Andrew Nevins*, Cilene Rodrigues and Kevin Tang

The rise and fall of the L-shaped morphome: diachronic and experimental studies

Abstract: It has been suggested that the Romance first person singular indicative constitutes a natural class with the present subjunctive paradigm for the purposes of stem selection (Maiden 2005), thus forming a kind of 'diagonal syncretism', as the latter shares no morphosyntactic features with the former. The existence of such patterns has been taken to be an argument for autonomous morphology and the existence of unnatural 'morphomes', in the sense of Aronoff (1994). Our experimental investigations with native speakers of Portuguese, Italian, and Spanish reveal that this pattern is underlearned, and that speakers do not generalize it to novel forms, instead preferring the 2nd person singular indicative to the 1st person as the base for the derivation of the subjunctive paradigm (and the 2nd person indicative as opposed to the 2nd person subjunctive as the base for the derivation of the 1st person indicative as well). The results implicate a role for naturalness biases in morphological structure, and an awareness that the first person singular is an unreliable and idiosyncratic base for productive inflectional identity. We then study the underlearning of the L-morphome in terms of historical change in the salience of these patterns. We demonstrate, through means of diachronic corpus studies spanning five centuries, a change in the ratio of first conjugation verbs to second & third conjugation verbs, and a resulting decrease in the relative type frequency of where morphomic verbs reside. If indeed learners need increased evidence in order to incorporate and actively uptake unnatural patterns, this lexical support has dwindled over time. Even though many of the morphomic verbs have maintained a very high token frequency (allowing them to survive as memorized), their productivity has diminished over time, and hence they go unlearned as a generalizable pattern. When the distribution of irregular alternations is overshadowed in the lexicon, a morphologically unnatural pattern may cease to maintain its status as part of the grammar.

***Corresponding author: Andrew Nevins,** University College London, London WC1N 1PF, UK, E-mail: a.nevins@ucl.ac.uk
Cilene Rodrigues, PUC-Rio, Rio de Janeiro, Brazil, E-mail: crodrigues@puc-rio.br
Kevin Tang, University College London, London WC1N 1PF, UK, E-mail: kevin.tang.10@ucl.ac.uk

Keywords: morphome, Romance languages, inflectional identity, diagonal syncretism, subjunctive, experimental linguistics

DOI 10.1515/probus-2015-0002

1 The case for autonomous morphomes

The question we examine in this paper is the extent to which the 'naturalness' of morphological generalizations interact with their productivity in the lexicon. Once a learner has observed a trend within the formation of a certain morphological category, to what extent do they generalize it beyond the static set of highly token-frequent examples? This question has been extensively asked for phonological patterns of varying types (e.g. phonotactic constraints and phonological processes; Zimmer (1969), Berent et al. (2012), Finley (2012), Becker et al. (2011) among many others). For example, if a Turkish word has a round vowel in the first syllable, will it automatically disallow a non-round vowel within the second syllable – and do native speakers pick up on this to the extent that they have well-formedness judgements about non-existing novel words? On the other hand, if a language shows evidence that high vowels and voiceless consonants do not occur word-finally, is that too obscure to generalize? In other words, do learners have a good set of 'hunches' about what kind of constraints are ones that, no matter how very well may hold within the lexicon, still look too accidental to enforce in new words and novel formations?

Within the realm of a predictive morphological generalizations, questions of these sort have received less attention, although they are very much ripe for the asking. For example, suppose that across the lexicon of verbs, the form of the second singular is predictable from the form of the first singular (as opposed to any other form within the inflectional paradigm) – is this reasonable enough to be applied with confidence to novel words for which the entire inflected paradigm is not known? What about, say, forming denominal verbs, given a choice of two suffixes, *M* and *N* – will the choice between *M* and *N*, even if it depends on a potentially obscure fact about whether the word ends in a mid-vowel or not, be salient enough and 'make enough morphological sense' to leverage in an experimental task when deeming the well-formedness of two potential competing candidates? Asking a phonologist if a pattern of seemingly pathological palatalization (e.g. k → č before [e] but not before [i]) exists or not according to a particular set of theories of phonological naturalness is pretty straightforward these days, as a voluminous set of substantively-biased and formally biased predictions have been formulated (Hayes et al. 2003; Moreton 2008; Heinz 2010). However, the parallel trajectory of 'Natural Phonology' in the 1970s (reflourishing in the 1990s with a host of increased dialogue between phonological typology and phonetic research) has not proceeded alongside much consensus as to what constitutes Natural Morphology, despite occasional valiant efforts in the 1980s (e.g. Dressler et al. 1987). Research in this area has been overall too sparse to have yet arrived at a real consensus anything like what we have in phonology today.

The dominant trend in many corners of morphological research is thus to say 'Nothing is natural'. Statements of this sort express a degree of nihilism for predictive morphological typology to an extent that doesn't even parallel Anderson's (1981), Blevins's (2004) or Hale and Reiss's (2008) eschewal of any systematic crosslinguistic patterns in phonology – as these authors directly embrace formal simplicity as an evaluation metric and thereby a predictive measure of the overall naturalness, acquirability, diachronic persistence, and productivity of certain logically possible patterns. Take one core tendency of morphological inflection: defectiveness in verbal or nominal paradigms, a phenomenon found time and again in languages as distinct as Russian, Mohawk, and the Iberian languages. Focusing on the latter group of languages, Maiden and O'Neill (2010, p. 112) state that "The major domain of defectiveness (present subjunctive and 1sg present indicative) seems, and is, irreducibly arbitrary if one seeks a motivation outside the morphological system itself".

This is an essential expression of the credo of 'autonomous morphology', as developed by Aronoff (1994) and adopted in many subsequent works – that morphology needn't answer to phonology or syntax; that may follow its own logic. On this view, learners are directly endowed with the ability to sift through and compute morphological productivity (or lack thereof, in the case of defectiveness) purely by problem-solving abilities exclusive to morphological geometries; a data-driven puzzle to be eventually mastered given enough internal and external reward in the learning process; a poker hand dealt by the exigencies of diachronic drift, in which an L-shaped pattern (borrowing a metaphor from the Roman alphabet) can be learned in a two-by-three geometric paradigm (e.g. all [−plural] cells and the [+ plural, −participant] cell, but not the [+ participant] plurals). Is such a pattern easier or harder to learn – i.e. more natural – than a π-shaped, ꓶ-shaped, or и-shaped pattern? The right answer is not yet in sight, and could come from many directions – among them extensive crosslinguistic typology in which database comparisons inform patterns of natural vs. unnatural syncretism and formal modeling of the computational primitives that favor one system over another (e.g Bobaljik (2004), Pertsova (2011), Graf (2012), and in the case of the present paper, adopting the tack of creating closely-controlled experimental conditions in which native speakers must choose between one of two potential, non-existing candidates, or rate the well-formedness of unseen inflected forms of wug words.

In the current paper we therefore examine a case study in the Romance languages, the *L-shaped morphome*, described in greater detail below, with the hopes of contributing to the above triangulation between theoretical modeling, typological sampling, and experimental comparison of the relative naturalness of certain generalizations that learners may adopt or neglect. Work within many (but not all schools) of morphological theory within the past half-century have proposed certain formal primitives, substantive constraints, and interface-derived properties that systems of inflectional and derivational morphology 'like to have' – and these result in delimiting patterns that are more 'natural' than others. Among the components in the potential naturalness toolkit investigated in the present work is the following:

(1) Featural decomposition – that an inflectional paradigm is constituted by 'natural classes', akin to the ones familiar from segmental phonology, that are directly formed by a small set of binary-valued features that characterize particular 'rows' or 'columns' of the paradigm, as opposed to others. While many distinct systems of featural decompositions are possible, research often employs relative attestedness vs. unattestedness as a means of comparison among them, along with a generalized dispreference for disjunctive statements (e.g. the morphological equivalent of 'all the voiced obstruents or the non-lateral, non-glide sonorants').

Turning to inflectional morphology, there are irregular stem alternants in Portuguese found in the 1sg present indicative that are also found throughout the present subjunctive, constituting what Maiden (2005) identifies as a 'morphome' (a term due to Aronoff (1994)) – an autonomous morphological pattern of shared identity, which he calls the L-shaped pattern, as shown for the following singular columns of the indicative and subjunctive in the present tense:

(2)

'to say'	IND	SBJ
1sg	**dig**-o	**dig**-a
2sg	diz-es	**dig**-as
3sg	diz	**dig**-a

Diachronically, the L-shape is essentially a consequence of the theme vowels that follow the stems causing palatalization: in the II/III conjugation, the 1SG.IND and all SBJ forms have in common a [+ back] vowel, which enjoys the velar alternant, while the others have a [–back] vowel, with the palatal/coronal alternant.

Long after the cessation of the process of palatalization in verb stems, this L-shaped pattern was apparently extended to verbs lacking a phonological reason for identity between the 1SG.IND and SBJ:

(3)

'to hear'	IND	SBJ
1sg	**ouç**-o	**ouç**-a
2sg	ouv-es	**ouç**-as
3sg	ouv-e	**ouç**-a

According to Maiden (2005) this L-shaped relation among unrelated forms is what constitutes a 'morphome', and shows the productive autonomy of the morphological component by itself: it needn't answer to phonology or syntax. Of course, the rule was in fact originally phonologically motivated within verbs from the *-er* and *-ir* class (the second and third conjugations), having to do with the fact that both the 1SG.IND and all of the SBJ forms had a [+ back] vowel following them, while the others had a [–back] vowel, at a time during which palatalization by front vowels was active in the language. However, long after palatalization ceased to be productive, the pattern was extended to forms with no such conditioning possible to state, e.g. *ouç-/ouv-* (phonetically [ows/owv]). Moreover, from a featural perspective, there is no reason these should form a natural class; they are disjunctive. The question, therefore is whether such patterns are synchronically productive within the Romance languages, a question all the more important to ask as the L-shaped pattern is trivially true for even all regular verbs in the language:

(4)

'to say'	IND	SBJ
1sg	**fal**-o	**fal**-e
2sg	fal-as	**fal**-es
3sg	fal-a	**fal**-e

Note that it is the case that the 1SG.IND and the SBJ forms, e.g. 2SG.SBJ, are in fact identical to each other in *fal-/fal-* in (4); if the learner includes regular verbs within the scope of their generalization, the L-shaped pattern potentially holds of every verb in the language. Although there may be potential theoretical differences in whether or not the L-morphome is taken to be a symmetrical syncretism (e.g. the 1SG.IND equally predicts the form of the SBJ and vice versa) or asymmetrical (learners will adopt the 1SG.IND as the base for the SBJ but will be less sure in the opposite direction), scholars such as Maiden (2005) take it to be symmetrical, as a generalization over a set of paradigm cells that form an 'unnatural' class for shared stem identity. While there are not many verbs in contemporary Portuguese that show the L-shaped pattern (perhaps twenty at most), many of them are quite frequent and salient. Their pattern of inflectional identity in cases like (3), however, raises what Bachrach and Nevins (2008) call the 'inclusion question': why is the 2SG.IND not included – in other words, why is the pattern 'L-shaped', rather than 'sideways T' shaped (namely including the 2SG.IND but *not* the 1SG.IND)? Would either be equally likely shapes for learners to deal with, given spontaneous chance to show a preference for one or the other?

2 Exploring biases: when unnatural patterns go underlearned

Certainly, the very existence of the term *morphome* is predicated on the fact that patterns such as (3) are neither phonologically nor morphosyntactically natural.

Phonological research enjoys a long tradition of testing whether certain patterns in the lexicon are actually generalized. The work of Zimmer (1969), Zhang et al. (2006), and Becker et al. (2011), for example, find that *unnatural patterns are memorized for the existing items, but underlearned*, and thereby not generalized to novel items.

The work reported here is a first step towards experimentally testing whether learners of Romance languages extend 'unnatural identity relations' in novel inflectional tasks. It finds consonance with Pertsova's (2010) recent results that Russian speakers 'underlearn' the unnatural relation between nom. sing. and gen. pl. in wug tasks, and with Tucker's (2000) work that Italian speakers underlearn the relation holding in the static lexicon between the participal form of a verb and the agentive, to the exclusion of the infinitive.

Morphomes are akin to units of 'Priscianic' syncretism (Matthews 1991), e.g. "For any verb, however irregular it may be in other respects, the Present Infinitive always predicts the Imperfect Subjunctive". But are these simply useful observations for grammarians (and indeed second-language learners) – or are they actually rules and principles that learners build into their grammars? Morphomic patterns are claimed to be "informative, because the deductions that they sanction reduce uncertainty about the paradigmatic structure of a system"

(Blevins 2010). Under this hypothesis, learners should readily shoehorn new verbs into the L-shaped pattern, regardless of what their phonological makeup is like: they are abstract statements over the geometry of inflectional paradigms that license implicational statements.

While the L-morphome emerged from a phonetically-motivated process, and it was later analogically extended to novel forms in great measure (as amply documented by Maiden 2005), peaking in the sixteenth century or so, it reached a point with no phonetic naturalness or morphosyntactic naturalness to support it. Still, the verbs in which it occurs are highly token frequent. But do learners take it to be a memorized part of the lexicon, or an active rule, used in acquisition, a principle of structuring inflectional paradigms of verbs more generally in the language? What, if anything, might have led to a cessation in its productivity today?

3 An overview of the experimental method used throughout

As one cannot learn much about the state of speakers' synchronic knowledge by examining the handful of existing L-morphome verbs alone, we launched an experiment on implicational generalization. We illustrate this first for European Portuguese, and replicated the same overall experimental method with Italian and Spanish.

In order to best test the predictions of whether the L-morphome is actively employed in structuring the inflectional paradigms for newly-learned verbs (and hence a principle that forms some detectable part of these speakers' morphological grammar), we created verbs with divergent forms for the 1SG.IND and the 2SG.IND, like the cases of (2)-(3) above. However, we used three novel morphophonological alternations, none of which are extant in Portuguese: *p~f, t~s, k~x* – but which are part of the phonology of other languages (e.g. spirantization in Hebrew). The motivation for avoiding existing alternations (e.g. those like *ouç-/ouv-* or *dig-/diz-*) was to specifically test the claim that 'L-shapes', once incorporated into the grammar of the language, form an autonomous kind of paradigm knowledge, independent and above any of the specific phonological forms themselves. We return to a discussion of this methodological step at various points throughout the rest of this paper.

Participants were divided into two groups, each of which saw 15 items. The methodology thus followed Wilson's (Wilson 2006) design for studying naturalness in phonology, in which each group has the opposite group's data points "held out" and are probed for generalization.

The *indicative predicting subjunctive* group (I → S) was exposed to frames such as (5–6), and prompted for the 2sg. subjunctive form. As participants have exposure to exactly two stem alternants, they can logically choose either to base the 2SG.SBJ on person (2SG.IND) or on the L-shape.

(5) Eu *nepo* muito rápido, e tu também *nefes*, embora tu ainda não ____ rápido.
'I *nepo*1SG.IND very quickly, and you *nefes*2SG.IND too, even though you don't ____2SG.SBJ quickly yet.'

(6) Tu me disseste que tu *zisas*. Eu também *zito*, mas espero tu não ____ amanhã de manhã porque temos visitas.
'You told me that you 2SG.IND. I also 1SG.IND, though I hope that tomorrow you'll not ____2SG.SBJ, because we have visitors.'

The *subjunctive predicting indicative* (S → I) group was exposed to frames such as (7–8), and prompted for the 1sg. indicative form. Again, as participants have exposure to exactly two stem alternants, they can logically choose either to base the 1SG.IND on mood (2SG.IND) or on the L-shape.

(7) Tu *nefes* muito bem. Caso tu *nepas* amanhã, eu também ____.
'You *nefes*2SG.IND very well. In case you *nepas*2SG.SBJ tomorrow, I'll ____ 1SG.IND too.'

(8) Antes que tu *plicas* todos os dias, primeiro tu *plirres* uma vez por semana. Eu fiz assim e agora eu ____ todos os dias.
'Before you can 2SG.SBJ every day, first you 2SG.IND once a week. I did so that way, and now I ____1SG.IND every day.'

