 2010), it would seem that the defectiveness in Portuguese should be given the same explanation. Spanish defective verbs are strikingly similar to those reported in Portuguese grammars: the lexemes are all of relatively low frequency, they belong overwhelmingly to the *-ir* class, they display similar or identical patterns of defectivity and, at times, the exact same lexemes are noted as defective in exactly the same way. For example, the Spanish verbs *abolir* ‘abolish’ and *brandir* ‘brandish’, have exactly the same patterns of defectivity as the Portuguese verbs above. I therefore focus my discussion on the aforementioned accounts for defectiveness in Spanish and assess these against the Portuguese data.

Table 1: The Portuguese verb *abolir* with a simplified imperative paradigm.^a

	PRS.IND	PRS.SBJ	FUT	COND	IPFV.IND
1SG	–	–	abolirei	aboliria	abolía
2SG	–	–	abolirás	abolirias	abolias
3SG	–	–	abolirá	aboliria	abolía
1PL	abolimos	–	aboliremos	aboliríamos	abolíamos
2PL	abolis	–	abolireis	aboliríeis	abolíeis
3PL	–	–	abolirão	aboliriam	abolíam
	PLUP.IND	IPFV.SBJ	PFV	FUT.SBJ	INFL.INF
1SG	abolira	abolisse	aboli	abolir	abolir
2SG	aboliras	abolisse	aboliste	abolires	abolires
3SG	abolira	abolisse	aboliu	abolir	abolir
1PL	abolíramos	abolíssemos	abolimos	abolirmos	abolirmos
2PL	abolíreis	abolísseis	abolistes	abolirdes	abolirdes
3PL	aboliram	abolissem	aboliram	abolirem	abolirem
	IMP (TU)	IMP (VOCÊ)	INF	GER	PST.PTCP
	–	–	abolir	abolindo	abolido

^aPortuguese imperatives are complex since they differ morphologically according to the polarity and the pronoun used, which, depending on the variety, can encode distinctions between (a) deferential and non-deferential address and (b) singular and plural. For simplicity, only the affirmative imperatives corresponding to singular pronouns 'tu' and 'você' are given.

Table 2: The Portuguese verb *brandir* 'brandish'.

	PRS.IND	PRS.SBJ	FUT	COND	IPFV.IND
1SG	–	–	brandirei	brandiria	brandia
2SG	brandes	–	brandirás	brandirias	brandias
3SG	brande	–	brandirá	brandiria	brandia
1PL	brandimos	–	brandiremos	brandiríamos	brandíamos
2PL	brandis	–	brandireis	brandiríeis	brandíeis
3PL	brandem	–	brandirão	brandiriam	brandíam
	PLUP.IND	IPFV.SBJ	PFV	FUT.SBJ	INFL.INF
1SG	brandira	brandisse	brandi	brandir	brandir
2SG	brandiras	brandisse	brandiste	brandires	brandires
3SG	brandira	brandisse	brandiu	brandir	brandir
1PL	brandíramos	brandíssemos	brandimos	brandirmos	brandirmos
2PL	brandíreis	brandísseis	brandistes	brandirdes	brandirdes
3PL	brandiram	brandissem	brandiram	brandirem	brandirem
	IMP (TU)	IMP (VOCÊ)	INF	GER	PST.PTCP
	brande	–	brandir	brandindo	brandido

These accounts can be classed as structural; they see synchronic gaps as not accidental but systematic, reflecting the organization of larger morphological patterns. Indeed, the missing forms are intimately related to parallel patterns of allomorphy within the language. These patterns are defined as autonomous morphological structures or *morphomes* (Aronoff 1994). Models that do not recognize an analogue of morphomic structure are largely compelled to describe these patterns in phonological terms, resulting in high levels of phonological abstraction and a loss of generalization (see O'Neill 2024 §2.2.1.2). As will be discussed, complex abstract phonological factors have also been put forward for defective verbs in Portuguese (Postma 2013), and this explanation has also received validation from an experimental study (Nevins et al. 2014).

The descriptive grammatical treatment of defectivity in Portuguese, however, is characterised by a lack of consensus and contradictions. As shown in Figure 1, there is a great disparity among grammarians as to the number of defective verbs in Portuguese and, as illustrated in Table 3, there is little agreement between scholars over which particular cells of particular lexemes are defective.

Such discrepancy invites one to pose the question as to whether defectivity in Portuguese constitutes a psychological reality for speakers or is just an invention of grammarians, possibly influenced by Spanish grammars. The present article answers this question via a statistical analysis of corpus data in order to validate the

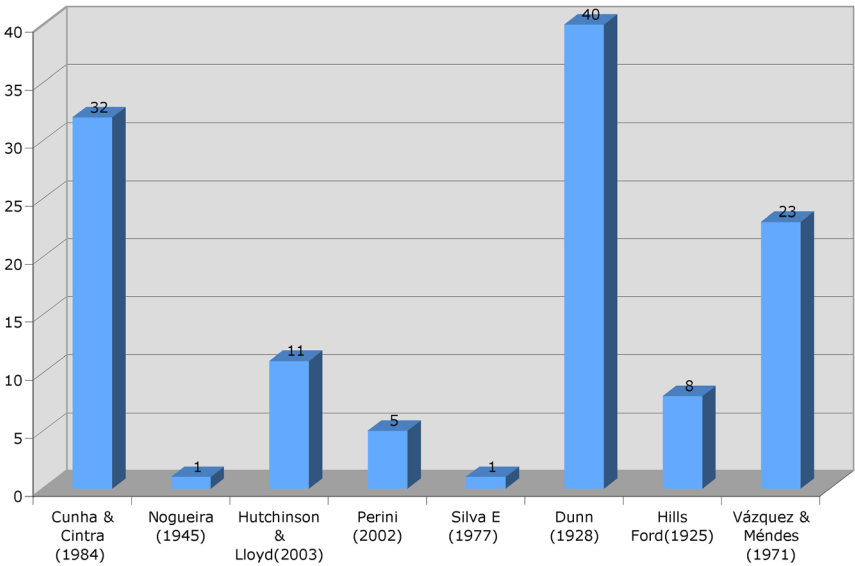


Figure 1: Bar chart of the number of alleged defective verbs in Portuguese according to different grammars. Note that the Cunha and Cintra edition consulted is from (Cunha and Cintra 1984).

Table 3: Different patterns of defectivity for different lexemes according to different Portuguese grammars.

	Cunha and Cintra (1984)	Perini (2002)	Dunn (1928)	Hills and Ford (1925)	Vázquez and Mendez (1971)
abolir 'abolish'	1sg				
	2sg	aboles			
	3sg	abole			
	1pl	abolimos	abolimos	abolimos	abolimos
	2pl	abolis	abolis	abolis	abolis
	3pl	abolem			
demolir 'demolish'	1sg	Not listed as defective		Not listed as defective	
	2sg	demoles			
	3sg	demole			
	1pl	demolimos	demolimos		demolimos
	2pl	demolis	demolis		demolis
	3pl	demolem			
emergir 'emerge'	1sg			Not listed as defective	
	2sg	emerges			emerges
	3sg	emerge			emerge
	1pl	emergimos	emergimos		emergimos
	2pl	emergis	emergis		emergis
	3pl	emergem			emergem
precaver-se 'be prepared'	1sg	Not listed as defective		Not listed as defective	
	2sg				precaves
	3sg				precave
	1pl	precavemos	precavemos		precavemos
	2pl	precaveis	precaveis		precaveis
	3pl				precavem

putative defective lexemes and patterns listed in the literature. As will be seen, the results invalidate the phonological-based theory of defectivity for Portuguese and support the idea of a common ‘morphomic’ explanation for defective verbs in Spanish and Portuguese.

3 The phonological account of defectivity in Portuguese

Postma (2013) has argued that defective verbs in Portuguese are phonologically conditioned by the appearance of a coronal sonorant after the vocalic root vowel. This phonological explanation of defectivity depends on the assumption, dating back to early versions of generative phonology, that the vocalic alternations in Portuguese *-ir* verbs (as illustrated in Table 9 further down) all derive from the same underlying root which undergoes different phonological rules (Harris 1974). The general strategy for assigning a phonological analysis to a morphological phenomenon involves reconstituting an environment that conditions a morphologized alternation (often mimicking historical origins) at a non-surface level. This idea is expressed in the treatment of Portuguese mid-open vowels in the 1/3SG.PRS.IND and the high-vowels in the 1SG.PRS.IND and all of the PRS.SBJ (e.g., *d[ɔ]rme*, *s[ɛ]rve* (3SG.PRS.IND) and *durma*, *sirva* (1/3SG.PRS.SBJ) respectively for *dormir* ‘sleep’ and *servir* ‘serve’) which are assumed to be the result of a processes of vowel harmony with an underlying thematic vowel. Note that this vowel is deleted in the surface representation (*dorm+i+e/a*, *serv+i+e/a*). The different vowel raisings in these forms are attributed to some type of auto-segmental spreading from the deleted theme vowel to the preceding rhizotonic vowel. The claim is that coronal sonorants which occur immediately after this proposed thematic vowel (e.g., in the verb *colorir* ‘colour’) compete with the theme vowel in the suprasegmental docking and the result is defective forms. In order to test the hypothesis that defectiveness arises through slot competition between a surfacing coronal and a postulated harmony-triggering theme vowel, Nevins et al. (2014) asked speakers to produce the 1PL.PRS.IND and 3SG.PRS.SBJ of the defective forms and to rate their confidence in their own productions. A statistical model was then created to establish, based on the confidence rates, what forms might be classed as defective. The results were that the majority of defective verbs, except the verbs *imergir* ‘emerge’ and *ruir* ‘collapse’, contained a coronal consonant, providing independent support for Postma’s (2013) phonologically conditioned account.

