Proceedings of the LSA

Reconciling diverging frequency effects in inflectional defectiveness: A study of Peninsular Spanish verbs

Charlie Rowe, Maria Copot & Andrea D. Sims*

Abstract. In his study of defective Spanish verbs, Albright (2003) found speakers' form judgments to be positively correlated with lexeme frequency. For defective French verbs, however, speakers' judgments are negatively correlated with lexeme frequency (Copot & Sims 2025). What is the source of these diverging observations? We test the role of methodological choices by using Copot & Sims's experimental design to investigate Spanish verbal defectiveness. The results confirm Albright's findings, but only for the subset of verbs drawn from his study. We briefly consider implications for the phenomenology of defectiveness.

Keywords. morphology; defectiveness; paradigm gaps; word frequency; Spanish

1. Introduction. In this paper, we examine the role that lexeme frequency plays in inflectional defectiveness, i.e., situations in which a lexeme has no inflected form that speakers will accept as the realization of some well-defined paradigm cell. Previous studies have observed opposite relationships between frequency and defectiveness. In a handful of languages, including Spanish (Albright 2003), speakers' judgments of wordforms that are candidates to fill defective cells (i.e. gap-filling forms) have been found to be positively correlated with frequency: gap-filling forms belonging to higher frequency lexemes are rated better. At the same time, French speakers' judgments of gap-filling forms have been found to be negatively correlated with frequency (Copot & Sims 2025): gap-filling forms belonging to higher frequency lexemes are rated worse. What is the source of this difference? Does the relationship between frequency and defectiveness simply vary from one case of defectiveness to another? While this seems possible, such a conclusion is premature. Previous studies differ not only in the language of investigation, but also in methodology. This makes it challenging to know whether the two frequency effects reflect something truly different about the relevant cases of defectiveness or are perhaps simply a methodological artifact.

To investigate this issue, we use Spanish verbs to test the role of methodological choices in the observed frequency effects. Specifically, we use the task and experimental design from Copot & Sims's French study to investigate Spanish verbal defectiveness. We draw experiment items from both Albright's original study and from the Real Academia Española (RAE) grammar (2009) and dictionary (2014), the latter following the item selection procedure in Copot & Sims's study. Defective Spanish verbs offer a good test case because there are different accounts of the causes of their defectiveness, with implications for the predicted relationship to lexeme frequency.

* This project was funded by an Undergraduate Research Scholarship awarded by the OSU College of Arts & Sciences to Charlie Rowe, an OSU President's Postdoctoral Scholarship and a British Academy International Fellowship (IF24\100591), the latter two awarded to Maria Copot. An OSU Arts & Humanities Undergraduate Research Small Grant made it possible for Charlie Rowe to present this paper at the LSA2026 meeting. We are grateful for this support. Thanks also to the OSU Linguistics Department, the Surrey Morphology Group, and the LSA2026 audience for their feedback on this work in progress. We thank Prof. Manuel Badal for his help in ensuring all the materials in the experiment conformed to Peninsular Spanish grammar. All errors remain our own. Authors: Charlie Rowe, The Ohio State University (charlieroweclm@gmail.com) & Maria Copot, Surrey Morphology Group, University of Surrey (m.copot@surrey.ac.uk) & Andrea D. Sims, The Ohio State University (sims.120@osu.edu).

The results show minimal overall effect of lexeme frequency. However, a post-hoc analysis shows that for the subset of items drawn from Albright's study, speakers' judgments of potentially defective words are positively correlated with lexeme frequency. This replicates Albright's findings and suggests that the different frequency effects found previously for Spanish and French are likely not a result of the experimental designs employed. In contrast, the items drawn from the grammar and dictionary sources show neutral to negative frequency effects, depending on the task. Thus, despite the defective verbs being structurally similar, their acceptability seems to be governed by different principles, and the results suggest that lexeme frequency plays different roles for different items. We argue that this is not as surprising a result as it might seem at first glance and that the contrasting frequency patterns for the two subsets of Spanish verbs can, in fact, be reconciled through attention to the sources of evidence that speakers have for a lexeme's defectiveness and the role of language authorities in shaping speakers' evaluations of their language. More broadly, the results in this paper point to the phenomenological variety that is to be found under the umbrella of defectiveness: even within one language, and with maximally similar structural conditions, different defective items diverge in the factors that underpin their acceptability to speakers.

2. Spanish verbal defectiveness. In this section we begin by laying out some relevant facts about stem allomorphy in Spanish verbs and its relationship to defectiveness.

2.1. STRUCTURAL FACTS. Spanish verbs exhibit two relevant patterns of stem alternation: mid-vowel diphthongization and velar alternation. As illustrated in Table 1, in verbs with an underlying /o/ or /e/ in the final syllable of the stem, the vowel may diphthongize under stress, in other words, in 1SG, 2SG, 3SG, and 3PL in both present indicative and present subjunctive.[1] If the vowel is /e/, it may raise to [i] instead. Maiden (2009) calls this distribution of alternants the "N-pattern." Third conjugation verbs with mid-vowels additionally have vowel raising in the 1PL and 2PL present subjunctive.[2] (The stem vowel in these cells is not stressed.)

	PRS.IND	PRS.SBJV	PRS.IND	PRS.SBJV	PRS.IND	PRS.SBJV
1SG	d[we]rmo	d[we]rma	s[je]nto	s[je]nta	p[i]do	p[i]da
2SG	d[we]rmes	d[we]rmas	s[je]ntes	s[je]ntas	p[i]des	p[i]das
3SG	d[we]rme	d[we]rma	s[je]nte	s[je]nta	p[i]de	p[i]da
1PL	d[o]rmimos	d[u]rmamos	s[e]ntimos	s[i]ntamos	p[e]dimos	p[i]damos
2PL	d[o]rmís	d[u]rmáis	s[e]ntís	s[i]ntáis	p[e]dís	p[i]dáis
3PL	d[we]rmen	d[we]rman	s[je]nten	s[je]ntan	p[i]den	p[i]dan

Table 1. Partial paradigms of DORMIR 'sleep,' SENTIR 'feel,' and PEDIR 'ask for,' showing N-pattern stem vowel diphthongization or raising (dark shading) and also raising in the remaining present subjunctive cells (light shading)

While N-pattern diphthongization – and to lesser extent, N-pattern raising – can in principle occur in any conjugation class, relatively few verbs in the first conjugation alternate: for example,

[1] Abbreviations in this paper follow the Leipzig Glossing Rules: IND = 'indicative', PL = 'plural', PRS = 'present', SBJV = 'subjunctive', SG = 'singular', 1/2/3 = 'first/second/third person.'

