

Spanish pseudoplurals: phonological cues in the acquisition of a syntax-morphology mismatch

Ricardo Bermúdez-Otero
University of Manchester

Forthcoming in:

Baerman, Matthew, Greville Corbett, Dunstan Brown & Andrew Hippisley (eds). *Deponency and morphological mismatches* (Proceedings of the British Academy). Oxford: Oxford University Press.

Abstract. Spanish has two types of nouns with homophonous singular and plural forms ending in /s/. In pseudoplural nouns like *Carl-o-s*, the number contrast is neutralized in the morphology: the syntactically singular form contains the plural marker /-s/. In athematic nouns like *virus*, the number contrast is neutralized in the phonology through a process of degemination. The distinction between pseudoplural and athematic nouns manifests itself only under stem-based evaluative suffixation: e.g. augmentative *Carl-ot-e* vs *virus-ot-e*. However, children are almost never exposed to these crucial forms, which have vanishingly low token frequencies; instead, learners of Spanish acquire the pseudoplural/athematic distinction on the basis of language-particular parsing preferences motivated by morphological and phonological properties of the Spanish lexicon. The emergence of the Spanish pseudoplural nouns shows that, during language acquisition, the phonological properties of words can trigger syntax-morphology mismatches. This finding has significant implications for the syntax-morphology interface.

Keywords: syntax-morphology mismatch, deponency, defectiveness, ineffability, pseudoplural, athematic, root, stem, stem formative, theme vowel, infix, learnability, underdetermination, phonotactics, variation, corpus, Google.

1. Introduction¹

In its broadest sense, the term DEPONENCY denotes a mismatch between the morphological shape and the syntactic behaviour of a word (Baerman this volume). Latin deponent verbs provide the eponymous and canonical example of this phenomenon: in (1), for instance, the verb *sequor* has a passive inflectional marker but is syntactically active —and indeed transitive.²

¹ This paper develops and expands aspects of the analysis of Spanish nominal classes presented in Bermúdez-Otero (forthcoming a), and has greatly benefited from James Harris's and Iggy Roca's detailed criticism of that article; however, neither of them endorses the proposals made here. Early versions of this work were presented at the conference on 'Deponency and morphological mismatches' (British Academy, London, 17 January 2006), at the Linguistics and English Language seminar of the University of Manchester (7 February 2006), and at a meeting of the Cambridge University Linguistic Society (27 April 2006); I am very grateful to the audiences on all three occasions for their stimulating questions.

² Glosses follow the Leipzig glossing rules (Comrie, Haspelmath & Bickel 2004). The following abbreviations are used: 1 (first person), 2 (second person), ACC (accusative), AUG (augmentative), DIM (diminutive), F (feminine), IND (indicative), IPFV (imperfective), M (masculine), N (neuter), PASS (passive), PL (plural), PRES (present), PRET (preterite), SF (stem formative), SG (singular). Other abbreviations include UR for 'underlying representation' and SR for 'surface representation'.

(3) *Pseudoplural nouns*

	sg.	pl.
UR	$[[\text{word } [\text{stem } \text{karl-o}]s]]$	$[[\text{word } [\text{stem } \text{karl-o}]s]]$
SR	[kárlos]	[kárlos]

The main challenge for Spanish learners is to acquire the distinction between athematic nouns like *virus* and pseudoplural nouns like *Carl-o-s*, for there is absolutely no phonological or syntactic difference between the two classes. The contrast emerges only under stem-based derivation, i.e. in the presence of derivational suffixes that subcategorize for stems (as opposed to roots or words); in Spanish these include most evaluative suffixes, like diminutive *-it-o* (f. *-it-a*) or augmentative *-ot-e* (f. *-ot-a*). In this environment the /s/ of *virus* occurs inside the derivational suffix because it is part of the stem (4a), whereas the /s/ of *Carl-o-s* disappears because, being an inflectional (plural) marker, it lies outside the stem (4b):

(4) *Augmentatives*

	a. with an athematic base	b. with a pseudoplural base
UR	$[[\text{word } [\text{stem } [\text{stem } \text{birus}]ot-\{e,\emptyset\}]]]$	$[[\text{word } [\text{stem } [\text{stem } \text{karl-o}]ot-\{e,\emptyset\}]]]$
SR	[birusóte]	[karlóte]
	‘virus’ aug.sg.	‘Charles’ aug.sg.

However, most evaluative derivatives of athematic or pseudoplural bases have vanishingly low token frequencies and are therefore extremely unlikely to be robustly represented in the learner’s primary linguistic data. This creates an obvious learnability puzzle: in the absence of any direct syntactic, morphological, or phonological cues, Spanish speakers producing the forms in (4) have managed to acquire a contrast between pseudoplural nouns and nonpseudoplural nouns with athematic stems in /s/. The existence of a learnability deficit in this area of Spanish grammar is independently confirmed by the observation that not all idiolects display the distinction shown in (4): e.g. although the most frequent diminutive form of *virus* is indeed *virus-it-o*, several other variants are attested (see §3.2 below for details). Nonetheless, the fact remains that a substantial proportion of Spanish learners do acquire a contrast between pseudoplural nouns and athematic nouns in /s/ even though the opposition is underdetermined in their trigger experience.

How is this possible? This paper suggests that learners of Spanish use a variety of indirect cues to estimate the probability that the final /s/ in a singular noun with a homophonous plural is part of the stem or is a pseudoplural affix. Some of these indirect cues are morphological. For example, the final /s/ is unlikely to be parsed as an inflectional affix if it is immediately preceded by any segment other than /o/, /a/, or /e/, for in that case the stem could not be a member of any of the three major classes to which the vast majority of Spanish nouns belong: viz. *o*-stems, *a*-stems, and *e*-stems (see note 4). On morphological grounds, therefore, learners prefer to parse singular *virus* as a nonpseudoplural form with an athematic stem in /s/, i.e. $[[\text{word } [\text{stem } \text{birus}]]]$, rather than as a pseudoplural form with an athematic stem in /u/, i.e. $[[\text{word } [\text{stem } \text{biru}]s]]$, because the syntax-morphology mismatch incurred by the pseudoplural parse affords no morphological advantages: in either case, the stem remains a member of the marginal athematic class. Crucially, however, learners also make extensive use of phonological cues. Among singular nouns with a closed ultima, for example, antepenultimate stress is supermarked (Roca 2005), occurring by and large only in partially assimilated loans (‘xenonyms’: see note 4). This leads learners to parse proparoxytonic nouns

like *Sócrat-e-s* ‘Socrates’ sg./pl. as pseudoplural rather than athematic (cf. dim. *Sócrat-it-o*, not **Socrates-it-o*), since otherwise they would be metrically deviant (Bermúdez-Otero forthcoming a: §2.3.2).

The way in which Spanish learners deal with nouns like *Sócrat-e-s* indicates that, in language acquisition, the phonological properties of a word can cause it to receive a morphological parse that fails to match its syntactic features. This result has important implications for our understanding of deponency and, more generally, for the theory of the syntax-morphology interface. As we shall see in §5 below, the syntax-morphology mismatches found in the languages of the world turn out to have a rather diverse ætiology. This raises the question whether the formal mechanisms used to model instances of semantically or pragmatically induced deponency will cast any light on syntax-morphology mismatches driven by phonological factors. More generally, the finding that phonological constraints on morphological categories can dramatically disrupt syntax-morphology correspondence lends new urgency to the task of ascertaining the universal limits to possible syntax-morphology mismatches.

The argument will proceed as follows. In §2 I present the evidence for distinguishing between pseudoplural nouns and nonpseudoplural nouns with athematic stems in /s/; a key step in the argument is the demonstration that Spanish evaluative suffixes such as diminutive *-it-o* and augmentative *-ot-e* attach to stems, rather than to roots or to words, and can therefore be used to establish whether a final /s/ belongs to the stem or to an inflectional suffix. Regrettably, the distinction between athematic and pseudoplural nouns has hitherto not been properly understood. In consequence, §3 is devoted to clearing up certain misconceptions concerning athematic stems in /s/: notably, I argue against the view that diminutive forms like *virus-it-o* are ungrammatical even for those native speakers who produce them spontaneously. Having settled the relevant descriptive issues, in §4 I chart the acquisition path that leads Spanish learners to the contrast between pseudoplural nouns and nondeponent nouns with athematic stems in /s/. The paper concludes in §5 with a survey of the general implications of our findings.

2. Pseudoplural nouns vs nonpseudoplural nouns with athematic stems in /s/

Nouns like *Carl-o-s* and *virus* share the property that their singular form is superficially homophonous with the plural. This is shown by their behaviour under agreement. In (5a), for example, *Carl-o-s* bears the plural marker /-s/ according to our analysis, but it must nonetheless be specified as singular in the syntax since the copula *es* and the predicative adjective *buen-o*, which are required to agree with it in number, are both unambiguously singular.⁶

- (5) a. Carl-o-s es buen-o
 Charles-SF-PL be.3SG.PRES.IND good-SF[SG]
 ‘Charles is good’
- b. Aquí hay much-o-s Carl-o-s
 here there be.PRES.IND many-SF-PL Charles-SF-PL
 ‘Here there are many Charleses’

⁶ For our purposes it does not matter whether number agreement takes place in the ‘narrow syntax’ (cf. Bobaljik 2006). Similarly, it is not essential that we identify the precise features involved; see Corbett (2000) for a typological survey of the possible values of the category number, and Harbour (2006) for a proposed universal inventory of number features.

- c. Est-e virus es muy peligros-o
 This-SF[SG] virus be.3SG.PRES.IND very dangerous-SF[SG]
 ‘This virus is very dangerous’
- d. Est-o-s virus son muy peligros-o-s
 This-SF-PL virus be.3PL.PRES.IND very dangerous-SF-PL
 ‘These viruses are very dangerous’

The main descriptive challenge posed by these nouns is to ascertain the morphological affiliation of the final /s/ in the singular form. Logically, two possibilities suggest themselves:

- (6) a. The final /s/ of the singular may be part of the stem. In that case, the homophony between the singular and plural forms can be explained by positing a phonological process of degemination applying to the underlying /ss/ sequence in the plural: see (2) above. This process is well motivated as Spanish does not tolerate geminates.
- b. Alternatively, the final /s/ of the singular form may lie outside the stem. In that case, this final /s/ will be phonologically identical with the plural marker /-s/ and will occupy the same morphological slot. We can therefore assume that it *is* indeed the plural marker, and that the singular form in question is deponent. This automatically explains the homophony between singular and plural, since both consist of the noun stem followed by the plural marker /-s/: see (3) above.

I have formulated the choice in (6) in terms of whether or not the final /s/ of the singular form is affiliated to the *stem*. Implicit in this formulation is the traditional trichotomy of root, stem, and word (see e.g. Benveniste 1935). In brief, a *ROOT* is defined as an uninflectable lexical item: roots cannot be inflected without first undergoing root-to-stem conversion, which may be either overt (involving some operation on the phonological form of the root) or covert. In contrast, a *STEM* is a unit that can provide the base for an inflectional operation, i.e. it is an inflectable lexical item. Finally, *WORDS* are fully inflected, i.e. syntactically free. These traditional definitions are fairly robust and can easily be transposed into a variety of theoretical frameworks. For example, the constituents identified in a traditional parse of the plural noun *caud-al-e-s* ‘capital’ (7a) can be straightforwardly mapped onto nodes in a Distributed Morphology representation (7b); for the latter, see Oltra-Massuet (1999, 2000), Oltra-Massuet & Arregi (2005), and Harris (1999).⁷

⁷ The symbols in (7b) are to be interpreted as follows: $\sqrt{}$ is a categoryless root (Marantz 1997); *n* is a category-giving functional head creating a noun; *Th* is the theme vowel position adjoined to *n* (Oltra-Massuet 1999); *#* is the functional head containing number features. I should point out that, although the traditional parse in (7a) and the Distributed Morphology representation in (7b) are equivalent for our purposes, the theories behind them are of course not identical: notably, the traditional parse does not carry the assumption that words are built by head adjunction in the syntax.

derivational suffix, it will be an inflectional marker. Finding the criterial suffixes for our test is no trivial matter, for a suffixation operation can in principle belong to any of the following four types: root-to-stem, stem-to-stem, stem-to-word, and word-to-word (see e.g. Kiparsky 2003). Fortunately, Spanish evaluative suffixes like diminutive *-it-o* (f. *-it-a*), augmentative *-ot-e* (f. *-ot-a*), and augmentative *-ón-Ø* (f. *-on-a*) can easily be shown to attach to stems rather than to roots or words.⁹

First, the evaluative suffixes do not attach to uninflectable bases, and so cannot be root-based. Consider for example the *e*-stem noun *caud-al-Ø* ‘capital’, which we first encountered in its plural form in (7) above. As we saw there, the stem of this noun is formed through the addition of the derivational suffix */-al-{e,Ø}/* to the base */kawd-/*, which can only be a root since it is not inflectable: cf. **caud-o(-s)*, **caud-a(-s)*, **caud-e(-s)*.¹⁰ In consequence, the evaluative suffixes are able to attach to the stem *[[kawd]al-{e,Ø}]*, but not to the root *[[kawd]]*.