In all of the experiments we conducted, participants were presented with instructions (in Portuguese, Spanish, or Italian, as appropriate) that were adapted translations of the following: "You will be presented with examples of invented verbs, such as *eu blino, tu blines*. Then you will see a sentence with a blank space. Your task is to fill in the blank with the appropriate form of the verb". Importantly, participants saw frames in which the linear order of 1SG.IND and 2SG.IND was the reverse in the I → S group, or in which 2SG.IND and 2SG.SBJ were the reverse in the S → I group, to ensure that primacy and/or recency effects would not bias the overall responses. This was also important insofar as a potential response pattern by our participants might be to treat the two verbal forms (e.g. *nepo, nefes* or *nefes, nepas*) with distinct semantics. The instructions that we provided, as well as the presence of 50% filler items with identical stems (e.g. *blino, blines* or *blines, blinas*) served to steer the participants towards treating both forms as the stems of a single verb (as indeed, we presuppose they treat both *ouç-* and *ouv-* in (3) as instances of the same verb). Let us suppose, for the sake of argument, however, that some participants treated *nefes* and *nepas* as distinct verbs, i.e. something along the lines of *You gabble, but you shouldn't talk so much. I myself often* ____. Which of these verbs would they be inclined to use to supply the third form? If they indeed treated these as distinct verbs, there is no linguistically-based hypothesis that we could formulate that would make consistent predictions across all four orders of presentation, but for precisely this reason, we counterbalanced the order of the two verbal forms, to investigate potential recency or primacy effects (e.g. pragmatic or attentional biases for always choosing the most recently presented, or the first-ever presented) form.

Summing up, the basic experimental design here (exemplifying just with the I → S group) is to present speakers with an 'incomplete' paradigm and ask for the 2sg subjunctive. In the S → I group, the pattern is the opposite. Both groups are thus always given the 2SG.IND. In addition, they are given the 1SG.IND and asked for the 2SG.SBJ (or vice-versa). More schematically, there is a paradigm of six cells, in which a participant is presented with only two of them, say C1 with form *nep-* and C2 with form *nef-*, each presented within appropriate morphosyntactic contexts. The participant is then asked to provide the form for what the morphosyntax of the frame sentence (e.g. the tense, mood, and person features of its pronominal subject) would require as C3. Whether they consistently choose C1 or C2 as the stem on which to base the unseen form – in principle a coin flip of 50%-50% choice all else being equal – potentially bears on whether the L-shape is alive and kicking or not. If morphomes – that is, morphological outlaws that do not correspond to the notion of *morpheme*, as they represent disjunctive distributions with no grounding in morphosyntax or phonology – have any explanatory or ontological force, then non-interface-grounded distributions should be chosen. After all, speakers of European Portuguese (with a healthy ongoing use of the subjunctive mood) have abundant evidence staring them in the face that irregular verbs in their language show such 'diagonal syncretisms'.

Given this incomplete paradigm, their choice of a base for the 2SG.SBJ, can be either based on the L-shaped morphome (in which case it would be the 1SG.IND stem *nep-*), or instead on choosing the stem for 2nd person across moods, whether indicative or subjunctive (in which case it would be *nef-*). These alternations do not exist in Portuguese, and that's why we used them: the clearest and most predictive aspects of the L-morphome theory say that it is about an abstract relation of complete identity between these cells of the paradigm without any reference to their phonological form or phonological naturalness. Using extant alternations would have potentially invited analogical relations with existing verbs, which would not have told us anything about whether the L-morphome is in fact generalizable to new forms.

In what follows, we discuss three such experiments, conducted with Portuguese (Section 4), Italian (Section 5), and Spanish (Section 6), respectively. After summing up the general results in Section 7, we turn to a diachronic explanation in Section 8 of why the results show that in fact, the L-morphome is synchronically underlearned, despite having been healthily productive 500 years ago. Section 9 concludes the paper.

4 Experimental method: European Portuguese

As mentioned above, the experimental task was relatively simple: given only two forms of an incomplete paradigm, each of which has a different base, to produce a third form. The dependent variable is thus whether the base chosen is the one that would be expected given the L-morphome hypothesis (e.g. the 1SG.IND or the 2SG.SBJ, constituting 'diagonal syncretism' in either direction), or whether the base chosen reflects a 'natural' response (namely, the 2SG.IND, which shares mood with the 1SG.IND ('vertical syncretism') and shares person with the 2SG.SBJ ('horizontal syncretism').

The experiment took into account a number of independent variables in order to see whether the choice of L-morphome vs. natural response was affected by these factors. One such factor is the *direction* of syncretism: is the putative principle that governs going from the 1SG.IND to the 2SG.SBJ the same in both directions? Half the participants were assigned to the I → S group, which pits a diagonal syncretism from 1SG.IND to 2SG.SBJ against a horizontal syncretism, namely one in which the base for 2nd person singular forms is identical across both indicative and subjunctive moods (call this 'Uniformity of Person'), as shown in the left of Figure 1. The other half of the participants were assigned to the I → S group, which pits a diagonal syncretism from 2SG.SBJ to 1SG.IND against a vertical syncretism, namely one in which the base for 2nd person singular forms and 1st person singular forms is identical across within a given mood (call this 'Uniformity of Mood'), as shown in the right of Figure 1.

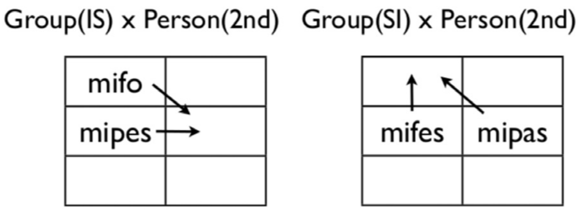

Figure 1: Schematic representation of the I → S condition (left) and S → I condition (right).

In order to control for potential primacy or recency effects, half of the experimental items showed the L-shaped congener as the first item in the frame sentence, whereas the other half showed it as the second item in the frame sentence. There were 15 target items (5 from each place of articulation; coronal, velar, labial), and 15 fillers (with nasal or liquid stem-final consonants, non-alternating). The total of target items and fillers yielded 30 novel verbs. There were 10 different sentential frames, like the ones shown in (5)-(8). On each run of the experiment, the software randomly slotted each item into one of these 10 frame sentences. Notably, the 10 frame sentences were always counterbalanced for whether the 2sg.ind or the L-shaped congener was presented as the first or last instance of this verb in the frame (e.g. (5) vs (6), and (7) vs (8)).

A final variable manipulated within the stimuli was the conjugation class of the novel verb. Half of the verbs ended in a thematic vowel *-a* in the indicative (e.g. *nepas*), and in a thematic vowel with *-e* in the subjunctive, thereby identifying them as 1st conjugation verbs. The other half ended in a thematic vowel *-e* in the indicative and in a thematic vowel with *-a* in the subjunctive, thereby identifying them as 2nd conjugation verbs.[1] The purpose of this manipulation was to test whether, say, L-morphome effects are restricted to the conjugation class in which they are found in the static lexicon, or generalized to all new verbs, across diverse conjugation classes.

The experiment took approximately 10 minutes to complete, was presented on a webpage using Experigen (Becker and Levine 2010), and participants were voluntary and recruited by word of mouth. At the beginning of each session, a participant was randomly assigned to the S → I or I → S group.

1 In fact, for rhizotonic present tense forms, one cannot tell the 2nd and 3rd conjugation apart – compare 2nd conjugation *bates* 2sg.ind, *batas* 2sg.sbj ‘to beat’ with *lates* 2sg.ind, *latas* 2sg.sbj ‘to bark’ – so we leave open which of these two participants treated them as, and take the contrast to be one between 1st conjugation and non-1st conjugation verbs, the latter in fact forming a natural class for the purposes of L-morphome productivity, as we return to in Section 8.

Summarizing the experimental design, the independent variables of interest are below:

(9) *Group* (I → S vs. S → I). The former are asked for 2SG.SBJ, while the latter are asked for 1SG.IND. The purpose of manipulating this factor is to check for the 'direction' of implication.
Frame (L-shape last vs. L-shape first): a mere psycholinguistic control, to examine potential primacy/recency effects due to frame presentation; no specific hypothesis about the potential natural of its effect.
Place (coronal, labial, velar). Largely included as a psycholinguistic control to ensure the diversity of items; no specific hypothesis about the potential natural of its effect.
Conjugation (1st vs. 2nd): Half of the experimental items were from the 1st conjugation, and half were from the second conjugation. Potentially one might find more L-shaped responses in the 2nd conjugation, as these are the locus of such forms in the present lexicon.

The dependent variable is whether the form produced by participants in the elicitation task shares a base with the L-shaped congener (an *L-shaped response*, be it the 1SG.IND or the 2SG.SBJ) or instead with the 2SG.IND, thereby constituting what we henceforth call a *Natural response*, where 'Natural' can be taken as a shorthand for 'morphosyntactically motivated'. Participants' written responses were analyzed as follows. Only those with the tense, person, and mood that were elicited were retained (as we are testing hypotheses about the formation of these specific forms, and not those of other tenses that are tangential to the L-morphome). Similarly, only those that contained a verbal base that was actually presented were analyzed. In the example at hand, therefore (e.g. given *nepo, nefes*, ...) only two responses were analyzed: *nepas* or *nefas*. These were then classified as L-shaped or Natural responses, respectively.

4.1 Results: European Portuguese

In order to see whether there were more L-shaped responses than Natural responses or not, and how this choice was modulated by our experimental design with a range of independent variables, we performed our statistical analyses in three stages. Following current standard practice in the field, we treated participants as random effects, and items as random effects, and every experimental manipulation in the design above as a fixed effect. We then carried out separate analyses with participants (commonly referred to as F1-analysis) and with items (commonly referred to as F2-analysis) with non-parametric paired t-tests. This allowed us to examine if the two responses (L-shaped or Natural) were significantly different and which response is preferred, as only statistical analyses of this sort can reveal whether the direction of the difference and its statistical significance reflect a Naturalness or L-shaped bias, or are instead all explainable in terms of other effects that are tangential to our research question (e.g. order of presentation in frame sentences, place of articulation, etc.). Secondly, we aimed to identify any significant predictors for the responses with mixed-effects modelling. Finally, we again performed participant analyses and item analyses on the subsets of the data, divided up by each significant predictor. This latter method allow us to identify if the preference of a particular response would still hold after taking into account other factors.

We removed trials that were fillers and those that contained uninterpretable typos, incorrect tenses and other erroneous entries. We removed participants who did not complete a majority of the trials, which left us with 249 participants with 3,343 responses.

In what follows, we classify response types as 'L-shaped responses' or 'Natural responses' (which could alternatively be called 'diagonal responses' or 'horizontal/vertical responses', respectively). An L-shaped response is one in which the form for the 1SG.IND is chosen for the 2SG.SBJ, or vice-versa. A natural response is one in which the 2SG.IND is chosen for the 1SG.IND (exhibiting Uniformity within Mood) or in which the 2SG.IND is chosen for the 2SG.SBJ (exhibiting Uniformity within Person). The classification of a response as 'L-shaped' or 'natural' is mutually exclusive for a given trial.

4.1.1 F1-F2 analyses: European Portuguese

A non-parametric paired t-test (two sided), the Wilcoxon signed rank test, was performed throughout the analyses, using the *wilcox.test* function in the *stats* package in *R* (R Core Team 2013). This test returned four outputs: the $W+$, $W-$, N and p-value. W+ is the sum of the ranks of those pairs for which the number of Natural responses is higher than that of L-shaped responses. Similarly, $W-$ is the sum of the ranks of those pairs for which the number of Natural responses is lower than that of L-shaped responses. N is the number of pairs examined. p-value is the level of statistical significance.

To visualise the response preference by participant and by item, we applied a log ratio metric with Laplace smoothing with the two response types, *log(the number of Natural responses/the number of L-shaped responses)*,

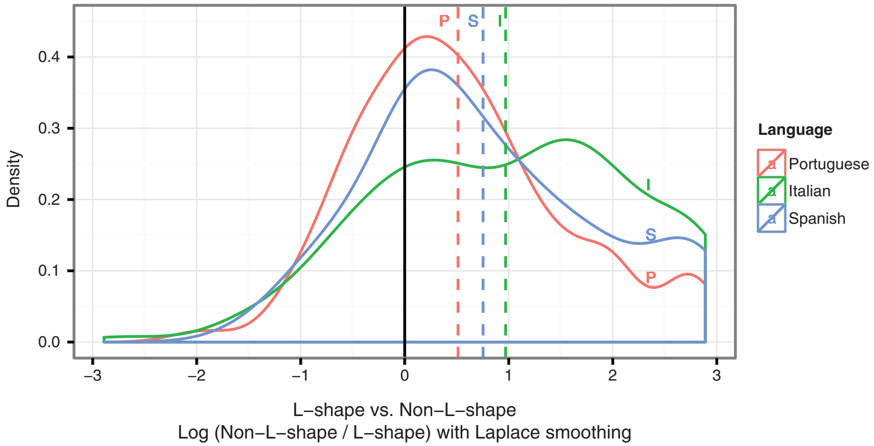

Figure 2: Density of Response Preference by Participant – Portuguese (Red), Italian (Green), Spanish (Blue); the solid line indicates the neutral response preference, and the dashed lines indicate the mean of the log ratio for each language.

with a resulting scale from −3 to 3, where 0 indicates no preference. Figure 2 (Portuguese) shows a by-participant plot, and Figure 3(a) shows a by-item plot; each of them show the distribution of Natural responses (by participant and by item) relative to the midline.

Figure 2 (Portuguese) shows that there is strong preference for Natural responses by participant. The majority of the participants (65%, 163 out of 249) have such a preference, as indicated by their positive log ratio. The average of the log-ratio across *all* participants is 0.51, which lies in the positive range (recall that 0 is neutral). The paired Wilcoxon signed rank test indicates that the two responses differed significantly, with more Natural responses than L-shaped responses ($W+$: 21,852.5, $W-$: 6,588.5, N: 249, $p = 6.582 \times 10^{-13***}$).

Figure 3(a) shows that there is a strong preference for Natural responses by item. All of the items (100%, 15 out of 15) have such a preference, as indicated by their positive log ratio. The average of the log-ratio across all items is 0.50, which lies in the positive range. The paired Wilcoxon signed rank test indicates that the two responses differed significantly, with more Natural responses than L-shaped responses ($W+$: 120, $W-$: 0, N: 15, $p = 0.000723^{***}$).

The F1 and F2 analyses together suggested that the L-shaped response is not preferred and that there is an overwhelming bias to the Natural response. Having evaluated our responses by participant and by item, the next section presents mixed-effects models to identify any potential predictors that have an effect on the response preferences we found.

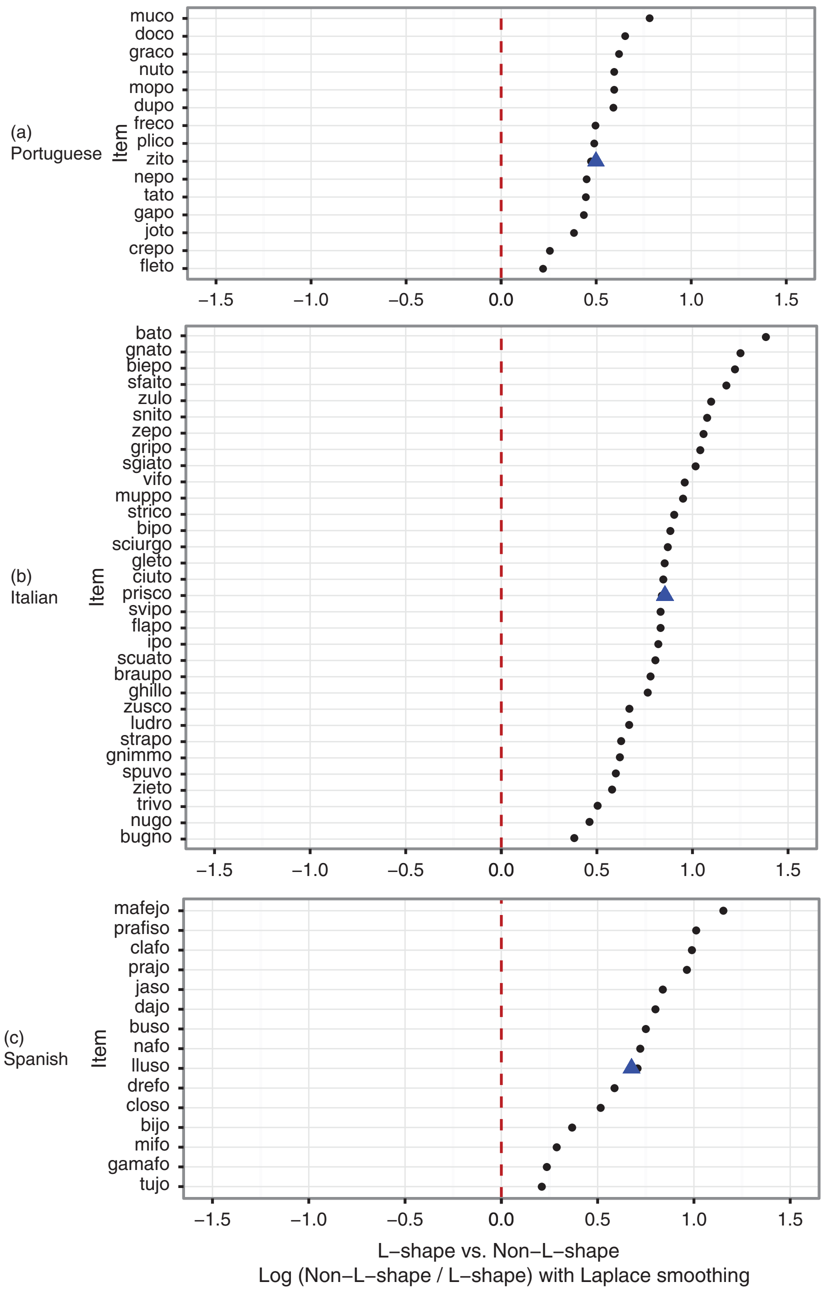

Figure 3: Response Preference by Item – (a) Portuguese, (b) Italian, (c) Spanish; the red line indicates the neutral response preference, and the blue triangle indicates the mean of the log ratio.

