When Nobody Wins[*]

Kyle Gorman and Charles Yang

[In Rainer, F., Gardani, F., Luschützky, H. C., and Dressler, W. U. (ed),
Competition in inflection and word formation. Dordrecht: Springer, in press.]

1 Introduction

As young children vividly illustrate with their performance on the *wug*-test (Berko 1958), the ability to generalize linguistic patterns to novel items is a core property of language. It thus comes as a surprise when we stumble into a dark and dusty corner of language where our boundless linguistic productivity has unexpectedly failed. For example, the English verb *forgo* 'to go without' is unexceptional in all ways but one; unlike nearly every other verb in the language, it lacks a well-formed simple past. There's nothing pragmatically odd about having gone without in the past—the past participle *forgone* is unobjectionable—but speakers recoil at the sound of **forwent*, and **forgoed* is no better (and perhaps a bit worse). It is possible to find occasional examples of the hated **forwent*, given a large enough corpus, but very many of these tokens are "mentions" of the gap rather than earnest "uses". This is not the only English verb exhibiting such an idiosyncracy; for instance, it has long been noted (e.g., Hill 1976:668, Pinker 1999:136f., Pullum and Wilson 1977:770, among others) that *stride*, which has an unobjectionable simple past, *strode*, lacks a standard past participle; both **stridden* and **strided* are ill-formed, and few speakers accept *strode* in a past participle context.[1] We refer to these unexpectedly ineffable forms as *lexical gaps*.[2]

[*]This chapter adapts and expands upon analyses first presented in chapter 5 of Yang 2016. We thank Margaret Borowczyk for assistance with Polish, and Jennifer Preys and Vitaly Nikolaev for assistance with Russian.

[1]For a quantitative analysis of the **stridden* gap and the failings of frequency and other measures of indirect direct negative as a solution to the gap problem, see Yang 2016, section 5.1.1.

[2]We avoid the competing term *paradigm* (or *paradigmatic*) *gap*, as we do not wish to make a commitment to the theoretical notion of the inflectional paradigm.

In one sense, there is nothing particularly novel about lexical gaps. They are found both in the languages of large and powerful nation-states—those with armies and navies, those with national language academies—-and in endangered languages spoken by a few toothless elders; for cross-linguistic samples, see papers in Baerman et al. 2010, Fanselow and Féry 2002, and Rice and Blaho 2009. But while lexical gaps caught the attention of both ancient grammarians (see, e.g., Neue 1866) and the founders of modern linguistic theory (see below), they have played remarkably little role in theories of productivity; they have been "ill incorporated into our theories of language" (Baerman and Corbett 2010:2). For instance, Optimality Theory (Prince and Smolensky 1993), Distributed Morphology (Halle and Marantz 1993), and dual-route models of word formation (e.g., Pinker 1999) all appear to assume the existence of a default form and must employ additional mechanisms to predict the illformedness of **forwent* and **stridden*.

In his *Prolegemona to a theory of word formation*, Halle (1973) draws attention to the existence of lexical gaps, illustrating this with an example from Russian (see section 5 below). Halle proposes two accounts of these gaps, which he regards as equivalent. Under the first—and better-known—proposal, words like **forwent* are generated by the word formation component of the grammar, but are marked with a feature [−Lexical Insertion] preventing their actual use. This idea is implicit in accounts of lexical gaps (e.g., Pertsova 2005, 2016) depending upon the notion of *lexical conservativism* (Steriade 1997, 2008), which holds that learners require explicit positive evidence before they will attempt to use a particular forms. This is easily encoded in classical Optimality Theory: assuming the "initial state" is one where all markedness constraints dominate all faithfulness constraints (Smolensky 1996), then for any novel input, a phonologically null—yet ineffable—candidate, the Null Parse, is optimal (e.g., Raffelsiefen 1996, 1999, Rebrus and Törkenczy 2009, Rice 2005, Wolf and McCarthy 2009).[3] However, this account is difficult to reconcile with the unbounded creativity of word formation: children do pass the *wug*-test, and the instant a novel verb like *google* entered the English lexicon, its past tense *googled* became available.

[3]Orgun and Sprouse (1999) propose a different, albeit roughly equivalent, mechanism for encoding defectivity in Optimality Theory; see the papers in Fanselow and Féry 2002 and Rice and Blaho 2009 for debate regarding these two mechanisms.

As an alternative, Halle (1973; fn 1) suggests an that only those items undergoing "nonproductive" word formation rules are eligible for the feature [−Lexical Insertion], all others being marked [+Lexical Insertion] by default. Hetzron (1975), arguing against Halle's first proposal, comes to the same conclusion: gaps arise in unproductive corners of the grammar. Many later authors have adopted some form of Halle's second proposal (e.g., Albright 2003, 2009, Anderson 2010, Baronian 2005, Hudson 2000, Maiden and O'Neill 2010, Pullum and Wilson 1977), linking lexical gaps to unproductivity (though putting forth divergent accounts of the factors contributing to productivity). We too endorse Halle's second proposal. But since words and rules do not wear their (un)productivity on their sleeves, we employ a model of linguistic processing, and more specifically, upon one of the consequences of this model, the *Tolerance Principle* (Yang 2005, 2016). This principle predicts, on strictly numerical grounds, to predict whether or not a linguistic generalization will be encoded as productive by the child learner. Under our account, lexical gaps arise in precisely those linguistic contexts for which the Tolerance Principle identifies no productive generalizations.

In the following section, we outline the Tolerance Principle and its predictions for lexical gaps, including a discussion of how they are related to the notion of competition in morphology, the theme of the present volume. We then provide case studies of lexical gaps in three languages. Note that our goal, in each case study, is to predict the linguistic *contexts* in which gaps arise, defining those contexts using distributional, and where available, behavioral, evidence. However, unlike some prior work (e.g., Albright 2003, Hudson 2000, Pertsova 2016), we do not attempt to predict exactly which complex words will be judged ineffable. We adopt this more modest goal due to our suspicion that ineffability of individual forms is synchronically arbitrary (see, e.g., Baerman 2008 on defectivity in Russian verbs), and as we show in section 3, the failure of an existing, explicit model of gaps making predictions at the word level. Furthermore, we do not model actuation or change of lexical gaps, or the strategies speakers use to compensate for gaps (e.g., periphrasis), though these represent fruitful topics for future work.

2 Productivity and gaps

How would one represent a rule (R) and the exceptions of that rule (e.g., a set of words w_1, w_2, …, w_n)? If one is committed to a mechanistic account—like a computer programmer, for instance—perhaps the only way to encode rules and exceptions is through a set of conditional statements:

(1) If $w = w_1$, then …

If $w = w_2$, then …

…

If $w = w_e$, then …

Apply R

(1) entails that in order for R—the productive rule—-to apply, one must first scan through a list of competing generalizations to ensure that it is not one of the exceptions (w_1, w_2, …, w_e). This immediately recalls the Elsewhere Condition, an idea dating back to Pāṇini and later adopted by generative grammarians (e.g., Anderson 1969, Kiparsky 1973). But there is something perverse about (1); to produce *walked*, for example, does one really need to scan a list of irregular verbs to confirm that *walk* is not among them?[4] A moment of reflection, however, suggests that the only other option—listing thousands of irregulars instead of a hundred-odd irregulars—is even worse. One could imagine assigning to each regular verb a special feature triggering the application of the "add *-d*" rule, but that would imply that the morphological status of *every* verb would need to be committed to lexical memory: both the regulars and the irregulars too, since they are by definition unpredictable.