- (9) a. *[[word [[stem [[stem [[root kawd]al-{e,Ø}]]it-o]]] → [kawðalító]* ‘capital’ dim.sg.
 b. **[[word [[stem [[root kawd]]it-o]]]*

The finding that evaluative suffixation cannot be root-based is confirmed by a slightly more intricate phonological argument (Bermúdez-Otero forthcoming a: §1.2.1). Spanish has a phonological process whereby certain lexically specified vowels are realized as diphthongs under primary stress and as monophthongs elsewhere: cf. (10a) and (10b) below (for surveys of the literature on this phenomenon, see Cole 1995: §6.2 and Eddington 2004: §6.1). However, this process misapplies in the presence of the evaluative suffixes: e.g. in (10c) primary stress falls on the diminutive suffix, but the vowel of the base nonetheless surfaces as diphthongal. The overapplication of diphthongization requires the cyclic derivation in (10d): the stem undergoes diphthongization in the first cycle, and then stress moves rightwards to the diminutive suffix in the second cycle.¹¹ However, if the diminutive were root-based, then it could not possibly have the domain structure shown in (10d): roots are commonly assumed not to constitute cyclic domains (Kiparsky 1982a: 144-5, 1982b: 32-3; Inkelas 1990: §3.5.5; Bermúdez-Otero forthcoming b); therefore, if the diminutive suffix attached itself directly to a root, then it would already be present in the first cycle, as shown in (10e).

- (10) a. *[pwért-a]* ‘door’ sg.
 b. *[port-ér-o]* ‘doorman’ sg.
 c. *[pwert-eθít-a]* ‘door’ dim.sg.

⁹ In this respect the Spanish evaluative suffixes make a remarkable contrast with one of the Portuguese diminutive suffixes, namely */-ziɲ-u/* (f. */-ziɲ-e/*), which is demonstrably word-based (Rainer 1996: §3, Bachrach and Wagner 2006, Bermúdez-Otero forthcoming b: ch. 2).

¹⁰ There is a learned noun *caud-a* (< Latin *cauda* ‘tail’), meaning ‘train of an episcopal robe of state’, which is synchronically and etymologically unrelated to *caud-al-Ø* (< Latin *capitālis*).

¹¹ A full analysis of *puert-ecit-a* would have to account for the selection of the augmented allomorph of the diminutive suffix. The details are not relevant here.

d. *correct derivation*

morphological structure	$[[_{\text{stem}} [[_{\text{stem}} p\{we/o\}rt-a]e\theta it-a]]$
domain structure	$[[p\{we/o\}rt-a]e\theta it-a]$
first cycle	pwérta
second cycle	pwerteθíta

e. *incorrect derivation*

morphological structure	$[[_{\text{stem}} [_{\text{root}} p\{we/o\}rt]e\theta it-a]]$
domain structure	$[p\{we/o\}rt-e\theta it-a]$
first cycle	*porteθíta

The evaluative suffixes cannot be word-based either. This is shown, first, by the fact that they cause stem-final vowel deletion, as shown in (8) above. In this respect, diminutive and augmentative forms contrast with denumeral partitives with the suffix *-av-o* (f. *-av-a*), which attaches to fully inflected bases, i.e. to words: (11a) shows *-av-o* occurring outside the plural marker *-s*, whilst (11b) shows that *-av-o* does not trigger stem-final vowel deletion (Pensado Ruiz 1999: 4461-4462; Bermúdez-Otero forthcoming a: §2.2).

- (11) a. $[[_{\text{word}} [[_{\text{word}} dos-\theta jent-o-s]ab-o]] \rightarrow [dos\theta jentosá\beta o]$ ‘200th’ (partitive)
 cf. $[dos\theta jéntos]$ ‘200’ (cardinal)

- b. $[[_{\text{word}} [[_{\text{word}} on\theta-e]ab-o]] \rightarrow [on\theta eá\beta o]$ ‘11th’ (partitive)
 cf. $[ón\theta e]$ ‘11’ (cardinal)

Relatedly, evaluative derivatives constitute single prosodic words. This opposes them to deadjectival adverbs formed with the word-based suffix *-ment-e*. In the latter, the base and the suffix surface as separate prosodic words. This complex prosodic structure permits stress clashes (12a) and noninitial lapses (12b), which are never found in evaluative forms.

- (12) a. $[_{\omega} ko.mún]$ ‘common’ m./f.
 $[_{\omega'} [_{\omega} ko.mùn][_{\omega} mén.te]]$ ‘commonly’
 cf. $[_{\omega} ka.mjón]$ ‘lorry’
 $[_{\omega} kà.mjón.θí.to]$ ‘lorry’ dim. cf. $*[_{\omega} ka.mjòn.θí.to]$
- b. $[_{\omega} pe.núl.ti.ma]$ ‘penultimate’ f.
 $[_{\omega'} [_{\omega} pe.nùl.ti.ma][_{\omega} mén.te]]$ ‘penultimately’

If the evaluative suffixes are neither root-based, as shown in (9) and (10), nor word-based, as shown in (11) and (12), then they can only be stem-based —*quod erat demonstrandum*. As suggested above, therefore, we can use the evaluative suffixes to test the morphological affiliation of the final /s/ in singular nouns with homophonous plural forms: if the final /s/ appears before the suffix in evaluative derivatives, the noun is athematic and the /s/ is part of its stem; if the final /s/ does not appear before the suffix in evaluative derivatives, the noun is pseudoplural and the /s/ is a plural marker, lying outside the stem.

The application of this morphological test confirms that nouns of both classes coexist in the mental lexicon of many Spanish speakers. In (13), for example, I present a sample of evaluative forms from my own Northern Peninsular idiolect. All of these forms are found in both Peninsular and Latin American dialects. Their occurrence in spontaneous language use

is documented in Bermúdez-Otero (forthcoming a), abbreviated as ‘RB-O’ in (13). In certain cases there are competing variants, but the forms given in (13) are clearly the most frequent: see §3.2 for further details and discussion.

(13)

noun type	sg.	pl.	evaluative derivative		gloss
athematic	bírus	bírus	birus-ít-o	dim. (§3.2 below)	‘virus’
	bríndis	bríndis	brindis-ít-o	dim. (RB-O: note 29)	‘toast’
pseudoplural	kárl-o-s	kárl-o-s	karl-ót-e	aug. (RB-O: note 38)	‘Charles’
	sókrat-e-s	sókrat-e-s	sokrat-ít-o	dim. (RB-O: note 30)	‘Socrates’
	krísi-s	krísi-s	kris-eθít-a	dim. (note 17 below)	‘crisis’

3. Apparent problems

3.1. Obstacles to knowledge

As we saw in §2, the literature on the morphological structure of Spanish nouns contains analyses that roughly match the categories ‘nondeponent athematic stem in /s/’ and ‘pseudoplural’ as defined above. However, previous scholarship has failed to recognize that both classes are needed and actually coexist in the mental lexicon of Spanish speakers. Generative linguists, in particular, have consistently misanalysed athematic nouns such as *virus* and *brindis*, parsing the final /s/ of the singular form as being outside the stem (e.g. Harris 1983, 1985, 1991, 1992, 1999). As a result, this work incorrectly conflates the athematic and pseudoplural classes. For recent instances of this error, see Bonet (forthcoming), Lloret & Mascaró (2005), Oltra-Massuet & Arregi (2005), Roca & Felíu (2003), and Roca (2005).

The chief obstacle in the way to a proper understanding of the structure of nouns like *virus* and *Carl-o-s* has lain in inaccurate reports of their behaviour under evaluative suffixation. In turn, these flaws in the empirical record have largely followed from the limitations of data-gathering methods such as introspection, form elicitation, and judgement elicitation. Ordinarily, these techniques can be relied upon to produce valid information about wide areas of grammar, but they prove inadequate for probing items of linguistic knowledge such as the contrast between athematic and pseudoplural nouns in Spanish, for here learners are forced to rely opportunistically on indirect, partially conflicting language-particular cues to overcome an underdetermination effect in their trigger experience (§4); tools such as introspection and elicitation are too blunt to access the resulting knowledge. It is only recently that it has become possible to overcome these obstacles by automatically searching the massive volume of data available on the Internet (cf. Rainer 2003). This new source of information proves crucial to evaluating the grammatical status of forms such as [birus-ít-o] and [brindis-ít-o], which occur with vastly greater frequency than competing variants. Smaller, purpose-built corpora such as the *Corpus de referencia del español actual* (CREA) fail to capture the relevant quantitative patterns: see (25) below and cf. also Eddington (2002: 409-410).

In §3.2 below I discuss the implications of the Internet data. Subsequent sections (§3.3 to §3.5) deal with other descriptive issues that bear on our understanding of nouns with athematic stems in /s/.

3.2. The grammaticality of [birus-ít-o] and [brindis-ít-o]

Evaluative derivatives such as [birus-ít-o] and [brindis-ít-o] are essential to my argument because they show that nouns like *virus* and *brindis* cannot be analysed as pseudoplural, despite frequent claims to the contrary (see §3.1). In §2 I noted that these forms are in competition with other variants, but the Internet data show that this variation is quantitatively structured: crucially, the frequency of [birus-ít-o] and [brindis-ít-o] sharply outstrips that of their competitors. In this section I shall take the diminutive of *virus* as a test case; for information about *brindis*, see Bermúdez-Otero (forthcoming a: note 29).

The diminutive of *virus* occurs in the five variants shown in (14). Frequencies are estimated on the basis of the number of Google hits on 23 July 2006; in each case this corresponds to the number of webpages identified by Google as being in Spanish and containing one or more tokens of the relevant form. These data are relatively noisy: webpages are sometimes duplicated, and Google occasionally misidentifies the language of a webpage (*a fortiori*, it cannot discriminate between native and nonnative writers). Nonetheless, frequency asymmetries are so large that reliable conclusions can be drawn despite the noise. Where the plural differs from the singular (e.g. sg. *virus-it-o* vs pl. *virus-it-o-s*, sg. *vir-it-o* vs pl. *vir-it-o-s*), the figures refer to the singular only. In general it is not possible to ascertain the geographical provenance of the data, but I have not come across any evidence of bias in the dialectal distribution of variants.

(14) Variation in the diminutive of *virus*

a. *virus-it-o*

1,060 Google hits. The data are relatively clean, with few instances in which the form is not the diminutive of *virus*; cf (14c). Speakers producing this form parse the base noun as having a nondeponent athematic stem ending in /s/.

Me imagino... pero te cuento que yo en MAC en los más de 9 años que llevo en la informática y usando MAC solo agarré un **virusito** inofensivo que vino metido en un CD de una revista que compré mucho que se llamaba MacFormat...