The *glmer* function from *lme4* (Bates et al. 2014) in *R* (R Core Team 2013) was used to construct logistic mixed-effects models, with the *bobyqa* optimizer. The predictee and predictors are listed below.

> Predictee: Response (fricative (Natural response) vs. stop (L-shaped response))
>
> Predictors of fixed effects: *Group* (I → S vs. S → I), *Frame* (L-shaped last vs. L-shaped first), *Conjugation* (1st vs. 2nd), and *Place* (coronal, labial, velar). All predictors were dummy-coded with the exception of *Place*, which was sum-coded since there were 3 places.
>
> Predictors of random effects: *Participant* and *Item*

Appendix 1 describes the procedure for arriving at the best model. Once this process was completed, R^2_{GLMM}, the percentage of variance explained by the final model, was calculated (Nakagawa and Schielzeth 2013; Johnson 2014; Barto 2014) with the marginal R^2_{GLMM} yielding 8.5% and conditional R^2_{GLMM} yielding 42.1%. Marginal R^2_{GLMM} represents the variance explained by fixed factors and conditional R^2_{GLMM} represents the variance explained by both fixed and random factors. Importantly, the value for conditional R^2_{GLMM} shows that a majority of the variance in the data, 57.9%, remains to be explained beyond the terms in (9) alone, and thus indicates that the global difference between the L-shaped and Natural responses cannot be fully captured by design-specific factors.

The complete summary of the final model is shown in Table 1. The model suggests the following significant predictors – *Group*, *Frame*, *Conjugation*, *Place* as well as an interaction term between *Group* and *Frame*.

Table 1 allows us to identify which predictors best explained variance in the results. We now turn to F1-F2 analyses within each a subset of the data by each of these predictors in turn.

4.1.2 F1-F2 analyses with subset data: European Portuguese

To evaluate whether the L-shaped pattern is preferred or not given these predictors, we applied a series of non-parametric paired t-tests on the data by *Participant* and *Item*, just as in Section 4.1.1, but crucially we split the data by the significant predictors found during mixed-effects modelling.

Table 2 summarises the results of the F1-analyses, and Table 3 summarises the results of the F2-analyses. The by-item analyses were conducted with a subset of each of the three places of articulation, with the α-level less than 0.05 considering a one-tailed test.

Table 1: Best mixed-effects model – Portuguese.

Fixed effects	Estimate	SE	z	$p(>\|z\|)$
(Intercept)	−0.24	0.13	−1.864	0.0624^{+}
Group[I-S]	−1.75	0.22	−7.830	4.88×10^{-15}***
Frame[L-last]	−0.44	0.19	−2.313	0.0207*
Conjugation[2nd]	0.34	0.09	3.811	0.000138***
Place[Cor. vs. All Places]	0.12	0.07	1.847	0.0648^{+}
Place[Lab. vs. All Places]	0.03	0.07	0.387	0.698
Group[I-S]:Frame[L-last]	1.69	0.30	5.630	1.80×10^{-8}***

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, $^{+}p < 0.1$

Random effects	Variance
Participant.(Intercept)	1.00
Participant.Group[I-S]	1.43
Participant.Frame[L-last]	2.82
Item.(Intercept)	0.01
Residual	1.00

Data size	N
Observations	3,343
Participants	249
Items	15

Table 2: F1-analyses with subset data – Portuguese.

Subset	N	W+	W–	p (two-tailed)	p (one-tailed)
Group[I-S]	121	6,225.5	1,034.5	1.03×10^{-11}***	5.15×10^{-12}***
Group[S-I]	128	4,532	2,489	0.00597**	0.00298**
Conjugation[1st]	248	20,804.5	5,301.5	6.17×10^{-15}***	3.09×10^{-15}***
Conjugation[2nd]	248	17,860	8,018	6.21×10^{-7}***	3.11×10^{-7}***
Frame[L-first]	248	20,057.5	6,048.5	1.96×10^{-12}***	9.80×10^{-13}***
Frame[L-last]	248	18,263	9,232	1.26×10^{-5}***	6.30×10^{-6}***
Place[Coronal]	247	19,154.5	8,106.5	5.63×10^{-8}***	2.82×10^{-8}***
Place[Labial]	246	18,878	7,918	5.10×10^{-8}***	2.55×10^{-8}***
Place[Dorsal]	248	21,621	6,820	2.34×10^{-11}***	1.17×10^{-11}***
Group[I-S] + Frame[L-first]	121	5,976.5	351.5	2.86×10^{-16}***	1.43×10^{-16}***
Group[I-S] + Frame[L-last]	121	4,153	2,288	0.00747**	0.00373**
Group[S-I] + Frame[L-first]	127	3,636.5	3,149.5	0.502(*n.s.*)	0.251(*n.s.*)
Group[S-I] + Frame[L-last]	126	5,079	2,302	0.000312***	0.000156***

Table 3: F2-analyses with subset data – Portuguese.

Subset	N	W+	W–	p (two-tailed)	p (one-tailed)
Group[I-S]	15	120	0	0.000721***	0.000361***
Group[S-I]	15	116	4	0.00161**	0.000805***
Conjugation[1st]	15	120	0	0.000723***	0.000362***
Conjugation[2nd]	15	120	0	0.000716***	0.000358***
Frame[L-first]	15	120	0	0.000714***	0.000357***
Frame[L-last]	15	105	0	0.00109**	0.000546***
Place[Coronal]	5	15	0	0.0625+	0.0313*
Place[Labial]	5	15	0	0.0625+	0.0313*
Place[Dorsal]	5	15	0	0.0625+	0.0313*
Group[I-S] + Frame[L-first]	15	120	0	0.000718***	0.000359***
Group[I-S] + Frame[L-last]	15	117	3	0.00132**	0.000659***
Group[S-I] + Frame[L-first]	15	63	28	0.233 (*n.s.*)	0.117 (*n.s.*)
Group[S-I] + Frame[L-last]	15	115	5	0.001937**	0.000969***

In this series of F1 and F2 analyses with different subsets of the data, we found that all but one very specific subset yielded a significant difference between the two responses, with more Natural responses than L-shaped responses. In Table 1, splitting the data by Groups [I-S] and [S-I], each group produced significantly more Natural responses. The same was true for both Conjugations, both Frame orders, and all three Places of articulation, as shown by the first four sets of rows in Table 2 and Table 3. In other words, no matter which conjugation participants saw, no matter which place of articulation, and no matter which order of presentation (e.g., 'primacy or recency') of the two possible bases of inflection in the frame sentence, participants showed more Natural responses than L-Shaped responses in all such subsets of the data.

In the interaction between Group and Frame Order, we split them into four such groups for analysis. Within three of these four, there were significantly more Natural responses, as shown in the last four rows of Table 2 and Table 3. We turn to the one non-significant subset, *Group[S-I] + Frame[L-first]*: this was a group that saw an example sentence like (8), where participant was asked for the 1SG.IND and saw the 2SG.SBJ first in the frame, followed by the 2SG.IND form. Although the difference in responses did not reach significance, the *W+* and *W–* values still showed a tendency for more Natural responses than L-shaped responses, and the fact that difference was smaller in this subset does not seem reducible to any interpretable hypothesis about the interaction between frame order or group (as there is no general primacy or recency effect). Overall, our within-factor analysis confirmed the initial analyses that demonstrate an overwhelming bias for Natural responses.

To conclude, participants showed more Natural responses than L-shaped responses globally, with consistent results across participants and across items (as shown in Figure 2 (Portuguese) and Figure 3(a)). There were significantly more Natural responses whether probed for the 1SG.IND or for the 2SG.SBJ, with the effect being even stronger in the latter group. There were significantly more Natural responses whether the wug verb was 1st conjugation or 2nd conjugation, with the effect being even stronger in the former group. There were significantly more Natural responses whether the L-shaped form was presented first or last in the frame, with the effect being even stronger in the former group. There were significantly more Natural responses across all three Places of articulation, with no real differences among these three groups. Finally, there were significantly more Natural responses across most combinations of Frame and Group, with a weak primacy effect in only specific combinations, which slightly lowered the rate of Natural responses but still retained their numerical primacy.

4.2 Interim conclusion: European Portuguese

Neither group displayed results compatible with the predictions of the L-morphome theory. Speakers given a chance to base the 2SG.SBJ form on the 1.IND form largely did not do so, preferring instead to maintain paradigmatic uniformity across Persons. Speakers given a chance to base the 1.IND form on the 2.SBJ also largely did not do so, showing instead a preference for uniformity within Mood. Put differently, given a choice between applying a diagonal syncretism pattern that is robustly present in the lexicon or a horizontal/vertical pattern that is favored from the point of view of morphological naturalness, participants as a whole opted for the latter.

We wish to remark here on the fact that as Figure 2 (Portuguese) shows, 35% of participants used the L-shape responses more than the Natural responses. What explains this subgroup of participants, as opposed to the majority? Were these people more metalinguistically aware – what made them more ‘in tune’ with the L-shaped pattern found in the lexicon? It is worth remarking that all the participants were recruited through a computer-based social network of educated adults, and therefore, although we do not have any way of knowing for sure, some of them might have explicitly been taught that one ‘builds the subjunctive on the basis of the first person singular’, as is formulated in traditional grammars. We will return to this point in the general discussion.

As shown in the detailed results in Table 3 (the results are entirely parallel in the F1 analysis in Table 2), even though a very wide range of experimental manipulations were conducted with the stimuli and their mode of presentation, none of these modulated the core finding: participants prefer the Natural response to the L-Shape response, thereby providing no evidence for the L-Morphome as having an effect in the online production of novel inflected forms. As Figure 3(a) shows, the items were very well-constructed and consistent, with no Place effect – it doesn't matter whether participants saw a phonological alternation involving labials, or coronals, or velars. It also doesn't matter whether they see the putative L-shaped base first or last, and even more importantly, it doesn't matter whether the novel verb was drawn from the 1st conjugation or the 2nd conjugation. For any of these cases, Natural responses won out – namely basing the form of the 2SG.SBJ on the 2SG.IND (thereby maintaining identity across Persons), or basing the 1SG.IND on the 2SG.SBJ (and thereby maintaining identity within Mood). This latter point deserves a bit more elaboration: one might think that the L-morphome is alive and kicking, but only works in one direction (say, from the 1SG.IND to the subjunctive forms, but not vice versa). This wasn't the case, however: in either direction, a Natural response was preferred – even when the Naturalness bias varied between being Person or Mood – in short, anything that 'makes sense' grammatically will be latched onto instead of an arbitrary L-morphome. However, the bias within Person (e.g. the IS group) was slightly stronger than that within Mood (the SI group). We will return to an interpretation of a stronger bias for Person than Mood in the context of the General Discussion in Section 7.

In terms of other comparisons, it should be noted that the Natural responses emerge even more easily in the 1st conjugation than the 2nd – which is to be expected – but crucially emerge in both. In other words, while one might expect the Morphome effect to be confined to the 2nd conjugation (as well as the 3rd conjugation), and hence be absent outside of this context, in fact it was absent in both. In short, whether the phonology (place of articulation), morphology (conjugation class), or morphosyntax (direction of implication, e.g. from indicative to subjunctive) are manipulated, nothing helps the L-morphome to get its head out over the parapet. Finally, the sentential frames had no confounding effect – no general theory of primacy or recency can explain the effects, which hold stably across combinations of order of presentation and item.

5 Replication in Italian

In the previous section, we reported the finding that the L-morphome was not sufficiently strong enough to show any consistent effect in experimental tasks with novel inflecting verbs in contexts where it is predicted that it should be used. But one could always counterargue that Portuguese may not be the 'right' Romance language for this test. Portuguese, one might say, doesn't have enough L-morphomes in its lexicon to provide support. Or perhaps the problem is that the alternations we showed were based on consonantal morphophonology, rather than vowels, as the latter are perhaps more entrenched than the former in contemporary Portuguese (John Charles Smith, pers. comm.). Perhaps there are other reasons having to do with the particulars of the subjunctive in Portuguese that led to less evidence for the L-shaped morphome as synchronically productive. To address all of these potential objections, we aimed to conduct the same basic task but to vary other independent variables. (Anticipating the results, however, the unproductiveness of the L-shaped morphome does not turn out to be unique to Portuguese). We chose Italian, another Romance language for which the L-morphome has been proposed, as shown below:

(10)	'to come'	IND	SBJ
	1sg	**veng**-o	**veng**-a
	2sg	vien-i	**veng**-a
	3sg	vien-e	**veng**-a

We maintained the same basic design and length as the Portuguese version of the experiment, but introduced a new factor, namely consonant-based versus vowel-based morphomic correspondence. Indeed, as some literature (e.g. Steriade 2008) has proposed, certain cases of 'unnatural' base selection may be more widespread with consonant-based factors (e.g. presence or absence of palatalization) as opposed to vowel-based ones, for reasons that we do not yet understand. For these reason, we included a high vowel/mid vowel alternation e.g. *mopp-o* ~ *mupp-i*) of the type found in Romance languages (e.g. Portuguese itself) – though not in Standard Italian. We contrasted this with a consonantal alternation between stops and nasals (e.g. *svip-o* ~ *svim-i*) (where stop/nasal alternations are used for inflectional distinctions in the Celtic languages) in order to see whether L-shaped responses fared better in one or the other of these. This factor, equally balanced in our experimental items (as we did with Place in Portuguese, above), is henceforth referred to as Alternation (Nasal vs Vowel). Conjugation class was discarded as a factor, as the 1st and 2nd person present indicative are not distinct among conjugation classes in Italian. Where distinguished (in the subjunctive), we employed unambiguously 2nd conjugation verbs, to favor the likelihood of L-shaped responses. Representative fillers and stimuli are thus as follows, where we balanced between presentation of frame (L-first vs. L-last) within these as well, as we did for Portuguese (see (5) vs (6) and (7) vs (8)).

(11) Io muppo ogni giorno, ma tu moppi soltanto una volta alla settimana. È meglio che anche tu ____ più frequentemente.
'I *muppo*1SG.IND every day, but you *moppi*2SG.IND only once a week. It's better if you, too, ____2SG.SBJ more frequently'

(12) Tu quando svimi? Spero che tu non svipa troppo tardi. Così ____ anch'io in contemporanea.
'When do you *svimi*2SG.IND? I hope you don't *svipa* too late. In that case, I'll ____1SG.IND at the same time.'

Before the experiment, participants were presented with instructions that were adapted translations of the following: "You will be presented with examples of invented verbs, such as *io marbo, tu marbi*. Then you will see a sentence with a blank space. Your task is to fill in the blank with the appropriate form of the verb".