In fact, there is wide-ranging behavioral evidence for the "irregulars first, regulars last" representation of morphology; see see Yang 2016, chap. 3, for a review. The key evidence comes from online processing of words in lexical decision and words production studies. When irregulars and regulars are suitably matched for various factors (e.g., stem and surface frequency) known

[4]We do not claim that the language user will *activate* the full lexical entries for the irregulars before accessing the regulars. It is conceivable that the computational search process deals with a hash list, as described in (1): the entries are "keys" indexed by the semantic and phonological properties of words, which point to full lexical entries in the long term memory, rather than lexical entries themselves.

to affect the speed of processing, irregulars are recognized and produced significantly *faster* than regulars—convergent evidence for an algorithmic interpretation of the Elsewhere Condition as a explicit model of language processing.

2.1 The Tolerance Principle

From this we can provide a cost-benefits analysis for any set of competing generalizations. Specifically, an application of a productive rule must "wait" until all the exceptional conditions have been checked, and the more exceptions there are, the longer one will have to wait. Under very general assumptions about word frequencies (Yang 2016:61f.), we can prove:

(2) *Tolerance Principle*

Suppose a rule R is applicable to N items in a learner's vocabulary, of which e are exceptions that do not follow R. The necessary and sufficient condition for the productivity of R is:

$$e \leq \theta_N \text{ where } \theta_N := \frac{N}{\ln N}$$

The Tolerance Principle requires two input values, N and e, and returns the productivity status of a rule. Its application requires a well-defined rule such that N and e can be measured, whether by the child learner or the researcher. To learn the structural description of a rule, typically in the form of "X $\rightarrow$ Y", one will need to invoke an inductive learning procedure for extracts generalizations from specific learning instances. For example, suppose two strong baseball batters are described with feature bundles [+red cap, +black shirt, +long socks] and [+red cap, +black shirt, +short socks]. Simply by intersecting the two feature bundles we can obtain a minimal, exceptionless generalization "[+red cap, +black shirt] $\rightarrow$ good hitter". Obviously, an appropriate inductive learner must encode the structural constraints on the human language faculty and other cognitive systems implicated in language acquisition (Chomsky 1965).

Table 1 provides some sample values of N and the associate threshold value θ_N.

These thresholds are significantly lower than a naïve "majority rules" principle, and this has

Table 1: The Tolerance Principle threshold for various values of N.

N	θ_N	%
10	4	40.0
20	7	35.0
50	13	26.0
100	22	22.0
200	38	19.0
500	80	16.0
1,000	145	14.5
5,000	587	11.7

some interesting implications for language acquisition. For instance, the principle predicts that a smaller vocabulary (i.e., smaller values of N) can tolerate a higher percentage of exceptions, all else held equal, so productive rules are *more* detectable in a learner who has encountered *less* primary linguistic data. This may explain children's remarkably early command of the main ingredients of language, and explain why maturational constraints seem to aid, rather than hinder, child language acquisition (Newport 1990; see Yang 2016, chap. 7, for extensive discussion).

Despite its simplicity, the Tolerance Principle has proved surprisingly effective. In Yang 2016, it is applied nearly 100 times, making accurate productivity predictions across numerous languages and domains from corpus statistics. Furthermore, artificial language studies conducted with young children have found near-categorical confirmation for the Tolerance Principle (Schuler et al. 2016). These robust results are somewhat unexpected; the derivation in (2) makes use of numerical approximations that only hold when N is large, but in the artificial language studies, the vocabulary, and thus N, is often quite modest (e.g., 8 or 9).

Finally, we note that the Tolerance Principle provides a discrete threshold for productivity, which conforms with the traditional conception discussed by Aronoff (1976) and reviewed by Bauer (2001) but are at odds with the more recent, gradient, perspective on productivity (e.g., Baayen 2009).[5] The child language acquisition evidence, however, unambiguously favors the categorical

[5] Which, unfortunately, and tautologically, uses an inherently gradient measure—a numerical ratio—to quantify productivity.

view; see Yang 2016, chap. 2, for a cross-linguistic review. In addition, it is interesting to observe that a threshold such as that provided by the Tolerance Principle, can only be analytically derived and is in principle empirically *undiscoverable*. Productive processes will lie above the threshold and unproductive processes will lie below, but with arbitrary "distance" from it in both cases: the threshold cannot be regressed out of the data.

2.2 Acquisition of the English past tense

Before returning to the question of lexical gaps, we first demonstrate the application of the Tolerance Principle to a well-known case, the acquisition of the English past tense. Any reasonable inductive procedure will identify the regular *-d* suffix (realized as [-d, -t, -əd]) as applying to stems of all phonological shape; thus, its productivity will be determined by the total number of verbs and the number of irregular verbs—*N* and *e*, respectively—in the learner's vocabulary. The same considerations apply to the various irregular rules. For instance, the irregular verbs *bring*, *buy*, *catch*, *fight*, *seek*, *teach*, and *think* all undergo a stem change replacing the rime with [ɔt]. But this "rime → ɔt" rule will fare terribly when applied to the remainder of the lexicon; its seven undergoers are easily swamped by thousands of exceptions, far exceeding the tolerance threshold. As a result, the rule must be lexicalized. Other irregular patterns can be analyzed similarly: as shown elsewhere (Yang 2016, chap. 4), all rules except the regular "add *-d*" are unproductive by this principle, accounting for the near-total absence of over-*ir*regularization errors in the speech of children acquiring English (Xu and Pinker 1995; see Yang 2016:34 for further discussion).

Following the same logic, we can see that the emergence of the "add *-d*" rule will require a long gestation period. Although children can quickly induce its structural description—perhaps using no more than a few dozen verbs (e.g. Yip and Sussman 1997)—irregulars are overrepresented in baby's first verbs. For instance, in a corpus of 5 million words of child-directed English drawn from the CHILDES (MacWhinney 2000) database, 76 of the 200 most common past tense verbs are irregular. As θ_{200} is only 37, we predict that children who only know the top 200 verbs cannot recognize the productivity of "add *-d*" despite the fact that it is the majority rule. For a learner with such a vocabulary, we posit that verbs produced with an *-d* suffix are in effect encoded as similarly to ("true") irregulars, so that the generalization does not extend beyond regular verbs encountered thus far.

Telltale evidence for productivity comes from the first attested overregularization errors, and when the developmental record is sufficiently rich, the Tolerance Principle helps us predict when this will occur with some precision. It certainly works for "Adam" (Brown 1973), one of the most-studied language learners in psycholinguistics, who produced his first recorded overregularization error at 2;11: "What dat feeled like?" Adam's earlier transcripts, going back nearly a year, do not contain a single instance of an incorrect irregular verb past tense. Tolerance predicts that by 2;11, Adam had acquired a sufficiently large number of regular verbs to overwhelm the irregulars. In Adam's transcripts up to age 2;11, there are $N = 300$ unique verbs in all, of which $e = 57$ are irregular. This is quite close to the predicted $\theta_{300} = 53$, and the discrepancy may be due to under-sampling of the regulars, which are less frequent on average and thus more likely to be left out of a sample. In summary, Adam appears to have acquired a productive "add *-d*" only once he acquired a filibuster-proof majority of regular verbs, consistent with the predictions of the Tolerance Principle.