<http://foros.softonic.com/showthread.phtml?t=13939>

b. *vir<it>us*

96 Google hits. Relatively clean data. This form is derived by means of a comparatively rare type of diminutive infixation; the diminutive infix usually subcategorizes for bases with unstressed /o/ or /a/ in the nucleus of the final syllable (see §3.3 below). It is not possible to determine whether speakers producing this form treat the base as pseudoplural or just as athematic.

Bueno a mi parecer yo conzco muchos q por hacer un **viritus** (lease virus pequeño) o abrir un hotmail ya se creen hackers [...]

<http://somos.lamerma.com.ve/mermachat/index.htm?TID=67&PN=1&TPN=10> (cached by Google on 9 August 2005)

c. *vir-it-o*

Approximately 40 genuine Google hits out of a total of 454. Very noisy data: most hits are morphologically unrelated to *virus*, but correspond to the personal name *Virito*. Speakers producing this diminutive of *virus* parse the singular form of the base noun as pseudoplural, i.e. [word [stem biru]s]. No

Google hits for a similar variant with the augmented allomorph of the diminutive suffix, i.e. **vir-ecit-o*; cf. (14d).

A veces son problemas de software, que se corrompen por apagones indebidos, a veces problemas de memoria y a veces algún **virito** pescado en internet.

http://www.yoreparo.com/foros/reparacion_de_computadoras/23017.html

d. *virus-cit-o*

13 Google hits. Clean data. Speakers producing this form parse the base noun as having a nondeponent athematic stem ending in /s/. This variant differs from (14a) only in the choice of the allomorph of the diminutive suffix.

Quizas Tomas haya estornudado sobre la pantalla sin tapar la boca, asi infectandole con un **viruscito**.

http://ndialectica.blogspot.com/2003_02_01_ndialectica_archive.html

e. \emptyset (defectiveness)

Elicitation tasks and linguistic introspection sometimes yield the report that *virus* does not have a diminutive form: e.g. “in many of these cases derivation is normally avoided” (Méndez-Dosuna & Pensado 1990: 101). This is likely to be true for at least some speakers, but the inherent unreliability of elicitation tasks makes it difficult to estimate the real proportion: see §4 for indirect evidence that the extent of defectiveness is in fact rather limited.

The quantitative evidence in (14) is conclusive. *Virus-it-o* occurs at least 11 times more frequently than its most frequent competitor (viz. *vir<it>us*).¹² Diminutive forms unambiguously implying a nondeponent parse of *virus* (viz. *virus-it-o* and *virus-cit-o*) are approximately 27 times more frequent than *vir-it-o*, the only diminutive form unambiguously implying a pseudoplural parse. Among those speakers for whom *virus* does have a diminutive form, an athematic parse with a stem in /s/ is somewhere between 8 and 27 times more likely than a pseudoplural parse.¹³ This finding roundly refutes the claim that, for all speakers, the final /s/ of sg. *virus* lies outside the stem; cf. §3.1 above.

The Internet data also enable us to draw firm conclusions regarding the grammatical status of evaluative derivatives such as *virus-it-o* and *brindis-it-o*. Iggy Roca (personal communication, 29 March 2006) asserts that, for most speakers, *virus* does not have a diminutive form. If so, one could argue that \emptyset is the core variant in (14), and that *virus-it-o* is at best marginal, along with *vir<it>us*, *vir-it-o*, and *virus-cit-o*. Can this claim be reconciled with the evidence of over 1,000 tokens of sg. *virus-it-o* in spontaneous language use across the Spanish-speaking world? The obvious way to advance Roca’s argument would be to appeal to the distinction between competence and performance (Chomsky 1965). One could thus suggest that the form *virus-it-o* is in fact an artificial creation, constructed by means of extragrammatical problem-solving strategies such as metalinguistic reasoning or *ad hoc* pattern-matching, and perhaps intended for special rhetorical effect. Roca certainly countenances this possibility, for in a different connection he speaks of ‘on-line performance

¹² The actual figure should be higher than 11 because (14) gives the number of Google hits for *vir<it>us* in both the singular and the plural, whereas the hit count for *virus-it-o* excludes 288 hits for plural *virus-it-o-s*.

¹³ The lower figure, i.e. 8, corresponds to a scenario in which every speaker producing *vir<it>us* parses *virus* as pseudoplural. The higher figure, i.e. 27, corresponds to a scenario in which every speaker producing *vir<it>us* parses *virus* as having an athematic stem in /s/.

trial outputs’ as opposed to ‘finished I-language products’ (Roca 2005: note 32). Relatedly, Méndez-Dosuna & Pensado (1990: 100-101) describe some of the diminutive forms produced by their informants in an elicitation task as ‘aberrant’. If we take this line, therefore, *virus-it-o* would be ungrammatical even for those speakers who produce it spontaneously.

However, there are powerful arguments against treating *virus-it-o* either as an artificial creation or as a speech error. First, the evidence in (14) reflect spontaneous language use; it is therefore free of the suspicion that attaches to data obtained by means of forced elicitation. In addition, most of the tokens originate in message boards and weblogs, which are characterized by their informality and low degree of linguistic monitoring; the evaluative derivatives that occur in such texts are therefore unlike the lexical innovations of poets or scientists. Moreover, the diminutive data match the evidence of other evaluative suffixes: thus, alongside *virus-it-o* we encounter the augmentatives *virus-ot-e* and *virus-ón-Ø* (Bermúdez-Otero forthcoming a: note 36).

Secondly, the figures in (14) reveal that *virus-it-o* sits at the top of an abrupt acceptability cline. It is improbable that the judgements of informants relying on extragrammatical problem-solving strategies would match this gradation. For example, such informants might favour *Carl<it>-o-s* ‘Charles’ dim.sg./pl. as an analogical model, since it occurs extremely frequently and is not subject to variation.¹⁴ However, this strategy would favour infixal *vir<it>us*, which, as we saw above, is more than 11 times less frequent than *virus-it-o*. In this light, the speakers whose usage is recorded on the Internet appear more likely to be guided in the main by their tacit knowledge of the morphological structure of *virus* than by metalinguistic reasoning or *ad hoc* pattern-matching.

Thirdly, the variability that affects the diminutive of *virus* arises from an underdetermination effect in language acquisition: as we have seen, the information contained in the inflectional paradigm of *virus* does not uniquely determine the morphological affiliation of the final /s/. This situation can resolve itself in either of two ways (see §4 below). On the one hand, the learner may fail to commit herself to a particular parse below the word node; in that case, she will not be able to apply stem-based suffixation to *virus*, which will consequently lack a (suffixal) diminutive. On the other hand, the learner may settle on a parse for the final /s/ by relying on the indirect morphological and phonological cues to which I referred in §1; in that case, the learner will acquire a grammar that generates one of the diminutive forms listed in (14a) to (14d). Given our limited knowledge of language acquisition, neither scenario can be rejected *a priori*, and the empirical evidence indicates that both are instantiated in practice. Therefore, if one wished to dismiss *virus-it-o* as the outcome of extragrammatical phenomena, one would have to accord an arbitrary and unwarranted privilege to the first scenario over the second. In fact, §4 below provides some quantitative data suggesting that defectiveness is a rare outcome.

I conclude that the evidence reported in (14) should be taken as face value. For native speakers who produce *virus-it-o* spontaneously, this form is grammatical. Spanish speakers are between 8 and 27 times less likely to parse *virus* as a pseudoplural noun than as having an athematic stem ending in /s/. Therefore, previous scholarship was in error when it failed to recognize the existence of athematic stems in /s/ and their opposition to pseudoplurals.

3.3. Diminutive infixation

It is easy to use augmentative suffixes such as *-ot-e* (f. *-o-ta*) and *-ón-Ø* (f. *-on-a*) for isolating noun stems, as shown in §2. In contrast, the diminutive presents additional

¹⁴ On the structure of *Carl<it>-o-s*, see §3.3 below. On 23 July 2006 *Carlitos* scored over a million hits in Google.

complications, which however, when properly understood, turn to the advantage of our analysis. The initial difficulty arises over the fact that, in certain cases, diminutive formation is unambiguously infixal (Jaeggli 1980; Lázaro Mora 1999: §71.6; Méndez-Dosuna & Pensado 1990).

(15) a. <i>sg.</i>	<i>pl.</i>
aθúkar-Ø sugar-SF[SG]	aθúkar-e-s sugar-SF-PL
b. <i>dim.sg. (suffixal)</i>	<i>dim.pl. (suffixal)</i>
aθukar-(θ)ít-o sugar-DIM-SF[SG]	aθukar-(θ)ít-o-s sugar-DIM-SF-PL
c. <i>dim.sg. (infixal)</i>	<i>dim.pl. (infixal)</i>
aθuk<ít>ar-Ø sugar<DIM>-SF[SG]	aθuk<ít>ar-e-s sugar<DIM>-SF-PL

In the noun [aθúkar-Ø] ‘sugar’, for example, the status of the string /ar/ is not in doubt: it cannot be some sort of exotic stem formative but must simply be part of the root, for the plural form [aθúkar-e-s] shows the stem to be /aθukar-{e,Ø}-/, with the formative /-{e,Ø}/ alternating predictably between [-Ø] in the singular and [-e-] in the plural (see §3.4 below). The diminutive variant [aθukar-(θ)ít-o] follows the expected pattern for stem-based suffixation: the stem formative of the base disappears before the diminutive suffix, and the latter in turn determines the stem class of the derived form (for a masculine stem, the *o*-class). In [aθuk<ít>ar-Ø], in contrast, the diminutive morpheme [-it-] appears inside the root and must therefore be an infix. From its infixal position, [-it-] does not interfere with the selection of the stem formative of the derivative, which remains an ordinary *e*-stem like the base: cf. the infixal plural [aθuk<ít>ar-e-s] with the suffixal plural [aθukar-(θ)ít-o-s].

Clearly, an infix is of no use for distinguishing between pseudoplural nouns and nouns with athematic stems in /s/. Fortunately, however, infixal and suffixal derivatives are easy to tell apart and, indeed, the evaluative forms adduced in (4) and (13) are all unequivocally suffixal. In the case of [birus-ít-o], the base [bírus] appears whole to the left of the diminutive suffix, which, in addition, imposes a different stem class on the derivative; the same is true of [brindis-ít-o]. In the case of [karl-ót-e], [sokrat-ít-o], and [kris-eθít-a], the pseudoplural ending of the base disappears before the stem-based suffix (as it is not part of the stem) and it does not reappear to the right of the suffix: crucially, the forms [karl-ót-e-s], [sokrat-ít-o-s], and [kris-eθít-a-s] exist, but they are not pseudoplural singulars; they are the genuinely plural counterparts of singular [karl-ót-e], [sokrat-ít-o], and [kris-eθít-a]. In addition, the evaluative suffixes in [karl-ót-e], [sokrat-ít-o], and [kris-eθít-a] again impose a different stem class on the derivative: e.g. *e*-stem [sókrat-e-s] changes into *o*-stem [sokrat-ít-o]. In this respect, the diminutive morpheme in [sokrat-ít-o] behaves entirely like the suffix in [aθukar-(θ)ít-o] (15b) and not at all like the infix in [aθuk<ít>ar-Ø] (15c). Therefore, the existence of diminutive infixation in no way challenges our conclusions concerning the opposition between pseudoplural nouns and athematic nouns with stems in /s/.