5.1 Results: Italian

The same analysis procedure was used as in the case of Portuguese. We performed a participant analysis (an F1-analysis) and an item analysis (an F2-analysis) with non-parametric paired t-tests. This allowed us to examine if the two responses (L-shaped or Natural) were significantly different and which response is preferred. Secondly, we aimed to identify any significant predictors for the responses with mixed-effects modelling. Finally, we again performed participant analyses and item analyses on the subsets of the data, divided by each significant predictor. This latter method allow us to identify if the preference of a particular response would still hold after taking into account other factors. After preprocessing the responses in the method described for Portuguese, we were left with 135 participants and 1,996 responses.

5.1.1 F1-F2 analyses: Italian

To visualise the response preference by participant and by item, we applied a log ratio metric with Laplace smoothing with the two response types, a log ratio between the number of Natural responses and the number of L-shaped responses. Figure 2 (Italian) shows a by-participant plot, while Figure 3(b) shows a by-item plot. Figure 2 shows that there is strong preference for Natural responses by participant. The majority of the participants (72.5%, 98 out of 135) have such a preference, as indicated by their positive log ratio. The average of the log-ratio across all participants is 0.97 which lies in the positive range. The paired Wilcoxon signed rank test suggested that the two responses differed significantly, with more Natural responses than L-shape responses ($W+$: 6,999.5, $W-$: 1,001.5, N: 135, $p = 2.76 \times 10^{-13***}$).

Figure 3(b) shows that there is strong preference for Natural responses by item. All of the items (100%, 31 out of 31) have such a preference indicated by their positive log ratio. The average of the log-ratio across all items is 0.86, which lies in the positive range. The paired Wilcoxon signed rank test suggested that the two responses differed significantly, with more Natural responses than L-shaped responses ($W+$: 528, $W-$: 0, N: 32, $p = 8.22 \times 10^{-7***}$).

Our F1 and F2 analyses together indicate that the L-shaped response is not preferred and there is an overwhelming bias to the more Natural response. Having evaluated our responses by participant and by item, in the next section, mixed-effects models are used to identify any potential predictors that have an effect on the response preferences we found.

The *glmer* function from *lme4* (Bates et al. 2014) in *R* (R Core Team 2013) was used to construct logistic mixed-effects models, with the *bobyqa* optimizer. The predictee and predictors are listed below.

Predictee: Response (L-shaped vs. Natural)

Predictors of fixed effects: *Group* (I → S vs. S → I), *Frame* (L-shaped last vs. L-shaped first), and Alternation (Nasal vs. Vowel). All predictors were dummy coded.

Predictors of random effects: *Participant* and *Item*

Appendix 1 describes the procedure for arriving at the best model. Once this process was completed, R^2_{GLMM}, the percentage of variance explained by the model, was calculated (Nakagawa and Schielzeth 2013; Johnson 2014; Barto 2014) with marginal R^2_{GLMM} yielding 3.65% and conditional R^2_{GLMM} yielding 38.9%. Marginal R^2_{GLMM} represents the variance explained by fixed factors and conditional R^2_{GLMM} represents the variance explained by both fixed and random factors. The conditional R^2_{GLMM} shows that a majority of the variance in the data, 61.1%, remained to be explained and thus indicates that the global difference between the L-shape and Natural responses cannot be fully captured by design-specific factors.

The complete summary of the final model is shown in Table 4. This model contains the following significant predictors: *Group* (namely S-I or I-S) and *Alternation* (Nasal vs. Vowel). The factor *Frame* (L-first or L-last within the frame) was not significant on its own, but there were significant terms between *Group* and *Frame* (L-first or L-last within the frame), and between *Frame* and *Alternation*, as well as a three-way interaction between *Group*, *Frame* and *Alternation*. In the next section, the F1-F2 analyses will be repeated with subsets

Table 4: Best mixed-effects model – Italian.

Fixed effects	Estimate	SE	z	$p(>\|z\|)$
(Intercept)	−1.84	0.27	−6.730	1.70×10^{-11}***
Group[I-S]	1.09	0.41	2.651	0.00804**
Frame[L-last]	0.32	0.22	1.479	0.139
Alternation[Vowel Lower]	0.81	0.40	2.026	0.0428*
Group[I-S]:Frame[L-last]	−0.54	0.32	−1.675	0.0940^{+}
Group[I-S]:Alternation[Vowel Lower]	−0.81	0.58	−1.389	0.165
Frame[L-last]:Alternation[Vowel Lower]	−1.07	0.32	−3.312	0.000925***
Group[I-S]:Frame[L-last]:Alternation[Vowel Lower]	1.26	0.46	2.758	0.00581**
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{+}p < 0.1$				
Random effects			**Variance**	
Participant.(Intercept)			1.90	
Item.(Intercept)			0.00	
Residual			1.00	
Data size			**N**	
Observations			1,996	
Participants			135	
Items			32	

of the data based on these predictors, which follows the procedure above of splitting the data into multiple groups and analysing the significance of the Natural preference within each.

5.1.2 F1-F2 analyses with subset data: Italian

To evaluate whether the L-shape pattern is preferred or not given these predictors, we applied a series of non-parametric paired t-tests on the data by *Participant* and *Item* with the same subsetting approach as before. In the F1 and F2 analyses of Portuguese, we performed two-tailed tests, which analyse statistical significance regardless of the direction of the relationship. This decision is justified by the fact that we had no a priori knowledge of the expected direction for each and every one of the predictors. Unlike the F1 and F2 analyses of Portuguese, for Italian we performed one-tailed tests, which analyse statistical significance in the one direction of interest, as the analyses of Portuguese provided an a priori expected direction of interest.

Table 5 summarises the results of the F1-analyses. Table 6 summarises the results of the F2-analyses. In this series of F1 and F2 analyses with different

Table 5: F1-analyses with subset data – Italian.

Subset	N	W+	W–	*p* (one-tailed)
Group[I-S]	61	1,241.5	354.5	0.000149***
Group[S-I]	74	2,332	153	8.92×10^{-11}***
Alternation[Nasal]	70	1,981	230	1.11×10^{-8}***
Alternation[Vowel]	65	1,548	282	1.57×10^{-6}***
Group[I-S] + Frame[L-first]	61	1,134.5	461.5	0.003**
Group[I-S] + Frame[L-last]	61	1,187.5	352.5	0.000231***
Group[S-I] + Frame[L-first]	74	2,255.5	372.5	6.08×10^{-8}***
Group[S-I] + Frame[L-last]	74	2,435	266	1.17×10^{-9}***
Frame[L-first] + Alternation[Nasal]	70	1,954	392	8.81×10^{-7}***
Frame[L-first] + Alternation[Vowel]	65	1,389	441	0.000239***
Frame[L-last] + Alternation[Nasal]	70	1,858.5	352.5	7.29×10^{-7}***
Frame[L-last] + Alternation[Vowel]	65	1,679	274	4.06×10^{-7}***
Group[I-S] + Frame[L-first] + Alternation[Nasal]	29	285.5	120.5	0.0306*
Group[I-S] + Frame[L-first] + Alternation[Vowel]	32	289.5	116.5	0.0247*
Group[I-S] + Frame[L-last] + Alternation[Nasal]	29	272	79	0.00724**
Group[I-S] + Frame[L-last] + Alternation[Vowel]	32	330.5	104.5	0.00738**
Group[S-I] + Frame[L-first] + Alternation[Nasal]	41	736	84	5.70×10^{-6}***
Group[S-I] + Frame[L-first] + Alternation[Vowel]	33	419	109	0.00189**
Group[S-I] + Frame[L-last] + Alternation[Nasal]	41	720	100	1.53×10^{-5}***
Group[S-I] + Frame[L-last] + Alternation[Vowel]	33	521	40	8.59×10^{-6}***

subsets of the data, we found that in every individual row in these tables there was a significant difference between the two responses, with more Natural responses than L-shaped responses. Summarizing, in Italian the Natural responses were chosen significantly more than L-shaped responses overall, and in individual subgroups, whether analysed separately within the consonantal items (stop ~ nasal alternations), within the vocalic items (mid~ high alternations), whether analysed separately by Group (probed for the 2SG.SBJ or proved for the 2SG.IND), whether analysed by Frame on its own or in two-way interactions, and in the overwhelming majority of three-way interactions between group, frame, and alternation type. Whether the phonology (place of articulation), morphophonology (vowel height alternation or consonantal manner alternation), or morphosyntax (direction of implication, e.g. from indicative to subjunctive) are manipulated, the L-morpheme is always trumped by the Natural response in Italian. Indeed, as Figure 2 (Italian) shows, 72.5% of the participants chose the Natural response more often than the unnatural morphomic one, and as Figure 3(b) shows, the consistency of Natural responses held

Table 6: F2-analyses with subset data – Italian.

Subset	N	W+	W–	*p* (one-tailed)
Group[I-S]	32	518	10	1.03×10^{-6}***
Group[S-I]	32	528	0	4.10×10^{-7}***
Alternation[Nasal]	16	136	0	0.000235***
Alternation[Vowel]	16	136	0	0.000238***
Group[I-S] + Frame[L-first]	32	453	12	2.87×10^{-6}***
Group[I-S] + Frame[L-last]	32	438.5	26.5	1.14×10^{-5}***
Group[S-I] + Frame[L-first]	32	523	4.5	6.26×10^{-7}***
Group[S-I] + Frame[L-last]	32	528	0	3.98×10^{-7}***
Frame[L-first] + Alternation[Nasal]	16	136	0	0.000233***
Frame[L-first] + Alternation[Vowel]	16	134.5	1.5	0.000316***
Frame[L-last] + Alternation[Nasal]	16	136	0	0.000237***
Frame[L-last] + Alternation[Vowel]	16	136	0	0.000240***
Group[I-S] + Frame[L-first] + Alternation[Nasal]	16	119	1	0.000425***
Group[I-S] + Frame[L-first] + Alternation[Vowel]	16	112.5	7.5	0.00152**
Group[I-S] + Frame[L-last] + Alternation[Nasal]	16	118	2	0.000805***
Group[I-S] + Frame[L-last] + Alternation[Vowel]	16	110	10	0.00240**
Group[S-I] + Frame[L-first] + Alternation[Nasal]	16	136	0	0.000235***
Group[S-I] + Frame[L-first] + Alternation[Vowel]	16	131.5	4.5	0.000551***
Group[S-I] + Frame[L-last] + Alternation[Nasal]	16	136	0	0.000235***
Group[S-I] + Frame[L-last] + Alternation[Vowel]	16	136	0	0.000234***

across items quite consistently. Interestingly, the bias within Person (e.g. the IS group) was less strong than that within Mood (the SI group) for the Italian participants.

5.2 Interim conclusion: Italian

Italian speakers were no more likely to prefer the L-morphome than a Natural response than European Portuguese speakers, despite the fact that L-morphomic examples run rampant throughout the history of Italian and in present day synchronic forms, as shown by Maiden (2005). Nonetheless, when given an opportunity to extend this pattern of diagonal syncretism to novel verbs that have a distinct 1SG.IND and 2SG.IND form (or a distinct 2SG.IND and 2SG.SBJ form), speakers did not do so, no matter the whether the forms differed in a consonantal or vocalic alternation.

6 Replication and test of person in Spanish

Beyond replication in a third language, we sought to investigate a new independent variable, given potential objections to our previous results, that often go along the following lines. Perhaps we have indeed shown that Uniform-Person trumps the L-shape – but maybe the L-shape is still present in speakers' minds, it's just overshadowed by other, stronger biases. For this reason, we compared generalization of 1SG.IND and 2SG.IND with 2SG.SBJ versus 1SG.IND and 2SG.IND with 3SG.SBJ, the idea being that the 3SG.SBJ is a different person from 2SG.IND and hence might allow the L-shape to exert its force. In other words, we decided to explore the L-morphome in a third language, Spanish, maintaining the same basic comparison of diagonal syncretism against natural syncretism:

(13)

'to hear'	IND	SBJ
1sg	**oig**-o	**oig**-a
2sg	oy-es	**oig**-as
3sg	oy-e	**oiga**-a

The new manipulation however was the following: when the 1SG.IND and 2SG.IND form are provided and, by contrast, the 3SG.SBJ is probed, Uniformity of PERSON makes no specific prediction, while the L-Morphome theory predicts use of the 1SG.IND base. In a sense, testing the I $\rightarrow S_3$ condition potentially gives the L-Morphome theory its best chance, as we are comparing one diagonal syncretism against another. We had 15 items and 15 fillers, much like the Portuguese experiment, with the distribution of stop and fricative alternants reversed (stop in 2SG.IND, fricative in L-shaped forms), to additionally test whether a blanket preference for fricatives was responsible for those results.

We compared I$\rightarrow S_2$ in frames such as (14) and I$\rightarrow S_3$ in frames such as (15).

(14) Tú llutes solamente con la mano derecha, pero yo lluso con cualquier mano. Es necesario que ____ con las dos manos para que así seas más productivo.
'You *llutes*2SG.IND only with your right hand, and but I *lluso*1SG.IND with either hand. It's necessary that you ____2SG.SBJ with both hands in order to be more productive.'

(15) El periódico local dice que yo mifo desastrosamente. Tú mipes bien, pero el periódico no dice nada de ti directamente. Pero dijeron que esperan que mi hijo ____ como tú en la competición.
'The local newspaper says that I *mifo*1SG.IND disastrously. You *mipes*2SG.IND well, but the newspaper did not mention you directly. However, they said they hope that my son ____3SG.SBJ like you in the competition.'

Before the experiment, participants were presented with instructions that were adapted translations of the following: "You will be presented with examples of invented verbs, such as *yo tulo, tú tules*. Then you will see a sentence with a blank space. Your task is to fill in the blank with the appropriate form of the verb".

6.1 Results: Spanish

The same analysis procedure was used as in the case of Portuguese and Italian. After preprocessing and cleaning the responses, we were left with 107 participants and 1,539 responses.

6.1.1 F1-F2 analyses: Spanish

To visualise the response preference by participant and by item, we applied a log ratio metric with Laplace smoothing with the two response types, *log (the number of Natural responses/the number of L-shape responses).* See Figure 2 (Spanish) for a by-participant plot, and Figure 3(c) for a by-item plot.

Figure 2 (Spanish) shows that there is strong preference for Natural responses by participant. The majority of the participants (71.9%, 77 out of 107) have such a preference indicated by their positive log ratio. The average of the log-ratio across all participants is 0.75 which lies in the positive range. The paired Wilcoxon signed rank test suggested that the two responses differed significantly, with more Natural responses than L-shape responses ($W+$: 4,306.5, $W-$: 844.5, N: 107, $p = 4.43 \times 10^{-9***}$).

Figure 3(c) shows that there is strong preference for Natural responses by item. All of the items (100%, 15 out of 15) have such a preference indicated by their positive log ratio. The average of the log-ratio across all participants is 0.68 which lies in the positive range. The paired Wilcoxon signed rank test indicated that the two responses differed significantly, with more Natural responses than L-shape responses ($W+$: 120, $W-$: 0, N: 15, $p= 0.0007229^{***}$).

Our F1 and F2 analyses together suggested that the L-shape response is not preferred and that there is an overwhelming bias to the more Natural response. Having evaluated our responses by participant and by item, in the next section, mixed-effects models were used to identify any potential predictors that have an effect on the response preferences we found.

The *glmer* function from *lme4* (Bates et al. 2014) in *R* (R Core Team 2013) was used to construct logistic mixed-effects models, with the *bobyqa* optimizer. The predictee and predictors are listed below.

Predictee: Response (stop (non-L-shaped) vs. fricative (L-shaped))

Predictors of fixed effects: *Group* (I → S vs. S → I), *Frame* (L-shaped last vs. L-shaped first), *Person* (2nd vs. 3rd), and *Place* (coronal, labial, velar). All predictors were dummy-coded with the exception of *Place* which was sum-coded since there were 3 places.

Predictors of random effects: *Participant* and *Item*

Appendix 1 describes the procedure for arriving at the best model. Once this process was completed, R^2_{GLMM}, the percentage of variance explained by the model, was calculated (Nakagawa and Schielzeth 2013; Johnson 2014; Barto 2014) with marginal R^2_{GLMM} yielding 5.8% and conditional R^2_{GLMM} yielding 49.4%. Marginal R^2_{GLMM} represents the variance explained by fixed factors and conditional R^2_{GLMM} represents the variance explained by both fixed and random factors. The conditional R^2_{GLMM} shows that a majority of the variance in the data, 50.6%, remained to be explained and thus indicates that the global difference between the L-shape and Natural responses cannot be fully captured by design-specific factors.