2.3 A condition on lexical gaps

Just as this principle predicts productivity, it also can be applied to predict cases where no productive generalization holds, as follows:

(3) Condition on lexical gaps (Yang 2016:142):

Let there be S rules, each affecting N_i lexical items ($1 \leq i \leq S$), and $N = \sum_i N_i$. Gaps arise if and only:

$$\forall i \, . \sum_{j \neq i} N_j > \theta_N$$

That is, gaps arise when no rule (N_i) in N applies to sufficiently many items to tolerate the remaining ($\sum_{j \neq i} N_j$) items as exceptions. The following three case studies, focusing on previouslydocumented lexical gaps in Spanish, Polish, and Russian, illustrate this prediction in action.[6] To reiterate: in each case, our goal is simply to predict the contexts in which lexical gaps arise, defining those contexts via independently-motivated linguistic analyses.

This is a good place to review the Tolerance Principle and its application to gaps in the context of competition, the main theme of the present volume. The centrality of competition can be observed in three aspects of the present work. First, and as long recognized, the Elsewhere Condition is a direct embodiment of competition between the lexicalization of lexicalization exceptions and the automatic application of rule-based computation. Second, the derivation of the Tolerance Principle is predicated on the assumption that the language learner favors efficiency: the grammar that lists the exceptions ahead of a rule and the grammar that lists all lexical items are in competition, and the faster one is chosen. Third, the condition on gaps claims that gaps arise from the competition over a morphological category: when no one generalization is sufficiently dominant, nobody wins.

3 Spanish verb stems

Harris (1969:114) observes that the Spanish verbs *agredir* 'to attack' and *aguerrir* 'to harden' are defective (i.e., have unpredictable lexical gaps) in certain inflectional forms. Below we review this pattern of defectivity, identify several other verbs with the same pattern of defectivity, and show that these patterns are consistent with the above condition on lexical gaps. But first, we review the competing generalizations.

3.1 Stem changes and defectivity

The majority of Spanish verbs conjugate without any stem change save for a shift of primary stress between (the final syllable of) the stem and the various syllables of the desisence (i.e., the inflectional suffixes). However, many verbs undergo stem changes targeting the vowel of the final syllable of the stem. All such verbs have a mid vowel—*e* or *o*—in the final syllable of their stem in

[6] Additional case studies can be found in Yang 2016, chap. 5.

citation form (the infinitive), and henceforth we refer to these as mid vowel stems. The most frequent such change is seen in those mid vowel stems in which the stem's final syllable vowel is *e* or *o* when primary stress falls on the desinence, as in the infinitive (e.g., *contar* 'to count'), but is realized as a diphthong—*ie* [je] or *ue* [we], respectively—when stress falls on the stem, as in the 1st person singular indicative (*cuento* 'I count').[7] Near-minimal pairs like *negar-niego* 'deny' and *pegar-pego* 'stick onto' show that one cannot easily predict whether or not a mid vowel stem will diphthongize or not.[8] Following Brame and Bordelois (1973:132f.) and Harris (1969:74f., 1985), we assume a lexically-specific rule which maps underlying /e/ to *ie* and /o/ to *ue* in certain stems.

(4) Diphthongization:

$$\begin{Bmatrix} \text{é} \\ \text{ó} \end{Bmatrix} \longrightarrow \begin{Bmatrix} \text{je} \\ \text{we} \end{Bmatrix} \text{ (Condition: } \sqrt{\text{cont-}},\ \sqrt{\text{neg-}},\ \ldots)$$

Some verbs of the third (*-i-*) conjugation exhibit a superficially-similar stem change, an alternation between *e* and *i* in the stem-final syllable (e.g., *pedir-pido* 'ask for', *gemir-gimo* 'groan').[9] Unlike diphthongization, however, this alternation is not directly conditioned by primary stress, and indeed, there does not seem to be any one positive condition for raising /e/ to *i*. Rather, we follow Harris (1969:111) and Lerner (2011) in modeling this alternation with a rule of high-vowel dissimilation lowering /i/ to *e* when followed by another *i*.[10] This rule must also be lexically conditioned so as to prevent overapplication to non-alternating stems such as *vivir-vivo* 'live'.

(5) Lowering:

$$\text{i} \longrightarrow \text{e} \;/\; \_\_\ \text{C}_0\ \text{i} \text{ (Condition: } \sqrt{\text{pid-}}, \sqrt{\text{gim-}},\ \ldots)$$

Albright (2003) and Maiden and O'Neill (2010) claim that, with rare exceptions, defective verbs are mid vowel stems of the third conjugation, like the examples noted by Harris (1969). Maiden

[7] Diphthongization alternations are also present in deverbal derivative (see, e.g., Eddington 1996, Harris 1969:106f., Lerner 2011), but these alternations do not transparently covary with surface stress.

[8] Albright et al. (2001) argue that there is some weak segmental conditioning on the application of diphthongization. Even if this is correct—and the evidence is not overwhelming—it does not obviate the need for lexical diacritics.

[9] These *e-i* alternations are also present in deverbal derivatives (Harris 1969:110, Lerner 2011).

[10] This correctly predicts that this pattern will be restricted to the third conjugation, since only this conjugation has a *-i-* theme vowel needed to condition the lowering rule.

and O'Neill further observe that gaps are present in exactly those inflectional forms alternating due to the diphthongization and/or lowering rules. For example, the 1st person singular present indicative of *agredir* could be **agredo*, **agriedo*, or **agrido*, but speakers are reluctant to produce or to accept any of the three. Examples of the four patterns are given below.[11]

(6) Third-conjugation mid vowel stem change patterns:

	sumergir 'to submerge' (no change):					
pres. indic.	sumerjo	sumerges	sumerge	sumergimos	sumergís	sumergen
pres. subj.	sumerja	sumerjas	sumerja	sumerjamos	sumerjáis	sumergen

	discernir 'to distinguish' (diphthongizing):					
pres. indic.	discierno	disciernes	discierne	discernimos	discernís	disciernen
pres. subj.	discierna	disciernas	discierna	discernamos	discernáis	disciernan

	desvestir 'to undress' (lowering):					
pres. indic.	desvisto	desvistes	desviste	desvestimos	desvestís	desvisten
pres. subj.	desvista	desvistas	desvista	desvistamos	desvistáis	desvistan

	agredir 'to attack' (defective):					
pres. indic.	*	*	*	agredimos	agredís	*
pres. subj.	*	*	*	*	*	*

To identify additional defective verbs, we consulted three Spanish dictionaries which mention verb defectivity (Butt and Benjamin 1988, Mateo and Rojo Sastre 1995, Real Academia Española 1992). The following verbs are listed as defective according to at least two of the three references.