There are a number of issues with both studies, however, I will concentrate on the experimental methodology of Nevins et al. (2014). The main problem with this study is that the list of defective verbs which formed the basis of their experiment

was extremely reduced. Firstly, these authors, who restricted their attention to Brazilian Portuguese, consulted a single source for the list of possible defective verbs in Portuguese, the 40 verbs listed in Cunha and Cintra (1984) and they only tested for L-pattern defectivity, not L&N pattern defectivity. This is problematic because, as noted, within Portuguese grammars, there is much discrepancy over the actual number of defective verbs and the types of defectiveness they display. Secondly, the authors eliminated many defective verbs from their study. They used the judgements of just two PhD students to vet the list of 40 defective verbs in Cunha and Cintra (2013). Almost half of these verbs were eliminated from the study because the students did not know the meaning of the verbs. Moreover, more verb forms were removed if the participants in the study rated their familiarity with the lexemes as below 2.5 on a 1–5 scale. This removal of data from the experiment is a serious issue, not only methodologically but also theoretically, since, cross-linguistically it has been noted that defectivity seems to occur in lexemes of a very low frequency (Sims 2019). Taken together, these extensive omissions and exclusions reduce the empirical coverage of the study to the point that no conclusions of any generality can be drawn from any results that it purports to obtain.

4 A purely-morphological/morphomic account of defectivity

A subset of cells in a paradigm or class constitute a morphome when they exhibit a high degree of cohesion and inter-predictability that allows them to serve a diagnostic function. A salient characteristic of morphomes is that the recurrent patterns of allomorphy that they display cannot be attributed to common semantic or phonological properties of the cells, but instead reflect a shared morphological structure.

The morphomes relevant to the present discussion are those termed the N-pattern and the L-pattern, both arbitrary¹ labels coined by Maiden (2004a). These labels denote an alternation within the verbal paradigm whereby an allomorph distinct from the rest of the paradigm is shared by the collection of semantically diverse cells defined by the pattern; those of the L-pattern are 1SG.PRS.IND and all of the PRS.SBJ and syncretic imperative forms, those of the N-pattern are all singular and 3PL forms of the PRS.IND and SBJ and 2SG.IMPER. Witness the examples of L-pattern allomorphy

¹ The label 'L-pattern', was conceived due to the similarity between a rotated orthographic letter 'L' of this paradigmatic set of cells in conventional representations of the verbal paradigm. The N-pattern was termed thus due to the perception that the pattern, as conventionally set out on paper, resembled the form of the letter 'N' in Morse Code.

Table 4: A selection of Spanish verbs displaying allomorphy according to the L-pattern: *decir* ‘say’, *hacer* ‘do’, *venir* ‘come’, *tener* ‘have’, *crecer* ‘grow’, *caer* ‘fall’.

	PRS.IND	PRS.SBJ	PRS.IND	PRS.SBJ	PRS.IND	PRS.SBJ
1SG	digo	diga	hago	haga	vengo	venga
2SG	dices	digas	haces	hagas	vienes	vengas
3SG	dice	diga	hace	haga	viene	venga
1PL	decimos	digamos	hacemos	hagamos	venimos	vengamos
2PL	decís	digáis	hacéis	hagáis	venís	vengáis
3PL	dicen	digan	hacen	hagan	vienen	vengan
	PRS.IND	PRS.SBJ	PRS.IND	PRS.SBJ	PRS.IND	PRS.SBJ
1SG	tengo	tenga	crezco	crezca	caigo	caiga
2SG	tienes	tengas	creces	crezcas	caes	caigas
3SG	tiene	tenga	crece	crezca	cae	caiga
1PL	tenemos	tengamos	crecemos	crezcamos	caemos	caigamos
2PL	tenéis	tengáis	crecéis	crezcáis	caéis	caigáis
3PL	tienen	tengan	crecen	crezcan	caen	caigan

Table 5: L-pattern allomorphy in the Portuguese verbs *ter* ‘have’, *ver* ‘see’, *fazer* ‘do’, *vir* ‘come’, *medir* ‘measure’, and *caber* ‘fit’.

	Indicative	Subjunctive	Indicative	Subjunctive	Indicative	Subjunctive
1SG	tenho	tenha	vejo	veja	faço	faça
2SG	tens	tenhas	vês	vejas	fazes	faças
3SG	tem	tenha	vê	veja	faz	faça
1PL	temos	tenhamos	vemos	vejamos	fazemos	façamos
2PL	tendes	tenhais	vedes	vejais	fazeis	façais
3PL	têm	tenham	vêem	vejam	fazem	façam
	Indicative	Subjunctive	Indicative	Subjunctive	Indicative	Subjunctive
1SG	venho	venha	caibo	caiba	meço	meça
2SG	vens	venhas	cabes	caibas	medes	meças
3SG	vem	venha	cabe	caiba	mede	meça
1PL	vimos	venhamos	cabemos	caibamos	medimos	meçamos
2PL	vindes	venhais	cabeis	caibais	medis	meçais
3PL	vêm	venham	cabem	caibam	medem	meçam

in Table 4 for Spanish and in Table 5 for Portuguese, and N-pattern allomorphy in Table 6 for Spanish and Table 7 for Portuguese; the shaded cells correspond to the morphomic cells. For a full discussion of the origins of these patterns, their different

Table 6: A selection of Spanish verbs displaying N-pattern allomorphy according to the N-pattern: *negar* ‘refuse’, *sentir* ‘feel’, *poder* ‘be able’, *morir* ‘die’, *medir* ‘measure’ and *pedir* ‘ask for’.

	Indicative	Subjunctive	Indicative	Subjunctive	Indicative	Subjunctive
1SG	niego	niegue	siento	sienta	puedo	pueda
2SG	niegas	niegues	sientes	sientas	puedes	puedas
3SG	niega	niegue	siente	sienta	puede	pueda
1PL	negamos	neguemos	sentimos	sintamos	podemos	podamos
2PL	negáis	neguéis	sentís	sintáis	podéis	podáis
3PL	niegan	nieguen	sienten	sientan	pueden	puedan
2SG.IMPER		niega	2SG.IMPER	siente	2SG.IMPER	puede
	Indicative	Subjunctive	Indicative	Subjunctive	Indicative	Subjunctive
1SG	muero	muera	mido	mida	pido	pida
2SG	mueres	mueras	muere	midas	pides	pidas
3SG	muere	muera	mide	mida	pide	pida
1PL	morimos	muramos	medimos	midamos	pedimos	pidamos
2PL	morís	muráis	medís	midáis	pedís	pidáis
3PL	mueren	mueran	miden	midan	piden	pidan
2SG.IMPER		muere	2SG.IMPER	mide	2SG.IMPER	pide

Table 7: A selection of Portuguese –ar verbs which display N-pattern allomorphy: *apegar* ‘attach’, *levar* ‘carry’, *nevar* ‘snow’, *jogar* ‘play’, *rogar* ‘request’, *lograr* ‘achieve’.

	Indicative	Subjunctive	Indicative	Subjunctive	Indicative	Subjunctive
1SG	ap[ɛ]go	ap[ɛ]gue	l[ɛ]vo	l[ɛ]ve	n[ɛ]vo	n[ɛ]ve
2SG	ap[ɛ]gas	ap[ɛ]gues	l[ɛ]vas	l[ɛ]ves	n[ɛ]vas	n[ɛ]ves
3SG	ap[ɛ]ga	ap[ɛ]gue	l[ɛ]va	l[ɛ]ve	n[ɛ]va	n[ɛ]ve
1PL	ap[ɛ]gamos	ap[ɛ]guemos	l[ɛ]vamos	l[ɛ]vemos	n[ɛ]vamos	n[ɛ]vemos
2PL	ap[ɛ]gais	ap[ɛ]gueis	l[ɛ]vais	l[ɛ]veis	n[ɛ]vais	n[ɛ]veis
3PL	ap[ɛ]gam	ap[ɛ]guem	l[ɛ]vam	l[ɛ]vem	n[ɛ]vam	n[ɛ]vem
	Indicative	Subjunctive	Indicative	Subjunctive	Indicative	Subjunctive
1SG	j[ɔ]go	j[ɔ]gue	r[ɔ]go	r[ɔ]gue	l[ɔ]gro	l[ɔ]gre
2SG	j[ɔ]gas	j[ɔ]gues	r[ɔ]gas	r[ɔ]gues	l[ɔ]gras	l[ɔ]gres
3SG	j[ɔ]ga	j[ɔ]gue	r[ɔ]ga	r[ɔ]gue	l[ɔ]gra	l[ɔ]gre
1PL	jogamos	joguemos	rogamos	roguemos	logramos	logremos
2PL	jogais	jogueis	rogais	rogueis	lograis	logreis
3PL	j[ɔ]gam	j[ɔ]guem	r[ɔ]gam	r[ɔ]guem	l[ɔ]gram	l[ɔ]grem

instantiations in the different Romance languages and justifications as to their autonomous morphological status, see Maiden (2018) and references therein; see O'Neill (2024) for arguments against the Spanish allomorphy being phonologically conditioned.