In fact, recognizing the existence of diminutive infixation is essential to understanding what would otherwise be an intractable puzzle (Bermúdez-Otero forthcoming a: §2.3.1). Compare the following two paradigms:

(16)		‘Socrates’	‘Charles’
a.	sg.	sókrat-e-s	kárl-o-s
	pl.	sókrat-e-s	kárl-o-s
b.	dim.sg.	sokrat-ít-o	karl<ít>-o-s
	dim.pl.	sokrat-ít-o-s	karl<ít>-o-s
c.	aug.sg.	sokrat-ót-e	karl-ót-e
	aug.pl.	sokrat-ót-e-s	karl-ót-e-s

Both [sókrat-e-s] and [kárl-o-s] are pseudoplural nouns, but they behave subtly differently under evaluative derivation. In the evaluative derivatives of [sókrat-e-s] we see all the expected effects of stem-based suffixation: the base’s pseudoplurality disappears, and the stem class of the derivative is determined by the evaluative suffix (*o*-stem in the case of the diminutive, *e*-stem in the case of the augmentative). The augmentative of [kárl-o-s] follows the same pattern: the pseudoplural /-s/ of the base is absent in the singular [karl-ót-e], and the derived stem switches to the *e*-class. The surprise comes in the diminutive of [kárl-o-s], which retains the pseudoplural /-s/ of the base in its singular form: [karl<ít>-o-s] can be both a pseudoplural singular or a genuine plural; cf. (5a,b) and (17).

(17) a.	Carl<ít>-o-s	es	buen-o
	Charles<DIM>-SF-PL	be.3SG.PRES.IND	good-SF[SG]
	‘Charley / little Charles is good’		
b.	Aquí hay	much-o-s	Carl<ít>-o-s
	here there	_be.PRES.IND many-SF-PL	Charles<DIM>-SF-PL
	‘Here there are many Charleys / little Charleses’		

This intriguing pattern can be explained by making two simple assumptions, extensively discussed in Bermúdez-Otero (forthcoming a: §2.3.1). First, diminutive /-it-/ can either be suffixed or infix, but augmentative /-ot-/ can only be suffixed. Secondly, the diminutive infix subcategorizes for bases in which the nucleus of the final syllable consists of unstressed /o/ or /a/. This subcategorization requirement can be explained diachronically: the diminutive infix arose by reanalysis from the diminutive suffix, which selects the stem formative /-o/ in masculine forms and /-a/ in feminine forms; the morphological selection properties of the suffix were simply reanalysed as phonological selection properties in the infix. Thus, /sókrat-e-s/ cannot take the diminutive infix, for, being an *e*-stem, it does not contain an unstressed /o/ or /a/ in the final syllable; accordingly, it takes the diminutive suffix, which attaches to the stem /sokrat-{e,Ø}-/, losing the pseudoplural /-s/. In contrast, /kárl-o-s/ fulfils the phonological selection requirements of the diminutive infix /-it-/, which is inserted to the left of the stem formative /-o/ without disrupting the adjacency of the stem and the pseudoplural /-s/: [karl<ít>-o-s]. Finally, both augmentatives are suffixal, and hence lose the pseudoplurality of the base.

Against this background, the infixal diminutive [bir<ít>us], which we encountered in (14b), appears problematic, since it violates the requirement that /-it-/ should be infix

immediately to the left of a final syllable nucleus consisting of unstressed /o/ or /a/. In this respect, we should observe that the form is unacceptable for many speakers, and is indeed more than 11 times rarer than the suffixal variant [birus-ít-o] (see note 12). I suspect that speakers who produce [bir<ít>us] take advantage of the fact that the diminutive infix /-it-/ is aligned with the left edge of the nucleus of the final syllable of the base, in contrast with the diminutive suffix /-it-o/, which aligns itself with the right edge of the base. Thus, by exceptionally relaxing the selectional requirements of the infix, these speakers are able to generate a diminutive form for /birus/ even if they do not know whether the right edge of the stem falls before or after the final /s/. If this suggestion is on the right track, the occurrence of [bir<ít>us] is another symptom of the underdetermination effect that besets nouns with homophonous singular and plural forms ending in /s/.

3.4. On *[birus-e-s]

Before turning to the explanatory task of showing how learners respond to this underdetermination effect, it remains still to clear up two descriptive problems that have occasionally hampered the analysis of nouns like *virus*.

Consider the paradigms in (18). The singular forms illustrate the full range of consonants permitted word-finally in Spanish, excluding xenonyms (see note 4). These legal word-final codas are all monosegmental: viz. /d, θ, s, n, l, r/ (see e.g. Harris 1999: 60).

(18)	a. <i>sg.</i>	b. <i>pl.</i>	
	pa.réð	pa.ré.ðes	‘wall’
	kruθ	krú.θes	‘cross’
	í.ris	í.ri.ses	‘iris’ ¹⁵
	θer.tá.men	θer.tá.me.nes	‘contest’
	áj.xel	áj.xe.les	‘angel’
	mu.xér	mu.xé.res	‘woman’

It is sometimes suggested that the [e] which appears in the plural forms before the inflectional marker /-s/ is inserted by a phonological process that avoids illegal coda clusters. If this view were correct, then one might expect the plural of *virus* to be *[bí.ru.ses], with an epenthetic [e] breaking up the illegal word-final cluster in /birus-s/. Since epenthesis does not in fact apply, this might be taken as an argument for treating *virus* as pseudoplural rather than as having an athematic stem ending in /s/. However, the premise of this argument is false. Spanish does not use vowel epenthesis to repair illegal word-final strings (Bermúdez-Otero forthcoming a: §1.2.2; Bonet forthcoming; Colina 2003a, 2003b; Harris 1999). When fully nativized, for example, English /stɒp/ becomes [es.tó], not *[es.tó.pe].

The behaviour of the nouns in (18) must be explained in terms of their morphological structure. They are all ordinary members of the *e*-class and, as such, their stems all end with the formative /-{e,Ø}/. The allomorphy of this stem formative is controlled by phonological factors in the singular form, but by morphological factors in the plural (Bermúdez-Otero forthcoming a: §1.2.2). In the singular, the phonology selects the null allomorph [-Ø] after

¹⁵ For some speakers this noun does not belong in the *e*-class, but is rather an athematic stem ending in /s/: sg. [íris], pl. [íris], dim.sg. [iris-ít-o]. See Bermúdez-Otero (forthcoming a: note 27).

permissible word-final sequences, and the nonnull allomorph [-e] elsewhere. In the plural, however, a morphological process preempts phonological selection:¹⁶

$$(19) \quad -\{e, \emptyset\} \rightarrow -e / \left[\begin{array}{c} \text{SF} \\ \text{stem} \end{array} \right] -S_{\text{word}}$$

In this sense, phonotactics do play a rôle in causing the null formative [-Ø] to appear in the singular forms in (18a). However, the appearance of the nonnull formative [-e] in the plurals in (18b) has nothing whatever to do with phonotactics, at least synchronically.

This assertion is confirmed by two arguments. First, [-e-] fails to appear in plural forms when the stem is not a member of the *e*-class. This can be observed in xenonyms such as the noun [klip] ‘clip’. For the native speaker, the fact that this word ends in a consonant other than [ð, θ, s, n, l, r] marks it as phonotactically deviant and excludes it from the core vocabulary. The noun is thereby disqualified from membership of the *e*-class, and so it is assigned to the marginal athematic class, even though this compounds the phonotactic problem. Athematic stems have no stem formative, and so the addition of /-s/ in the plural turns an illegal monosegmental coda into an illegal coda cluster: [klip-s], not *[klip-e-s].

Conversely, [-e-] appears in the plural form of *e*-stems even when it is not required to attain phonotactic well-formedness. In my idiolect, for example, there is a contrast between the *e*-stem adjective /indú- $\{e, \emptyset\}$ -/ ‘Hindu’ and the athematic noun /menú/ ‘menu’. The contrast manifests itself in the plural, where the former has an [-e-] but the latter does not:

(20)	<i>e-stem</i>	<i>athematic stem</i>
sg.	indú-Ø	menú
pl.	indú-e-s	menú-s

The [-e-] in [indú-e-s] is phonotactically superfluous, as [ús] is a legal word-final sequence: cf. [awtoβús] ‘bus’, [xesús] ‘Jesus’, etc.

In conclusion, parsing the final /s/ of *virus* as part of the stem does not predict the ungrammatical plural *[bírus-e-s] because Spanish does not apply vowel epenthesis to illegal word-final clusters.

3.5. On *[birus-ál-Ø]

Harris (1992) adduces data like (21) in support of the claim that the final /s/ of *virus* is not part of the stem.

(21)	bir-ál-Ø	‘viral’	cf. *birus-ál-Ø
	bír-ik-o	‘viral’	cf. *birús-ik-o

However, these forms yield to a different analysis (Bermúdez-Otero forthcoming a: §2.3.1). The lexical item VIRUS has two root allomorphs: (22a). The first, $[\text{root bir-}]$, is bound, occurring only in root-to-stem constructions with overt derivational suffixes: e.g. (22c), surfacing as (21). The second, $[\text{root birus-}]$, is homophonous with an athematic stem, (22b),

¹⁶ Observe that allomorphy rule (19) is triggered by the morph /-s/, not by the syntactic feature [+pl] (or its equivalent; see note 6). Accordingly, the rule does apply to the singular form of *e*-stem pseudoplural nouns like [sókrat-e-s], which contain the morph /-s/ but not the feature [+pl].

which provides the base for stem-based constructions such as the evaluative derivatives in (22d).

(22) *VIRUS*

- | | |
|---------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| a. root allomorphs | $\llbracket_{\text{root}} \text{bir-}\rrbracket$
$\llbracket_{\text{root}} \text{birus-}\rrbracket$ |
| b. stem | $\llbracket_{\text{stem}} \llbracket_{\text{root}} \text{birus}\rrbracket\rrbracket$ |
| c. root-based derivatives | $\llbracket_{\text{stem}} \llbracket_{\text{root}} \text{bir}\rrbracket \text{al-}\{e, \emptyset\}\rrbracket$
$\llbracket_{\text{stem}} \llbracket_{\text{root}} \text{bir}\rrbracket \text{ik-o}\rrbracket$ |
| d. stem-based derivatives | $\llbracket_{\text{stem}} \llbracket_{\text{stem}} \llbracket_{\text{root}} \text{birus}\rrbracket\rrbracket \text{it-o}\rrbracket$
$\llbracket_{\text{stem}} \llbracket_{\text{stem}} \llbracket_{\text{root}} \text{birus}\rrbracket\rrbracket \text{ot-}\{e, \emptyset\}\rrbracket$ |

This situation is not at all unusual. For example, the lexical item DRAMA ‘drama’ has two root allomorphs too: (23a). The first, $\llbracket_{\text{root}} \text{dramat-}\rrbracket$, is again bound and occurs only in root-to-stem constructions with overt derivational suffixes: (23c). The second, $\llbracket_{\text{root}} \text{dram-}\rrbracket$, supplies the base for an *a*-class masculine stem, (23b), which is in turn the base of stem-based evaluative derivatives such as the *e*-class augmentative in (23d).

(23) *DRAMA*

- | | |
|--------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| a. root allomorphs | $\llbracket_{\text{root}} \text{dramat-}\rrbracket$
$\llbracket_{\text{root}} \text{dram-}\rrbracket$ |
| b. stem | $\llbracket_{\text{stem}} \llbracket_{\text{root}} \text{dram}\rrbracket \text{a}\rrbracket$ |
| c. root-based derivative | $\llbracket_{\text{stem}} \llbracket_{\text{root}} \text{dramat}\rrbracket \text{ik-o}\rrbracket$ ‘dramatic’ |
| d. stem-based derivative | $\llbracket_{\text{stem}} \llbracket_{\text{stem}} \llbracket_{\text{root}} \text{dram}\rrbracket \text{a}\rrbracket \text{on-}\{e, \emptyset\}\rrbracket \rightarrow \llbracket \text{dramón} \rrbracket$ ‘drama’ aug. |

Two arguments confirm that the derivational suffixes which appear in (21), $/\text{-al-}\{e, \emptyset\}/$ and $/\text{-ik-o}/$, can attach to roots (cf. §2). First, we have seen plenty of evidence that they occur with uninflectable bases: cf. (7) and (9), (22c), (23c). Secondly, they do not cause diphthongization to overapply. This shows that their base does not constitute a separate cyclic domain, which is precisely how roots are supposed to behave: cf (10) and (24).