(7) abolir 'abolish'

[11] Note that the *g-j* alternation in the conjugation of *sumergir* is purely orthographic, both are [x] throughout.

agredir ‘attack’

aguerrir ‘harden’

arrecir(se) ‘freeze’

aterir(se) ‘freeze’

colorir ‘color, dye’

descolorir ‘discolor, bleach’

despavorir ‘fear’

empedernir ‘harden’

preterir ‘ignore’

tra(n)sgredir ‘transgress’

As Albright and Maiden and O’Neill note, the strong association between mid vowel stem changes and defectivity suggest that the latter results from speaker uncertainty. Imagine a Spanish speaker who has encountered the infinitive *agredir*, but (for whatever reason) has not encountered any other form of this verb. To produce the 1st person singular present indicative (or any of the inflectional forms which may differ from the infinitive by stem change) they must know whether or not it is is no change, diphthongizing, or lowering. Since speakers are unwilling to extend any of the three patterns to a novel stem, we predict that none of the three competing generalizations is productive. More concretely, we predict that the competing rules satisfy the condition on lexical gaps. We now show this prediction is supported both by numerical and behavioral evidence.

3.2 Productivity of stem changes

Table 2 gives the frequencies of the competing mid vowel stem generalizations in the LEXESP database (Sebastián et al. 2001) by conjugation.[12] In both the first (-*a*-) and second (-*e*-) conjuga-

[12]Many Spanish verb stems have multiple prefixal variants; for instance, *desvestir* ‘to undress’, *investir* ‘to invest’, *revestir* ‘to decorate’, and *tra(s)vestir(se)* ‘to cross-dress’ are all plausibly derived from *vestir* ‘to dress’. Without exception, all verbs derived from the same stem undergo the same set of stem changes, and therefore we collect counts over verb stems rather than verbs. Verbs were manually grouped by stem, with an etymological dictionary (Roberts 2014) used to adjudicate unclear cases.

1st (-*a*-):		
	no change	1,200
	diphthongizing	105
2nd (-*e*-):		
	no change	144
	diphthongizing	23
3rd (-*i*-):		
	no change	3
	diphthongizing	13
	lowering	20

Table 2: Counts of stem changes in Spanish mid vowel verb stems by conjugation.

tions, the “no change” pattern emerges as productive: it has fewer exceptions (105 and 23, respectively) than the Tolerance Principle thresholds (181 and 32).[13] In the third conjugation, lowering represents a slim majority, but there are too many exceptions to the three competing generalizations for any of the three to satisfy the Tolerance threshold (10), consistent with our predictions above.

These generalizations are supported by results from two studies of child speech errors. Clahsen et al. (2002) analyze verb inflection errors in a sample of fifteen children acquiring Spanish, constituting roughly 6,000 tokens. They report that nearly all errors involving verb stems (116 out of 120) involved underaplication of diphthongization: that is, the child failed to apply the rule to a diphthongizing verb. Mayol (2007) analyzes verb stem errors in a sample of six children acquiring Spanish (constituting roughly 2,000 tokens). Mayol identifies several hundred instances where diphthongization was underapplied, but not a single case where it was overapplied to a no change verb. Statistical analyses find a significant effect of conjugation on correct usage of irregular verbs, with the highest error rates occurring in the first and second conjugations. Of the 345 errors identified by Mayol, the only errors in the third conjugation were two overregularizations of the diphthongizing stems *dormir* ‘to sleep’ and *venir* ‘to go’, respectively. This suggests not only that children acquiring Spanish are sensitive to conjugation, but it is consistent with our predictions that

[13] We analyze the three conjugations separately because it has long been theorized that the productivity of the various stem changes may be a function of conjugation. Note, however, that the Tolerance predictions are the same if the first two conjugations are grouped together.

"no change" rule is productive in the first and second conjugations, and this results in underapplication of diphthongization in these conjugations. Somewhat counterintuitively—but as predicted by our model—the absence of a productive default for mid vowel stems of the third conjugation is responsible in part for the rarity of errors in that conjugation.

In summary, a straightforward application of the Tolerance Principle predicts the context in which Spanish lexical gaps are found, as well as patterns of inflection errors made by children acquiring Spanish.

3.3 Evaluation of the Minimum Generalization model

Albright (2003) proposes a computational model intended to predict exactly which Spanish verbs will be defective. In an earlier study, Albright et al. (2001) uses a probabilistic rule induction system, the Minimal Generalization Learner (MGL; Albright et al. 2001, Albright and Hayes 2003) to model stem changes in Spanish verbs. This model takes as input pairs of strings and induces a set of rules to map from the former to the latter. Albright et al. report that this model accurately predicts human performance on a *wug*-test conducted with adult Spanish speakers, though they do not evaluate their model on real word data. In the MGL model, each induced rule is associated with a confidence score, a function of the number of undergoers and exceptions to that rule. Albright (2003) extends the *wug*-task by asking speakers to rate, on a seven-point Likert scale, their "uncertainty" as to the correct inflection of various Spanish verbs. This measure of uncertainty, averaged across participants, is negatively correlated with (subjective) frequency and with model confidence scores, and Albright implies these measures predict defectivity of specific verbs.

We share with Albright the intuition that productivity of a rule is a function of the number of that rule's undergoers and exceptions. However, our theory makes no predictions about the effect of frequency, nor does it make any predictions about the defectivity of individual verbs. Furthermore, it is not obvious to us that Albright's dependent variable—average speaker "uncertainty"—is an appropriate operationalization of lexical gaps, particularly since Albright notes it is not strongly correlated with dictionary data on defectivity. We therefore we attempted to estimate the degree to which the MGL model confidence score and frequency could act as predictors of the lexical gaps listed in (7). Following Albright (2003), we train the MGL model on pairs of candidate words—non-defective third-conjugation mid-vowel verbs—from the LEXESP database (Sebastián et al. 2001), "holding out" the defective verbs (7) for evaluation.[14] Each input pair consists of the infinitive minus *-ir* and the 1st person singular indicative minus *-o*. This latter form is chosen because it exhibits a diphthong in diphthongizing stems, and the high vowel alternant in lowering stems. For example, for the lowering verb *pedir-pido*, the input form is [ped] and the output is [pid]. As in Albright 2003, the confidence score associated with each pair is that of the highest-scoring applicable rule.

In figure 1, the confidence score is plotted on the x-axis and lemma frequency on the y-axis; defective verbs are indicated with circles. While some defective verbs have a relatively low confidence score, and while they are of relatively low frequency on average, there are also many non-defective verbs which have even lower confidence scores and/or frequencies.[15] In other words, neither of the features proposed by Albright appear to correctly predict gaps, at least without further information. This result confirms our hypothesis that while Albright's model may produce scores closely correlated with speakers' intuitions about "uncertainty", these intuitions do not adequately operationalize defectivity. We note that similar results are reached by Boyé and Cabredo Hofherr (2010) in an evaluation of Albright's model on similar verb stem phenomenon in French. We also note that in other languages, lexical gaps are found even in high-frequency words (e.g., Greek *kopéla* 'girl'; Sims 2007), further undermining the hypothesis that low frequency contributes to lexical gaps.

4 Polish masculine genitives

Polish declension has been discussed in the language acquisition literature as a serious challenge to the dual-route model of morphology, according to which learners must identify a default rule. In Polish, masculine nouns in the genitive singular (gen.sg.) either take *-a* or *-u* suffix, but Dąbrowska

[14]Whereas in table 2 we collapse prefixal variants of a stem into a single entry, Albright does not mention any such practice, so inputs to the MGL system in this experiment are verb stems plus any prefix(es), rather than stems.