[2] The raising pattern of PEDIR 'ask' can also be included in the N-pattern, albeit as a stealth case, since raising in the N-pattern cells coincides with the raising expected in the 1PL and 2PL present subjunctive.

CONTAR ‘count’ and PENSAR ‘think’ diphthongize (1SG.PRS.IND *cuento*, *pienso*), but TOMAR ‘take’ and ESPERAR ‘hope’ do not (*tomo*, *espero*). In contrast, second conjugation and even more so third conjugation verbs (as in Table 1) have a high rate of alternation.

Table 2 illustrates the velar alternation pattern. In the second and third conjugations, verbs with stems ending in a sonorant or /θ/ (in Peninsular Spanish, /s/ in Latin American Spanish) may either have a velar stop inserted at the end of the stem (as in CRECER ‘grow’) or have it replace the stem-final consonant (as in DECIR ‘say’) in the 1SG present indicative and all present subjunctive cells. Maiden (2009) calls this the “L-pattern.” However, again, not all relevant verbs undergo velar alternation (cf. MECER ‘rock, cradle,’ with 1SG.PRS.IND *me*[θ]*o*, not **me*[θk]*o* or **me*[g]*o*).

	PRS.IND	PRS.SBJV	PRS.IND	PRS.SBJV
1SG	cre[θk]o	cre[θk]a	di[g]o	di[g]a
2SG	cre[θ]es	cre[θk]as	di[θ]es	di[g]as
3SG	cre[θ]e	cre[θk]a	di[θ]e	di[g]a
1PL	cre[θ]emos	cre[θk]amos	de[θ]imos	di[g]amos
2PL	cre[θ]éis	cre[θk]áis	de[θ]ís	di[g]áis
3PL	cre[θ]en	cre[θk]an	di[θ]en	di[g]an

Table 2. Partial paradigms of CRECER ‘grow’ and DECIR ‘say,’ showing velar alternation (shaded cells)

The defective verbs that are the focus of this paper follow the distribution of these alternations. As Table 3 shows, the defective cells in the group represented here by ABOLIR ‘abolish’ can be interpreted as the union of the N-pattern and L-pattern (Maiden 2009).[3] Another group of verbs, represented by BALBUCIR ‘stammer,’ follows (only) the L-pattern. In both cases, the defective verbs fall into the third conjugation (RAE 2009).

	PRS.IND	PRS.SBJV	PRS.IND	PRS.SBJV
1SG	–	–	–	–
2SG	–	–	balbu[θ]es	–
3SG	–	–	balbu[θ]e	–
1PL	ab[o]limos	–	balbu[θ]imos	–
2PL	ab[o]lís	–	balbu[θ]ís	–
3PL	–	–	balbu[θ]en	–

Table 3. Partial paradigms of ABOLIR ‘abolish’ and BALBUCIR ‘stammer,’ showing defective cells (shaded)

[3] While ABOLIR is widely cited as the prototypical example of a defective Spanish verb, it appears to be in the process of becoming non-defective, with the traditionally defective cells filled by non-alternating forms, e.g., 1SG.PRS.IND *abolo* (RAE 2009, Zacarías Ponce de León 2021). This ongoing shift, and the processes behind it, are interesting in themselves, but they are beyond the scope of the present study other than to observe that all of the absent cells are seemingly becoming available together as a stem form becomes conventionalized. This further demonstrates that defectiveness is related to stem alternation. At the same time, speakers’ judgments about ABOLIR did not notably differentiate this verb from other defective verbs in our study.

In the following section we review what previous studies of defectiveness suggest about the role of lexeme frequency in defectiveness.

3. The relationship between defectiveness and lexeme frequency. Albright (2003) treats the Spanish pattern of defectiveness as a byproduct of the uncertainty involved in predicting whether a verb has stem alternation. In Albright's Minimal Generalization Learner (MGL, Albright & Hayes 2003), the reliability of some inflectional rule is the basis for its productivity. Rules that apply more reliably within their conditioning environment and which have broader scope (i.e. apply to more lexemes) are predicted to be more productive. However, the probabilistic nature of the competition this creates between rules also means that it is possible for there to be morphophonological contexts in which *no* rule is sufficiently productive as to apply with confidence. Albright proposes that such cases produce defectiveness,[4] as part of a story tying defectiveness to speaker uncertainty: since memory strength is highly frequency sensitive, weak memory representations of low frequency forms require speakers to predict these, rather than drawing them from memory. When the need to predict a form coincides with indeterminate grammar, speakers may be uncertain about what the form should be and may choose to avoid using any form at all. For Spanish specifically, Albright argues that for some verbs, the rule system does not reliably predict whether the verb will have diphthongization/raising, velar insertion, both alternations (see DECIR in Table 2 above), or neither of them, and defectiveness is the result.

Albright's account thus predicts that defectiveness will disproportionately affect low frequency verbs, and this is what he found for Spanish verbs. In a production and rating study of 1SG present indicative forms (e.g. *abolo* 'abolish.1SG'), he found that when the MGL predicts the grammar to be indeterminate – i.e. when it predicts that for some combination of lexeme and paradigm cell, no potential inflected form is sufficiently productive – speakers were less likely to agree about what the inflected form for the cell should be, and they reported lower confidence in their own productions. However, these judgments were subject to a strong effect of lexeme frequency, with items belonging to low frequency lexemes receiving lower ratings. Interpreting this lack of both agreement and confidence as indicative of defectiveness, Albright concludes that when defectiveness is structurally possible, the likelihood of a Spanish verb having paradigm gaps is higher for low frequency verbs. Other studies have made similar observations, including for at least Russian verbs (Pertsova 2016; Pertsova & Kuznetsova 2015; Sims 2006), Modern Greek nouns (Sims 2015), and Finnish verbs (Nikolaev & Bermel 2022).