The complete summary of the final model is shown in Table 7. The model suggests the following significant predictors: *Group, Place*, *Group:Place*, *Frame: Place*, *Group:Frame:Place* and *Group:Place:Person*. In the next section, the F1-F2 analyses will be repeated with a subset of the data by each of these predictors in turn, which follows the procedure above of splitting the data into multiple groups and analysing the significance of the Natural preference within each.

6.1.2 F1-F2 analyses with subset data: Spanish

To evaluate whether the L-shaped pattern is preferred or not given these predictors, we applied a series of non-parametric paired t-tests on the data by *Participant* and *Item* with the same subsetting approach as before.

Table 8 summarises the results of the F1-analyses. Table 9 summarises the results of the F2-analyses. In this series of F1 and F2 analyses with different subsets of the data, we found that a majority of the tests yielded a significant difference between the two responses, with more Natural responses than L-shaped responses. For example, in Table 8, the preference for Natural responses holds significantly in both directions of generalization (1SG.IND to a subjunctive form and vice-versa), across all places of articulation, and across all interactions of these two factors. It also holds across all interactions of Frame type (L-first vs. L-last) and Place. As for the three-way interactions, the

Table 7: Best mixed-effects model – Spanish.

Fixed effects	Estimate	SE	z	$p(>\|z\|)$
(Intercept)	−0.76	0.27	−2.836	0.00456**
Group[I-S]	−0.90	0.42	−2.129	0.0333*
Frame[L-last]	−0.14	0.25	−0.556	0.578
Place[Cor. vs. All Places]	−0.63	0.27	−2.344	0.0191*
Place[Lab. vs. All Places]	0.33	0.25	1.325	0.185
Person[3rd]	0.47	0.31	1.549	0.121
Group[I-S]:Frame[L-last]	−0.18	0.38	−0.473	0.637
Group[I-S]:Person[3rd]	0.25	0.55	0.450	0.653
Group[I-S]:Place[Cor. vs. All Places]	1.09	0.37	2.920	0.00350**
Group[I-S]:Place[Lab. vs. All Places]	−0.57	0.37	−1.556	0.120
Frame[L-last]:Place[Cor. vs. All Places]	0.44	0.26	1.698	0.0895+
Frame[L-last]:Place[Lab. vs. All Places]	−0.61	0.26	−2.351	0.0187*
Place[Cor. vs. All Places]:Person[3rd]	0.41	0.30	1.393	0.164
Place[Lab. vs. All Places]:Person[3rd]	−0.06	0.29	−0.216	0.829
Group[I-S]:Frame[L-last]:Place[Cor. vs. All Places]	−0.77	0.39	−2.004	0.0451*
Group[I-S]:Frame[L-last]:Place[Lab. vs. All Places]	0.98	0.39	2.517	0.0118*
Group[I-S]:Place[Cor. vs. All Places]:Person[3rd]	−1.00	0.44	−2.295	0.0217*
Group[I-S]:Place[Lab. vs. All Places]:Person[3rd]	0.59	0.44	1.335	0.182

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{+}p < 0.1$

Random effects	Variance
Participant.(Intercept)	1.70
Participant.Group[I-S]	4.15
Participant.Frame[L-last]	1.58
Participant.Place[Cor. vs. All Places]	0.28
Participant.Place[Lab. vs. All Places]	0.28
Item.(Intercept)	0.08
Residual	1.00

Data size	N
Observations	1,539
Participants	107
Items	15

preference holds numerically across all of them, and holds significantly across 9 of the 12 Group:Frame:Place interactions and significantly across 8 of the 12 Group:Person:Place interactions. For those for which it does not hold significantly, the $W+$ and $W-$ values still showed that there was a tendency for more Natural responses than L-shaped responses, and there is no interpretable trend among the ones that do not – for example, why should there be a weaker

Table 8: F1-analyses with subset data – Spanish.

Subset	N	W+	W−	*p* (one-tailed)
Group[I-S]	54	1,233.5	197.5	2.28×10^{-6}***
Group[S-I]	53	938.5	237.5	0.000160***
Place[Coronal]	107	4,334	1,022	1.90×10^{-8}***
Place[Labial]	107	3,836	1,417	2.38×10^{-5}***
Place[Dorsal]	107	3,948	1,102	4.06×10^{-7}***
Group[I-S] + Place[Coronal]	53	1,184	301	6.41×10^{-5}***
Group[I-S] + Place[Labial]	54	1,062	369	0.00100**
Group[I-S] + Place[Dorsal]	54	1,188	138	3.61×10^{-7}***
Group[S-I] + Place[Coronal]	53	1,008	217	3.48×10^{-5}***
Group[S-I] + Place[Labial]	53	868	357	0.00515**
Group[S-I] + Place[Dorsal]	53	787	438	0.0400*
Frame[L-first] + Place[Coronal]	103	3,245.5	940.5	1.99×10^{-6}***
Frame[L-first] + Place[Labial]	106	2,846	1,340	0.00130**
Frame[L-first] + Place[Dorsal]	104	2,856	972	2.66×10^{-5}***
Frame[L-last] + Place[Coronal]	106	3,676.5	1,174.5	3.37×10^{-6}***
Frame[L-last] + Place[Labial]	105	3,002.5	1,002.5	1.58×10^{-5}***
Frame[L-last] + Place[Dorsal]	103	2,869	1,047	6.09×10^{-5}***
Group[I-S] + Frame[L-first] + Place[Coronal]	52	935	290	0.000615***
Group[I-S] + Frame[L-first] + Place[Labial]	54	870	258	0.000541***
Group[I-S] + Frame[L-first] + Place[Dorsal]	52	884.5	150.5	1.41×10^{-5}***
Group[I-S] + Frame[L-last] + Place[Coronal]	54	987.5	287.5	0.000293***
Group[I-S] + Frame[L-last] + Place[Labial]	52	711.5	323.5	0.0134*
Group[I-S] + Frame[L-last] + Place[Dorsal]	54	1,060.5	265.5	8.35×10^{-5}***
Group[S-I] + Frame[L-first] + Place[Coronal]	51	712.5	190.5	0.000470***
Group[S-I] + Frame[L-first] + Place[Labial]	52	577	413	0.168(*n.s.*)
Group[S-I] + Frame[L-first] + Place[Dorsal]	52	565.5	337.5	0.0746^{+}
Group[S-I] + Frame[L-last] + Place[Coronal]	52	868.5	307.5	0.00181**
Group[S-I] + Frame[L-last] + Place[Labial]	53	814	176	6.41×10^{-5}***
Group[S-I] + Frame[L-last] + Place[Dorsal]	49	445.5	257.5	0.0754^{+}
Group[I-S] + Person[2nd] + Place[Coronal]	27	280	98	0.0138*
Group[I-S] + Person[2nd] + Place[Labial]	27	280.5	70.5	0.00369**
Group[I-S] + Person[2nd] + Place[Dorsal]	27	293	32	0.000206***
Group[I-S] + Person[3rd] + Place[Coronal]	27	321.5	56.5	0.000680***
Group[I-S] + Person[3rd] + Place[Labial]	27	254	124	0.0587^{+}
Group[I-S] + Person[3rd] + Place[Dorsal]	27	314	37	0.000206***
Group[S-I] + Person[2nd] + Place[Coronal]	31	406	29	1.96×10^{-5}***
Group[S-I] + Person[2nd] + Place[Labial]	31	336	129	0.0163*
Group[S-I] + Person[2nd] + Place[Dorsal]	31	262.5	143.5	0.0865^{+}
Group[S-I] + Person[3rd] + Place[Coronal]	22	138	72	0.110(*n.s.*)
Group[S-I] + Person[3rd] + Place[Labial]	22	128.5	61.5	0.0890^{+}
Group[S-I] + Person[3rd] + Place[Dorsal]	22	146.5	84.5	0.143(*n.s.*)

Table 9: F2-analyses with subset data – Spanish.

Subset	N	W+	W–	*p* (one-tailed)
Group[I-S]	15	120	0	0.000358***
Group[S-I]	15	109.5	10.5	0.00266**
Place[Coronal]	5	15	0	0.0313*
Place[Labial]	5	15	0	0.0313*
Place[Dorsal]	5	15	0	0.0313*
Group[I-S] + Place[Coronal]	5	15	0	0.0313*
Group[I-S] + Place[Labial]	5	15	0	0.0290*
Group[I-S] + Place[Dorsal]	5	15	0	0.0290*
Group[S-I] + Place[Coronal]	5	15	0	0.0290*
Group[S-I] + Place[Labial]	5	14	1	0.0625^{+}
Group[S-I] + Place[Dorsal]	5	11.5	3.5	0.172(*n.s.*)
Frame[L-first] + Place[Coronal]	5	15	0	0.0290*
Frame[L-first] + Place[Labial]	5	14	1	0.0625^{+}
Frame[L-first] + Place[Dorsal]	5	10	0	0.0502^{+}
Frame[L-last] + Place[Coronal]	5	15	0	0.0313*
Frame[L-last] + Place[Labial]	5	15	0	0.0313*
Frame[L-last] + Place[Dorsal]	5	15	0	0.0313*
Group[I-S] + Frame[L-first] + Place[Coronal]	5	15	0	0.0313*
Group[I-S] + Frame[L-first] + Place[Labial]	5	15	0	0.0313*
Group[I-S] + Frame[L-first] + Place[Dorsal]	5	15	0	0.0313*
Group[I-S] + Frame[L-last] + Place[Coronal]	5	15	0	0.0290*
Group[I-S] + Frame[L-last] + Place[Labial]	5	13.5	1.5	0.0681^{+}
Group[I-S] + Frame[L-last] + Place[Dorsal]	5	15	0	0.0313*
Group[S-I] + Frame[L-first] + Place[Coronal]	5	15	0	0.0290*
Group[S-I] + Frame[L-first] + Place[Labial]	5	11	4	0.219(*n.s.*)
Group[S-I] + Frame[L-first] + Place[Dorsal]	5	11	4	0.219(*n.s.*)
Group[S-I] + Frame[L-last] + Place[Coronal]	5	15	0	0.0313*
Group[S-I] + Frame[L-last] + Place[Labial]	5	15	0	0.0290*
Group[S-I] + Frame[L-last] + Place[Dorsal]	5	8.5	1.5	0.135(*n.s.*)
Group[I-S] + Person[2nd] + Place[Coronal]	5	15	0	0.0290*
Group[I-S] + Person[2nd] + Place[Labial]	5	15	0	0.0290*
Group[I-S] + Person[2nd] + Place[Dorsal]	5	15	0	0.0290*
Group[I-S] + Person[3rd] + Place[Coronal]	5	15	0	0.0313*
Group[I-S] + Person[3rd] + Place[Labial]	5	10	0	0.05015^{+}
Group[I-S] + Person[3rd] + Place[Dorsal]	5	15	0	0.0290*
Group[S-I] + Person[2nd] + Place[Coronal]	5	15	0	0.0290*
Group[S-I] + Person[2nd] + Place[Labial]	5	15	0	0.0290*
Group[S-I] + Person[2nd] + Place[Dorsal]	5	11	4	0.219(*n.s.*)
Group[S-I] + Person[3rd] + Place[Coronal]	5	15	0	0.0502^{+}
Group[S-I] + Person[3rd] + Place[Labial]	5	12.5	2.5	0.112(*n.s.*)
Group[S-I] + Person[3rd] + Place[Dorsal]	5	12	3	0.156(*n.s.*)

preference for Natural responses when the alternating consonant is labial, and the direction of implication is from subjunctive to 1SG.IND but not in the other Labial conditions and not when the L-shaped congener appears last? Overall, our subsetting analyses confirmed our initial analyses that there is an overwhelming bias to the more natural response. The same remarks are true of the F2 analyses in Table 9.

As the Spanish experiment included a new experimental manipulation specifically aimed at the effects of Person in the subjunctive, namely comparing the amount of L-shaped responses with the 2.sg subjunctive vs. 3sg. subjunctive, we were interested in subset-based analyses of this comparison, even though the main effect of Person was not significant to begin with. Specifically, we wanted to examine the amount of L-shaped vs. Natural responses graphically depicted in the four scenarios illustrated in Figure 4. As shown in Tables 10 and 11, the Natural response was significantly chosen in all four, in both F1 and F2 analyses.

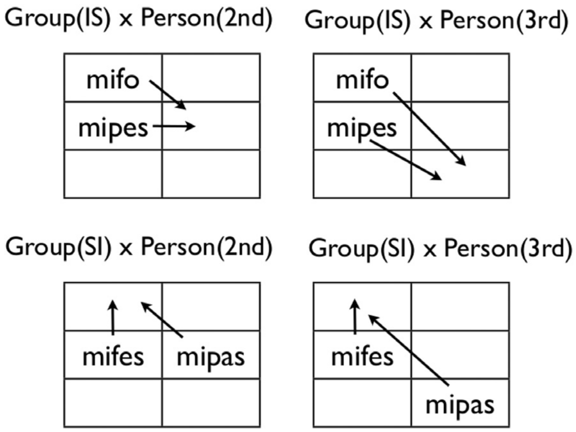

Figure 4: Schematic representation of Group:Person combinations.

Table 10: F1-analyses with data subsetted by Group and Person – Spanish.

Subset	N	W+	W–	*p* (one-tailed)
Group[I-S] + Person[2nd]	27	301.5	49.5	0.000708***
Group[I-S] + Person[3rd]	27	331.5	46.5	0.000318***
Group[S-I] + Person[2nd]	31	320	58	0.000822***
Group[S-I] + Person[3rd]	22	167	64	0.0379*

Table 11: F2-analyses with data subsetted by Group and Person – Spanish.

Subset	N	W+	W–	*p* (one-tailed)
Group[I-S] + Person[2nd]	15	120	0	0.000357***
Group[I-S] + Person[3rd]	15	105	0	0.000539***
Group[S-I] + Person[2nd]	15	111.5	8.5	0.00187**
Group[S-I] + Person[3rd]	15	90.5	14.5	0.00918**

6.2 Interim conclusion: Spanish

The overall results from Spanish are largely similar to those of the preceding two experiments: speakers did not productively extend the L-morphome to novel verbs. The present results, however, have an added twist particularly in the $\mathrm{I} \rightarrow S_3$ condition, which reveal that speakers do not go for the L-morphome pattern, even when given their best chance to extend identity between the 1SG.IND and 3SG.SBJ, as its 'competitor' base for derivation, the 2SG.IND, is also, on the face of it, a diagonal syncretism. There are two possible interpretations of this result. The first interpretation is that Uniformity of Person is more fine-grained, and makes reference to identity of [-author] persons. Under this view, 2SG.IND and 3SG.SBJ still constitute a 'horizontal' syncretism, as they share a featural value that defines the inflectional space, namely [-author]. Of course, this involves a deliberate commitment to features such as [± participant] and [± author] as dividing the featural space of Romance inflection, an assumption amply justified elsewhere (see Nevins 2007, among others).

The second interpretation for why speakers do not go for the 1SG.IND as a base for the 3SG.SBJ is that speakers are implicitly aware of the fact that the 1SG.IND is an unreliable form, and deliberately avoid using it as a base of inflection, even when the other option is not necessarily morphosyntactically more natural. Formalizing language-particular *avoidance* biases of this sort presents particular analytical challenges: speakers would seem to know that the 1SG.IND constitutes an odd-man out, to be avoided as a base for any 2nd/3rd person forms. Evidence from *defective* verbs such as Spanish *asir* 'to grasp' (Albright 2003) may provide exactly this evidence, as verbs of this sort are defective in the 1SG.IND but not in the 2SG.IND or 3SG.IND (unlike other defective verbs, e.g. *abolir* 'to abolish', which are defective in all rhizotonic (e.g. present tense indicative) and in the present subjunctive forms).