(24) a. *root-based derivation*

- | | |
|-------------------------|-------------------------------------------------------------------------------------------------------------------------------|
| | ‘gate’ |
| morphological structure | $\llbracket_{\text{stem}} \llbracket_{\text{root}} \text{p}\{we/o\} \text{rt}\rrbracket \text{al-}\{e, \emptyset\}\rrbracket$ |
| domain structure | $\llbracket \text{p}\{we/o\} \text{rt-al-}\{e, \emptyset\} \rrbracket$ |
| first cycle | portál |

b. *stem-based derivation*

	‘door’ dim.
morphological structure	$[[_{\text{stem}} [[_{\text{stem}} [[_{\text{root}} p\{we/o\}rt]a]e\theta it-a]$
domain structure	$[[[p\{we/o\}rt-a]e\theta it-a]$
first cycle	pwérta
second cycle	pwerteθíta

I conclude that derivatives such as [bir-ál-Ø] and [bír-ik-o] tell us nothing about the stem of *virus* because they are based on a bound root allomorph.

4. Acquiring the contrast between athematic stems in /s/ and pseudoplurals

Our descriptive work is done: now, at last, we have a precise grasp of the facts to be explained. Most Spanish speakers tacitly know that *Carl-o-s* is a deponent noun whose final /-s/ lies outside the stem not only in the plural but also in the singular form. They also tacitly know that *virus* is a nondeponent noun with an athematic stem ending in /s/ and subject to phonological degemination in the plural. However, they cannot have acquired their knowledge of this morphological contrast by inspection of the inflectional paradigms of the two nouns, since these simply show that both nouns have homophonous singular and plural forms ending in /s/. The only direct morphological cue to the distinction comes from suffixal evaluative derivatives: e.g. *Carl-ot-e* vs *virus-ot-e*. Only suffixal evaluative derivatives are of use: neither root-based constructions (e.g. *vir-al-Ø*, *vir-ic-o*) nor evaluative forms derived by infixation (e.g. *Carl<it>-o-s*, *vir<it>us*) can determine the location of the right edge of the stem in *Carl-o-s* and *virus*. We, as linguists, can learn about the contrast by diligently searching for the crucial data on the Internet; but children acquiring Spanish natively cannot. Yet the criterial forms have such low token frequency that we must assume them to be largely unavailable to learners of Spanish. Therein lies the learnability puzzle.

To appreciate just how infrequently the crucial forms occur, consider the figures in (25). In (25a) I give the number of Google hits scored by each datum in (4) and (13) on 23 July 2006. (25b) gives the number of tokens of the same forms in the corpora CREA and CORDE, which together contain 387,488,848 words.

(25)	a. <i>Google hits</i>	b. <i>CREA+CORDE</i>
virus-it-o	1,060	0
virus-ot-e	56	0
brindis-it-o	62	1
Carl-ot-e	549	0
Socrat-it-o	13	0
cris-ecit-a ¹⁷	0 (5 irrelevant hits)	1

This evidence confirms that Spanish learners have a negligible chance of hearing suffixal evaluative forms such as *Socrat-it-o* or *cris-ecit-a* in spontaneous discourse. Those who

¹⁷ The five irrelevant Google hits for *cris-ecit-a* ‘crisis’ dim. correspond to tokens of the diminutive of *Cris* ‘Chris’, which is in turn the hypocoristic of *Cristin-a* ‘Christine’. For the one token of *cris-ecit-a* in CREA, see Bermúdez-Otero (forthcoming a: note 31). Additionally, Iggy Roca (personal communication, 24 March 2006) reports hearing a speaker on a Spanish radio channel produce the augmentative *cris-ón-Ø* ‘fully on line and with utter spontaneity’. This datum supports our conclusion that the base *crisi-s* is a pseudoplural noun with an athematic stem in /i/.

eventually produce such forms must therefore be supposed to have relied on other cues for the knowledge that *Sócrat-e-s* and *crisi-s* are pseudoplural nouns.

Before proceeding to search for those cues, however, we should ask why the forms in (25) occur so infrequently. It might be suggested that their rarity is in fact a symptom of defectiveness: for most speakers, the base nouns would simply lack suffixal evaluative forms (cf. §3.2 above). However, this is not true; the forms in (25) are infrequent mainly because the base nouns occur only rarely in pragmatic circumstances favouring the use of evaluative derivation. To establish this point, let us again take *virus* as a test case. If the low frequency of *virus-it-o* were caused by defectiveness, then we would expect nondefective nouns with similar meanings to occur much more frequently in the diminutive. The facts prove otherwise: although deriving the diminutives of nouns such as *o*-stem *organism-o* ‘organism’ and *e*-stem *germen-Ø* ‘germ’ is morphologically unproblematic, forms like *organism-it-o* and *germen-cit-o* are not proportionally more frequent than *virus-it-o*.

(26)		Google hits ¹⁸	% of diminutives
a.	<i>virus</i>	‘virus’ sg.	16,300,000
	<i>virus-it-o</i>	‘virus’ dim.sg.	1,060
			0.0065
b.	<i>germen-Ø</i>	‘germ’ sg.	1,180,000
	<i>germen-cit-o</i>	‘germ’ dim.sg.	112
			0.0095
c.	<i>organism-o</i>	‘organism’ sg.	19,500,000
	<i>organism-it-o</i>	‘organism’ dim.sg.	17
			0.000087

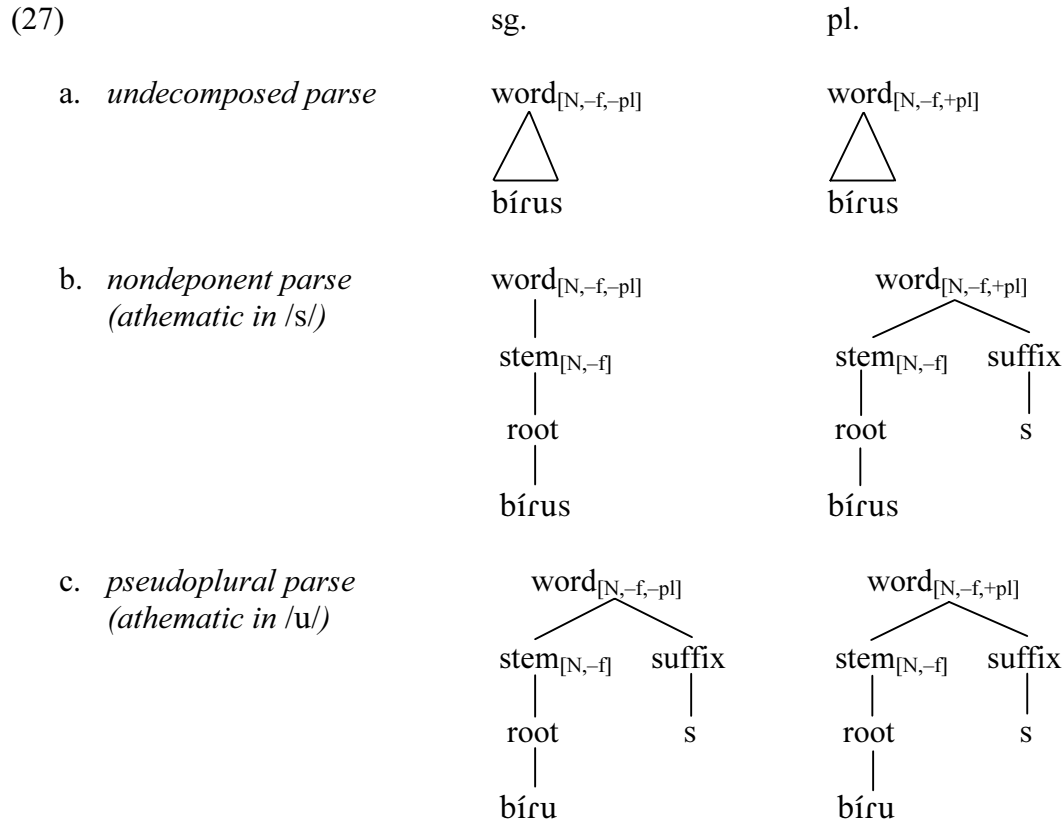
The evidence in (26) strongly suggests that the frequency of a diminutive form depends mainly on the pragmatic connotations of the base, rather than on its morphological structure. Anecdotal observation supports this conclusion: *virus-it-o* has a comparatively healthy number of tokens on the Internet because contributors to message boards concerned with information technology often have cause to use the noun *virus* in situations of heightened emotion; in the same way, the poor showing of *cris-ecit-a* is explained by adverse pragmatic conditions. However, I do not wish to deny that some speakers avoid constructing suffixal evaluative derivatives like those in (25): as we saw in §3.3, the occurrence of unexpected infixal forms such as *vir<it>us* (14b) suggests that this is sometimes the case.

Indeed, of all the possible outcomes of the learnability problem posed by nouns with homophonous singular and plural forms ending in /s/, the easiest to explain is ineffability, i.e. the situation where adult speakers are unable to generate suffixal evaluative derivatives. Consider, for example, a child who has been exposed to both the singular and plural forms of *virus* early in life. Since in the general course of language acquisition inflection is acquired before derivation, we are compelled to assume that, at some developmental stage, the child will store the singular and plural forms of *virus* in her protolexicon in a format resembling (27a). The representation in (27a) encodes the child’s knowledge that *virus* can be either a singular masculine noun or a plural masculine noun, but it does not provide any information concerning the structure of either form below the word node. We can therefore call (27a) the ‘undecomposed parse’. To progress beyond (27a) the child must choose between two decomposed parses: the nonpseudoplural parse in (27b) and the pseudoplural parse in (27c).¹⁹

¹⁸ Google seems to automatically round up the number of hits when it reaches the tens of thousands.

¹⁹ The representations in (27b,c) can easily be transposed into a Distributed Morphology format by observing the equivalences noted in (7). I leave this exercise to the interested reader.

But at this point, as we have seen, the child comes up against an underdetermination effect: the inflectional paradigm of *virus* cannot determine the morphological affiliation of the final /s/. Therefore, unless the child resorts to other indirect sources of information, she will remain stuck at (27a).



A speaker whose lexical representation of *virus* looks like (27a) will be unable to apply evaluative suffixation to this noun. As we saw in §2, evaluative suffixes subcategorize for stems rather than for words, and demand to be aligned with the right edge of the stem. In (27a), however, there is no stem node and, *a fortiori*, no right edge of a stem. In conclusion, a lexical entry like (27a) fails to provide a suitable input for stem-based suffixation (see further §5.1 for some technical considerations).

Nonetheless, *virus* is robustly attested in the diminutive: see again (14) and (26). This indicates that most learners do eventually overcome the underdetermination of the choice between (27b) and (27c). However, the variation reported in (14) also suggests that learners achieve this result by means that are not fully deterministic, but rather probabilistic. Therefore, our account of the learning path should have two properties. First, it should explain the most frequent outcome: namely, the knowledge of mature native speakers for whom *virus* and *brindis* are nondeponent nouns with athematic stems in /s/, whereas *Carl-os*, *Sócrat-es*, and *crisi-s* are pseudoplural; see (13). Secondly, it should have a certain degree of in-built flexibility, so that minority outcomes are not ruled out in principle.

My proposal is that learners choose between parses like (27b) and (27c) by relying on a number of indirect cues. These consist of parsing preferences grounded on morphological and phonological properties of the Spanish lexicon. These parsing preferences are violable and partially conflicting, so that the outcome of their interaction depends on their relative weights. As we shall see below, the weighting that leads to the majority outcome in (13) can be at least partially explained by lexical statistics: the rarer a particular structure in the

lexicon, the stronger the preference to avoid it. This approach possesses the required flexibility: variation in the size and composition of the lexicon across individuals and across developmental stages may significantly affect the weight of parsing preferences.