While there is strong evidence that speaker uncertainty plays a role in some cases of defectiveness, it is equally clear that the factors underpinning defectiveness are diverse (Baerman & Corbett 2010; Sims 2023). It is therefore interesting that Copot & Sims (2025) find the opposite frequency effect for defective French verbs (e.g. CLORE 'close'). In a task in which French speakers were asked to rate gap-filling verb forms either normatively (i.e., according to "correctness") or according to how normal it would be to hear the form used in conversation, Copot & Sims found that for verbs conventionally cited in French dictionaries and grammars as defective, and thus proscribed in use, speakers gave lower ratings to gap-filling items belonging to higher frequency lexemes. Additionally, speakers who reported higher rates of agreement with prescriptive language discourses rated defective items lower, especially in the normative task. Interpreting higher frequency as a likely correlate to higher metalinguistic awareness of the items' prescriptive stigmatization, Copot & Sims conclude that anti-variationist pressures in French language

[4] See also Gorman & Yang (2019) for a variant of this approach based on the Tolerance Principle.

culture cause speakers who align themselves with standard language ideologies to disprefer forms that are (or may be) prescriptively taboo, especially in situations where the risk of social sanction is high.

Turning back to Spanish, there are hints that a similar pressure may also be at work in this language. Boyé & Cabredo Hofherr (2010:39) argue that the Spanish verbal gaps cannot be entirely reduced to uncertainty regarding stem alternation, in part because mid-vowel raising in the 1PL and 2PL present subjunctive (see Table 1 above) is predictable for third conjugation verbs. Since the defective verbs have gaps also in these cells, a purely uncertainty-based account is not fully sufficient.[5] They instead posit that the required (and suppletive) stem allomorphs are lexicalized as being absent.[6] Moreover, they suggest that form uncertainty is not a problem inherently, and that avoidance "...seems to be caused by a resistance to regularization for sociolinguistic reasons" (Boyé & Cabredo Hofherr 2010:49). They do not consider implications for how defectiveness might interact with lexeme frequency. However, this suggestion echoes the account that Copot & Sims (2025) offer for French verbal defectiveness. So, to the extent that "sociolinguistic reasons" refers to social stigmatization, we might expect the same kind of frequency effect that Copot & Sims found, namely, one in which speakers' judgments of gap-filling forms are lower for high frequency verbs. In short, despite Albright's findings, there are hints that the relationship between lexeme frequency and defectiveness is not a fully settled question in Spanish.

Opposite frequency effects, each with its own explanatory story connecting frequency to defectiveness, have thus been documented cross-linguistically, and different predictions can be derived for Spanish. As Copot & Sims (2025:1330) observe, both types of frequency effect could be valid, but for different kinds of defectiveness. Yet it is premature to conclude that this *is* the source of the differing frequency effects. While both Albright's and Copot & Sims's studies involved speaker judgments, they differed in the language of investigation, the specifics of the tasks (e.g. the criteria that speakers were asked to use to rate gap-filling forms), and perhaps most notably, the method for identifying potentially defective items: the items in Albright's study were verbs predicted to have no high confidence inflected form, while Copot & Sims's items were verbs listed as defective in grammars and dictionaries. Did any of these differences impact the study results, and in particular, the direction of the frequency effect? More generally, what factors determine whether there is a positive or negative frequency effect in any given case of defectiveness? To get some clarity on this issue, we sought to put the studies on more comparable footing by using Copot & Sims' methods to study Spanish verbal defectiveness.

4. Research questions. This study investigates the source of diverging frequency effects documented in the defectiveness literature. A first set of questions concerns potential methodological explanations: does the direction of the frequency effect vary as a function of experimental task type or the criterion used to identify defective items – specifically, selection based on modeling of grammar uncertainty versus dictionary-based classification? A related question is whether the diverging effects reported by Albright (2003) and Copot & Sims (2025) reflect genuine cross-linguistic differences, methodological incompatibility, or rather construct heterogeneity – that is, whether the two research traditions have been studying phenomenologically distinct objects under a shared label.

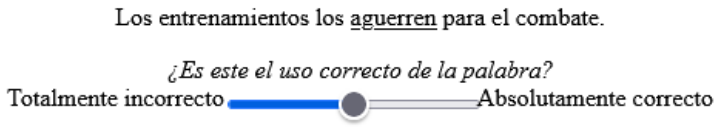
[5] For more discussion, see Sims (2015:65-69, 196-206, inter alia).

[6] See also Arregi & Nevins (2014:327) for a similar idea that the suppletive stem allomorphs required for the defective cells are undefined.

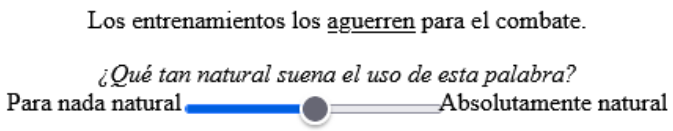
A second set of questions concerns the role of prescriptivism. We ask whether the effect found by Copot & Sims (2025) is unique to French, or whether it also emerges in Spanish when defective items are selected via dictionary classification. More broadly, we ask whether individual prescriptivist orientation modulates frequency effects in Spanish as it does in French, or whether the two languages differ in how individual speakers orient to institutional norms.

5. Methodology. To tackle the research questions, we performed an experiment on Spanish defective words replicating the methodology that Copot & Sims (2025) used for French. The experiment uses some items from Albright's (2003) original set, alongside other words identified as defective by the RAE (2009; 2014). The combination allows us to tease out whether the diverging frequency effects are to be attributed to the experimental design, the language, or the method for identifying defective items.