It should also be noted that, in Portuguese, the N-pattern interacts with the L-pattern, which effectively dominates it and reduces the N-pattern to the 2SG, 3SG, 3PL and relevant imperative forms, thereby creating a new pattern, which I have termed the Prolific-Portuguese Pattern. This name is due to this pattern being extremely prominent in the Portuguese verb; nearly all² *-er* and *-ir* verbs which display an orthographic mid-vowel as the root-vowel exhibit allomorphy which alternates in accordance with this pattern. In *-er* verbs, as illustrated in Table 8, the L-pattern cells display a high-mid vowel in the root which alternates with an open-mid vowel in the reduced N-pattern cells. In *-ir* verbs, the root of the reduced N-pattern cells also displays an open-mid vowel, but the vowel in the L-pattern is a high vowel; witness the examples in Table 9. Remember also that a number of lexemes are reported as defective exclusively in the cells of this morphome (see *abolir* ‘abolish’ in Table 1).

Such morphomes have been used to argue that systematic form-form correspondences should be reflected in any theory of grammar (Aronoff 1994: 25) especially since there is much diachronic evidence to suggest that they are psychologically real for speakers (Maiden 2018) That is, the historical perseverance and resilience of morphomic patterns suggest that they perform a function that speakers seem willing to preserve even when languages undergo significant restructuring. This function is that they are informative or ‘meaningful’ system-internally.

Within this context, Maiden and O’Neill (2010: 121) noted for Spanish that although the cells of the morphomes have, amongst themselves, a high interpredictive function, the particular allomorphy within these cells is not readily predictive from cells outside the morphome. They proposed that the defectivity in

Table 8: The Portuguese *-er* verbs *dever* ‘owe’, *mover* ‘move’, *beber* ‘drink’ (L>N).

	Indicative	Subjunctive	Indicative	Subjunctive	Indicative	Subjunctive
1SG	d[e]vo	d[e]va	m[o]vo	m[o]va	b[e]bo	b[e]ba
2SG	d[ɛ]ves	d[ɛ]vas	m[o]ves	m[o]vas	b[ɛ]bes	b[ɛ]bas
3SG	d[ɛ]ve	d[ɛ]va	m[ɔ]ve	m[o]va	b[ɛ]be	b[ɛ]ba
1PL	devemos	devamos	movemos	movamos	bebemos	bebamos
2PL	devis	devais	moveis	movais	bebeis	bebais
3PL	d[ɛ]vem	d[ɛ]vam	m[ɔ]vem	m[o]vam	b[ɛ]bem	b[ɛ]bam

2 According to Cunha and Cintra (1984: 416) the exceptions to this rule are: verbs whose root vowel is nasalized due to a following heterosyllabic consonant (*encher* ‘fill up’, *romper* ‘break’); Brazilian Portuguese verbs whose root vowel is followed by a nasal consonant (*temer* ‘fear’, *comer* ‘eat’); the verbs *querer* ‘want’ and *poder* ‘be able’.

Table 9: The Portuguese *-ir* verbs *servir* ‘serve’, *dormir* ‘sleep’, *vestir* ‘dress’.

	Indicative	Subjunctive	Indicative	Subjunctive	Indicative	Subjunctive
1SG	sirvo	sirva	durmo	durma	visto	vista
2SG	s[ε]rves	sirvas	d[ɔ]rmes	durmas	v[ε]stes	vistas
3SG	s[ε]rve	sirva	d[ɔ]rme	durma	v[ε]ste	vista
1PL	servimos	servamos	dormimos	durmamos	vestimos	vistamos
2PL	servis	servais	dormis	durmais	vestis	vistais
3PL	s[ε]rvem	servam	d[ɔ]rmem	durmam	v[ε]stem	vistam

Spanish is due to speakers being ‘paranoid’³ about the allomorphy in these particular cells and they therefore avoid the forms even where there are no reasonable grounds to expect allomorphy to occur. This inexact notion of ‘paranoia’ is clarified and substantiated in a later study by O'Neill (2010). It is claimed that due to the high number of *-ir* verbs that display N-pattern and L-pattern allomorphy, it has become a ‘rule’ for the language that all *-ir* verbs have a lexicalized root exclusively for these patterns. Essentially, the dominant pattern of frequent verbs in an inflectional class has been analogically extended to all verbs of the class. Defective verbs are therefore verbs which have not had the roots for these verbs committed to memory due to the lexemes being low frequency and the particular idiosyncratic history of the verbs. According to a diachronic corpus study (O'Neill 2009), these verbs are historically those which (i) did have a full paradigm but which were increasingly used in the past-participle form as verbal adjectives (e.g., *compungido* ‘remorseful’, *embutido* ‘stuffed’, *tullido* ‘crippled’) and so their N&L-pattern forms were lost and (ii) borrowings from Latin or French which entered the language in the infinitive or past participle (*abolir* ‘abolish’, *blandir* ‘brandish’, *precaver* ‘guard against’) and which never possessed forms for the N&L patterns and which still today have not been created.

The following sections investigate whether a parallel morphomic analysis can be applied to Portuguese. The first step towards developing a viable analysis of the general phenomenon of defectivity in Portuguese involves a careful evaluation of the competing claims about the number and type of defective cells in Portuguese. By grounding an analysis in a principled enumeration of cases, it is also possible to evaluate more extensively the proposed phonological account for defectivity: the influence of coronal sonorants. In what follows, therefore, I propose a list of defective verbs based on a corpus study.

3 This idea of paranoia should not be confused with Albright’s (2003, 2010) explanation for defectivity related to uncertainty and morphological variability. As explained in Maiden and O'Neill (2010: 121) and O'Neill (2010) this theory makes predictions which are not borne out by the Spanish data.

5 Corpus study and statistical analysis of defective verbs in Portuguese

To establish the defective verbs of Portuguese, I followed a similar methodology used for Spanish (O'Neill 2009). I carried out corpus searches on the list of possible defective verbs and established whether these lexemes could be categorized as defective based on a predictive statistical model. This model was based on the frequency distribution of the forms of 100 non-defective lexemes and compared their distribution of forms with that of the putative defective forms. The statistical model provides a nuanced, gradient analysis of defectivity, in place of an 'all or nothing' account. Lexemes are classified as defective if they contain cells that are realized by forms that show a frequency of occurrence that is significantly below the expected frequency. I concentrated exclusively on European Portuguese and used the Centempúblico corpus, with some 180 million words, comprised of texts published in the Portuguese newspaper *El Público*.⁴

Firstly, a list was compiled of all the verbs classed as defective in the different Portuguese grammars (Figure 1). Secondly, these lexemes were then subject to corpus searches in which, in addition to the overall frequency of each lexeme, the frequency of the following inflectional forms for these verbs were noted:⁵ (a) 3SG.PRS.IND, (b) 1PL.PRS.IND, (c) all the PRS.SBJ, (d) PST.PTCP. The 3SG.PRS.IND was used as a diagnostic of the modified N-pattern and the forms of the PRS.SBJ as a diagnostic for the L-pattern. The forms of the 1PL.PRS.IND were noted to check whether it was the case that a verb was simply defective in the present tense, as opposed to it being defective in accordance with the different morphomic patterns. The forms of the PST.PTCP were logged since for Spanish it had been noted (O'Neill 2009) that a number of the supposed defective verbs were not verbs but adjectives whose forms coincided with the PST.PTCP forms of the putative verbs. The overall frequency of the purported defective verbs was noted in order to make testable predictions based on a mathematical model extracted from corpus data of verbs which were not classed as defective.

4 As noted, Nevins et al. (2014) focused exclusively on the Brazilian variety of Portuguese but designed their experiment on the basis of a modified list of the verbs classed as defective for all varieties of Portuguese by Cunha and Cintra (2013). Since these verbs are a subset of those in the European Portuguese Centempúblico, any results obtained in the current study cover the patterns considered in the phonological accounts, eliminating any potential concerns about dialectal variation as a confounding factor.