These parsing preferences should not be mistaken for optimality-theoretic constraints. First, I am not suggesting that they are components of the grammar of Spanish; rather, they are a linguist's description of the way in which properties of the Spanish lexicon guide learners' parsing choices. Secondly, their relative weights can be expressed as a ranking, but I am not claiming that they are arranged in a strict dominance hierarchy in the sense of Prince & Smolensky (1993: §1.1): indeed, we shall encounter instances in which two low-weighted preferences are able to gang up together and defeat a high-weighted preference.²⁰ Thirdly, I am assuming that the relative weighting of parsing preferences essentially emerges from the lexicon; thus it differs from the ranking of constraints in an optimality-theoretic grammar, which must be inferred from the primary linguistic data by using specific ranking algorithms.

The first parsing preference that I shall consider is the following:

(28) *Avoid pseudoplurals*

Prefer parses in which singular nouns do not contain $_{\text{stem}}\text{Sword}$.

(28) needs little justification. The only function of /-s/ in Spanish nouns is to mark plural number; it has no other inflectional or derivational rôle, and it is not involved in any kind of syncretism. In this light, (28) simply expresses the fact that, by default, learners assume forms not to be deponent.

Next I shall consider a parsing preference grounded on another morphological property of the Spanish lexicon. Spanish nouns fall into exactly four stem classes: *o*-stems, *a*-stems, *e*-stems, and athematic stems (Bermúdez-Otero forthcoming a). The *e*-stems comprise two subclasses: ordinary *e*-stems with the stem formative /-{e, Ø}/, and *e*-only stems with the stem formative /-e/. Because of allomorphy rule (19), however, ordinary *e*-stems and *e*-only stems differ only in the singular; the plural ends in /-e-s/ in both subclasses.

(29)		sg.	pl.	
a.	<i>o</i> -stem	grúp-o	grúp-o-s	'group'
b.	<i>a</i> -stem	més-a	més-a-s	'table'
c.	<i>e</i> -stem { ordinary <i>e</i> -stem {	luθ-Ø	luθ-e-s	'light'
		mónt-e	mónt-e-s	'hill'
		krúθ-e	krúθ-e-s	'crossing'
d.	athematic stem	menú	menú-s	'menu'

The athematic class is very small, containing only a tiny fraction of Spanish nouns. Some of these nouns are phonologically well-formed. One such instance is *menú* (pl. *menú-s*): its segmental and prosodic structure is no bar to membership of the core classes, as shown by the *e*-stem *hindú*-Ø (pl. *hindú-e-s*); see §3.4. However, phonologically deviant forms are consistently assigned to the athematic class: recall our discussion of *clip* in §3.4. In this sense, the athematic class is marginal both numerically and structurally. Accordingly, we can assume that learners of Spanish observe the following parsing preference:

²⁰ Admittedly, versions of Optimality Theory incorporating local constraint conjunction (Smolensky 1993) can replicate this effect by means of rankings of the format $A \& B \gg C \gg A, B$.

(30) *Avoid athematic stems*

Prefer parses in which nouns are not members of the athematic class.

So far we have not established the relative weight of preferences (28) and (30), but we can already achieve a crucial result: *virus* must be parsed as a nondeponent noun with an athematic stem ending in /s/. Recall that, for this noun, the learner faces a choice between structures (27b) and (27c). However, both parses put *virus* in the athematic class: (27b) makes it an athematic stem in /s/, and (27c) makes it an athematic stem in /u/. Accordingly, *Avoid athematic stems* does not discriminate between the two possibilities. The choice is left to *Avoid pseudoplurals*, which favours the nonpseudoplural analysis in (27b). At this point, therefore, we can legitimately claim to have explained—in the strong sense of the word—why most Spanish speakers prefer diminutive *virus-it-o* to *vir-it-o*: see (14a,c).

The case of *Carl-o-s* is more complex. This noun could be parsed either as a nonpseudoplural athematic stem in /s/, as in (31a), or as a pseudoplural *o*-stem, as in (31b).

	‘Charles’ sg.	‘Charles’ pl.	
a. <i>nonpseudoplural athematic</i>	[[word [[stem kárlos]]]	[[word [[stem kárlos]]s]	✕
b. <i>pseudoplural o-stem</i>	[[word [[stem kárl-o]]s]	[[word [[stem kárl-o]]s]	✓

In this case, the two preferences we have considered so far conflict: *Avoid pseudoplurals* favours (31a); *Avoid athematic stems* favours (31b). We therefore would obtain the right result, namely (31b), if we gave greater weight to the avoidance of athematic stems than to the avoidance of pseudoplurals. However, since we have no independent means of estimating the weight of preference (28), this would be a circular argument—descriptive, but not explanatory.

Fortunately, it is possible to break the circle, for there is another parsing preference which, though relatively low-weighted, favours treating *Carl-o-s* as a pseudoplural *o*-stem. This preference is grounded on phonological facts. In brief, the position of primary stress in Spanish singular nouns with a closed final syllable follows a markedness cline: final stress is unmarked, penultimate stress is marked, and antepenultimate stress is supermarked (see e.g. Roca 2005). This markedness cline corresponds directly to a scale of lexical frequency: Núñez-Cedeño & Morales-Front (1999: 211), quoted in Roca (2005: 350), provide the percentage frequencies in table (32), based on an electronic search of a 91,000-word list.

(32)

antepenultimate stress	penultimate stress	final stress
0.05%	2.03%	97.80%
e.g. [ré.xi.men] ‘diet’	e.g. [xér.men] ‘germ’	e.g. [sos.tén] ‘support’
(supermarked)	(marked)	(unmarked)

The cognitive effects of this markedness cline have been amply demonstrated in psycholinguistic experiments: for example, Waltermire (2004: 185) reports that “[...] the weight of the final syllable is consistently the most powerful factor in the placement of stress for nonce words [...]”. Crucially, however, the domain of this metrical generalization excludes the plural marker /-s/; in other words, the presence of a final consonant affects the

stress profile of a noun only if the final consonant is part of the stem.²¹ We are therefore justified in deducing the following two parsing preferences:

(33) *Avoid óō stems*

For a paroxytonic noun with a final syllable closed by /s/, prefer a parse in which the final /s/ lies outside the stem.

(34) *Avoid óóō stems*

For a proparoxytonic noun with a final syllable closed by /s/, prefer a parse in which the final /s/ lies outside the stem.

Given the lexical frequencies shown in (32), we can confidently predict that (34) will have greater weight than (33). This will be independently confirmed in (39a) below.

Let us now focus on *Avoid óō stems*. Paroxytonic nouns with closed final syllables like *ángel*-Ø [áj.xel] ‘angel’ and *germen*-Ø [xér.men] ‘germ’ are a minority in the Spanish lexicon, but a sizable minority. Moreover, their membership of the *e*-class places them squarely within the core vocabulary: recall that phonologically deviant loanwords (‘xenonyms’) are excluded from this class and forced to join the athematic set. We therefore predict that *Avoid óō stems* will be rather weak. This is confirmed by the behaviour of *virus*: the parsing bias against pseudoplural nouns favours structure (27b), whereas the avoidance of paroxytonic stems with closed final syllables favours (27c), and it is (27b) that most speakers select. Thus, *Avoid óō stems* proves weaker than *Avoid pseudoplurals*.

We can now return to *Carl-o-s*. It now emerges that there are two independent parsing preferences militating against the nondeponent athematic parse in (31a): *Avoid athematic stems* and *Avoid óō stems*. The two combined prevail against *Avoid pseudoplurals*, which favours (31a). Thus, *Avoid pseudoplurals* is strong enough to defeat *Avoid óō stems* by itself, but it yields to the combined forces of *Avoid athematic stems* and *Avoid óō stems*. This kind of ‘ganging-up’ effect should not surprise us: since we are not dealing with optimality-theoretic constraints but rather with parsing biases emerging from lexical statistics, we should expect them to interact additively.

In fact, an example we have not so far considered provides another interesting instance of an additive interaction: it is [á^l.las] ‘atlas/Atlas’ sg./pl. This noun has two conceivable parses: as a nondeponent athematic stems in /s/, or as a pseudoplural *a*-stem. Given what we know so far, we would expect the preferred parse to be pseudoplural, like *Carl-o-s*. Yes this is not

²¹ The plural marker /-s/ does not affect the position of stress. One can capture this fact by assuming that Spanish has the following extrametricality rule (or its equivalent in terms of ranked alignment constraints):

(i) $C \rightarrow \langle C \rangle / \text{stem}] (\dots) \text{word}]$

(i) says that a word-final consonant becomes extrametrical if it belongs to an inflectional marker.

This statement has the advantage of applying both to nouns and to verbs. In verbs, stress is morphologically determined in the majority of tenses: in the imperfective, for example, the stem formative is accented even if this gives rise to a metrical structure which is unattested in nonverbs (e.g. *am-a-ba-is* [a.má.bajs] ‘love’ 2pl.pret.ipfv.ind; cf. *convoy* [kom.bój] ‘convoy’, not *[kóm.boj]). In the present indicative and present subjunctive, however, stress alternates between the root and the stem formative in response to syllable structure (see e.g. Bermúdez-Otero forthcoming a: note 22; cf. Oltra-Massuet & Arregi 2005). However, although the default stress pattern for words with a closed ultima is oxytonic as shown in (32), closed final syllables do not attract stress in the present tenses: e.g. *am-a-s* [á.mas] ‘love’ 2sg.pres.ind. This shows that the final consonant, supplied by the person and number marker, is extrametrical.

A cyclic account of the facts would be inappropriate because all the inflectional suffixes of a verb are included in its innermost cyclic domain together with the verb stem (Harris 1995: 879). Notably, verbal inflection markers do not cause diphthongization to overapply: e.g. *encuentr-a-s* [enj.kwén.tras] ‘find’ 2sg.pres.ind, but *encontr-á-is* [enj.kon.trájs] 2pl.pres.ind; cf. (10) and (24).

what the evidence shows. As it turns out, [a^l.las] has two attested diminutives: infixal [a^l<ít>tas] and suffixal [a^llas-ít-o].²²

(35) a. [a^l<ít>tas]

Algo para un fic de esos dos...bufff con lo bestias que los imagino. Un duro entrenamiento, un pequeño fallo de 'Bel(porque atlas es jefazo jejeje) y el limalimón(**atlit**as) se lo hace pagar duro duro al pobre de Bel...que pese a todo disfruta con el castigo...

<http://miarroba.com/foros/ver.php?foroid=695410&temaid=3727597>

(accessed on 27 July 2006)

b. [a^llas-ít-o]

jajajajajaj te afe'ctó bastante ese SI de **atlasito** XDDDD jajajajaja pero bueno me alegro por ti

<http://miarroba.com/foros/ver.php?foroid=174926&temaid=2146952>

(accessed on 27 July 2006)

Like all infixal diminutives, [a^l<ít>tas] tells us nothing about the morphological affiliation of the final [s], but [a^llas-ít-o] can only be derived from the athematic stem /atlas-/. Why does this noun not behave like *Carl-o-s*? The answer is that both nouns are masculine, but *Carl-o-s* belongs in the *o*-class whilst *Atlas/atlas* belongs in the *a*-class. As is well-known, however, the *a*-class is the default for feminine stems, whereas the general default for nonfeminines is the *o*-class. On this basis, we may assume the existence of the following parsing preference:

(36) *Avoid masculine a-stems*

Prefer parses in which nouns are not simultaneously masculine and *a*-stem.

What we see in the case of *Atlas/atlas*, then, is *Avoid pseudoplurals* and *Avoid masculine a-stems* ganging up together against *Avoid athematic stems*.