5.1. TASKS. Two between-subjects conditions manipulated whether participants applied prescriptivist filters when judging sentences with underlined target words. The normative condition elicited prescriptive judgments by invoking formal register and standard language ideology. Participants judged whether teachers would mark the usage as incorrect or whether it would appear in dictionaries/books. The possibility condition minimized prescriptivist framing by anchoring judgments in informal contexts (conversations at bars, students socializing). Participants judged whether they could plausibly hear the usage in casual speech. Both conditions used continuous scales with endpoints/prompts matched to their framing (Figure 1 shows identical stimuli under both framings). Instructions explicitly defined the judgment criterion and provided heuristics and examples.



(a) normativity formulation



(b) possibility formulation

Figure 1. The same practice item in the two task conditions

5.2. ITEMS. All participants saw the same items in randomized order. Each item consisted of a minimal sentence with an underlined verb form. The experiment features three item types: ungrammatical, slang, and defective. *Ungrammatical items* contained subject-verb agreement errors, providing a baseline for how speakers rate ill-formed structures. *Slang items* were colloquialisms marked as informal in major dictionaries, providing a baseline for proscribed but grammatically licit forms. *Defective items* were verbs of two types. First, we took Albright's

stimulus list and attempted to identify which lexemes from that list were intended to be defective. This entailed a degree of educated guesswork, since while Albright's paper gave a list of items used in the experiment, he treats defectiveness as gradient and the paper does not break out defective from non-defective items; nor does it provide item-level measures. We assumed a lexeme counted as defective if we could identify metalinguistic discourse surrounding difficulty with some of the word's forms. This yielded eight defective lexemes. We aimed for a total of fifteen defective lexemes, so we supplemented this set with items cited as having paradigm gaps in the RAE's new grammar (2009) and dictionary (2014). A number of defective types are cited; we focused on third conjugation i-stem verbs that are believed to be defective for morphophonological reasons, as described in the RAE grammar (§4.14d),[7] in order to keep the items maximally comparable to the set of defective items found in Albright. This produced seven lexemes that were not in Albright's list of test items, as well as four lexemes listed in both Albright's stimulus list and the RAE sources. The final set comprised fifteen items per condition. Of these items, three exhibited L-pattern defectiveness, and twelve exhibited N-pattern defectiveness. Lexemes were matched on log frequency across item conditions, since frequency effects follow power-law distributions. We chose to present these lexemes in present indicative or present subjunctive cells in which they were defective according to the RAE dictionary.[8] Information about individual items, their source, and their frequency can be found in the OSF repository. See the data availability statement at the end of this paper.

5.3. PROCEDURE. The experiment was administered in Spanish on PCIbex (Schwarz & Zehr 2018). After consent, participants were asked to answer nine questions about their personal orientation towards prescriptivist discourse in Spain. These questions targeted three kinds of prescriptivism identified by Curzan (2014): stylistic, standardizing, and restorative. Each question consisted of a prescriptivist statement, followed by a 5-point Likert scale where participants were instructed to signal their degree of agreement with the statement. These scores were averaged by participant, resulting in the participant's prescriptivism score (range 0-4). A full list of the prescriptivism questions is available on OSF. Participants were then randomly assigned to one of the two task conditions. Following three practice items, participants judged forty-five test items (fifteen per item condition) in randomized order. A post-task familiarity check presented infinitives of all experimental verbs plus an equal number of pseudoverbs generated with Wuggy (Keuleers & Brysbaert 2010), matching the real verbs' phonotactic structure, bigram frequency and length distribution. Participants were asked to select all words they knew. Pseudoverbs served both as fillers and reliability checks: participants reporting more than 20% pseudoverb familiarity were excluded. Participants then provided demographics and received debriefing.

The experiment median time was 13m30s. Participants were paid €2.50 for their time.

5.4. PARTICIPANTS. In total, 119 Spanish speakers were recruited on Prolific.[9] The participant pool was filtered for people currently residing in Spain and who identified as L1 Spanish speakers. In the experiment, participants were asked which cities had they lived in, and when in their daily lives they spoke Spanish (with friends? with family? at work?). No participants listed cities

[7] https://www.rae.es/gramtica/morfologa/verbos-irregulares-vi-verbos-defectivos - last accessed 03/23/2026.

[8] Eight cells feature in the experiment: PRS.IND.1PL (3), PRS.IND.1SG (6), PRS.IND.3PL (6), PRS.IND.3SG (11), PRS.SBJV.1PL (4), PRS.SBJV.2SG (4), PRS.SBJV.3PL (9), PRS.SBJV.3SG (2)

[9] https://www.prolific.com/ - last accessed 3/23/2026.

outside of Spain or reported speaking Spanish in less than two out of the three contexts given, so no participants were excluded based on Spanish language proficiency. However, we removed from analysis all responses from thirteen participants who claimed to know more than 20% of the pseudolexemes in the lexeme knowledge checklist, yielding a total of 106 participants included in the analysis (N = 53 for each of the tasks).

Of these, the gender breakdown of the sample is as follows: 51% identified as male, 46% as female and 3% as non-binary. The educational background of the sample was distributed as follows: 8% had not completed secondary education, 37% held a high school diploma, 13% possessed a vocational or associate-level qualification, 16% held a bachelors degree, and 26% held a masters degree or above.[10] The median prescriptivism score (possible range: 0-4) was 2.25, with an inter-quartile range of 1.56-2.67, a lowest value of 0 and a highest value of 3.87.

5.5. ANALYSES. In addition to the participant exclusions described above, we excluded responses to real verbs that participants did not report knowing (22% of trials). The excluded judgments were particularly concentrated within the defective lexemes: 45% of defective judgments had to be excluded, compared to 6% of slang judgments and 16% of ungrammatical judgments. The most well-known defective lexemes were ABOLIR 'abolish' and AGREDIR 'attack', which were known by everyone. The least well-known were EMBAÍR 'swindle' (only known by 8%) and PRETERIR 'omit' (known by 22%).