5 N.B all these data were manually checked to ensure that they corresponded to *bona fide* forms of the lexeme and not misclassifications by the corpus.

For this model, a sample of 100 verbs was taken from the 1,000 most frequent verbal lexemes according to the same corpus; 50 from the top end of the list and 50 from the bottom end. For these lexemes, in addition to their overall frequency in the corpus, the frequency for the following forms was noted: (a) 3SG.PRS.IND, (b) 1PL.PRS.IND and (c) all the PRS.SBJ.⁶ The numerical values for these forms were then converted into logarithmic values and a number of linear regressions were calculated using the software R (R Core Team 2021). The result was a statistical model which, from the overall frequency of a non-defective Portuguese verbal lexeme, could predict not only the expected values for (a)–(c) above but also a maximum and minimum value for each form according to the natural variance of the model.

This statistical model was subsequently applied to the attested overall frequency values of the purported defective verbs in order to generate a range of values for the forms (a)–(c) above. These values were then compared with the actual attested values of these forms in the corpus. If the frequency of a particular verb form fell below the minimum value predicted by the statistical model, then this form was classed as defective, converting raw patterns of continuous variation into a discrete classification of lexemes and cells. To illustrate how this process worked, compare in Table 10 the values generated by the mathematical model for two supposed defective verbs: *abolir* ‘abolish’ and *punir* ‘punish’.

From this table, it can be concluded that the verb *abolir* is defective in 3SG.PRS.IND and all the PRS.SBJ forms, but not in the 1PL.PRS.IND. This verb was therefore classed as defective according to the N&L-pattern. Contrastively, the verb *punir* was not classed as defective since its attested forms in the corpus fell within the predicted maximum and minimum values for this verb, given its overall frequency. As with other statistical analyses (including those describing Zipfian power-law distributions), this

Table 10: Values for two Portuguese verbs.

Verb	Overall frequency	Verb form	Frequency	Expected value	Minimum value	Maximum value	Defective
<i>abolir</i> ‘abolish’	1,404	3SG.PRS.IND	13	83	21	334	YES
		1PL.PRS.IND	2	1	0.07	8	NO
		PRS.SBJ	2	18	4	85	YES
<i>punir</i> ‘punish’	3,443	3SG.PRS.IND	163	245	61	978	NO
		1PL.PRS.IND	3	3	0.3	32	NO
		PRS.SBJ	82	49	11	226	NO

6 These forms were also checked to ensure they belonged to the corresponding lexeme.

Table 11: Values, to the second decimal point, for the Portuguese verb *polir* ‘polish’. YES means that the verb form is defective according to the mathematical model while NO means that it is not.

Verb	Overall frequency	Verb form	Freq.	Expected value	Min. value	Max. value	Defective
<i>polir</i> ‘polish’	361	3SG.PRS.IND	2	20.34	5.01	82.58	YES
		1PL.PRS.IND	0	0.10	0.01	1.07	NO
		PRS.SBJ	0	5.07	1.08	23.93	YES

methodology does not, in general, induce appropriate thresholds for statistical outliers, such as hapaxes and low-frequency lexemes. Witness in Table 11, the data for another alleged defective verb, the verb *polir* ‘polish’.

From this data it is unclear whether the verb is defective according to the N&L-pattern or just defective in the entire present tense since, on the basis of the verb’s overall frequency, the probable range of values for the 1PL.PRS.IND form was between 0.01 and 1.07, with a predicted expected value of 0.10. To compensate in cases that exhibit a skewed or ambiguous distribution the Poisson distribution was used to calculate if the form should be classed as defective, since equation $P(0) = G0e^{-G}/0!$ can predict, from the minimum occurrence value (G), the probability (P) that the actual frequency in the corpus is zero. The result for *polir* was a p -value of 0.99 that the 1PL.PRS.IND should produce a result of zero in the corpus. This form was therefore considered not defective and the lexeme, overall, was considered defective according to the N&L-pattern.

The general problem of extrapolating from the distribution of sparsely-attested lexemes is, however, intrinsic to the kind of statistical methodology adopted in this study. Extremely infrequent lexemes will be predicted as having a threshold value of zero in the corpus, and, counterintuitively, will be classified as non-defective items with full paradigms. Therefore, for the purposes of the present study, and on the basis of results from the Spanish data, two types of lexemes were excluded from the statistical analysis: (a) those whose overall frequency in the corpus was less than 20 and (b) those which were just attested in the PST.PTCP/ADJ form. The former were classed as infrequent and their inflectional forms were noted, the latter were classed as PST.PTCP /adjectives.

5.1 Results

A summary of the results can be found in Table 12, divided into different categories.

Table 12: Summary of results for the number of defective verbs in Portuguese.

Portuguese	
53	Number of alleged defective verbs
6	Not attested in the corpus: <i>buir</i> ‘polish’ (0), <i>delir</i> ‘dissolve’(0), <i>emolir</i> ‘soften’ (0), <i>exinanir</i> ‘empty’ (0), <i>embair-se</i> ‘deceive’ (0), <i>cernir</i> ‘sift’ (5 INF),
5	Infrequent but not attested in PRS.IND: <i>brunir</i> ‘shine’ (1 PRET. 1 INF., 2 IPFV.IND, 2 PST.PTCP, 1 GER) <i>haurir</i> ‘drain’ (1 PST.PTCP, 1 PRET), <i>jungir</i> ‘yoke’(2 INF, 1 PST.PTC), <i>latir</i> ‘bark’ (5 PST.PTCP, 5 INF. 1 GER); <i>remir</i> ‘redeem’ (5 PST.PTCP, 5 INF., 1 PRET.);
9	Only/mainly attested in past participle/adjective: <i>aguerrir</i> ‘to train in war’(522), <i>com-balir-se</i> ‘to become weak/deteriorate’ (76), <i>comedir-se</i> ‘to show restraint’(459), <i>descomedir-se</i> ‘be rash’ (5), <i>empedernir(se)</i> ‘to become hardened’ (139), <i>foragir-se</i> ‘to emigrate’ (144), <i>fornir</i> ‘provide’(4), <i>renhir</i> ‘to argue, to dispute’ (523); <i>aturdir</i> ‘stun’ (72 PST.PTCP, 1 INF. 1 IPFV.IND)
9	Not defective: <i>advir</i> ‘to come after’, <i>delinquir</i> ^a ‘commit a crime’, <i>explodir</i> ‘to explode’, <i>ganir</i> ‘to howl’, <i>adequar</i> ‘to adjust, to accommodate’ <i>punir</i> ‘to punish’, <i>fremir</i> ‘to roar’, <i>fulgir</i> ‘to flash’, <i>submergir</i> ‘submerge’
10	Defective in N & L-pattern: <i>abolir</i> ‘abolish’, <i>banir</i> ‘banish’, <i>carpir</i> ‘mourn/weep’ <i>colorir</i> ‘colour’, <i>demolir</i> ‘demolish’, <i>falir</i> ‘fail’, <i>florir</i> ‘flower’, <i>polir</i> ‘polish’, <i>precover-se</i> , ‘prepare against’, <i>reaver</i> ‘regain’.
12	Defective in L-pattern: <i>brandir</i> ‘brandish’, <i>compelir</i> ‘compel’, <i>discernir</i> ‘discern’, <i>emergir</i> ‘emerge’, <i>exaurir</i> ‘drain’, <i>extorquir</i> ‘extort’, <i>feder</i> ‘stink’, <i>fruir</i> ‘enjoy’, <i>gerir</i> ‘digest’, <i>imergir</i> ‘immerse’, <i>retorquir</i> ‘reply’, <i>ungir</i> ‘to anoint’
2	Defective in reduced N-pattern: <i>escapulir</i> ‘slip off’; <i>munir</i> ‘furnish’

^aDespite this verb being very infrequent (6 tokens only); it was attested in the 3PL.PRS.IND).

The first thing to note is that Portuguese does have a number of defective verbs. Significantly, many of these verbs do not contain a coronal sonorant after the root vowel, demonstrating that coronal sonorants cannot be treated as conditioning defectivity. This larger pattern invalidates any phonological account that adopts this assumption, including the analysis of Postma (2013), and makes the type of experimental study conducted by Nevins et al. (2014) irrelevant to an understanding of the phenomenon of defectivity in Portuguese.

The verbs classified as defective in Portuguese, are remarkably similar, and in some cases identical, to those reported for Spanish by O'Neill (2010). For comparison, the Spanish defective verbs are listed below in Table 13.

The striking similarity between defective verbs in Spanish and Portuguese cannot reasonably be treated as accidental. There are three obvious types of explanations for their convergence. The least interesting would dismiss the similarity as an artifact of applying a common method to the two languages. However, this possibility can readily be discounted, given that the method will determine different classes when applied to languages with less congruent patterns of defectivity. A second explanation would disregard the convergence as a historical relic

Table 13: Proposed list of defective verbs in Spanish according to O'Neill (2010).