We can now proceed to discuss *Avoid óōō stems*. The way in which Spanish speakers treat proparoxytonic stems with closed final syllables indicates that this parsing preference is extremely strong. First, items of this sort are numerically very rare: see table (32). Secondly, virtually all of them are assigned to the athematic class: a xenonym like [bé.ne.ton] 'Benetton', for example, is predictably assigned the plural [béneton-s] (Bermúdez-Otero forthcoming a: note 43). This indicates that native speakers see these nouns as phonologically deviant and incapable of entering the core vocabulary as members of the *e*-class. There is only one potential exception in common use: the noun [ré.xi.men] 'diet, régime', which has a learned plural [rexímen-e-s] where stress shifts to the right in order to keep within the final three-syllable window. Nonetheless, it is not at all unusual even for this noun to move into the athematic class: thus, *pace* Roca (2005: note 32), the plural [réximen-s] is not an 'on-line performance trial output' but a manifestation of morphophonological knowledge on the part of native speakers (see also §3.2 above).

²² Both tokens in (32) come an Internet forum on Japanese animé. The contributors are discussing scenarios involving two characters called Bel and Atlas.

(37) [réximen-s]

Debería hacernos pensar que los **regimens** autoritarios acorralados (Irán y Siria) quieren ganar tiempo y harán todo lo posible para crear confusion en Occidente hacia la moralidad de un enfrentamiento ...

<http://www.infomedio.org/2006/02/16/echando-lena-al-fuego>
(accessed on 27 July 2006)

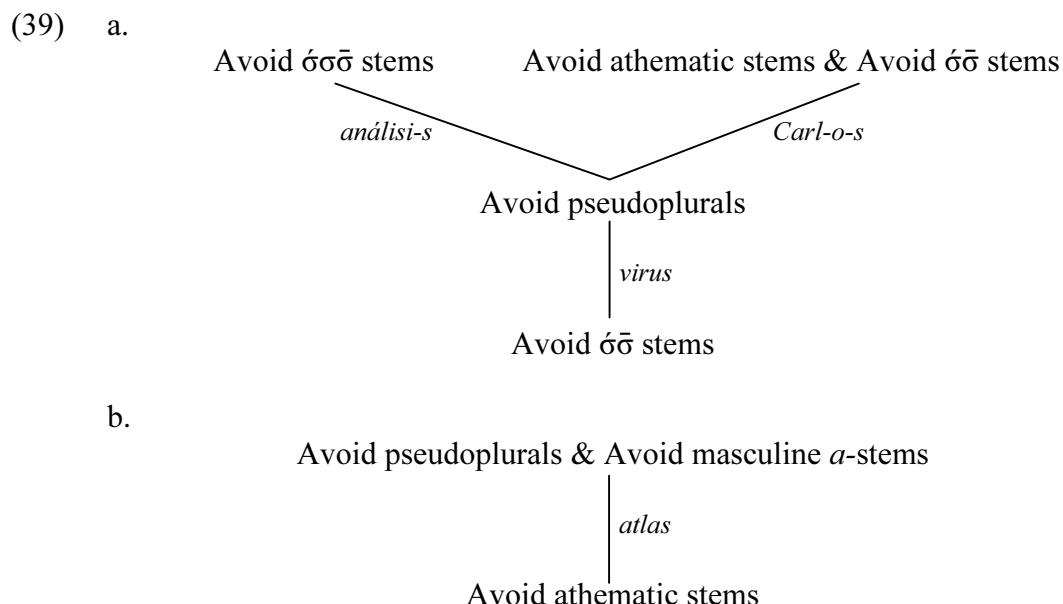
Finally, the judgement of deviancy attached to $\acute{o}\sigma\bar{o}$ stems is so strong that Spanish speakers sometimes impose antepenultimate stress upon loanwords with closed final syllables as a marker of exoticism and un-Spanishness: thus, Dutch *Amsterdám* becomes Spanish *Ámsterdam* (cf. Roca 2005: 368).

Given this evidence, we should expect that *Avoid $\acute{o}\sigma\bar{o}$ stems* weighs heavier than *Avoid pseudoplurals*. This is indeed what we find:

(38)	sg.	pl.	dim.	
a.	[[word [[stem sókrat-e]s]]	[[word [[stem sókrat-e]s]]	[sokrat-ít-o]	‘Socrates’
b.	[[word [[stem análisis]s]]	[[word [[stem análisis]s]]	[analís-ít-o]	‘analysis’

In the case of *e*-stem *Sócrat-e-s*, the pseudoplural parse is favoured by both *Avoid $\acute{o}\sigma\bar{o}$ stems* and *Avoid athematic stems*. However, *análisi-s* remains athematic whether the final /s/ is placed inside or outside the stem; hence, the fact that the noun is parsed as pseudoplural shows that *Avoid $\acute{o}\sigma\bar{o}$ stems* alone suffices to defeat *Avoid pseudoplurals*. This case is highly significant: it illustrates a phonologically grounded preference single-handedly triggering a syntax-morphology mismatch. The implications of this finding are discussed in §5.2 below.

In sum, we have so far motivated the relative weights shown in (39). In these diagrams, higher preferences are stronger and can defeat lower preferences. The ampersand denotes the additive combination of two preferences involved in a ganging-up effect.



There is only one group of nouns for which the weighted parsing preferences in (39) cannot account: paroxytonic pseudoplural athematic nouns like *crisi-s* (dim. *cris-ecita*, aug. *cris-on-a/cris-ón-Ø*: see note 17). These should behave like *virus* and *brindis*, which are

nondeponent athematic stems in /s/. I suspect that native speakers perceive a close morphological relationship between *crisi-s* and proparoxytonic items in the scientific, technical, or learned vocabulary like *análisi-s* ‘analysis’, *hipótesi-s* ‘hypothesis’, *parálisi-s* ‘paralysis’, *síntesi-s* ‘synthesis’, and so forth. Some speakers may even isolate a recurrent empty element /- $\{e/i\}$ si-s/ —a kind of *cranberry* morph. If that is the case, then the pseudoplural parse will be imposed by highly weighted *Avoid óóō stems* through its effects on the parsing of the proparoxytones.

In conclusion, the acquisition of the contrast between nouns with athematic stems in /s/ and pseudoplural nouns in Spanish raises a stark instance of Plato’s Problem (Chomsky 1986). The key to its solution, however, is to be sought not so much in Universal Grammar as in language-particular parsing preferences motivated by morphological and phonological properties of the Spanish lexicon. The basis of these parsing preferences and of their weighting is ultimately statistical, but our account in no way replaces ‘structure’ with ‘statistics’: it is only in the context of a grammatical system distinguishing between autonomous syntactic, morphological, and phonological levels of representation and incorporating the trichotomy of root, stem, and word that the problem can be meaningfully formulated, let alone solved.

5. Implications for the syntax-morphology interface

Our findings have general implications for two important phenomena at the syntax-morphology interface: defectiveness (morphological ineffability) and deponency.

5.1. Ineffability

To account for the fact that some Spanish speakers avoid constructing suffixal evaluative derivatives of nouns like *Carl-o-s* and *virus*, I have proposed that the underdetermination of the choice between pseudoplural and nonpseudoplural parses can literally deprive the learner of a suitable input for stem-based suffixation. This explanation opens up a promising avenue for research into defectiveness and, more widely, into the general problem of ineffability. It suggests that at least some instances of ineffability are caused by learnability deficits: speakers need to make a crucial syntactic, morphological, or phonological choice in order to generate the intended expression, but the choice cannot be made because it is permanently underdetermined in the trigger experience of learners. A case that appears to fit this pattern is the diminutive of German nouns whose last syllable contains an unstressed full vowel: e.g. *Monat* ‘month’ (Féry & Fanselow 2003, cited in an unpublished manuscript by Marc van Oostendorp). German speakers seem not to know which vowel should carry the umlaut triggered by the diminutive suffix, and so some avoid the construction altogether:

- (40) *Underdetermination of the diminutive of Monat ‘month’*
 ?Monatchen, ?Mönatchen, ?Monätchen, , ?Mönätchen, ?Ø.

I should point out, however, that the implementation of this insight about ineffability is not without technical difficulties. Notably, an excessively powerful theory of Universal Grammar is likely to make the learner’s mind up for her. In phonology, for example, most optimality-theoretic approaches to acquisition assume strong ranking biases, so that it is possible in certain circumstances for the relative ranking of two constraints to be determined in the absence of all experience; relatedly, failing to rank two crucially interacting constraints in Optimality Theory results in optionality, not ineffability. In Spanish, however, the technical implementation of our proposal is straightforward, provided that the stem-based

character of evaluative suffixes is expressed as a categorical subcategorization requirement rather than as a violable constraint (Bye forthcoming). In that case, undecomposed lexical entries like those in (27a) will fail to feed stem-based suffixation; the relevant evaluative derivatives become ineffable simply because they lack a suitable input.

5.2. Deponency

The main implications of our results concern the issue of deponency. In this connection, Corbett (this volume) seeks to catalogue the possible departures from ideal agglutination in inflectional morphology. His typological survey suggests that it is actually not possible to draw a sharp demarcation between deponency and other phenomena such as syncretism, defectiveness, and suppletion (see also Matthews this volume, Spencer this volume). This is shown, among other things, by the existence of several instances of syntax-morphology mismatches that diverge to a greater or lesser degree from the canonical pattern of deponency illustrated by Latin verbs like *sequor*: see (1). Canonical deponency is defined by the following three properties (Baerman this volume, Stump this volume):²³

(41) a. *Mismatch of form and function*

Sequor is passive in form but active in function.

b. *Concurring defectiveness*

The paradigm of *sequor* has no truly passive form meaning ‘I am followed’.

c. *Lexical exceptionality*

The deponency of *sequor* is unpredictable and lexically stipulated.

However, Bobaljik (this volume) exhibits a case of deponency, the spurious antipassive of Chukchi, which involves neither concurring defectiveness nor lexical exceptionality. Spanish pseudoplural nouns lie somewhere in between: they are lexically exceptional, and indeed constitute a minute proportion of the Spanish vocabulary, but they fail to exhibit concurring defectiveness insofar as the same morphologically plural form can behave syntactically as singular or as plural; see (5a,b) and (17).

As regards the causes of deponency, the main conclusion of my examination of Spanish pseudoplurals is that, during language acquisition, the phonological properties of a word can cause it to receive a morphological parse that fails to match its syntactic features: e.g. whether a learner parses a particular segment as belonging to the stem or to an inflectional marker may depend upon such morphosyntactically irrelevant properties as the weight of syllables and the arrangement of feet within prosodic words. If this is true, then we should expect at least some instances of deponency to have arisen diachronically through processes of reanalysis that were phonologically motivated. This prediction is strikingly confirmed by Good (this volume): in the Bantu language Ganda, for instance, all verbs with roots ending in /s/ have been reanalysed as having the morphological structure of causatives regardless of their syntactic properties, even if intransitive (see Hyman 2003). These instances of phonologically induced deponency contrast with cases whose ultimate causes are to be sought in semantics, pragmatics, and/or general cognition. In Latin, for example, membership in the class of deponent verbs appears to be partially —though by no means completely— governed by lexical semantics (Xu, Aronoff & Anshen 2006). Similarly, the spurious antipassive of Chukchi is motivated by a constraint against the morphological expression of certain

²³ In (41) I follow Stump’s wording.

combinations of subject and object in which the object outranks the subject in the person hierarchy (Bobaljik this volume).

The diverse aetiology of deponency poses a significant challenge to the theory of grammar, for it is unlikely that the mechanisms proposed to deal with semantically or pragmatically triggered deponency will adequately deal with all the cases where the motivation is essentially phonological. In Bobaljik's analysis of the spurious antipassive of Chukchi, for example, the syntax carries the main burden of explanation: the appearance of nonactive, intransitive agreement morphology on the verb in the relevant environment is argued to fall out automatically from the syntactic structure of the clause, given a simple morphological operation of deletion which removes the features of the object in a disallowed subject-object combination within the T domain. In contrast, Spencer's (2000) analysis places the burden of explanation on the morphology itself, invoking rules of referral (Stump 1993). Be that as it may, it is not immediately clear that the use of either feature deletion operations or rules of referral will particularly illuminate any aspect of the behaviour of Spanish pseudoplural nouns. One's response to this state of affairs may be simply to assert, with Bobaljik (this volume), that the term 'extended deponency' does not denote a natural class of phenomena and that it would therefore be wrong to look for a unified theory of deponency. However, the way in which Spanish pseudoplural nouns show phonological constraints riding roughshod over the requirements of syntax-morphology correspondence must surely give us pause and cause us to ask just what limits there are to syntax-morphology mismatches in the languages of the world (Corbett this volume, Matthews this volume, Spencer this volume).