The analysis was performed in R (R Core Team 2021) using the `glmmTMB` package (Brooks et al. 2017). A maximal beta regression was fit to the responses. The dependent variable is the raw scores from the judgment task, all values between 0.00 and 1.00. Two models were fit, a core model and a second model to statistically test the difference in frequency effects between the different defective items as a function of their source.

6. Core analysis. The first analysis investigated the relationship between item condition, task condition, frequency, and prescriptivism, to gain a sense of whether the same dynamics found by Copot & Sims (2025) for French are at work in Spanish.

6.1. MODEL. The core model contained the following fixed effects.

Task condition Whether the participant was assigned to the normativity or possibility condition. The factor is treatment-coded (possibility: 0, normativity: 1).

Item condition Whether the item is defective, slang or ungrammatical. The factor is treatment-coded with the system of contrasts shown in Table 4.

	2	3
Ungrammatical	0	0
Slang	1	0
Defective	0	1

Table 4. Treatment coding for item factor levels

[10] Gender, age, and education were not included in the analyses for this paper, but see Rowe (2026) for additional analyses.

Lexeme frequency The frequency of the lexeme in CORPES, the RAE's corpus (RAE 2024); when lexemes were unattested in CORPES, the VeLeSpa corpus (Herce 2024) was used. The variable was first log-transformed (word frequency is known to have a power law distribution) then standardized.

Prescriptivism score A continuous variable, ranging from 0 to 4, indicating the extent to which the participant subscribes to prescriptive ideology. The variable was standardized.

The model included random intercepts for participant and item and has the following formula:

```
judgment ~ ItemCondition * TaskCondition * LogFreq * Prescriptivism
+ (1|participant) + (1|item)
```

6.2. RESULTS. Figure 2 shows raw trends by item and task condition. Ungrammatical items are rated at floor in both tasks conditions, slang items are rated at ceiling in the possibility condition but highly variably in the normative condition, showcasing the behavior of proscribed items. Defective words show overall low ratings in both task conditions, but with substantial variability, including a non-trivial number of ratings at ceiling. This matches what was observed in Copot & Sims (2025), who ascribed this variability to both differences between participants and differences between items.

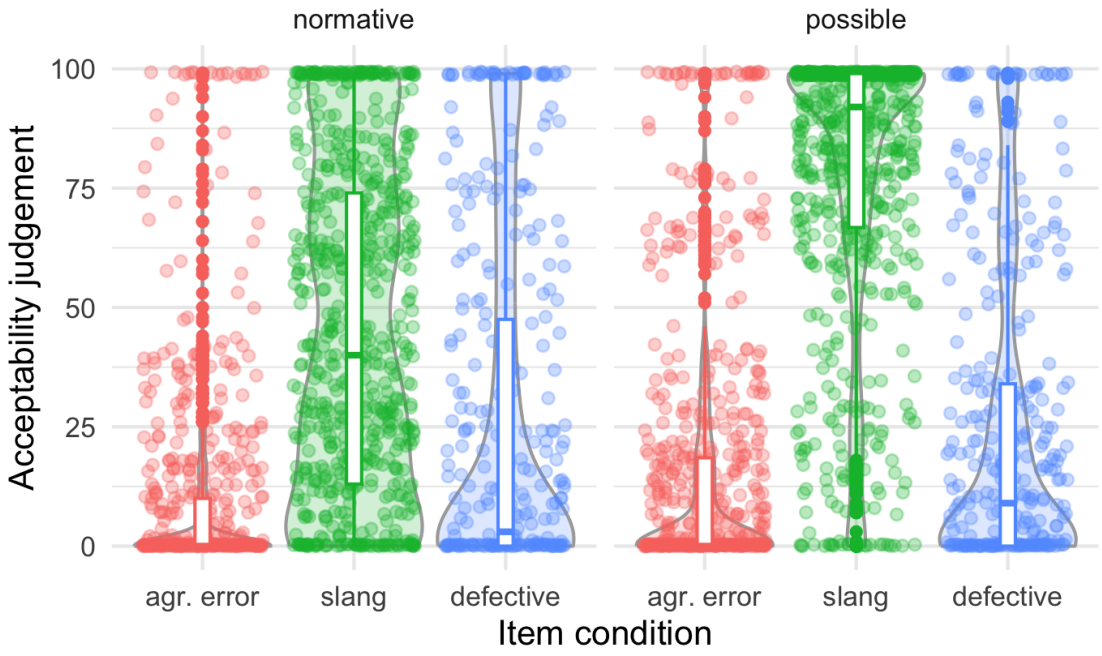


Figure 2. Raw scores for item and task condition

The results of the core model broadly match what was observed by Copot & Sims (2025). As shown in our model output (Table 5), ungrammatical items received low ratings throughout, and while slang items were rated substantially better (slang: $\beta = 2.3$, $p < 0.001$), average ratings for defective items were not distinguishably better than those for ungrammatical items ($\beta = 0.34$, $p = 0.11$). The normative task did not influence the ratings of ungrammatical items compared to

Term	Estimate	Std. Error	p	
(Intercept)	-1.69	0.16	<.001	***
ItemCond2	2.30	0.21	<.001	***
ItemCond3	0.34	0.22	0.119	
TaskCond2	-0.16	0.11	0.144	
Freq	0.14	0.14	0.312	
Prescriptivism	-0.17	0.08	0.028	*
ItemCond2:TaskCond2	-0.94	0.09	<.001	***
ItemCond3:TaskCond2	0.23	0.11	0.032	*
ItemCond2:Freq	0.30	0.20	0.144	
ItemCond3:Freq	-0.05	0.20	0.815	
ItemCond2:Prescriptivism	0.03	0.06	0.683	
ItemCond3:Prescriptivism	0.03	0.07	0.736	
TaskCond2:Freq	-0.01	0.06	0.870	
TaskCond2:Prescriptivism	0.24	0.11	0.035	*
Freq:Prescriptivism	-0.12	0.04	0.004	**
ItemCond2:TaskCond2:Freq	0.07	0.09	0.450	
ItemCond3:TaskCond2:Freq	0.21	0.11	0.049	*
ItemCond2:TaskCond2:Prescriptivism	-0.04	0.09	0.695	
ItemCond3:TaskCond2:Prescriptivism	-0.13	0.11	0.216	
ItemCond2:Freq:Prescriptivism	0.18	0.06	0.002	**
ItemCond3:Freq:Prescriptivism	0.06	0.07	0.416	
TaskCond2:Freq:Prescriptivism	0.13	0.06	0.028	*
ItemCond2:TaskCond2:Freq:Prescriptivism	-0.14	0.09	0.119	
ItemCond3:TaskCond2:Freq:Prescriptivism	-0.03	0.11	0.799	