Type	Verbs
Defective in the N&L-pattern	10 <i>abolir</i> ‘abolish’, <i>asir</i> ‘grasp’, <i>balbucir</i> ‘babble’, <i>bruñir</i> ‘polish’, <i>compungir</i> ‘feel remorseful’, <i>curtir</i> ‘tan (leather)’, <i>embutir</i> ‘stuff’, <i>precar</i> ‘provide against’, <i>raer</i> ‘scrape’, <i>ungir</i> ‘anoint’
Defective only in the L-pattern	3 <i>blandir</i> ‘brandish’, <i>estreñir</i> ‘cause constipation’, <i>erguir</i> ‘erect’

that has merely been inherited in the two languages. While this explanation is presumably relevant to the origins of the patterns, it does not account for their preservation in both of the descendant languages. A viable explanation of the preservation (and resilience) of these patterns must identify some kind of function that has contributed to their survival in independent languages. The null hypothesis in this case is that the same factors proposed for Spanish in Section 4 apply as well to Portuguese, and that the learning-facilitating predictive value of morphomic patterns were largely responsible for the preservation of a shared innovation.

6 Purely morphological/morphomic account of defectivity II

At first glance, the Spanish and Portuguese data seem confusing since it makes little sense that the Spanish verb *ungir* ‘anoint’ and the Portuguese *banir* ‘banish’ should be defective, since on the model of regular word formation for *-ir* verbs in these languages the 3SG.PRS.IND and PRS.SBJ forms should be *un[x]e*, *un[x]a* and *bane*, *bana* respectively; they are entirely predictable. However, the morphomic explanations for defectivity in Spanish (Maiden and O'Neill 2010: 121; O'Neill 2009, 2010) can explain this conundrum. The basic argument is the following: the cells of the morphomes were, for historical reasons, the locus of different types of allomorphy in a core set of frequently occurring *-ir* verbs. This allomorphy could not be predicted based on verb-forms outside the morphome and thus the allomorphs were lexically listed. This lexically listed nature of the morphomic forms of a number of verbs with a high token frequency was then analogically extended to all verbs of this class. The defective verbs are those verbs for which speakers have not heard and thus memorized a form for the different patterns.

The claim that the functional value of morphomic N&L patterns contributes to the preservation of patterns in Spanish makes testable predictions for closely-related

languages. If defectivity is a product of the N&L-pattern forms of *-ir* verbs being largely unpredictable due to them being the locus of different types of allomorphy, then the prediction would be that defective verbs should exist in related varieties in which the N&L-patterns are attested and they are also the locus of different types of lexically bound allomorphy. Likewise, defective verbs should not be attested in those varieties in which either the N&L-patterns are not active or the allomorphy is largely predictable. In this context, it is instructive to analyse Catalan, a closely related language to Spanish but one which has no listed defective verbs in the N&L-patterns.

Catalan has two different sub-paradigms for the present tense forms of *-ir* verbs. The first, traditionally termed conjugation IIIa, exemplified by the verb *sentir* 'feel', has rhizotonic stress and historically would have been the locus of extensive vowel allomorphy in the N&L-Pattern cells but, for a number of complex historical reasons, this allomorphy was levelled (see Wheeler (2015) for an overview). In the second subclass of *-ir* verbs, traditionally termed conjugation IIIb and exemplified by *servir* 'serve' and *abolir* 'abolish' in Table 14, the stress falls on the augment *-eix* (Maiden 2004b) which occurs after the root and is present for all verbs of this sub-paradigm. Therefore, whilst the Catalan verbs do display allomorphy in the N-pattern, unlike Spanish, for the IIIb conjugation this allomorphy is always predictable on the basis of any other form of the verb.

It appears to be no coincidence, that the Catalan cognate forms of the attested Spanish defective verbs⁷ all belong to this IIIb conjugation, which display the augment *-eix* in the N-pattern cells. These verbs **all** have a full paradigm even if they are extremely infrequent or, used mostly in the past participle form as adjectives. The reason why Catalan, unlike Spanish, has no similar defective verbs is because, for the Catalan IIIb class, knowledge of inflected forms outside the N-pattern predicts the forms of the N-pattern for all verbs.

Portuguese, on the other hand, does have defective verbs as attested in section 5.1. And, as expected, Portuguese has extensive L-pattern and (modified) N-pattern allomorphy as can be appreciated in Table 5, Table 8 and Table 9 and summarized in Table 15 below for *-ir* verbs. The explanation proposed for the Spanish data thus extends straightforwardly to the Portuguese patterns. The allomorphy in these cells was so unpredictable that it became a generalisation for this conjugation that the N&L-pattern forms were listed: the defective verbs are low frequency lexemes which, for historical reasons, have either lost or not acquired their N&L-pattern forms.

⁷ *abolir* 'abolish', *acolorir* 'colour', *brunir* 'burnish/polish', *compelir* 'compel', *compungir* 'make remorseful', *demolir* 'demolish', *discernir* 'discern', *embotir* 'stuff', *emergir* 'emerge', *extorquir* 'demand/impose fines', *florir* 'flower', *fruir* 'enjoy', *garantir* 'guarantee', *polir* 'polish', *retorquir* 'reply', *submergir* 'submerge', *ungir* 'anoint'.

Table 14: The Catalan verbs *sentir* ‘feel’, *servir* ‘serve’ and *abolir* ‘abolish’.

	Indicative	Subjunctive	Indicative	Subjunctive	Indicative	Subjunctive
1SG	sento	senti	serveixo	serveixi	aboleixo	aboleixi
2SG	sents	sentis	serveixes	serveixis	aboleixes	aboleixis
3SG	sent	senti	serveix	serveixi	aboleix	aboleixi
1PL	sentim	sentim	servim	servim	abolim	abolim
2PL	sentiu	sentiu	serviu	serviu	aboliu	aboliu
3PL	senten	sentin	serveixen	serveixen	aboleixen	aboleixen

Table 15: A selection of Portuguese *-ir* verbs classed in accordance with the root vowel and type of allomorphy in the 1SG present indicative (representative of the L-pattern) and 3SG present indicative (representative of reduced N-pattern).

Root vowel	Verb	Gloss	1SG PRS.IND = L-pattern	Type of allomorphy	3SG PRS.IND = modified N-pattern	Type of allomorphy
<a>	<i>sair</i>	go out	<i>saio</i>	consonantal	<i>sai</i>	irregular desinence
	<i>partir</i>	leave	<i>parto</i>	none	<i>parte</i>	none
<e>	<i>medir</i>	measure	<i>meço</i>	consonantal	<i>m[ɛ]de</i>	vocalic
	<i>servir</i>	serve	<i>sirvo</i>	vocalic	<i>s[ɛ]rve</i>	vocalic
	<i>submergir</i>	submerge	<i>submerjo</i>	none	<i>subm[ɛ]rge</i>	vocalic
	<i>agredir</i>	assault	<i>agrido</i>	vocalic	<i>agride</i>	vocalic
<i>	<i>frigir</i>	fry	<i>frijo</i>	none	<i>fr[ɛ]ges</i>	vocalic
	<i>permitir</i>	permit	<i>permito</i>	none	<i>permite</i>	none
<o>/<ou>	<i>ouvir</i>	hear	<i>ouço/oiço</i>	consonantal	<i>ouve</i>	none
<ou>	<i>dormir</i>	sleep	<i>durmo</i>	vocalic	<i>d[ɔ]rme</i>	vocalic
<u>	<i>cumprir</i>	fulfill	<i>cumpro</i>	none	<i>cumpre</i>	none
	<i>acudir</i>	help	<i>acudo</i>	none	<i>ac[ɔ]des</i>	vocalic
	<i>instruir</i>	instruct	<i>instruo</i>	none	<i>instrui</i>	irregular desinence

Confirmation of the lexical nature of this allomorphy in *-ir* verbs comes from experimental studies conducted with entirely independent grounds for distinguishing regular from irregular patterns, both in Portuguese (Veríssimo and Clahsen (2009) and Spanish (Linares et al. 2006: 113; Rodriguez-Fornells et al. 2002; Yang 2016: 149). Yang (2016: 150) specifically notes for Spanish that ‘for every item in the third conjugation, learners must have positive evidence for its stem alternation because it is lexically arbitrary. And they will be at a loss for verbs that do not have attested forms, resulting in gaps.’.

In sum, the essential particularity of *-ir* verbs in both Spanish and Portuguese can be correlated with the high token and type frequency of lexemes in this class which display allomorphy precisely in these patterns. These lexemes forge out a structure for this class of verbs whereby the N&L pattern forms are not predictable by the other forms of the paradigm, but must be memorized. This pattern, which is only explicitly present in a number of 'irregular' lexemes, is then adopted by all verbs of this class, even those classed as 'regular'.