[15]However, it may still be the case that subjective frequencies used by Albright are better predictors of defectivity than the corpus-based frequency norms used here.

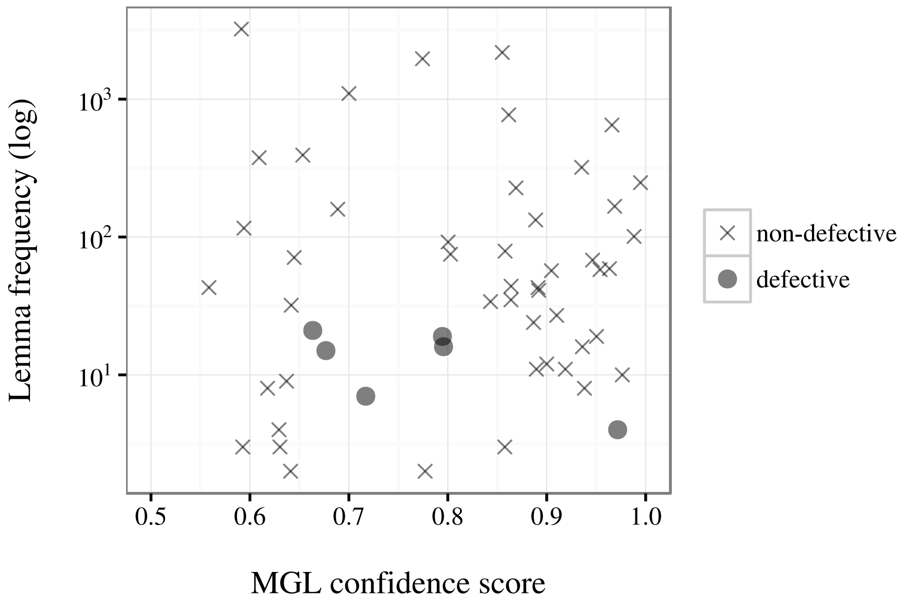

Figure 1: Confidence scores and frequencies of Spanish third-conjugation mid vowel stem verbs.

(2001) shows that neither is the default according to conventional criteria (Marcus 1995). For instance, unlike the English past tense *-d*, which is always extended to novel verbs (e.g., *googled*), some novel Polish masculine nouns take *-a* whereas others take *-u*. The distribution of these two suffixes is widely studied but has consistently defied philologists' attempts to provide a systematic classification; the choice appears to be largely arbitrary (Maunsch 2003). While some masculine nouns show *-a* and *-u* in apparent free variation (e.g., *deseni* 'design'), for others the genitive singular is ineffable:

> A Pole is often uncertain as to the correct genitive form of cities like *Dublin* 'Dublin' and *Göteborg* 'Gothenburg'. […] My informants were uncertain as to the correct genitive form of *Tarnobrzeg* 'Tarnobrzeg' which, according to the dictionaries, takes *-a*. (Kottum 1981:182f.)

We add to this list the following masculine nouns, identified as defective in the genitive singular by our informants.

(8) drut 'wire'

	suffix	*n*	avg. freq.	child error rate
gen.sg.:				
	-a	837	7.2	1.28%
	-u	516	8.8	0.24%
gen.pl.:				
	-ów	551	6.5	0.41%
	-i/-y	61	11.4	15.53%

Table 3: Distributions of genitive suffixes on Polish masculine nouns, the productivity predictions of the Tolerance Principle, average frequency (mean tokens per million words), and children's error rates (adapted from Dąbrowska 2001).

rower 'bike'

balon 'balloon'

karabin 'rifle'

autobus 'bus'

lotos 'lotus flower'

These facts collectively suggest the absence of a productive rule; which suffix a verb selects must be learned by rote, a process which continues well into the teenage years (Dąbrowska 2001, 2005). In contrast, the genitive plural (gen.pl.) for masculines is unproblematic: the default suffix is *-ów* with a small number of exceptional nouns taking *-i/-y*. Children acquiring Polish make very few errors in the genitive singular (gen.sg.), and frequently overextend *-ów* in the gen.pl., just as this description leads us to expect (Dąbrowska 2001).

Applied to noun stems found in child-directed Polish, the Tolerance Principle provides a straightforward account of both patterns. The data comes from a word frequency list compiled by Ewa Haman from CHILDES transcripts (Smoczyńska 1985, Weist et al. 1984, Weist and Witkowska-Stadnik 1986). Masculine nouns were manually classified into four classes according to their genitive case endings (sg.: *-a* or *-u*; pl.: *-ów* or *-i/-y*).[16] These results are summarized in table 3.

[16]This calculation is complicated slightly by a small number of masculine nouns which take *-a* in the gen.sg. and *-u* in the dative singular. Since this word list does not provide a way to determine the case of individual tokens, the count of *-u* nouns (516) is likely be a slight overestimate, but it is unlikely this would change the results, as the number of nouns potentially affected is rather small and both classes are quite far from the productivity threshold.

For the singulars, there are $837 + 516 = 1353$ stems. By the Tolerance principle, a productive suffix can tolerate no more than $\theta_{1353} \approx 188$ exceptions: *-a* and *-u* are both too numerous to meet this threshold, and the absence of a productive suffix is thus predicted. For the plurals, however, *-ów* attaches to 551 stems whereas *-i/-y* attaches only to 61: the productivity threshold is easily met ($\theta_{612} \approx 95$). These predictions accord well with errors made by children acquiring Polish. For the singulars, the lack of a productive process offers no opportunity for overregularization and children's performance on both suffixes is consistently high. For the plural, however, the existence of the *-ów* productive default serves as the attractor for overregularization: to wit, *-i/-y* nouns have the highest error rates by far, even though they have a higher average frequency in child-directed Polish. To summarize, the presence of a productive rule for generating the gen.pl. leads to overregularization errors, whereas the lack of such a rule in the gen.sg. leads to defectivity.

5 Russian verb stems

We now return to lexical gaps in Russian verbs, the phenomenon which occasioned some of the earliest consideration of productivity in generative grammar. Halle (1973) estimates that there are roughly 100 defective Russian verbs, but does not provide a list; Sims (2006) develops such a list by inspecting multiple Russian dictionaries annotating defectivity, and identifies 70 defective verbs in all; Pertsova (2016) further refines this list down to 63 verbs. A few examples are given below.

(9) oščutit' 'sense'
šerstit' 'search vehemently'
pobedit' 'win'
čudit' 'act strangely'
forsit' 'swagger'
pylesosit' 'vacuum'
derzit' 'be rude'
lazit' 'climb'

With one exception, all defective verbs are dental stems of the 2nd (*-i-*) conjugation,[17] and are gapped in the first person singular (1sg.) of the so-called "non-past" tense. Yang (2016:152f.) provides a brief Tolerance-based account of this gap, as follows. Second conjugation verb stems ending in *t* undergo one of two mutations in the 1sg. nonpast (in addition to the *-u* suffix); they may either become the affricate *č* [tʃʲ] (e.g., *metit'-meču* 'mark') or a alveopalatal fricative *šč* [ɕː] (*sokratit'-sokrašču* 'reduce in number'). While the former is the majority pattern, neither emerges as a productive default under the Tolerance Principle, and thus *t*-stems of the second conjugation appear to satisfy the condition on gaps.