Table 5. Coefficients of the core model

the possibility task. The interaction of item condition and task condition showed a depression of scores in the normative task for slang (β = 0.94, p $<$ 0.001), confirming that the task manipulation is operating as intended. Conversely, defective items were rated comparatively better in the normative condition relative to ungrammatical items (β = 0.23, p = 0.032), consistent with Copot & Sims's finding that defective items do not fully parallel slang, probably because defective lexemes tend to be inherently more formal (Copot et al. In Prep). More prescriptive participants also gave worse ratings than less prescriptive participants to ungrammatical items in the possibility condition (β = -0.17, p = 0.03), and the lack of significance of the interaction of item condition and prescriptivism tells us that the effect of prescriptivism on slang and defective items is not distinguishable from that on ungrammatical items. The model reveals an interaction of frequency and prescriptivism (β = -0.12, p = 0.004), showing that in the possibility condition, for ungrammatical items, the more frequent the lexeme, the harsher the judgment by more prescriptivist participants. The interaction of frequency and prescriptivism is indistinguishable from this for defective items (β = 0.06, p = 0.4), but is flipped for slang (β = 0.18, p = 0.002), where more frequent lexemes are rated better by more prescriptive participants.

Crucial for our research question, there is no significant interaction for either Freq or Item-

Cond3:Freq, signaling that the frequency slope for defective items is flat. The interaction with task condition (β = 0.21, p = 0.04) suggests a mildly positive slope for defective items in the normative condition only. Overall, little effect of lexeme frequency is detectable for defective items.

7. Analysis by item source. The absence of clear effects of frequency for defective items in the core analysis raises the question of whether this has to do with the heterogeneous nature of the defective class: items selected on the basis of low rule reliability (Albright 2003) and items selected from dictionary classifications (RAE 2009; 2014) may reflect fundamentally different underlying processes, whose distinct frequency signatures cancel out in a combined analysis. A second model was built to statistically test whether defective items from different sources exhibit differences in frequency effects.

7.1. MODEL. The second model is restricted to judgments of defective words. The task condition and prescriptivism variables are identical to the previous model. The log frequency variable is standardized within this set of data points. A variable coding the source of the defective items is added, with the treatment coding scheme shown in Table 6.

	Both	RAE
Albright	0	0
Both	1	0
RAE	0	1

Table 6. Treatment coding for the source of defective items

The model formula is:

```
judgment ~ ItemSource * TaskCondition * LogFreq * Prescriptivism
+ (1|participant) + (1|item)
```

7.2. RESULTS. Figure 3 previews the heterogeneity of the effect of frequency on the rating of defective items: lexemes taken from Albright's study show a positive frequency effect (especially in the normative task condition), confirming Albright (2003)'s observation, while lexemes marked as defective by the RAE show a flat (or possibly negative, in the normative task condition) effect of frequency.[11]

Table 7 shows the coefficients for the model predicting defective item judgments as a function of their source. No effects involving prescriptivism came out as significant, so the factor was removed from the model. The non-significant main effects of item source (ItemSourceboth: β = -0.03, p = 0.96; ItemSourcerae: β = 0.27, p = 0.65) and their interactions with frequency indicate that in the possibility condition, Albright, RAE-only, and overlap items do not differ significantly in their frequency slopes; as is visible in Figure 3, the divergence between item sources is localized to the normative condition. For Albright-only items, the frequency slope steepens substantially under normative framing (TaskCond2:Freq: β = 0.5, p = 0.003), replicating Albright's (2003) positive frequency effect: Albright items belonging to more frequent lexemes are judged more acceptable, but only when asked normatively. This slope is inverted for items present in

[11] Items that came from both types of sources are included in both panels in Figure 3. The regression lines shown are derived from a simple regression of judgments on log lexeme frequency. However, substantially the same pattern emerges from the full statistical model presented in this section.

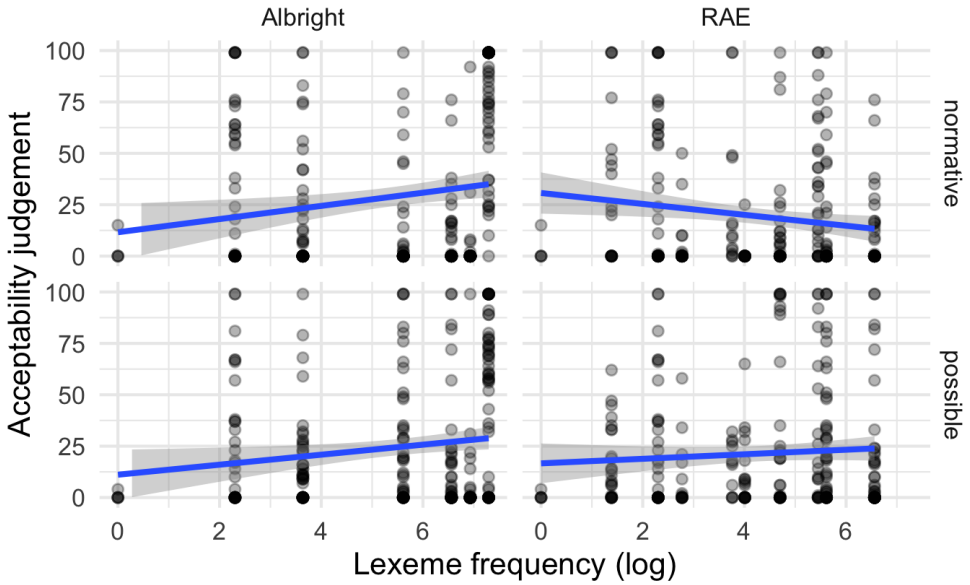


Figure 3. Differing effects of frequency on defective items from different sources (darker points = more responses)

both sources (ItemSourceboth:TaskCond2:Freq: β = -0.67, p = 0.003) with RAE items showing an even more pronounced inversion (ItemSourcerae:TaskCond2:Freq: β = -0.72, p = 0.009), consistent with Copot & Sims's (2025) negative familiarity effect. Together, these results indicate that the diverging frequency effects in the literature reflect not the language of investigation or methodology, but the classificatory criterion used to identify defective items.