In fact, the two interpretations above need not be exclusive: the formalization in terms of identity of [-author] persons might be the specific encoding of the untrustworthiness of 1sg forms as a base for 2sg/3sg forms. For example, in

Portuguese some 1SG.IND forms (e.g. *dou* 'I give', *sei* 'I know') are completely anomalous, having a stem that is not fully shared among the subjunctive forms. Thus, contra the way the L-morphome is formulated, in fact many speakers might be actively avoiding the 1SG.IND as a base, precisely because it is a locus of irregularity. We will return to this possibility in the general discussion. For now, assuming that the person features underlying Romance inflection are in fact [$\pm$ author] and [$\pm$ participant] (extensively justified on the basis of Person Case Constraints in Romance in Nevins (2007)), the I$\rightarrow S_3$ condition in fact involves a diagonal syncretism (the L-morphome) against a *horizontal* syncretism (two identical values of the feature [-author], whether 2nd or 3rd person), and given naturalness considerations, the latter wins out.

We also wish to highlight the potential here for alternative, grammatically-based explanations for our results that may not have to do with Uniformity of Mood *per se*. It may be the case, for example, that the 2SG.IND was always chosen as the base of derivation instead of the 1SG.IND or the 2SG.SBJ because of general markedness considerations, whereby the 2SG.IND is always the unmarked choice. As such, it may be possible to formulate the grammatical pressure underlying the choice of base in generalization tasks in terms of markedness, rather than identity/faithfulness. Continued exploration of this alternative hypothesis would require a forced choice among bases that pitted the 2SG.IND against a more marked but non-L-shaped congener (e.g. 1PL.IND) or a less-marked but non-L-shaped congener (e.g. 3SG.IND); as our results at present conduct only the direct comparison between diagonal syncretism and horizontal/vertical syncretism (i.e. identity), we cast our discussion in the latter terms.

7 Summary of experiments: general discussion

These experimental results reflect an underlearning of the 'diagonal syncretism' broadly found in Romance languages between the 1SG.IND and the 2SG.SBJ, and is consistent with the conclusion that the aforementioned pattern is unnatural, precisely in the sense that these two categories do not form a natural class to the exclusion of, say, the 2SG.IND. As such the results are convergent with findings in phonology that learners favour generalizations based on natural classes (see Cristia et al. (2013) for a recent study in phonology), and indeed may be related to broader results in cognitive science about the relative difficulty of learning disjunctive generalizations (Bourne 1970; Feldman 2000).

In this section we will focus on similarities and differences in the experimental results across the three languages, as a way of providing the reader with an overall discussion of whether the morphome is ‘equally unproductive’ in all three languages, as well as some more general conclusions that are made possible by considering the results in conjunction.

As noted at the end of Section 4, the overall trend in responses was for Natural responses (e.g. those manifesting horizontal/vertical syncretisms) rather than L-shaped responses (manifesting diagonal syncretisms) – a pattern that held up whether one cuts the pie in terms of place of articulation, conjugation, direction of prediction, and order of presentation. Moreover, analyses of responses by item showed that no items were particularly more likely to undercut this trend than others. However, analyses of responses by participants revealed that 35% of the European Portuguese respondents had a log ratio with more L-morphome responses than Natural responses (where again, ‘Natural’ can be taken as a shorthand for ‘morphosyntactically motivated’). What to make of the participants who tended towards more L-morphome responses? As mentioned earlier, our hunch is that these participants may have been ones with a greater amount of metalinguistic knowledge, perhaps due to the demographics of our recruitment pool. In this respect, therefore, it may be of interest to compare the percentage of such respondents in Italian at 27.5%, and in Spanish, at 28.1% – although we do not have any direct evidence that bears on this point, it may be that, contrary to the original responses to our initial work (Martin Maiden, pers. comm. and John Charles Smith, pers. comm.) as mentioned in Section 5, European Portuguese in fact *is* the language in which the L-morphome maintains its highest degree of productivity – diminished though it may be. It is also instructive, therefore, to compare the variances among participants in each language.

To compare the variances among participants in each language, the following measure was used. The variances of the log-ratios of the overall responses were calculated using the same raw values previously shown in Figure 2. We found that European Portuguese has the lowest variance in the log-ratio responses by participant with the value 1.002, Spanish has higher variance than Portuguese with the value 1.237, and finally Italian has the highest variance with the value 1.478. While it is perhaps possible to extrapolate conclusions from these results about the speech communities in which these languages are spoken (as reasonable as it may seem to argue that, say, Italian has greater dialect variation as a whole than Portuguese), we take the comparison at the very least to reflect the composition of the homogeneity of our participant pools. In any case, Portuguese speakers across a range of experimental manipulations within the F1 analyses chose Natural responses over L-shaped ones. Our recruitment efforts were targeted towards large numbers of participants (indeed, 250 participated in the Portuguese experiment). Nonetheless, future work on the synchronic productivity of the L-morphome may wish to focus on fine-grained differences between specific populations (e.g. diverse levels of educational attainment, elementary-school children, and so forth.[2]) to see whether the number of L-morphome responses is even lower among such groups. If in fact, studies of production or elicitation corpora were to reveal that younger children make errors in which they employ the 2SG.IND stem within the subjunctive, this would constitute complementary evidence against the synchronic status of the L-morphome.

A second dimension in which the comparison of the results across these different languages is potentially revealing for further research concerns differences in the strength of Natural vs L-shaped responses in the different directions of implication. The relative strength was estimated using a ratio of W+ divided by W– obtained in the F1 analyses (Table 2, Table 5 and Table 8) – the higher the W+/W– ratio, the stronger the naturalness bias (the overall results are largely similar if conducted based on the F2 analyses). As a reminder, W+ is the sum of the ranks of those pairs for which the number of Natural responses is higher than that of L-shaped responses; *W–* is the sum of the ranks of those pairs for which the number of Natural responses is lower than that of L-shaped responses. These ratios are summarised in Table 12, which show that while Portuguese and Spanish had a stronger Naturalness bias in the case of I → S than S → I, Italian had a stronger Naturalness bias in the case of S → I. Putting it differently, is Uniformity of Person stronger than Uniformity within Mood in Spanish/Portuguese, while Uniformity of Mood is stronger in Italian? While the grouping of Portuguese and Spanish as alike in this respect is reassuring in terms of their status as sister Iberian languages, the exact reasons for why this may hold merit further investigation. One potential line of explanation may be that in Italian, the 1SG.SBJ, 2SG.SBJ, and 3SG.SBJ forms are all identical even in their desinences (while Portuguese and Spanish as 'sigmatic plural' languages have a distinct *-s* in the 2SG.SBJ form), and that this identity across all form in the subjunctive strengthens the grammatical pull of Uniformity within Mood in Italian.

Table 12: Strength of Naturalness bias between I → S and S → I.

Language	**I→ S (W+/W–)**	**S→ I (W+/W–)**
Portuguese	6.018	1.821
Italian	3.502	15.242
Spanish	6.246	3.952

2 We note however that the aim of exploring the L-morphome involves subjunctive forms, and complete acquisition of the morphosyntactic environments that trigger the subjunctive may impose a certain limit on how young participants can be.

Our experiments were conducted across a range of segmental alternations (spirantization, vowel raising, stop/nasal alternations), and the comparison between Spanish and Portuguese was specifically designed in order to exclude any low-level bias to the effect that participants simply preferred fricative-final stems over stop-final stems, as the 2SG.IND had fricatives in Portuguese but stops in Spanish. We emphasize here the importance of conducting these wug-tests with novel segmental alternations as a specific test of the predictiveness of the L-morphome – it is a specific hallmark of autonomous morphology approaches that 'morphomes' are relations of abstract identity holding among unrelated cells of a paradigm, and as such the specifics of whether the segmental alternations in question are extant or not in the language is tangential to the predictions of such theories. Given that the two stems in extant morphomic forms can be quite divergent (e.g. *ouç-* ~ *ouv-*), there is no reason to restrict experimental tests of the productivity of the morphome to phonological processes already found in the language, and in fact, from a methodological point of view, employing alternations already present in the language would invite analogy with existing irregular forms rather than a true test of the productivity of diagonal syncretism in the language – a point made in different terms in work on the irregular past tense in English, such as Prasada and Pinker (1993). Now that we have established, however, the unproductivity of the L-morphome, potential future experiments could compare extant vs non-extant stem alternations in these languages – a possibility to which we will return in the general conclusion in Section 9.

Turning to specific results from each language and their broader importance, we note the interest of the Spanish experiment probing for 3SG.SBJ forms, in which speakers still actively avoided the 1SG.IND form as a base of derivation. The predictions of a Romance inflectional identity bias among [-author] persons now enables a series of experiments specifically designed to look at such phenomena, some entirely within the indicative. As the experiments in this paper have demonstrated the kind of unnatural identity relations that speakers *do not* look for, we can now begin to investigate the ones they do. For example, potential variants on our experiments could present speakers with the 1SG.IND and the 2SG.IND and ask for the 1SG.SBJ form. If the 1SG.IND is truly avoided as a base of derivation, one might expect a failure of the L-morphome to show up here as well, even when its competitor is a different, unnatural diagonal syncretism.

In the experiment conducted with Portuguese, we compared the extent to which speakers employed an L-shaped morphome with novel verbs from the 1st conjugation versus novel verbs from the 2nd (or 3rd) conjugation. A possible pattern of results one might have expected would be that L-shaped responses would be found in the 2nd conjugation (reflecting where they reside in the lexicon), and not in the 1st. On the other hand, as shown in example (4), the L-shaped pattern (i.e. the 1SG.IND stem predicts the 2SG.SBJ stem and vice-versa) trivially holds in the 1st conjugation as well, so strictly speaking, if it is an abstract grammatical principle of syncretism, we might expect it to hold there as well. In the end of the day, as we saw, speakers in fact do not productively extend the L-morphome pattern to novel verbs from *either* the 1st or 2nd conjugation, thereby rendering its synchronic status as a principle (as opposed to a list of memorized exceptional forms) highly tenuous. Importantly, while the specific experimental manipulation of conjugations (1st or 2nd) in modulating the amount of L-shaped responses turned out to have no predictive effect in the mixed-model analysis of the results, the status of (non-trivial) L-morphomes as residing specifically in 2nd and 3rd conjugations has a major role to play in understanding the history of this pattern in Romance, and its ultimate decline, as we describe at length in the next section.

8 What led to the current decline of the L-morphome?

Taken together, the results of these experiments conducted across three different Romance languages with a variety of experimental manipulations and large numbers of participants demonstrate that the L-shaped morphome is no longer productively extended to the subjunctive. This would in turn support a view of learning biases in morphology, whereby unnatural and disjunctive patterns are biased against, and only can be learned with overwhelming amounts of evidence to the learner that such biases should be overturned. While proposals about the existence of ‘morphomes’ merit extremely strong evidence (just as proposed phenomena in phonology which counter ‘natural’ trends do), and while we have demonstrated above that the L-morphomes of Romance seem to have little to no synchronic reality, these patterns did arise in the language at some point in their history. As Maiden (2005) shows, the L-morphome pattern was productive around 500 years ago, as it involved cases in which historically non-morphomic verbs acquired this disjunctive distribution of stems. Accepting for the sake of argument that the L-morphomes were semi-productive 500 years ago (perhaps given overwhelming relative evidence to the learner) our experimental results demonstrate that something has changed since then. In the next subsection, we report on a historical corpus study conducted in order to examine what might have led to the decline of the L-morphome.

When the L-shaped morphome was productive, new verbs were ushered into this pattern. By contrast, in the twenty-first century, the diagonal syncretism found in the morphomic verbs of the 2nd and 3rd conjugations (and trivially true for even every regular verb in the language) is not extended to new verbs – speakers instead impose a pattern of Uniformity of Person or Uniformity of Mood that has less lexical support than the L-morphome, as the latter seems to have lost its productivity. What changed since then? The answer we wish to explore is, at first blush, deceptively simple: there are simply more verbs in the Romance languages now than there were in the sixteenth century. This estimation of the number of verbs is henceforth referred to as *verb vocabulary size estimation*. Of course, having more verbs alone is not enough – what seems to have crucially changed is the ratio of 1st conjugation verbs to 2nd/3rd conjugation verbs. This estimation of the ratio of 1st to 2nd/3rd conjugation is henceforth referred to as *productivity estimation*. While 1st conjugation verbs have always been more productive, the difference between them and the 2nd/3rd conjugation verbs has become more extreme over time – rather than having stayed at a stable ratio of, say, 55% 1st conjugation over time, successive generations of learners have allowed the rich to get richer, and the 1st conjugation to grow in proportion over time at a rate beyond that of the 2nd/3rd conjugation. We conducted corpus studies in Portuguese, Italian and Spanish, and found that for all three of these languages, the proportion of 1st conjugation verbs to 2nd/3rd conjugation verbs has increased over time. As a result, the salience of the L-morphomic patterns found within the 2nd/3rd conjugation have ceased to receive the amount of lexical support they need to survive, particularly in the face of the fact that it is an unnatural distribution to begin with. Details of these corpus studies that cannot be included within scope of the present article are to be found in Tang and Nevins (2013), to which the reader is referred for a more extended report and discussion. The main findings are summarised below.

8.1 Diachronic study: corpora

We examined the largest diachronic corpora available for the three languages, with two corpora for each language. For Portuguese, they were *Corpus do Português* (45 million words, spanning the 1300s to the 1900s) (Davies and Ferreira 2006) and *Colonia* (5.1 million words, spanning the early 1500s to the early 1900s) (Zampieri and Becker 2013). For Italian, they were *Google Italian Ngram* (40 billion words, spanning the early 1500s to the early 2000s) (Lin et al. 2012) and *DiaCoris* (20 million words, spanning the late 1800s to the early 2000s) (Onelli et al. 2006). For Spanish, they were *Google Spanish Ngram* (84 billion words, spanning the early 1500s to the early 2000s) (Lin et al. 2012) and *IMPACT-es* (8 million words, spanning the late 1400s to the mid 1700s) (Sánchez-Martínez et al. 2013).

To estimate both the verb vocabulary size and productivity, the entire corpus (either tokenized texts with frequency information or untokenized texts) is required for offline processing, as opposed to being only searchable through an online platform. This criterion ruled out one corpus per language, leaving us with Colonia (Portuguese), Google Italian Ngram (Italian) and Google Spanish Ngram (Spanish).[3] The three corpora that were ruled out for the verb vocabulary size estimation (Corpus do Português, DiaCoris and IMPACT-es) were used to estimate productivity, since it is possible to exhaustively search for all the *-ar*, *-er* and *-ir* words with wildcards. This allowed us to examine the consistency of the trends obtained from the other corpora. Although not reported here, we found that the same productivity trends remained robust across corpus type and size.

8.2 Diachronic study: method

To estimate the verb vocabulary size, we counted the words which were Part-Of-Speech tagged as being a verb ending with *-ar*, *-er* or *-ir* (for Italian, they were *-are*, *-ere* and *-ire* respectively). To estimate the productivity, we took the ratio of *-ar* to *-er*/*-ir*.

When comparing verb vocabulary size across different periods, we must consider the fact that the estimated size is a function of sample size (Baayen 2001), such that the larger the sample, the larger the estimated vocabulary size. To tackle this artefact, first we divided the corpus by a fixed temporal interval (an epoch) and the volume of each of these epochs, and performed random simulations. Therefore, it is not the case that our findings are simply the result of the fact that there are more texts available in more recent periods. The details of this procedure are explained in the following sections.

Firstly, to estimate changes over time, we compared the changes every *N* years. In this study, we fixed the width of the epoch to 25 years. A 25-year epoch is chosen based on two reasons – the time unit of a linguistic generation and the size of the time span of interest. Generally, the size of

3 For Spanish, although the entire corpus is available for IMPACT-es, it was not selected on the basis of both its small corpus size (8 million words compared to that of Google Spanish Ngram, 84 billion words) and narrow temporal span (1400s–1700s compared to that of Google Spanish Ngram, 1500s–2000s).

one generation is around 18 years, which is sufficient to provide indications about the stability or change in the linguistic behaviour of an individual and of the speech community (Paiva and Duarte 2013, p. 186). Given that the time span of interest was around 500 years, we set the interval to a divisor of 500.

Secondly, to compare across epochs, we equalised their volume to that of one of the smallest epochs. For each epoch, we matched the overall number of words (regardless of grammatical classes) when estimating verb vocabulary size, and similarly we matched the overall number of verbs when estimating productivity. Concretely, say that for a given corpus, the epoch with the smallest volume is 1500–1524, which contains 8,000 words/verbs, while the epoch with the second smallest volume is 1525–1549, but contains an order of magnitude more words, say 80,000 words/verbs. In this case, the 1525–1549 period would be taken as the base epoch volume for all other periods, while the period 1500–1524 would be removed in the analyses.