Note also, that the general idea that irregular forms with a high token frequency can act as a type of cognitive anchor or prototype around which to organize less-frequent forms, is also the conclusion for a number of studies in CL (Goldberg et al. 2004; Strack and Mussweiler 1997) and is supported by historical data (Fertig 2013: §7.3.6, §8.6.1; Sims-Williams 2022). Additionally, there is historical (O'Neill 2014: 58–65) and computational evidence (Pirrelli et al. 2007) to support the specific claim that the patterns of inflection of frequently occurring and irregular verbs can be adopted by/are valid for verbs classed as regular, thus challenging the regular/irregular dichotomy.

7 Discussion

The present analysis of defective verbs in Spanish and Portuguese assumes a notion of the morpheme that plays a central role in the structure of these Romance languages and assumes that they constitute a psychological reality for their speakers. Morphemes in this sense are intra-morphological meaningful structures, providing information about other forms, rather than about language-external elements. Hence, morphemes lack counterparts in models that approach the description of language in terms of associations between 'units of meaning' and 'units of form', and have provoked resistance from advocates of these types of 'constructive' models (see Blevins (2006) for the constructive/abstractive distinction and Luis and Bermúdez-Otero (2016) and O'Neill (2022, 2024) for the polemic nature of morphemes), which have the morpheme as the basic unit of storage and processing, and derive word-forms based on symbolic rules associated with these underlying meaning-form pairings.

Models of morphology within CL are not constructive but, rather, abstractive: the minimal meaningful unit and basic element of lexical storage is the word, and complex word forms are stored in their entirety in the lexicon, where interconnections between words 'provide[s] generalisations and segmentation at various degrees of abstraction and generality whereby units such as morpheme, arise from the relations of identity and similarity that organize representation' (Bybee 2001: 7). Abstractive models of morphology have been argued as being more

suitable to model morphemes (Blevins 2010, 2016; O'Neill 2014), and, in fact, the historical diagnostic for the psychological reality of morphemes (the fact that they act as a domain within which morphophonemic alternations are levelled) was directly borrowed from Bybee's model of morphology. Nevertheless, morphemes have largely not been recognized in the CL literature due to their lack of reference to system-external information, i.e., meaning. However, as noted throughout this article, morphemes are meaningful in the broader sense of being informative: they exhibit systematic patterns of form that have a high predictive value within the morphological system.

Meaning is central to CL in which morphological structure is conceived as being symbolic in nature (Langacker 2019), and as the union of semantic and phonological connections. These two types of connections do not have equal status, however; semantic relations always take precedence over purely formal ones. This point is reiterated and stressed on a number of occasions by Bybee (2001: 117, 1985: 118), in whose model of morphology, a fundamental tenet of the lexical organisation of whole word forms is that form is subordinate to meaning to the extent that "the formal organisation of a paradigm **diagrams** the semantic organisation so that forms that are more similar semantically will be more similar in morphophonemic shape." (Bybee and Brewer 1980: 58). The central point of morphemes, however, is that one can have consistent patterns of similar form which do not diagram any type of system-external semantic organisation but have an essential system-internal predicative function.

At this point, it is informative to review Bybee's views and interpretation of velar allomorphy in Spanish (defined as the L-pattern morpheme in Table 4). Bybee (1985: 68) admits the lack of semantic relations between the L-pattern cells and defines them as corresponding to a 'coherent morphological set' (Bybee 1985: 71). However, whilst she acknowledges that such purely morphological patterns were 'originally morphologically arbitrary' (Bybee 1985: 71), she is reluctant to accept that they could endure in the language without any reference to system-external information and claims such forms 'can be preserved if they coincide or can be made to coincide with the morphological relations among forms.' (Bybee 1985: 68). Thus, the purely morphological distribution is justified with reference to issues relating to markedness, but in a way very specific to Spanish since she acknowledges (Bybee 1985: 78n9) that markedness alone cannot explain the distribution of this allomorphy. This markedness-based motivation for the L-pattern suffers from a number of problems (O'Neill 2015: 493n4), however, especially regarding the actual data.⁸

⁸ Bybee's argument is the following: the PRS.SBJ is more marked with respect to mood than the PRS.IND, thus, in Spanish, the 1SG.PRS.IND is in a basic-derived relationship with the PRS.SBJ. The author provides quantitative data which shows that, of all the forms of the PRS.IND, the 3SG is the most frequent,

Abstracting away from the specific data, however, what we have here is an acknowledgement that morphophonemic alternations need not always diagram semantic ones but a reluctance to accept that purely morphological relations can structure morphology without any reference to system-external factors. This stance is typical of CL approaches to morphology whereby there is a tacit assumption that form-form relations are secondary and a tendency to (a) try to explain what seems to be purely morphological phenomena in terms of semantics⁹ and (b) to only recognize purely morphological factors as a last-resort option when all other possible analyses have been exhausted. For example, one anonymous reviewer suggested that the judgements of relatedness and unrelatedness between the Romance morphemes are impressionistic and pointed out that there are more objective methods for quantifying relatedness of paradigm cells via distributional semantics (e.g., Chuang et al. (2022)).

CL should be open to the possibility of the importance of form-form relationships. Two central tenets of cognitive linguistics are that language is a system whose primary function is to convey meaning and that grammar emerges through usage. As noted, however, usage has produced various purely morphological phenomena which have no relation to established semantic meanings. The classic example is inflectional class. It is widely acknowledged that alongside languages with classes whose membership is correlated with semantic meaning, there are numerous languages in which membership is totally arbitrary and different lexemes have different, but systematic, morphological alternations depending on their random classification. Enger (2019) notes that ‘Langacker (1987: 422) is completely unfazed by arbitrary distributional classes, whose existence he simply acknowledges; they do not violate his “Content Requirement”’. Also, within the broader understanding of meaning as adopted in this article, these arbitrary different classes *can* be classed as

followed by the 1sg. Her claim is therefore that ‘The 3s[ingular] Present Indicative serves as the base for all the Present Indicative, except 1s[ingular], and the 1s[ingular] form serves as the base for all of the Present Subjunctive.’ (Bybee 1985: 72). To my knowledge, there is no comparative dialectal or historical evidence to corroborate this conclusion but there does exist a wealth of evidence to the contrary, i.e. which suggest that the subjunctive forms are not derived from the 1sg.PRS.IND. (see O'Neill 2011b: §5.1). Indeed the conclusion from the historical survey by Wheeler (2011) of the spread of velar allomorphy in Catalan is that the process of morphological replacement of an etymological form by a velar allomorph actually respects markedness principles. That is, less marked forms were not replaced before more marked ones. Such a conclusion contradicts the idea that the more marked forms of the present subjunctive were derived from the less marked 1sg.PRS.IND since the former generally adopted the velar allomorphy before the latter.

⁹ See for example the semantic interpretation in Schäfer and Pankratz (2018) of what has long been considered as purely morphological linking elements, or *Fugenlaut*, which appear in numerous German compounds (Wohnungsmiete ‘flat rent’, Hochzeitskleid ‘wedding dress’ Seitenzahl ‘page number’, Schafskopf ‘mutton head’).

meaningful: they are informative system-internally and have a high predictive value despite their lack of ‘system-external information’.

Additionally, whilst it is true that other methods for quantifying relatedness of paradigm cells via distributional semantics exist, and these, in fact, have even been used to motivate defectivity in Russian nouns semantically (Chuang et al. 2022), one must question, however, such methods being classified as objective and also their suitability as models of human cognition. As is typical of many mathematical and statistical models, their results are often an artefact of the researchers’ choice of representation, and so they are not therefore objective. A case in point is the distributional semantic analysis of morphomic structures in Spanish by Herce and Tang (2023) and Herce (2022), whose methodology and experimental design are flawed.¹⁰

Even for models with a sound methodology, the ecological validity of these mathematical models remains to be established, in the form of external evidence that supports the claim that they model human cognition, their mechanisms are cognitively viable and their conclusions are linguistically relevant for speakers. The fact that a computer can detect semantic and syntactic generalisations in linguistic data, does not mean (a) that humans necessarily have detected the same patterns and (b) that such patterns are relevant to linguistic structure in that there exists a cause and effect relationship (see also Divjak (2015)). Indeed, the conclusions of the present analysis of defective verbs in Spanish and Portuguese are similar to those of Yang (2016). His conclusions were based on a mathematically defined principle, the *Tolerance Principle* (Yang 2016: §73), which claims that speakers have a certain level of tolerance to the amount of exceptions to a rule; once this level is passed it is

10 These authors compare the cells of the different Spanish morphomic structures with 1,000 random samples of the same size and conclude that the L-pattern and PYTA morphome, but not the N-pattern ‘tend to be characterised by significantly greater syntactic/semantic similarity’ (Herce and Tang 2023: 9). There are serious problems, however, with the study’s methodology. Specifically, the study does not compare like with like. Their methodology was to count how many cells the different morphemes comprised and then compare a distributional semantic analysis of the morphomic cells with 1,000 samples of the same number of randomly selected cells. Thus, whilst the L-pattern comprises six semantically coherent cells + one semantically unrelated cell, it was being compared with 1,000 samples of seven randomly selected cells across the paradigm. It is unsurprising, therefore, that the seven inflectional forms of the L-pattern were characterised by ‘significantly greater syntactic/semantic similarity’ than a randomly selected set of another seven paradigmatic cells since six of these forms correspond to exactly the same tense and mood combination. What is syntactically and semantically ‘incoherent’ about the L-pattern with respect to the rest of the paradigm, is that these six paradigm cells are grouped exclusively with the 1SG.PRS.IND. Their question should have been whether the grouping of all the persons of the PRS.SBJ + 1SG.PRS.IND is characterised by significantly greater syntactic/semantic similarity than PRS.SBJ + the other possible single-paradigm cell combinations in Spanish. That is the key question but crucially this was not the one analysed. Similar criticisms are valid for their analysis of and conclusions for the PYTA morphome.

hypothesized that, instead of searching for other abstract generalisations, speakers simply store forms.