Term	Estimate	Std. Error	p	
(Intercept)	-1.19	0.47	0.011	*
ItemSourceboth	-0.03	0.60	0.959	
ItemSourcerae	0.27	0.61	0.655	
TaskCond2	0.10	0.19	0.583	
Freq	0.08	0.46	0.855	
ItemSourceboth:TaskCond2	-0.30	0.23	0.191	
ItemSourcerae:TaskCond2	-0.37	0.26	0.163	
ItemSourceboth:Freq	0.04	0.53	0.935	
ItemSourcerae:Freq	0.22	0.62	0.722	
TaskCond2:Freq	0.50	0.17	0.003	**
ItemSourceboth:TaskCond2:Freq	-0.67	0.22	0.003	**
ItemSourcerae:TaskCond2:Freq	-0.72	0.27	0.009	**

Table 7. Coefficients modeling judgment of defective items by source

8. Discussion. The central finding of this study is that the direction of the frequency effect for defective verbs is determined by how defectiveness is operationalized, not by the object language or the experimental methodology. Albright-only items – selected on the basis of the MGL's low

confidence in inflected forms – show a positive frequency effect under normative framing, replicating Albright (2003). Items classified as defective by the RAE show an attenuated or negative frequency slope in the same condition, replicating Copot & Sims (2025). Since both sets of items are Spanish verbs tested with the same participants and the same task, the divergence cannot be attributed to cross-linguistic differences or methodological incompatibility between studies. The two experiments were, in effect, studying different constructs. This has implications for the notion of defectiveness as a unified category.

Sims's (2015) definition of defectiveness – a lexeme is defective when it is missing a form it would be expected to have – places the weight of the analysis on "missing": the assumption is that there is a determinate fact of the matter about whether a form exists. Our results suggest that "missing" covers at least two phenomenologically distinct situations. For RAE-type defectives, speakers appear to have explicit metalinguistic knowledge that a form is proscribed: a knowing avoidance of the kind documented for French by Copot & Sims. For Albright-type defectives, the issue seems to be uncertainty: speakers lack sufficient evidence about the allomorphic behavior of a low-confidence lexeme, leading to unease. The positive frequency effect for Albright items is consistent with this; more familiar lexemes provide more evidence, reducing uncertainty and increasing acceptability. The negative or null frequency effect for RAE items reflects the opposite logic: frequency increases exposure to the prescriptive taboo, entrenching avoidance.

Whether these differences warrant separate theoretical treatment in a model of morphology is an open question. At minimum, the finding suggests that defectiveness is a descriptive label applied to the surface symptom of gap-filling avoidance, which may have heterogeneous causes even within a single language. Future work might productively examine metalinguistic discourse around specific defective items to test whether the phenomenological distinction between taboo-type and uncertainty-type defectiveness is detectable in how speakers talk about these words.

Finally, the role of individual speakers' degree of prescriptivism was less clear in Spanish than in French. While participants' prescriptivism scores modulated the frequency slope for dictionary-classified items, it did not emerge as a significant effect, in contrast to Copot & Sims's strong prescriptivism effect. This difference can perhaps be explained in terms of the standard languages' cultures.[12] Spain has a language authority in the RAE, which is overall conservative in its approach. However, it takes a somewhat less prescriptive approach than the French Academy (Ross 2004), with the RAE's guidance more likely to reflect language usage. Moreover, Bonnin (2012:1) argues that "...speakers decenter the RAE norm, locating it among many different sources of normative discourse but, at the same time, they claim a similar authoritative and central place for the norms they produce."[13] Thus, whereas a pervasive standardizing culture produces strong prescriptivism effects in French speakers' evaluation of defective items, weaker prescriptivism effects in our study are unsurprising, given the overall less rigidly prescriptive culture in Spain. The language culture of each country may thus be implicated as a factor in how speakers judge normatively defective lexemes.

9. Conclusions. In defective Spanish verbs, the direction of the frequency effect is determined by how defective items are selected, with different criteria – based on morphophonological rule reliability versus standard language norms as encoded in dictionaries and grammars – identifying

[12] See also Copot et al. (In Prep) on the role of standard language ideology in how speakers evaluate defective and overabundant forms.

[13] In the standardology literature, this decentering is known as *destandardization* (Coupland & Kristiansen 2011).

items whose low acceptability reflects fundamentally different information about defectiveness. This calls for greater methodological transparency in the study of defectiveness and suggests that a theoretically adequate account of paradigm gaps will need to distinguish between gap-filling avoidance rooted in morphological uncertainty and avoidance rooted in prescriptive stigmatization.

Data availability statement. The full set of experiment items, the prescriptivism questionnaire, the experiment instructions, the analysis script, and the raw, deidentified data are available on OSF at https://osf.io/n5pvq/overview. This study was approved as exempt research by the Institutional Review Board at The Ohio State University (protocol 2024E1095).

References

Albright, Adam. 2003. A quantitative study of Spanish paradigm gaps. In Gina Garding & Mimu Tsujimura (eds.), *West Coast Conference on Formal Linguistics 22 proceedings*, 1–14. Somerville, MA: Cascadilla Press.

Albright, Adam & Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90. 119–161. https://doi.org/10.1016/S0010-0277(03)00146-X.