Finally, as each sample of the lexicon potentially contains a different number of verbs (in the case of verb vocabulary estimations) or a different number of verb types (in the case of productivity), we ran 100 random samples on each of the Spanish and Italian corpora, and 1,000 random samples on the Portuguese corpus, due to its smaller size.[4]

8.3 Diachronic study: results

As shown in Figure 5, in the period from 1525–1899 for European Portuguese, not only has the number of verb types (the solid line) grown over time with fixed vocabulary samples, but the ratio of *-ar* to *-er/-ir* has increased as well, as indicated by the 'mean productivity ratio' (the dotted line). As the sample size is fixed across epochs, but nonetheless the ratio of types within these equal sample sizes, the results thus demonstrate a growing disparity between the ratio of verbs in the first conjugation relative to those of the second and third conjugation since the sixteenth century. Similar results are presented for Italian from 1550–1974 in Figure 6 and Spanish from 1522–1996 in Figure 7.

In order to establish whether the trend of productivity is related to that of verb vocabulary size or not, a non-parametric Kendall τ rank correlation analysis was performed on Colonia, Google Italian Ngram and Google Spanish Ngram.[5]

4 We found that the results with only 100 random samples are consistent with those with 1,000 random samples.

5 A non-parametric correlation was chosen over a parametric correlation because of small N.

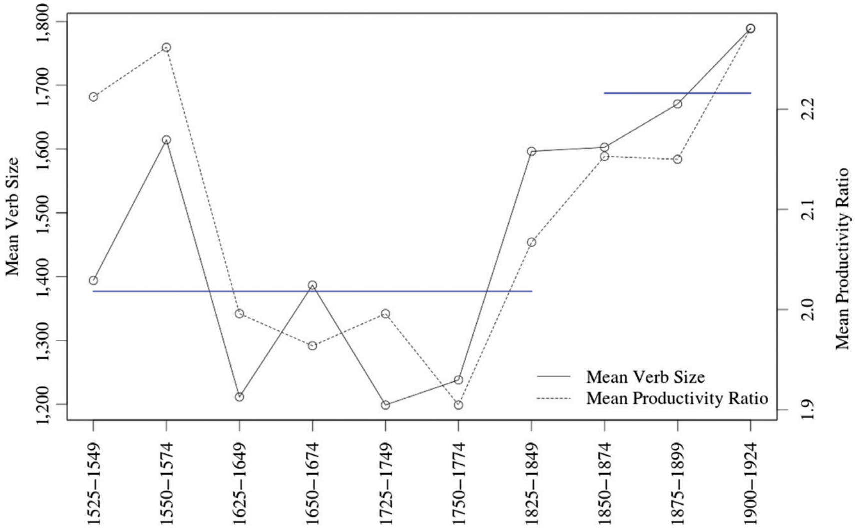

Figure 5: Temporal trends of verb vocabulary size and productivity with changepoint analysis; Language: Portuguese, Corpus: Colonia, Size: 5.1 million words, Epoch volume: 114,173 words (for verb vocabulary size estimation) and 1,252 verbs (for productivity estimation).

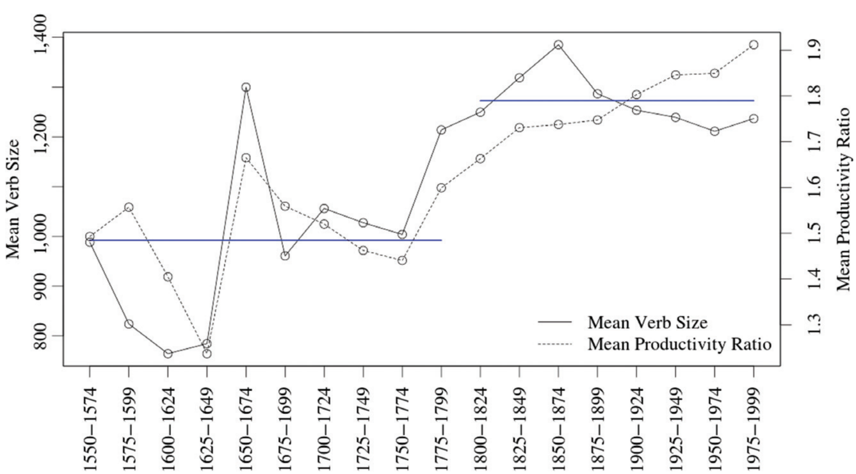

Figure 6: Temporal trends of verb vocabulary size and productivity with changepoint analysis; Language: Italian, Corpus: Google Italian Ngram, Size: 40 billion words, Epoch volume: 633,911 words (for verb vocabulary size estimation) and 5,456 verbs (for productivity estimation).

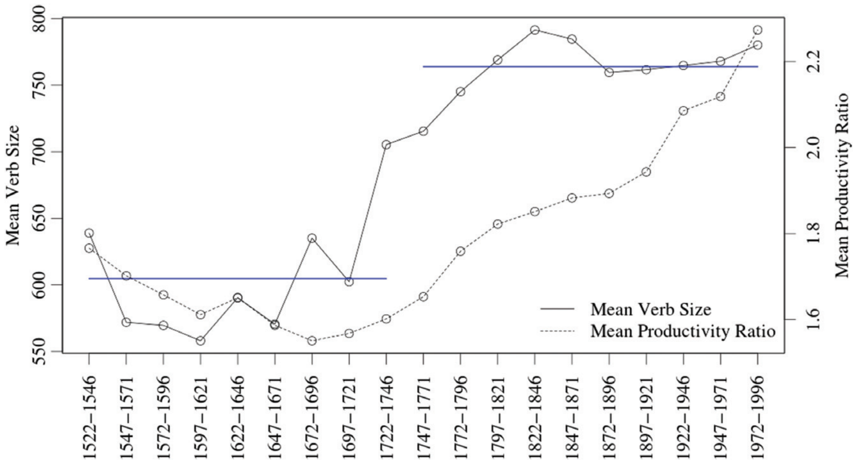

Figure 7: Temporal trends of verb vocabulary size and productivity with changepoint analysis; Language: Spanish, Corpus: Google Spanish Ngram, Size: 84 billion words, Epoch volume: 242,466 words (for verb vocabulary size estimation) and 2,646 verbs (for productivity estimation).

The mean value (of all the random samples) was used. The results are as follows: 1) Portuguese (Colonia): $r_\tau = 0.42$, $p = 0.0866$; 2) Italian (Google Ngram): $r_\tau = 0.49$, $p = 0.00393$; 3) Spanish (Google Ngram): $r_\tau = 0.54$, $p = 0.000793$. There was a strong and significant correlation between verb vocabulary size and productivity with the exception of Portuguese. Given the relatively small corpus size of Colonia, it was not expected to be as revealing, but nevertheless the correlation reached near-significance ($p < 0.1$). Visual examination of the trends in Colonia showed that the epoch 1675–1699 is likely to be an outlier in the estimation of productivity, and we found that the correlation would reach significance ($p < 0.05$) after removing this outlier epoch, yielding $r_\tau = 0.6$, $p = 0.0167$.

Having established the relationship between verb vocabulary size and productivity, the natural question to ask is when the amount of lexical support began to have a substantial impact on productivity. We did so by examining the trend in verb vocabulary further. Firstly, we observed the largest and most consistent sudden increase in verb vocabulary size at around 1750 across the three Romance languages. Secondly, we confirmed this sudden increase by performing an objective changepoint statistical analysis.

The R package *changepoint* (Killick and Eckley 2011) was used. Changepoint detection estimates the point(s) at which the statistical properties of a sequence of observations change. On the whole, there are two kinds of algorithms: single or multiple changepoint detection. Due to the relatively small number of epochs, the multiple changepoint detection method is not meaningful, since every epoch would be treated a changepoint, and we therefore employed the single changepoint detection method, which determines at most one change. Since the corpus data does not show a normal distribution, we selected Cumulative Sum (CUSUM) test statistics (Page 1954), which have no distributional assumptions.

We applied the single change detection method with CUSUM statistics to the mean values of the verb vocabulary size simulations for the Romance languages. We found there was a statistical significant change in each of the corpora, with the epochs at which this took place as follows. In the corresponding figures (Figures 5–7), a change in mean is indicated by horizontal lines depicting the mean value in different segments, where the disjunctures are the changepoints. It is tempting to speculate that that the reason for this sudden jump in the diversity of types in the lexicons of these languages is the result of the technological, medical, and educational innovations that followed in the wake of the Industrial Revolution (although we note that the changepoint appears later for Portuguese than Spanish and Italian, perhaps related to specific events in the histories of these countries).

- Portuguese (Colonia) (Figure 5): 1825–1849[6]
- Italian (Google Ngram) (Figure 6): 1775–1799
- Spanish (Google Ngram) (Figure 7): 1722–1746

Whatever the socioeconomic, cultural, and historical reasons for the changes across the lexicon of verb types over these nearly six hundred years, the result is a decreased representation of the 2nd and 3rd conjugations within the lexicon – precisely the conjugations in which the irregular and unnatural morphomic patterns reside. As a consequence, if indeed learners need increased evidence in order to incorporate and actively uptake unnatural patterns, this lexical support has dwindled over time. Even though many of the morphomic verbs have maintained a very high token frequency (allowing them to survive as memorized), their type frequency has diminished over time, and hence they go unlearned as a generalizable pattern. When the distribution of irregular alternations of this sort comes to be underrepresented and less salient in the lexicon (cf. Yang 2005) through the intrusion over time of new verbs in a different class, a morphologically unnatural pattern may cease to be productive.

To conclude, our diachronic corpus studies found that, while the 1st conjugation has always been more productive than the 2nd and 3rd conjugations,

6 The changepoint for Portuguese is the same with or without the exclusion of the outlier epoch, 1675–1699, mentioned previously in the correlation analysis.

over the last 500 years, it has become even *more* productive. Even though the verbs that show 'morphomic' patterns maintain a very high token frequency to this day, with the increase in the overall size of the lexicon and the 1st conjugation in particular, these verbs' *type* frequency has taken a plummet, and we contend that for this reason, the diagonal syncretisms present among this pocket of verbs has come to be treated as a list of memorized forms, rather than the result of an active principle of inflectional paradigm formation.

9 Conclusions, consequences, and further directions

The overall interpretation of our experiments bears directly on whether indeed, learning the patterns of inflectional paradigms is more than a data-driven puzzle, and one which involves an observable role for preferences of some types of syncretism over others. While the specific focus of the present investigations was the L-morphome of Romance and its synchronic productivity, in fact the question under study is not limited to the historical question of the L-morphome per se, but to the unnaturalness (and hence underlearning) of diagonal syncretisms more generally. The principle of featural decomposition, presented in (1) at the outset of this paper, is a way of delimiting natural classes: two (or more) inflectional forms that share a feature in common will constitute a natural class, and thereby a natural domain for syncretism. Thus, 2SG.IND and 2SG.SBJ share the same set of person features, while 2SG.IND and 1SG.IND share the same set of mood features. By contrast, 1SG.IND and 2SG.SBJ don't have any features in common (to the exclusion, say, of all the other cells in the paradigm). Once the set of forms that can potentially be identified as a natural class by means of inflectional features are identified, the mechanisms that deliver 'vertical' (within the same mood, in this case) or 'horizontal' (within the same person, in the case at hand) syncretism can be formalized in a number of ways. A variety of realizational morphological theories employ *underspecification* as a means of formalizing syncretism (e.g. Halle and Marantz 1993; Stump 2001), whereby, say, identical forms of 2nd person across moods would be formalized as exponents that simply do not realize a specific feature of mood, remaining underspecified along this dimension and thus compatible with either mood. By contrast, output-oriented theories that evaluate the structure of entire paradigms also have means of ensuring inflectional identity across certain dimensions (e.g. Kenstowicz 1997; McCarthy 2005) by means of constraints relativized to dimensions such as Person and Mood. As mentioned in Section 7, the statement of the L-morphome distribution would involve a disjunctive fractionation, referring to either the 1SG.IND or the entire SBJ column (alternatively, the 1SG forms across moods *or* the entire SBJ), and disjunctive statements are by hypothesis harder to learn and generalize.

It is important to underscore while the models above prescribe a metric for morphological naturalness (in terms of the number of constraints, features, or logical operators needed to specify the domain of syncretism), none of them would assert that diagonal syncretism is ruled out by the theory; it is simply biased-against in terms of complexity and naturalness. As we have stated above, the L-morphome was, at one point in the history of the Romance languages, thriving, despite its complexity and unnaturalness, and this is arguably because of the high token frequency and salience of this pattern in the lexicon at this point in time.

Such a case represents the kind of 'Bayesian' tradeoff between the prior probability of the hypothesis (e.g. diagonal syncretisms will be biased against) versus the likelihood of the data given the model (e.g. the amount of raw evidence for such a pattern in the data). Apparently, the token frequency (and the type frequency of the verbs displaying non-trivial L-morphomes) was sufficiently high enough in the sixteenth century (as early as our diachronic corpora stretch back to) so that even though the pattern was unnatural then too, its 'low prior' was balanced out by significant representativeness in the lexicon. Putting it more fancifully, if we had access to a time machine and ran the exact same set of stimuli and task in the experiments above in the sixteenth century (a time at which Maiden's (2005)'s observation about the L-morphome's productivity had held), it is entirely possible that participants *would* generalize the L-shaped morphome at that moment in time in the experiment, as by hypothesis, the L-morphome *was* an active principle of the grammar then, generalizable to novel forms at that stage of the language. However, as shown in Section 8, the successive over-representation of the 1st conjugation over the following five centuries left these verbs as a handful of exceptional forms, which we contend had to become memorized. In a sense, we are likening the L-morphome case to the English irregular past tense under the model of Prasada and Pinker (1993) – with the additional complexity that what is irregular is not the phonological relation between the two stem forms, but their *distribution* within the paradigm. As such, a dual-route mechanism would be entirely appropriate in describing the set of experimental results we found for these synchronic Romance languages – a mechanism of sporadic analogy and memorization is responsible for the extant L-shaped forms, and rule-based generalization reveals itself only with productive extension to novel forms, regardless of their phonological resemblance to the existing L-shaped forms. Indeed, one reason to believe that the L-morphome was originally abstract (and not purely analogical) in nature has to do with 'defective' forms, such as the Portuguese and Spanish verbs *abolir* 'to abolish', which lack a form in both the 1SG.IND and throughout the SBJ. As the 'shared' form in question in such cases is actually *ungrammaticality* (i.e. a null, or ineffable allomorph), this represents a high degree of abstract identity among the cells in question, as opposed to analogical extension based on existing forms such as *tengo* ~ *tienes* or anything of the sort. We therefore contend that at the time that 3rd conjugation verbs such as *abolir* became 'defective' in the language, the L-morphome was still grammaticality active in the language as a principle above and beyond the individual form of verbs themselves.[7]

A potential future experimental manipulation consistent with the present line of discussion would be one that contrasted wug forms of the type we employed above (e.g. *mipo* ~ *mifes*) with ones that bear high analogical resemblance to existing L-morphomic forms (e.g., say, *mengo* ~ *mienes*). If indeed the amount of L-shaped responses was modulated by the degree to which the wug forms resembled existing memorized forms, such that Natural responses continued to prevail in the former but gave way to L-shaped reponses in the latter, this would constitute potentially strong confirmation for an extension of the dual-route approach as a model of the synchronic state of L-shaped forms in Romance. Very much like their kin with the English past tense, these once-productive (and phonologically natural) forms became increasingly less representative of processes within the lexicon as a whole, and despite their quotidian familiarity, they have come to be treated as a set of memorized forms. When it comes to the relation between the 1SG.IND and the present subjunctive forms, Romance learners have demonstrated that inflectional morphology, too, coheres to the adage "Diachrony proposes, Naturalness disposes". Nonetheless, we must emphasize that the case we have examined here is but one of many 'morphomic' patterns with Romance, and we wish to contend that experimental studies of the kind we have conducted here should be ideally applied to each of them in order to arrive at a better understanding of their synchronic status (and if nil, what led to their decline). As such, experimental research into naturalness biases can

7 While experimental work on defective verbs is limited (see Albright 2003 on Spanish for a pioneering study) and cries out for much further investigation, recent work by Nevins et al. (2014) on Portuguese suggests that the set of defective forms in synchronic Portuguese may in fact be shrinking, as younger speakers rate the SBJ of many forms listed as defective as perfectly acceptable.

continue to contribute to a broader typology of the types of identity relations possible within morphological categories, including categories such as the formation of participles, agentives, and imperatives, ideally conducted with a range of participants that includes children as well as adults in order to see whether the effects of developing vocabulary size may influence the extent to which such generalizations represent a balance between inherent naturalness of a syncretic pattern and the amount of representativeness they bear within an individual's lexicon.