Pertsova (2016) notes that this account is incomplete, for it fails to account for gaps in *d*, *s*, *z*-stems of the second conjugation. For these verbs, it is more difficult to construct a competition-based account simply because all three are generally thought to each undergo a exactly one mutation: *d* and *z* to *ž* [ʐ] (e.g., *sidet'-sižu* 'sit', *vonzit'-vonžu* 'pierce, stab') and *s* to *š* [ʂ] (*kosit'-košu* 'scythe'). However, we will argue there is some evidence for residual competing generalizations.

Sims (2006:197f.) and Baerman (2008) provide a history of the 1sg. non-past mutations. The modern mutations to *š* and *ž* represent the expected Russian reflexes of Common Slavic *tʲ, *dʲ, respectively. Christianization, beginning at the close of the first millennium, brought about a period of substantial contact with southern Slavic speakers, and their liturgical language, Old Church Slavonic (OCS), contributed novel reflexes of *tʲ, *dʲ, namely *č* [tʃʲ] and *žd* [ʐd]. The OCS reflexes were found in, among other contexts, the 1sg. non-past—where they competed with the native mutations—and the past passive participle, where they were largely entrenched. Ultimately, *č* persisted in the 1sg. non-past but *žd* was driven out sometime in the early 20th century (Baerman 2008:85); however, the latter persists in past passive participles (e.g., *rodit'-roždënny* 'give birth').

We propose, novelly, that the OCS *žd* is still a grammatical competitor for the 1sg. non-past mutation, and thus there is a morphophonological competition in *d*-stems just as in in *t*-stems. Beyond the aforementioned fact that this mutation is present elsewhere in the grammar (i.e., in certain past passive participles), there is some behavioral evidence that it is synchronically active, despite the

[17] The only clear exception is the labial stem *zatmit'* 'to eclipse'. Moskvin (2015, cited in Pertsova 2016) argues this gap is due to phonotactic illformedness.

fact it is rarely if ever found in the written standard. Sims (2006: chap. 6) administers a cloze task in which native speakers are asked to produce the 1sg. of a (defective) verb shown in the infinitive: several participants select the supposedly-extinct OCS mutation (e.g., *ubedit'*-**ubeždu* 'convince'). Slioussar and Kholodilova (2011) find that this OCS mutation is used with verbs borrowed into the 2nd conjugation in a corpus of web text. For instance, the English borrowing *zafrendit'* 'to include on one's friend list' exhibits variation between the expected 1sg. *zafrenžu* and *zafrenždu*. Together, these suggest that the OCS alternation is synchronically active in the 1sg. non-past, even if it is not prescribed for any native verbs. We propose the following two rules for *t*, *d*-stem mutations.[18]

$$(10)\quad \begin{Bmatrix} \text{t} \\ \text{d} \end{Bmatrix} \longrightarrow \begin{Bmatrix} \text{ɕ} \\ \text{ʐ} \end{Bmatrix} \;/\; \_\_\; [+1, -\text{Pl}, -\text{Past}] \;\; (\text{Condition: } \sqrt{\text{sokrat-}}, \sqrt{\text{sid-}}, \ldots)$$

$$(11)\quad \begin{Bmatrix} \text{t} \\ \text{d} \end{Bmatrix} \longrightarrow \begin{Bmatrix} \text{tʃʲ} \\ \text{ʐd} \end{Bmatrix} \;/\; \_\_\; [+1, -\text{Pl}, -\text{Past}] \;\; (\text{Condition: } \sqrt{\text{met-}}, \sqrt{\text{lit-}}, \ldots)$$

The condition on lexical gaps can then be directly applied to the *t*, *d*-stems. Table 4, adapted from table 3 of Pertsova 2016, gives frequencies of the competing dental stop mutations.[19] The first mutation (10) is the majority pattern but it cannot tolerate the 39 exceptions represented by OCS-mutation stems ($\theta_{172} \approx 33$).

This account cannot be stretched any further to cover 2nd conjugation stems ending in the dental fricatives *s* and *z*, which also exhibit gaps, simply because there is no inherited OCS competitor. However, there is some evidence that a competitor has arisen. Slioussar and Kholodilova (2011) identify a substantial number of English verbs nativized as *s*, *z*-stem 2nd conjugation verbs; a few examples are given below.

[18]Pertsova (2016) distinguishes between *t*-stems and *st*-stems, presumably because the *s* of the latter is never present in the 1sg. non-past; e.g., *vyrasit'*-*vyrašču* 'raise, cultivate', but **vyra*[stʃʲ]*u*. However, this can be handled with an additional—surface-true—phonological rule simplifying the resulting sibilant cluster, as [stʃʲ] is not a valid onset in Russian (Vitaly Nikolaev, p.c.).

[19]As in Spanish, many Russian verbs are derived from prefixation to stems, but with rare exceptions, all derivatives of a stem undergo the same mutation in the 1sg. non-past. Therefore, Pertsova (2016) collects counts over stems rather than full verbs.

(s)tʲ			
	⟶	ɕː	39
	⟶	tʃʲ	58
dʲ			
	⟶	ʐ	75
	⟶	ʐd	0

Table 4: Counts of competing stem-final mutations in Russian 2nd conjugation dental verb stems.

(12) fiksit' 'fix'

kapsit' 'use capital letters'

brauzit' 'browse'

rejzit' 'raise'

Slioussar and Kholodilova find that for many of these borrowed verbs (including those examples above), the expected mutation (e.g., *kapšu*) is rarely applied in a corpus of Russian web text; instead, speakers apply no mutation at all (*kapsju*). We do not have any explanation for the emergence of this pattern (though see Sims 2006:200 for the suggestion that non-alternation in dental stems was also present in pre-revolutionary Russian), and we do not have comparable lexical counts for this pattern. However, we do note that the conventional 1sg. non-past mutations ($s \rightarrow š$, $z \rightarrow ž$) are themselves sufficiently rare that it would not take many non-alternating borrowings for this context to satisfy the condition on lexical gaps.[20] We tentatively propose that may be exactly what has occurred.

To summarize, we argued that gaps in the 1sg. non-past of *t*, *d*-stems are predicted by the condition on lexical gaps. While it is more difficult to relate this to well-known synchronic processes—at least those documented by standard registers—we also suggest that gaps in *s*, *z*-stems may also result from competition between two morpholexical rules. However, these conclusions must remain tentative until confirmed by other sources of evidence, e.g., errors in child language acquisition.

[20] Pertsova (ibid.) counts 38 non-defective $s \rightarrow š$ stems and 36 non-defective $z \rightarrow ž$ stems. Assuming these are related by a single rule of retroflexion, as seems likely, they can tolerate no more than 20 non-alternating exceptions.

6 Conclusions

Above we have shown how the Tolerance Principle can, from a numerical basis, detect productive generalizations as well as predict the existence of lexical gaps in those cases for which no productive generalizations can be found. However, we note that the lack of a productive generalization is only one of many pre-existing conditions by which a word can be afflicted; the principle has nothing to say about, e.g., taboo avoidance, phonotactic illformedness, or why one language realizes in a single phonological word what requires periphrasis in another. It can only be invoked in cases where gaps are associated with competing, lexically-specific linguistic generalizations. While the cases above are all essentially morphophological, the principle has also been applied in syntax (Yang 2016:191f.) and phonology (Yang 2017).