Arregi, Karlos & Andrew Nevins. 2014. A monoradical approach to some cases of disuppletion. *Theoretical Linguistics* 40. 311–330. https://doi.org/10.1515/tl-2014-0014.

Baerman, Matthew & Greville G. Corbett. 2010. Introduction: Defectiveness: Typology and diachrony. In Matthew Baerman, Greville G. Corbett & Dunstan Brown (eds.), *Defective paradigms: Missing forms and what they tell us*, 1–18. London: British Academy Press, in coordination with Oxford University Press. https://doi.org/10.5871/bacad/9780197264607.003.0001.

Bonnin, Juan Eduardo. 2012. Decentralization of the linguistic norm online: the Royal Spanish Academy challenged on the Internet. *Tilburg Papers in Culture Studies* 36. 1–25.

Boyé, Gilles & Patricia Cabredo Hofherr. 2010. Defectiveness as stem suppletion in French and Spanish verbs. In Matthew Baerman, Greville G. Corbett & Dunstan Brown (eds.), *Defective paradigms: Missing forms and what they tell us*, 35–52. London: Oxford University Press, in coordination with British Academy Press. https://doi.org/10.5871/bacad/9780197264607.003.0003.

Brooks, Mollie E., Kasper Kristensen, Koen J. van Benthem, Arni Magnusson, Casper W. Berg, Anders Nielsen, Hans J. Skaug, Martin Maechler & Benjamin M. Bolker. 2017. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal* 9(2). 378–400. https://doi.org/10.32614/RJ-2017-066.

Copot, Maria, Mari Aigro, Andrea D. Sims & Virve-Anneli Vihman. In Prep. Defectiveness, overabundance, and standard language ideologies. ms., University of Surrey.

Copot, Maria & Andrea D. Sims. 2025. The influence of prescriptivism on French verbal defectiveness. *Linguistics* 63. 1293–1338. https://doi.org/10.1515/ling-2023-0028.

Coupland, Nikolas & Tore Kristiansen. 2011. SLICE: Critical perspectives on language (de)standardisation. In Tore Kristiansen & Nikolas Coupland (eds.), *Standard languages and language standards in a changing Europe*, 11–35. Oslo: Novus.

Curzan, Anne. 2014. *Fixing English: Prescriptivism and language history*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139107327.

Gorman, Kyle & Charles Yang. 2019. When nobody wins. In Franz Rainer, Francesco Gardani, Hans Christian Luschützky & Wolfgang U. Dressler (eds.), *Competition in inflection and word formation*, 169–194. Dordrecht: Springer. https://doi.org/10.1007/978-3-030-02550-2_7.

Herce, Borja. 2024. VeLeSpa: An inflected verbal lexicon of Peninsular Spanish and a quantitative analysis of paradigmatic predictability. *Language Resources and Evaluation* 59. 1705–1718. https://doi.org/10.1007/s10579-024-09776-2.

Keuleers, Emmanuel & Marc Brysbaert. 2010. Wuggy: A multilingual pseudoword generator. *Behavior Research Methods* 42. 627–633. https://doi.org/10.3758/BRM.42.3.627.

Maiden, Martin. 2009. From pure phonology to pure morphology: The reshaping of the Romance verb. *Recherches linguistiques de Vincennes* 38. 45–82. https://doi.org/10.4000/rlv.1765.

Nikolaev, Alexandre & Neil Bermel. 2022. Explaining uncertainty and defectivity of inflectional paradigms. *Cognitive Linguistics* 33. 585–621. https://doi.org/10.1515/cog-2021-0041.

Pertsova, Katya. 2016. Transderivational relations and paradigm gaps in Russian verbs. *Glossa* 1. 1–13. https://doi.org/10.5334/gjgl.59.

Pertsova, Katya & Julia Kuznetsova. 2015. Experimental evidence for lexical conservatism in Russian: Defective verbs revisited. In Yohei Oseki, Masha Esipova & Stephanie Harves (eds.), *Formal Approaches to Slavic Linguistics 24 proceedings*, 418–437. Ann Arbor, MI: Michigan Slavic Publications.

R Core Team. 2021. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. https://www.R-project.org/.

Real Academia Española. 2024. CORPES XXI. Corpus del Español del Siglo XXI. https://www.rae.es/banco-de-datos/corpes-xxi. Version 1.2.

Real Academia Española & Asociación de Academias de la Lengua Española. 2009. *Nueva gramática de la lengua española*. Madrid: Espasa Libros. https://www.rae.es/gram%C3%A1tica/.

Real Academia Española & Asociación de Academias de la Lengua Española. 2014. *Diccionario de la lengua española, 23a edición*. Madrid: Espasa Libros. https://dle.rae.es/.

Ross, Nigel J. 2004. Academies and attitudes. *English Today* 20(3). 23–28. https://doi.org/10.1017/S0266078404003049.

Rowe, Charlie. 2026. *Missing or mixed up? What judgments to Spanish verbs reveal about defectiveness*. Columbus, OH: The Ohio State University B.A. thesis.

Schwarz, Florian & Jérémy Zehr. 2018. PennController for internet based experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD832. Last accessed 23 March 2026.

Sims, Andrea D. 2006. *Minding the gaps: Inflectional defectiveness in a paradigmatic theory*. Columbus, OH: The Ohio State University dissertation.

Sims, Andrea D. 2015. *Inflectional defectiveness*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781107053854.

Sims, Andrea D. 2023. Defectiveness. In Peter Ackema, Sabrina Bendjaballah, Eulàlia Bonet & Antonio Fábregas (eds.), *The Wiley Blackwell companion to morphology*, Malden, MA: Wiley-Blackwell. https://doi.org/10.1002/9781119693604.morphcom022.

Zacarías Ponce de León, Ramón. 2021. Verbos defectivos, entra la regularidad y la irregularidad: El caso de *abolir*. *Nueva Revista de Filología Hispánica* 69(1). 261–277. https://doi.org/10.24201/nrfh.v69i1.3714.