On a more positive note, however, my analysis of Spanish pseudoplural nouns suggests that morphological categories such as *stem* and *morphological word* possess a considerable degree of autonomy (see Aronoff 1994). Although the definitions of *stem* and *morphological word* in §2 make reference to syntax, the evidence of Spanish pseudoplurals shows that both categories can define domains for phonological constraints and, moreover, that the satisfaction of these phonological requirements may be attained through rather dramatic mismatches between syntax and morphology. In many other cases, however, the combinatorial demands of syntax cause stems and words to violate phonotactic requirements. Ultimately, then, it is because morphological categories are subject to simultaneous and contradictory pressures from syntax and phonology that they are reducible to neither.

Manchester, 27 July 2006

References

- Aronoff, Mark (1994). *Morphology by itself*. Cambridge, MA: MIT Press.
- Bachrach, Asaf & Michael Wagner (2006). Syntactically driven cyclicity vs. output-output correspondence: the case of adjunction in diminutive morphology. Handout of paper presented at the 29th GLOW Colloquium, Barcelona, 6 April 2006. Available at <http://glow.uvt.nl/GLOW2006/BachrachWagner173.pdf>
- Baerman, Matthew (this volume). Morphological typology of deponency.
- Benveniste, Émile (1935). *Origines de la formation des noms en indo-européen*. Paris: Librairie Adrien-Maisonneuve.
- Bermúdez-Otero, Ricardo (forthcoming a). Morphological structure and phonological domains in Spanish denominal derivation. In Sonia Colina & Fernando Martínez-Gil (eds), *Optimality-theoretic studies in Spanish phonology*. Amsterdam: John Benjamins. Available from the Rutgers Optimality Archive, ROA 766, <http://roa.rutgers.edu/>
- Bermúdez-Otero, Ricardo (forthcoming b). *Stratal Optimality Theory*. Oxford: Oxford University Press.

- Bobaljik, Jonathan David (2006). Where's Φ ? Agreement as a postsyntactic operation. Ms., University of Connecticut. <http://bobaljik.uconn.edu/papers/Phi.pdf>
- Bobaljik, Jonathan David (this volume). The limits of deponency: a Chukotko-centric perspective.
- Bonet, Eulàlia (forthcoming). Gender allomorphy and epenthesis in Spanish. In Sonia Colina & Fernando Martínez-Gil (eds), *Optimality-theoretic studies in Spanish phonology*. Amsterdam: John Benjamins.
- Bye, Patrick (forthcoming). Allomorphy —selection, not optimization. In Martin Kramer, Patrik Bye, & Sylvia Blaho (eds). *Freedom of Analysis*. Available at http://www.hum.uit.no/a/bye/Papers/allomorphy_book.pdf
- Chomsky, Noam (1965) *Aspects of the theory of syntax*. Cambridge, MA : The MIT Press.
- Chomsky, Noam (1986). *Knowledge of language: its nature, origin, and use*. Westport, Conn.: Praeger.
- Cole, Jennifer (1995). The cycle in phonology. In John A. Goldsmith (ed) *The handbook of phonological theory*. Oxford: Blackwell. 70-113.
- Colina, Sonia 2003a. The status of word-final [e] in Spanish. *Southwest Journal of Linguistics* 22(1).87-107.
- Colina, Sonia 2003b. Diminutives in Spanish: A morphophonological account. *Southwest Journal of Linguistics* 22(2).45-88.
- Comrie, Bernard, Martin Haspelmath & Balthasar Bickel (2004). The Leipzig glossing rules. <http://www.eva.mpg.de/lingua/pdf/LGR04.09.21.pdf>
- Contreras, Heles (1977). Spanish epenthesis and stress. *University of Washington Working Papers in Linguistics* 3: 9-33.
- Corbett, Greville (2000). *Number*. Cambridge: Cambridge University Press.
- Corbett, Greville (this volume). Deponency, syncretism and what lies between.
- CORDE = Real Academia Española. *Corpus diacrónico del español*. <http://www.rae.es>
- CREA = Real Academia Española. *Corpus de referencia del español actual*. <http://www.rae.es>
- Den Os, Els and René Kager. 1986. Extrametricality and stress in Spanish and Italian. *Lingua* 69.23-48.
- Eddington, David (2002). Spanish diminutive formation without rules or constraints. *Linguistics* 40: 395-419.
- Eddington, David (2004). *Spanish phonology and morphology: experimental and quantitative perspectives*. Amsterdam: John Benjamins.
- Féry, Caroline & Gisbert Fanselow (2003). 'Ineffability in Grammar'. *Linguistische Berichte, Sonderheft* 11: 288–310.
- Good, Jeff (this volume). Slouching towards deponency: A family of mismatches in the Bantu verb stem.
- Harbour, Daniel (2006). *Morphosemantic number: from Kiowa noun classes to UG number features*. Dordrecht: Kluwer.
- Harris, James W. (1983). *Syllable structure and stress in Spanish: a nonlinear analysis*. Cambridge, MA: MIT Press.
- Harris, James W. (1985). Spanish word markers. In Frank H. Nuessel (ed) *Current issues in Hispanic phonology and morphology*. Bloomington, IA: Indiana University Linguistics Club. 34-54.
- Harris, James W. (1991). The exponence of gender in Spanish. *Linguistic Inquiry* 22: 27-62.
- Harris, James W. (1992). The form classes of Spanish substantives. In Geert Booij & Jaap van Marle (eds) *Yearbook of Morphology 1991*. Dordrecht: Kluwer. 65-88.

- Harris, James W. (1995). Projection and edge marking in the computation of stress in Spanish. In John A. Goldsmith (ed) *The handbook of phonological theory*. Oxford: Blackwell. 867–887.
- Harris, James W. (1999). Nasal depalatalization *no*, morphological wellformedness *sí*; the structure of Spanish word classes. *MIT Working Papers in Linguistics* **33**: 47–82.
- Hooper, Joan Bybee and Tracy Terrell (1976). Stress assignment in Spanish: a natural generative analysis. *Glossa* **10**: 64–110.
- Hyman, Larry M. (2003). Sound change, misanalysis, and analogy in the Bantu causative. *Journal of African Languages and Linguistics* **24**: 55–90.
- Inkelas, Sharon (1990). *Prosodic constituency in the lexicon*. New York : Garland.
- Jaeggli, Osvaldo (1980). Spanish diminutives. In *Contemporary Studies in Romance Languages*, F. H. Nuessel (ed), 142–158. Bloomington, IA: Indiana University Linguistics Club.
- Kiparsky, Paul (1982a). From cyclic phonology to lexical phonology. In Harry van der Hulst & Norval Smith (Eds). *The structure of phonological representations*, Vol 1. (pp. 131–75). Dordrecht: Foris.
- Kiparsky, Paul (1982b). Lexical morphology and phonology. In I.-S. Yang (Ed). *Linguistics in the morning calm: selected papers from SICOL-1981*. (pp. 3–91). Seoul: Hanshin.
- Kiparsky, Paul (2003). Finnish noun inflection. In D. Nelson and S. Manninen (eds), *Generative approaches to Finnic and Saami linguistics: case, features and constraints*, 109–161. Stanford, California: CSLI Publications.
- Lázaro Mora, Fernando (1999). La derivación apreciativa. In *Gramática descriptiva de la lengua española*, Ignacio Bosque and Violeta Demonte (eds), 4645–4682. Madrid: Espasa Calpe.
- Lloret, Maria-Rosa & Joan Mascaró (2005). Depalatalization in Spanish revised. Ms., Universitat de Barcelona & Universitat Autònoma de Barcelona. Available at ROA 708-0105, Rutgers Optimality Archive, <http://roa.rutgers.edu/>
- Marantz, Alec (1997). No escape from syntax: don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics* **4.2**: 201–225.
- Matthews, Peter (this volume). How safe are our analyses?
- Méndez-Dosuna, Julián & Carmen Pensado (1990). How unnatural is Spanish *Víctor* → *Vict-ít-or*? Infixes diminutives in Spanish. In *Naturalists at Krems: Papers from the Workshop on Natural Phonology and Natural Morphology (Krems, 1-7 July 1988)*, Julián Méndez-Dosuna and Carmen Pensado (eds), 89–106. Salamanca: Ediciones Universidad de Salamanca.
- Núñez-Cedeño, Rafael A. & Alfonso Morales-Front (1999). *Fonología generativa contemporánea de la lengua española*. Washington, D.C.: Georgetown University Press.
- Oltra-Massuet, Isabel (1999). On the constituent structure of Catalan verbs. *MIT Working Papers in Linguistics* **33**: 279–322.
- Oltra-Massuet, Isabel (2000). On the notion of theme vowel: a new approach to Catalan verbal morphology. *MIT Occasional Papers in Linguistics* **19**.
- Oltra-Massuet, Isabel & Karlos Arregi (2005). Stress-by-structure in Spanish. *Linguistic Inquiry* **36**: 43–84.
- Pensado Ruíz, Carmen (1997). On the Spanish depalatalization of /ɲ/ and /ʎ/ in rhymes. In Fernando Martínez-Gil & Alfonso Morales-Front (eds) *Issues in the phonology and morphology of the major Iberian languages*. Washington DC, Georgetown University Press. 595–618.
- Prince, Alan & Paul Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar*. Ms, Rutgers University & University of Colorado, Boulder. Published with revisions (2004). Oxford: Blackwell.

- Rainer, Franz (1996). Inflection inside derivation: evidence from Spanish and Portuguese. In Geert Booij & Jaap van Marle (eds) *Yearbook of morphology 1995*. Dordrecht: Kluwer. 83-91.
- Rainer, Franz (2003). Studying restrictions on patterns of word-formation by means of the Internet. *Rivista di Linguistica*, **15**(1): 131 - 139
- Roca, Iggy (2005). Saturation of parameter settings in Spanish stress. *Phonology* **22**: 345-394.
- Roca, Iggy and Elena Feliú (2003). Morphology in truncation: the role of the Spanish desinence. In *Yearbook of Morphology 2002*, Geert Booij and Jaap van Marle (eds), 187-243. Dordrecht: Kluwer.
- Smolensky, Paul (1993). Harmony, markedness, and phonological activity. Paper presented at Rutgers Optimality Workshop 1, Rutgers University, New Brunswick, New Jersey, October 1993. Handout available from the Rutgers Optimality Archive, ROA 87, <http://roa.rutgers.edu/>
- Spencer, Andrew (2000). Agreement morphology in Chukotkan. In Wolfgang U. Dressler, Oskar E. Pfeiffer & Markus A. Pochtrager (eds.), *Morphological analysis in comparison*, 191-222. Amsterdam: John Benjamins.
- Spencer, Andrew (this volume). Extending deponency: implications for morphological mismatches.
- Stump, Gregory T. (1993). On rules of referral. *Language* 69: 449-479.
- Stump, Gregory T. (this volume). A non-canonical pattern of deponency and its implications.
- Waltermire, Mark (2004). The effect of syllable weight on the determination of spoken stress in Spanish. In Timothy L. Face (ed), *Laboratory approaches to Spanish phonology*, 171-191. Berlin: Mouton de Gruyter.
- Xu, Zheng, Mark Aronoff & Frank Anshen (2006). Deponency in Latin. Paper presented at the conference on 'Deponency and morphological mismatches', British Academy, London, 16 January 2006.