The Tolerance Principle is extremely general, but the principle, and the child language learner it models, need precise statements of these generalizations and their structural properties so as to tabulate the sets of target lexical items (N) and exceptions (e). Both depend upon a well-defined hypothesis space to determine what is to be counted. Furthermore, the quantitative analyses above depend in part on the lexical resources used. Had we operationalized other analyses, or used other lexical resources, we might have reached slightly different conclusions. Therefore we encourage others to reproduce of our results and to apply the principle to novel cases.

We also stress that the Tolerance Principle is simply the consequence of a theoretically and empirically motivated processing model, and is itself not a model of how productivity is encoded in the grammar. While the mere existence of lexical gaps is seen as an uncomfortable fact for many theories of grammar, the proposal here is applicable to any grammatical architecture in which the application of generalizations conforms to the Elsewhere Condition. Morphological productivity has, for example, often been linked to the distinction between inflectional and derivational morphology. Under one account (see Spencer 1991 *passim* for a review) the former is the result of a productive "narrow" syntactic computation whereas the latter is the result of a less-productive pre-syntactic ("lexical") computation. But as Halle (1973) already notes, the very existence of gaps in inflection show that productivity cannot be a definitive characteristic of the inflection/derivation dichtomy once we recognize the connection between unproductivity and gaps. Under the the present account, productivity is a property inferred from the primary linguistic data by the child language learner. While the grammar induced by the child must encode this distinction, a theory of this inference can be developed independently from theories of the encoding of productivity.

Learning under the Tolerance Principle proceeds conservatively, requiring considerable evidence before a generalization is ensconced. But by no means is this the only trajectory in language acquisition. In syntactic development, in particular, the learner may spontaneously, and productively, access non-target but linguistically possible options such as those in the parametric space of variation, which are then gradually eliminated as disambiguating evidence from the input accumulates over time (Yang 2002, 2010). The learning of morphology decidedly does not proceed this way: the hypothesis space for past tense formation is **not** one which allows the child to pick from either the English *-ed* or the French *-é*. These different patterns of acquisition may reflect familiar dichotomies such as functional vs. lexical heads, and innate parameter spaces vs. induced language-specific generalizations.

The Tolerance Principle, while applied narrowly here, is fundamentally an evaluation metric using processing time to rank competing grammars. Thus it represents an concrete proposal for the role of computational efficiency—the "third factor" (Chomsky 2005)—in the design of language.

References

Albright, Adam. 2003. A quantitative study of Spanish paradigm gaps. In *Proceedings of the 22th West Coast Conference on Formal Linguistics*, 1–14. Somerville, MA: Cascadilla.

Albright, Adam. 2009. Lexical and morphological conditioning of paradigm gaps. In *Modeling ungrammaticality in Optimality Theory*, ed. Sylvia Blaho and Curt Rice, 117–164. London: Equinox.

Albright, Adam, Argelia Andrade, and Bruce Hayes. 2001. Segmental environments of Spanish diphthongization. In *UCLA Working Papers in Linguistics 7: Papers in Phonology 5*, 117–151. University of California, Los Angeles.

Albright, Adam, and Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90:119–161.

Anderson, Stephen R. 1969. West Scandinavian vowel systems and the ordering of phonological rules. Doctoral dissertation, MIT.

Anderson, Stephen R. 2010. Failing one's obligations: Defectiveness in Rumantsch reflexes of DĒBĒRE. In *Defective paradigms: Missing forms and what they tell us*, ed. Matthew Baerman, Greville G. Corbett, and Dunstan Brown, 19–34. Oxford: Oxford University Press.

Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge: MIT Press.

Baayen, R. Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In *Corpus linguistics: An international handbook*, ed. Anke Ludeling and Merja Kyto, 899–919. Mouton de Gruyter.

Baerman, Matthew. 2008. Historical observations on defectiveness: The first singular non-past. *Russian Linguistics* 32:81–97.

Baerman, Matthew, and Greville G. Corbett. 2010. Defectiveness: Typology and diachrony. In *Defective paradigms: Missing forms and what they tell us*, ed. Matthew Baerman, Greville G. Corbett, and Dunstan Brown, 19–34. Oxford: Oxford University Press.

Baerman, Matthew, Greville G. Corbett, and Dunstan Brown, ed. 2010. *Defective paradigms: Missing forms and what they tell us*. Oxford: Oxford University Press.

Baronian, Luc. 2005. North of phonology. Doctoral dissertation, Stanford University.

Bauer, Laurie. 2001. *Morphological productivity*. Cambridge: Cambridge University Press.

Berko, Jean. 1958. The child's learning of English morphology. *Word* 14:150–177.

Boyé, Gilles, and Patricia Cabredo Hofherr. 2010. Defectivity as stem suppletion in French and Spanish verbs. In *Defective paradigms: Missing forms and what they tell us*, ed. Matthew Baerman, Greville G Corbett, and Dunstan Brown, 35–52. Oxford: Oxford University Press.

Brame, Michael K, and Ivonne Bordelois. 1973. Vocalic alternations in Spanish. *Linguistic Inquiry* 4:111–168.

Brown, Roger. 1973. *A first language: The early stages*. Cambridge: Harvard University Press.

Butt, John, and Carmen Benjamin. 1988. *A new reference grammar of Modern Spanish*. London: Edward Arnold.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge: MIT Press.

Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36:1–22.

Clahsen, Harald, Fraibet Aveledo, and Iggy Roca. 2002. The development of regular and irregular verb inflection in spanish child language. *Journal of Child Language* 29:591–622.

Dąbrowska, Ewa. 2001. Learning a morphological system without a default: The Polish genitive. *Journal of Child Language* 28:545–574.

Dąbrowska, Ewa. 2005. Productivity and beyond: Mastering the Polish genitive inflection. *Journal of Child Language* 32:191–205.

Eddington, David. 1996. Diphthongization in Spanish derivational morphology: An empirical investigation. *Hispanic Linguistics* 8:1–13.

Fanselow, Gisbert, and Caroline Féry. 2002. *Resolving conflicts in grammars: Optimality Theory in syntax, morphology, and phonology*. Hamburg: Helmut Buske.

Halle, Morris. 1973. Prolegomena to a theory of word formation. *Linguistic Inquiry* 4:3–16.

Halle, Morris, and Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, ed. Kenneth Hale and Samuel Keyser, 111–176. Cambridge: MIT Press.

Harris, James W. 1969. *Spanish phonology*. Cambridge: MIT Press.

Harris, James W. 1985. Spanish diphthongisation and stress: A paradox resolved. *Phonology* 2:31–45.

Hetzron, Robert. 1975. Where the grammar fails. *Language* 51:859–872.

Hill, Archibald A. 1976. [Obituary:] Albert Henry Marckwardt. *Language* 52:667–681.

Hudson, Richard. 2000. *I amn't. *Language* 76:297–323.

Kiparsky, Paul. 1973. “Elsewhere” in phonology. In *A festschrift for Morris Halle*, ed. Stephen R. Anderson and Paul Kiparsky, 93–106. New York: Holt, Rinehart and Winston.

Kottum, Steiner S. 1981. The genitive singular form of masculine nouns in Polish. *Scando-slavica* 27:179–186.