The central idea behind the morphome is that speakers have detected systematic, system-internal patterns of form across stored words and these patterns are psychologically real for speakers, as attested by the way that they can condition historical change (Maiden 2018). I would like to emphasize the importance of patterns independently of the actual form which expresses these patterns and that the importance of such patterns has not been fully considered in abstractive theories of morphology. For example, Bybee makes specific reference to the morphomic vowel alternations of Spanish in Table 6 (these are of different types but the paradigmatic pattern is the same across lexemes) and whilst she does not reject the idea that lexical connections can be made on the basis of formal identity alone, she states: 'In many cases, lexical connections among paradigms with similar alternations are not justified' (Bybee 1985: 131). Bybee is not wrong to note that consistent form in a semantically unmotivated set of paradigmatic cells could be a mere coincidence, thus invalidating a justification of lexical connections. However, as referenced in the introduction, there is much synchronic and diachronic evidence for morphomes and the current analysis of defective verbs in Portuguese adds to this evidence.

Morphomes are essentially abstract patterns and schemas. Schemas play a central role in morphology in CL and are defined as 'output-based generalizations over forms occurring in utterances' (Nesset 2019). Such schemas, which are internally predictive, offer much potential to model morphomes. This is especially true of 'second-order schemas' (Booij 2016; Kapatsinski 2013; Nesset 2008), which represent the implicative relationships between different paradigmatic forms, often across different inflectional classes. Such second order schemas have been used to model the purely paradigmatic relationship that exists between the 3_{PL.PRS} and the active participle in Russian, which, cross-conjugationally, always share the same root and thematic vowel: the forms X[ut] and X[at] in the 3_{PL.PRS} alternate with X[uʃʲ:] and X[aʃʲ:] in the participle (e.g., *razgovarivajut* 'they converse' vs. *razgovarivajuščij* 'conversing'; *prosjať* 'they ask' vs. *prosjaščij* 'asking' (Nesset 2008: 60–61, 2019: §9)). In fact, the relationship between these two inflectional forms, has been related to the concept of the morphome by Nesset (2008: 60), who discusses it in the context of other formalizations of how to deal with such phenomena and states that "An evaluation of the merits of these proposals is beyond the scope of this study, but it is interesting to notice that cognitive linguistics does not need any additional machinery in order to accommodate basic-derived relationships". The question is, however, whether such schema can handle the central insight of the morphome regarding the psychological reality of phonologically heterogeneous allomorphy with the same paradigmatic distribution/pattern. I leave this to future research and hope that in this article I have

made a case for the validity of morphemes and the potential they have to be modelled in and inform discussions about models of morphology within CL.

8 Conclusions

In this article a statistical analysis of corpus data from European Portuguese has been used to establish that defective verbs do indeed exist in this variety of Portuguese and are extremely similar to those attested for Spanish. The morphomic account of defectivity for Spanish verbs has been extended to Portuguese and this account has also been expanded and developed further. The claim is that the verbs in Portuguese and Spanish which are defective are so due to them not having a stored form for the N&L patterns, and the morphological generalization of *-ir* verbs is that the forms of these patterns need must be lexically retrieved and usually not created via knowledge of inflectional forms outside these patterns. This latter idea is also independently supported by a number of experimental and theoretical studies on Spanish and Portuguese *-ir* verbs (Linares et al. 2006: 113; Rodríguez-Fornells et al. 2002; Veríssimo and Clahsen 2009; Yang 2016: 149) and the overall view of defectivity in these languages is compatible with the views of Gorman and Yang (2019), who see defectiveness as the result of when a number of rules are in competition and none of them can be defined as productive.¹¹

The hypothesis put forward here is that, in acquisition, the frequent lexemes of the *-ir* class provided a blue-print for the geometry of the entire class whereby the N&L-pattern forms had to be memorized. This was due to the fact that, in a number of frequent lexemes, the different types of allomorphy was unpredictable based on other inflectional forms of the verbs. This tendency for the paradigmatic patterns of a high-frequency group of irregular verbs to be relevant to regular verbs is also attested diachronically for other morphemes (O'Neill 2014) and has been modelled computationally (Pirrelli et al. 2007: 285) for the acquisition of morphology.

In CL little attention has been paid to morphomic structures on account of their lack of reference to meaning and the central role which meaning plays in CL. Here it has been argued that morphemes are meaningful in a broader sense of them being informative: the forms in a morpheme exhibit a high degree of inter-predictability that allows them to serve a mutual diagnostic function. That is, they may not be informative (=meaningful) in extra-morphological ways but they are crucial for the organisation of the morphology as a complex inter-predictive network which

¹¹ Although, note that I espouse a fully abstractive model of morphology in which rules are not thought to exist independently of stored forms.

enables speakers to extrapolate from their linguistic input and thus produce forms they have never heard before.

Interestingly, in the case of Spanish and Portuguese, however, the organisation of this network is such that it also encourages speakers **not** to produce forms that they have not heard before, resulting in the attested defective verbs in these languages.

Acknowledgements: The author would like to thank Neil Bermel and Jim Blevins for reading over a draft of this paper and their extremely useful comments and suggestions.

Data availability: The datasets generated and analysed during the current study are available in the TROLLing repository at <https://doi.org/10.18710/TVYCZL>.

References

- Albright, Adam. 2003. A quantitative study of Spanish paradigm gaps. In Gina Garding & Mimu Tsujimura (eds.), *West Coast Conference on Formal Linguistic 22 Proceedings*, 117–151. Somerville: Cascadilla.
- Albright, Adam. 2010. Lexical and morphological conditioning of paradigm gaps. In Sylvia Blaho & Curt Rice (eds.), *Modeling ungrammaticality in OT*, 33–63. London: Equinox Publishing.
- Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes*. Cambridge: MIT Press.
- Baerman, Matthew. 2004. Directionality and (un)natural classes in syncretism. *Language* 80(4). 807–827.
- Bermel, Neil & Dunstan Brown. 2023. Introduction: Feast and famine /overabundance and defectivity. [Special Edition]. *Word Structure* 16(2–3). 147–153.
- Blevins, Jim. 2006. Word-based morphology. *Journal of Linguistics* 42(3). 531–573.
- Blevins, Jim. 2010. The morpheme as a unit of predictive value. Paper presented at the 14th International Morphology Meeting, Budapest. Available at: <http://www.nytud.hu/imm14/abs/blevins.pdf>.
- Blevins, Jim. 2016. *Word and paradigm morphology*, 1st edn. Oxford: Oxford University Press.
- Booij, Geert. 2016. Construction morphology. In Andrew Hippisley & Gregory Stump (eds.), *The Cambridge handbook of morphology*, 424–448. Cambridge: Cambridge University Press.
- Bybee, Joan. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam; Philadelphia: J. Benjamins.
- Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.
- Bybee, Joan & Mary Alexandra Brewer. 1980. Explanation in morphophonemics: Changes in provençal and Spanish preterite forms. *Lingua* 52(3). 201–242.
- Cappellaro, Chiara, Nina Dumrukic, Isabella Fritz, Francesca Franzon & Martin Maiden. 2024. The cognitive reality of morphemes. Evidence from Italian. *Morphology* 34(1). 33–71.
- Chuang, Yu-Ying, Dunstan Brown, Harald Baayen & Roger Evans. 2022. Paradigm gaps are associated with weird “distributional semantics” properties. *The Mental Lexicon* 17(3). 395–421.
- Clahsen, Harald, Claudia Felser, Kathleen Neubauer, Mikako Sato & Renita Silva. 2010. Morphological structure in native and nonnative language processing. *Language Learning* 60(1). 21–43.
- Cunha, Celso & Luis Lindley Cintra. 1984. *Nova gramática do português contemporâneo*. Lisboa: Edições João Sá da Costa.