Appendix 1: construction of best mixed-models

This section describes the procedures used to construct the best mixed models for each of the languages analyzed in the experiments, before proceeding to overall analyses of the best predictors of variance and the ratio of Natural responses to L-shaped responses in each, as detailed in Sections 4.1.2, 5.1.2, and 6.1.2, respectively.

Beginning with European Portuguese, we followed the modelling strategy as documented in Barr et al. (2013): we began with a saturated model, with fully crossed and fully specified random effects. This kind of model has an interaction term for all the predictors as fixed effects, with random intercepts and slopes:

(16) Saturated model for Portuguese results:
*Response ~ Group * Frame * Conjugation * Place + (1 + Group * Frame * Conjugation * Place|Participant) + (1 + Group * Frame * Conjugation * Place|item)*

Due to non-convergence, we simplified the model until it partially-converged. We followed two principles for choosing which term to exclude for the purpose of simplification, 1) hierarchically: most complex (the largest interaction terms) to the least complex (single terms), and 2) by-item slopes before by-participant slopes. The latter principle is justified by the fact that our data were collected from controlled experiments – item variations tend to be smaller than participant variations. A model was deemed to have converged if the relative gradient is below 0.002, as recommended by Ben Bolker (one of the developers of *lme4* (R-sig-ME mailing list n.d.)). By inspecting the partially-converged model, we excluded the slope associated with smallest variance. This process was repeated until the model converged, and the resulting converged model is shown below.

(17) Converged model for Portuguese results:
*Response ~ Group * Frame * Conjugation * Place + (1 + Group + Frame + Conjugation + Place|Participant) + (1 + Frame + Place|item)*

Next, we followed a data-driven approach to determine the random effect structure of our model, using the backward best-path algorithm, guided by which step of removal of a predictor would lead to the best next model. The model comparison was performed using an *anova* (test $=\chi^2$, $\alpha = 0.1$). The reason for α to be set at a liberal threshold of 0.1 is to be as conservative as possible with the detection of any potential predictors. If there were multiple subset models that resulted in p-values exceeding the α-level in their nested model comparisons with the superset model, the subset model with the strongest evidence (the highest p-value) was selected. Both of the random intercepts were kept by default, as it is a common practice to include them. The resulting model has *the maximal effect structure supported by the data*, following the terminology of Jaeger (2010):

(18) Model with the maximal effect structure supported by the data for Portuguese results
*Response ~ Group * Frame * Conjugation * Place + (1 + Group + Frame + Conjugation|Participant) + (1|item)*

We then performed a series of nested model comparison using *anova* (test $=\chi^2$, $\alpha = 0.1$). The addition or removal of terms was justified by whether a significant improvement to the model was made. We adhered to the principle of marginality, which does not allow for models containing an interaction without its respective main effects and all lower order terms. The model selection algorithm was again the best-path algorithm; the direction of comparisons was first forward (inclusion) then backward (exclusion), and this pattern was repeated until no terms could be further included or excluded. When excluding terms, we excluded from the most complex (the largest interaction terms) to the least complex (single terms), with the reverse being true when including terms. The comparison process was alternated between random effects and fixed effects. The resultant model has *the maximal effect structure justified by model comparison*, following the terminology of Jaeger (2010):

(19) Model with the maximal effect structure justified by model comparison for Portuguese results:
Response ~ Group + Frame + Conjugation + Place + Group:Frame + (1 + Group + Frame|Participant) + (1|item)

This model thereby included main effect terms for Group, Frame, Conjugation, and Place, an interaction term for Group x Frame as well as random effect terms such as a random intercept term for Item, and a random intercept for Participant with random slopes for Group and Frame.

For Italian, the same model selection procedure was used as for Portuguese, reported above. We began with a saturated model as shown below.

(20) Saturated model for Italian:
*Response ~ Group * Frame * Alternation + (1 + Group * Frame * Alternation|Participant) + (1 + Group * Frame * Alternation|item)*

The converged model is as shown below.

(21) Converged model for Italian:
*Response ~ Group * Frame * Alternation + (1 + Group * Frame * Alternation − Group:Frame:Alternation|Participant) + (1 + Group + Frame + Alternation|item)*

The model with *the maximal effect structure supported by the data* is as shown below.

(22) Model with the maximal effect structure supported by the data for Italian:
*Response ~ Group * Frame * Alternation + (1|Participant) + (1|item)*

We then performed a series of nested model comparison using *anova* (test $=\chi^2$, $\alpha = 0.1$) as before. The resultant model with *the maximal effect structure justified by model comparison* is shown below.

(23) Model with the maximal effect structure justified by model comparison for Italian results:
*Response ~ Group * Frame * Alternation + (1|Participant) + (1|item)*

In summary, this model included an interaction term for Group x Frame x Alternation as well as random effect terms such as a random intercept term for Item and for Participant.

For Spanish, the same model selection procedure was used as with the other two languages. We began with a saturated model as shown below.

(24) Saturated model for Spanish:
*Response ~ Group * Frame * Person * Place + (1 + Group * Frame * Person * Place|Participant) + (1 + Group * Frame * Person * Place|item)*

The converged model is as shown below.

(25) Converged model for Spanish:
*Response ~ Group * Frame * Person * Place + (1 + Group + Frame + Person + Place|Participant) + (1 + Group + Frame|item)*

The model with *the maximal effect structure supported by the data* is as shown below.

(26) Model with the maximal effect structure supported by the data for Spanish:
*Response ~ Group * Frame * Person * Place + (1 + Group + Frame + Place|Participant) + (1|item)*

We then performed a series of nested model comparison using *anova* (test $= \chi^2$, $\alpha = 0.1$) as before. The resultant model with *the maximal effect structure justified by model comparison* is shown below.

(27) Model with the maximal effect structure justified by model comparison for Spanish:
Response + Group + Frame + Person + Place + Group:Frame + Group:Person + Group:Place + Frame:Place + Place:Person + Group:Frame:Place + Group:Place:Person + (1 + Group + Frame + Place|Participant) + (1|item)

In sum, the model for Spanish included main effect terms for Group, Frame, Person, and Place, a number of interaction effects, and random effect terms such as a random intercept term for Item, and a random intercept for Participant with random slopes for Group, Frame, and Place.

Acknowledgement: Many thanks to our interlocutors along the course of this research: Asaf Bachrach, Michael Becker, Ricardo Bermúdez-Otero, Ana Castro, Maria Garraffa, Thomas Graf, Kyle Gorman, Martin Maiden, Gertjan Postma, Erica Rodrigues, Leticia Sicuro Corrêa, Donca Steriade, Leo Wetzels, and Marcos Zampieri.

References

Albright, Adam. 2003. A quantitative study of Spanish paradigm gaps. In *Proceedings of WCCFL 25*, 1–14.

Anderson, Stephen. 1981. Why phonology isn't 'natural'. *Linguistic Inquiry* 12. 493–539.

Aronoff, Mark. 1994. *Morphology by itself*. Cambridge, MA: MIT Press.
Baayen, R. Harald. 2001. *Word frequency distributions*, vol. 18. Cambridge, MA: MIT Press.
Bachrach, Asaf. & Andrew Nevins. 2008. Introduction: Approaching inflectional identity. In Asaf Bachrach and Andrew Nevins (eds.), *Inflectional identity*, 1–28. Oxford: Oxford University Press.
Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language* 68. 255–278.
Barto, Kamil. 2014. Mumin: Multi-model inference. R package version 1.10.0. http://CRAN.R-project.org/package = MuMIn. Accessed 16 February 2015.
Bates, Douglas, Martin Maechler, Ben Bolker & Steven Walker. 2014. lme4: Linear mixed-effects models using Eigen and s4. R package version 1.0-6. http://CRAN.R-project.org/package = lme4. Accessed 16 February 2015.
Becker, Michael, Nihan Ketrez & Andrew Nevins. 2011. The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* 87(1). 84–125.
Becker, Michael & Jonathan Levine. 2010. Experigen – An online experiment platform. https://github.com/tlozoot/experigen. Accessed 16 February 2015.
Berent, Iris, Tracy Lennertz, Paul Smolensky & Vered Vaknin-Nusbaum. 2012. Listeners' knowledge of phonological universals: Evidence from nasal clusters. *Phonology* 26. 1550–1562.
Blevins, Juliette. 2004. *Evolutionary phonology*. Cambridge: Cambridge University Press.
Blevins, James P. 2010. The morphome as a unit of predictive value. Paper presented at the 14th International Morphology Meeting, Budapest.
Bobaljik, Jonathan. 2004. *Universals in comparative morphology*. Cambridge, MA: MIT Press.
Bourne, Lyle E. 1970. Knowing and using concepts. *Psychological Review* 77. 546–556.
Cristia, Alejandrina, Jeff Mielke, Robert Daland & Sharon Peperkamp. 2013. Constrained generalization of implicitly learned sound patterns. *Journal of Laboratory Phonology* 4. 259–285.
Davies, Mark & Michael Ferreira. 2006. Corpus do português: 45 million words, 1300s–1900s. 2006. http://www.corpusdoportugues.org. Accessed 16 February 2015.
Dressler, Wolfgang, Willi Mayerthaler, Oswald Panagl & Wolfgang Wurzel. 1987. *Leitmotifs in natural morphology*. Amsterdam: Johns Benjamins.
Feldman, Jacob. 2000. Minimization of Boolean complexity in human concept learning. *Nature* 407. 630–633.
Finley, Sara. 2012. Typological asymmetries in round vowel harmony: Support from artificial grammar learning. *Language and Cognitive Processes* 27. 1550–1562.
Graf, Thomas. 2012. An algebraic perspective on the person case constraint. In *UCLA Working Papers in Linguistics 17*, 85–90.
Hale, Mark. & Charles Reiss. 2008. *The phonological enterprise*. Oxford: Oxford University Press.
Halle, Morris. & Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In Kenneth Hale and Samuel Jay Keyser (eds.), *The view from building 20*, 111–176. Cambridge, MA: MIT Press.
Hayes, Bruce, Donca Steriade & Robert Kirchner. 2003. *Phonetically-Based Phonology*. Cambridge: Cambridge University Press.
Heinz, Jeffrey. 2010. Learning long-distance phonotactics. *Linguistic Inquiry* 41. 623–661.

Jaeger, T. Florian. 2010. Random effect: Should I stay or should I go? http://hlplab.wordpress.com/2009/05/14/random-effect-structure/. Accessed 16 February 2015.

Johnson, Paul C. D. 2014. Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in Ecology and Evolution* 5: 944–946. http://dx.doi.org/10.1111/2041-210X.12225.

Kenstowicz, Michael. 1997. Base-identity and uniform exponence: Alternatives to cyclicity. In Jacques Durand & Bernard Laks (eds.), *Current trends in phonology: Models and methods*, 363–394. University of Salford.

Killick, Rebecca. & Idris A. Eckley. 2011. *Changepoint: An R package for changepoint analysis*. Lancaster: Lancaster University.

Lin, Yuri, Jean-Baptiste Michel, Erez Lieberman Aiden, Jon Orwant, Will Brockman & Slav Petrov. 2012. Syntactic annotations for the Google books Ngram corpus. In *Proceedings of the ACL 2012 System Demonstrations*, 169–174. Association for Computational Linguistics.

Maiden, Martin. 2005. Morphological autonomy and diachrony. *Yearbook of Morphology 2004*, 137–175.

Maiden, Martin & Paul O'Neill. 2010. On morphomic defectiveness: evidence from the Romance languages of the Iberian Peninsula. In Matthew Baerman, Greville G. Corbett, and Dunstan Brown (eds.), *Defective Paradigms: Missing forms and what they tell us*, 103–124. Oxford: Oxford University Press.

Matthews, P. H. 1991. *Morphology*. Cambridge: Cambridge University Press.

McCarthy, John. 2005. Optimal paradigms. In Laura Downing, Tracy Alan Hall, and Renate Raffelsiefen (eds.), *Paradigms in phonological theory*, 170–210. Oxford: Oxford University Press.

Moreton, Elliott. 2008. Analytic bias and phonological typology. *Phonology* 25. 83–127.

Nakagawa, Shinichi & Holger Schielzeth. 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4. 133–142.

Nevins, Andrew. 2007. The representation of third person and its consequences for person-case effects. *Natural Language and Linguistic Theory* 25. 273–313.

Nevins, Andrew, Gean Damulakis & Maria Luisa Freitas. 2014. Phonological regularities among defective verbs. *Cadernos De Estudos Lingüísticos* 56(1). 11–21.

Nevins, Andrew & Cilene Rodrigues. 2012. Naturalness biases, 'morphomes', and the Romance first person singular. In Paper presented at Univ. Coimbra, Newcastle University, University of York, and Univ. Paris VII. Handout available online: http://ling.auf.net/lingbuzz/001469. Accessed: 16 February 2015.

Onelli, Corinna, Domenico Proietti, Corrado Seidenari & Fabio Tamburini. 2006. The DiaCORIS project: A diachronic corpus of written Italian. In *Proceedings of LREC-2006, The Fifth International Conference on Language Resources and Evaluation*, 1212–1215.

Page, E. S. 1954. Continuous inspection schemes. *Biometrika* 41. 100–115.

Paiva, Maria da Conceição & Maria Eugênia Lamoglia Duarte. 2013. Mudança lingüstica: observações no tempo real. In M. C. Mollica and M. L. Braga (eds.), *Introdução à socio-lingustica: o tratamento da variação*, 4th ed., 179–190. São Paulo: Contexto.

Pertsova, Katya. 2010. Learning biases in the acquisition of Russian genitive plural allomorphy. Ms., Univ. North Carolina.

Pertsova, Katya. 2011. Grounding systematic syncretism in learning. *Linguistic Inquiry* 42. 225–266.

Prasada, Sandeep & Steven Pinker. 1993. Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes* 8. 1–56.
R Core Team. 2013. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/. Accessed 16 February 2015.
R-sig-ME mailing list. n.d. https://stat.ethz.ch/pipermail/r-sig-mixed-models/2014q2/021993.html (accessed 16 February 2015).
Sánchez-Martnez, Felipe, Isabel Martínez-Sempere, Xavier Ivars-Ribes & Rafael C. Carrasco. 2013. An open diachronic corpus of historical Spanish. *Language Resources and Evaluation* 47. 1327–1342. doi:10.1007/s10579-013-9239-y.
Steriade, Donca. 2008. A pseudo-cyclic effect in Romanian morphophonology. In Asaf Bachrach and Andrew Nevins (eds.), *Inflectional identity*, 313–359. Oxford: Oxford University Press.
Stump, Gregory. 2001. *Inflectional morphology: A theory of paradigm structure*. Cambridge: Cambridge University Press.
Tang, Kevin & Andrew Nevins. 2013. Quantifying the diachronic productivity of irregular verbal patterns in Romance. *UCL Working Papers in Linguistics* 25. 289–308.
Tucker, Emily. 2000. Multiple allomorphs in the formation of the Italian agentive. Master's thesis, UCLA.
Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30. 945–982.
Yang, C. 2005. On productivity. *Linguistic Variation Yearbook* 5. 265–302.
Zampieri, Marcos & Martin Becker. 2013. Colonia: Corpus of historical Portuguese. In Marcos Zampieri and Sascha Diwersy (eds.), *Special volume on non-standard data sources in corpus-based research*. Volume 5 of *ZSM Studien*, 77–84. Köln: Shaker.
Zhang, Jie, Yuwen Lai & Craig Turnbull-Sailor. 2006. Wug-testing the "Tone Circle" in Taiwanese. In *Proceedings of WCCFL 25*, 453–461.
Zimmer, Karl. 1969. Psychological correlates of some Turkish morpheme structure conditions. *Language* 45(2). 309–321.

Note: The authors' names are listed in alphabetical order. Earlier descriptions of the experimental research were presented in handout form as Nevins and Rodrigues (2012), which the present work supersedes.