Lerner, Marielle. 2011. Spanish 3rd conjugation verb stem alternations. Ms., University of Pennsylvania.

MacWhinney, Brian. 2000. *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates, 3rd edition.

Maiden, Martin, and Paul O’Neill. 2010. On morphomic defectiveness: Evidence from the Romance languages of the Iberian peninsula. In *Defective paradigms: Missing forms and what they tell us*, ed. Matthew Baerman, Greville G. Corbett, and Dunstan Brown, 103–124. Oxford: Oxford University Press.

Marcus, Gary. 1995. Children’s overregularization of English plurals: A quantitative analysis. *Journal of Child Language* 22:447–459.

Mateo, Francis, and Antonio Rojo Sastre. 1995. *El arte de conjugar en español*. Paris: Hatier.

Maunsch, Hanna. 2003. Current alternations in inflection of Polish masculine inanimate nousn in the singular: A pilot study. *Investigationes Linguisticae* 9:4–21.

Mayol, Laia. 2007. Acquisition of irregular patterns in Spanish verbal morphology. In *Proceedings of the Twelfth ESSLLI Student Session*, 1–11.

Moskvin, Vadim. 2015. Nedostatochnye glagoly: k utochneniju pon’atija. *Russian Linguistics* 39:209–232.

Neue, Friedrich. 1866. *Formenlehre der lateinischen Sprache*. Stuttgart: H. Lindemann.

Newport, Elissa L. 1990. Maturational constraints on language. *Cognitive Science* 14:11–28.

Orgun, Orhan C., and Ronald Sprouse. 1999. From MPARSE to CONTROL: Deriving ungrammaticality. *Phonology* 16:191–224.

Pertsova, Katya. 2005. How lexical conservatism can lead to paradigm gaps. In *UCLA Working Papers in Linguistics 11: Papers in Phonology 6*, 13–30. University of California, Los Angeles.

Pertsova, Katya. 2016. Transderivational relations and paradigm gaps in Russian verbs. *Glossa* 1:13.

Pinker, Steven. 1999. *Words and rules: The ingredients of language*. New York: Basic Books.

Prince, Alan, and Paul Smolensky. 1993. Optimality theory: Constraint interaction in generative grammar. Technical Report TR-2, Rutgers University Center For Cognitive Science and Technical Report CU-CS-533-91, University of Colorado, Boulder Department of Computer Science. Published by MIT Press, Cambridge, 2004.

Pullum, Geoffrey K, and Deidre Wilson. 1977. Autonomous syntax and the analysis of auxiliaries. *Language* 53:741–788.

Raffelsiefen, Renate. 1996. Gaps in word formation. In *Interfaces in phonology*, ed. Ursula Kleinhenz, 194–209. Berlin: Akademie-Verlag.

Raffelsiefen, Renate. 1999. Phonological constraints on English word formation. In *Yearbook of Morphology 1998*, ed. Geert Booij and Jaap van Marle, 225–287. Dordrecht: Kluwer.

Real Academia Española. 1992. *Diccionario de la lengua española*. Madrid: Real Academia Española, 21st edition.

Rebrus, Péter, and Miklós Törkenczy. 2009. Covert and overt defectiveness in paradigms. In *Modeling ungrammaticality in Optimality Theory*, ed. Sylvia Blaho and Curt Rice, 195–234. London: Equinox.

Rice, Curt. 2005. Optimal gaps in optimal paradigms. *Catalan Journal of Linguistics* 4:155–170.

Rice, Curt, and Sylvia Blaho, ed. 2009. *Modeling ungrammaticality in Optimality Theory*. London: Equinox.

Roberts, Edward A. 2014. *A comprehensive etymological dictionary of the Spanish language with families of words based on Indo-European roots*. Bloomington, IN: Xlibris. 2 volumes.

Schuler, Kathryn D., Charles Yang, and Elissa L. Newport Newport. 2016. Testing the Tolerance Principle: Children form productive rules when it is more computationally efficient to do so. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society*, 2321–2326.

Sebastián, Núria, Fernando Cuetos, Antònia Martí, and Manuel Carreiras. 2001. *LEXESP: Léxico informatizado del español*. Barcelona: Edicions de la Universitat de Barcelona.

Sims, Andrea. 2006. Minding the gap: Inflectional defectiveness in a paradigmatic theory. Doctoral dissertation, Ohio State University.

Sims, Andrea. 2007. Why defective paradigms are, and aren't, the result of competing morphological patterns. In *Proceedings of the 43rd Annual Meeting of the Chicago Linguistic Society*, 267–281. University of Chicago.

Slioussar, Natalia, and Maria Kholodilova. 2011. Paradigm leveling in non-standard Russian. In *Proceedings of the 20th meeting of Formal Approaches to Slavic Linguistics*, 243–258.

Smoczyńska, Magdalena. 1985. The acquisition of Polish. In *The crosslinguistic study of language acquisition. Volume 1: The data*, ed. Daniel I. Slobin, 595–686. Hillsdale, NJ: Erlbaum.

Smolensky, Paul. 1996. The initial state and 'richness of the base' in Optimality Theory. Technical Report JHU-CogSci-96-4, Department of Cognitive Science, Johns Hopkins University.

Spencer, Andrew. 1991. *Morphological theory: An introduction to word structure in generative grammar*. Oxford: Basil Blackwell.

Steriade, Donca. 1997. Lexical conservatism. In *Linguistics in the morning calm: Selected papers from SICOL 1997*, 157–179. Seoul: Hanshin.

Steriade, Donca. 2008. A pseudo-cyclic effect in Romanian morphophonology. In *Inflectional identity*, ed. Asaf Bachrach and Andrew Nevins, 313–360. Oxford: Oxford University Press.

Weist, Richard M., and Katarzyna Witkowska-Stadnik. 1986. Basic relations in child language and the word order myth. *International Journal of Psychology* 21:363—381.

Weist, Richard M., Hanna Wysocka, Katarzyna Witkowska-Stadnik, Ewa Buczowska, and Emilia Konieczna. 1984. The defective tense hypothesis: On the emergence of tense and aspect in child Polish. *Journal of Child Language* 11:347—374.

Wolf, Matthew, and John McCarthy. 2009. Less than zero: Correspondence and the null output. In *Modeling ungrammaticality in Optimality Theory*, ed. Sylvia Blaho and Curt Rice, 17–66. London: Equinox.

Xu, Fei, and Steven Pinker. 1995. Weird past tense forms. *Journal of Child Language* 22:531–556.

Yang, Charles. 2002. *Knowledge and learning in natural language*. Oxford: Oxford University Press.

Yang, Charles. 2005. On productivity. *Linguistic Variation Yearbook* 5:333–370.

Yang, Charles. 2010. Three factors in language variation. *Lingua* 120:1160–1177.

Yang, Charles. 2016. *The price of linguistic productivity*. Cambridge: MIT Press.

Yang, Charles. 2017. How to wake irregular (and speechless). In *On looking into words (and beyond): Structures, relations, analyses*, ed. Claire Bowern, Laurence Horn, and Raffaela Zanuttini, 211–232. Language Science Press.

Yip, Kenneth, and Gerald J. Sussman. 1997. Sparse representations for fast, one-shot learning. In *Proceedings of the National Conference on Artificial Intelligence*, 521–527.