- Divjak, Dagmar. 2015. Four challenges for usage-based linguistics. In Jocelyne Daems, Eline Zenner Kris Heylen, Dirk Speelman & Hubert Cuyckens (eds.), *Change of paradigms: New paradoxes: Recontextualizing language and linguistics* (Applications of Cognitive Linguistics), 297–311. Berlin: De Gruyter.
- Dunn, Joseph. 1928. *A grammar of the Portuguese language*. District of Columbia: National Capital Press.
- Enger, Hans-Olaf. 2019. In defence of morphomic analyses. *Acta Linguistica Hafniensia* 51(1). 31–59.
- Fertig, David. 2013. *Analogy and morphological change*. Edinburgh: Edinburgh University Press.
- Goldberg, Adele, Devin Casenhiser & Nitya Sethuraman. 2004. Learning argument structure generalizations. *Cognitive Linguistics* 14(3). 289–316.
- Gorman, Kyle & Charles Yang. 2019. When nobody wins. In Franz Rainer, Francesco Gardani, Wolfgang Dressler & Hans Christian Luschutzky (eds.), *Competition in inflection and word-formation*, 169–193. Cambridge: Springer International Publishing.
- Harris, James. 1974. Evidence from Portuguese for the elsewhere condition in phonology. *Linguistic Inquiry* 4. 61–80.
- Hawkins, Jeff & Sandra Blakeslee. 2005. *On intelligence*. New York: Henry Holt and Company.
- Herce, Borja. 2022. Quantifying the importance of morphomic structure, semantic values, and frequency of use in Romance stem alternations. *Linguistics Vanguard* 81. 53–68.
- Herce, Borja. 2023. *The typological diversity of morphemes: A cross-linguistic study of unnatural morphology*. Oxford: Oxford University Press.
- Herce, Borja & Marc Tang. 2023. The meaning of morphemes: Stem alternation patterns in Spanish and their distributional semantic profile. *Linguistics Vanguard*. Available at: https://www.degruyter.com/document/doi/10.1515/lingvan-2023-0010/html?lang=en&srsId=AfmBOoqY3pnCYFGSS_zpwBi_SDTFpY8aKxCqchDjuOWSsH_Dj14rL5p.
- Hills, Elijah C. & Jeremiah D. M. Ford. 1925. *A Portuguese grammar*. Massachusetts: D. C. Heath and Company.
- Hinzelin, Marc-Olivier. 2012. Verb morphology gone astray: Syncretism patterns in Gallo-Romance. In Sascha Gaglia & Marc-Olivier Hinzelin (eds.), *Inflection and word formation in Romance languages*, 55–81. Amsterdam: John Benjamins.
- Kapatsinski, Vsevelod. 2013. Conspiring to mean: Experimental and computational evidence for a usage-based harmonic approach to morphophonology. *Language (Baltimore)* 89(1). 110–148.
- Koch, Christof. 2004. *The quest for consciousness: A neurobiological approach*. Englewood: Roberts and Company Publishers.
- Langacker, Ronald. 1987. *Foundations of cognitive grammar, Vol. 1: Theoretical prerequisites*. Stanford: Stanford University Press.
- Langacker, Ronald. 2019. Morphology in cognitive grammar. In Jenny Audring & Francesca Masini (eds.), *The Oxford handbook of morphological theory*, 1st edn., 346–364. Oxford: Oxford University Press.
- Linares, Rafael Enrique, Antoni Rodríguez-Fornells & Harald Clahsen. 2006. Stem allomorphy in the Spanish mental lexicon: Evidence from behavioral and ERP experiments. *Brain and Language* 97(1). 110–120.
- Luís, Ana & Ricardo Bermúdez-Otero. 2016. *The morpheme debate*. Oxford: Oxford University Press.
- Maiden, Martin. 2004a. Morphological autonomy and diachrony. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology*, 137–175. Dordrecht: Springer.
- Maiden, Martin. 2004b. Verb augments and meaninglessness in Romance morphology. *Studi di Grammatica Italiana* 22. 1–61.
- Maiden, Martin. 2004c. When lexemes become allomorphs. On the genesis of suppletion. *Folia Linguistica* 38(3–4). 227–256.

- Maiden, Martin. 2018. *The romance verb: Morphomic structure and diachrony*. Oxford: Oxford University Press.
- Maiden, Martin & Paul O'Neill. 2010. Morphomic defectiveness. In Matthew Baerman, Greville Corbett & Dunstan Brown (eds.), *Defective paradigms: Missing forms and what they tell us*, 103–124. London: OUP/British Academy.
- Neisser, Ulric. 2014. *Cognitive psychology: Classic edition*. New York: Psychology Press.
- Nesset, Tore. 2008. *Abstract phonology in a concrete model: Cognitive linguistics and the morphology-phonology interface*. Berlin: De Gruyter Mouton.
- Nesset, Tore. 2019. *Morphology in cognitive linguistics*. Oxford: Oxford University Press.
- Nevins, Andrew, Gean Damulakis & Maria Luísa Freitas. 2014. Phonological regularities among defective verbs. *Cadernos de Estudos Linguísticos* 561. 1–21.
- O'Neill, Paul. 2009. Los verbos defectivos en la lengua española: estudio sincrónico y diacrónico descriptivo basado en datos de corpus. *Boletín de la Real Academia Española* 89. 255–287.
- O'Neill, Paul. 2010. Una explicación teórica de la defectividad verbal en la Lengua Española. *Boletín de la Real Academia Española* 90. 265–289.
- O'Neill, Paul. 2014. The morphome in constructive and abstractive models of morphology. *Morphology* 24(1). 25–70.
- O'Neill, Paul. 2015. The origin and spread of Velar allomorphy in the Spanish verb: A morphomic approach. *Bulletin of Hispanic Studies* 92(5). 489–518.
- O'Neill, Paul. 2022. Morphologically 'autonomous' structures in the Romance languages. In Francesco Gardani & Michele Loporcaro (eds.), *The Oxford Encyclopedia of Romance Linguistics*. Oxford: Oxford University Press.
- O'Neill, Paul. 2024. Morphologization and the boundary between morphology and phonology in the Romance languages. In Francesco Gardani & Michele Loporcaro (eds.), *The Oxford Encyclopedia of Romance Linguistics*. Oxford: Oxford University Press.
- Perini, Mário A. 2002. *Modern Portuguese: A reference grammar*. New Haven: Yale University Press.
- Pirrelli, Victor, Ivan Herreros & Basilio Calderone. 2007. Learning inflection: The importance of starting big. *Lingue e Linguaggio* 6. 175–199.
- Postma, Gertjan. 2013. Metaphonic blocking in Portuguese as a Linearization Deadlock. Paper presented at the Workshop on Metaphony, Meertens Institute.
- R Core Team. 2021. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rodriguez-Fornells, Antoni, Thomas F. Münte & Harald Clahsen. 2002. Morphological priming in Spanish verb forms: An ERP repetition priming study. *Journal of Cognitive Neuroscience* 14(3). 443–454.
- Schäfer, Roland & Elizabeth Pankratz. 2018. The plural interpretability of German linking elements. *Morphology* 28(4). 325–358.
- Schmid, Hans-Jörg. 2016. Why Cognitive Linguistics must embrace the social and pragmatic dimensions of language and how it could do so more seriously. *Cognitive Linguistics* 27(4). 543–557.
- Shepard, Roger N. 1987. Toward a universal law of generalization for psychological science, science. *American Association for the Advancement of Science* 237. 1317–1323.
- Sims, Andrea. 2019. *Inflectional defectiveness*. Cambridge: Cambridge University Press.
- Sims-Williams, Helen. 2022. Token frequency as a determinant of morphological change. *Journal of Linguistics* 58(3). 571–607.
- Strack, Fritz & Thomas Mussweiler. 1997. Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology* 73. 437–446.
- Vázquez Cuesta, Pilar & María Albertina Mendes da Luz. 1971. *Gramática portuguesa*. Madrid: Editorial Gredos.

- Veríssimo, João & Harald Clahsen. 2009. Morphological priming by itself: A study of Portuguese conjugations. *Cognition* 112(1). 187–194.
- Wheeler, Max. 2011. The evolution of a morpheme in Catalan verb inflection. In Martin Maiden, John Charles Smith, Maria Goldbach & Marc-Olivier Hinzelin (eds.), *Morphological autonomy: Perspectives from romance inflectional morphology*, 182–209. Oxford: Oxford University Press.
- Wheeler, Max. 2015. L'evolució dels paradigmes del temps present de la conjugació III en castellà i català: alternances vocàliques metafòniques i la morfologia autònoma. In Àlex Martín Escribà, Adolf Piquer Vidal & Fernando Sánchez Miret (eds.), *Actes del Setzè Col·loqui Internacional de Llengua i Literatura Catalanes. Universitat de Salamanca, 1–6 juliol de 2012*, vol. 2, 93–145. Barcelona: Publicacions de l'Abadia de Montserrat.
- Yang, Charles. 2016. *The price of linguistic productivity: How children learn to break the rules of language*. Cambridge, Massachusetts: MIT Press.