Naturalness Biases, 'Morphomes', and the Romance First Person Singular

Andrew Nevins and Cilene Rodrigues

Draft of March 14, 2012
Comments welcome!

Abstract: It has been suggested that the Romance first person singular indicative forms a base of derivation for the subjunctive paradigm, even though the latter shares no morphosyntactic features with the former (Maiden, 2004). 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). 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 as a base of derivation. 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.

1. Apparent autonomous morphomes

Irregular stem alternants in Portuguese found in the 1sg present indicative are also found throughout the present subjunctive (Schwindt, 2007), constituting what Maiden (2004) identifies as a 'morphome' (a term due to Aronoff (1994))– an unnatural, autonomous morphological pattern of shared identity, which he calls the L-shaped pattern:

(1)

'to say'	Indic	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.indic and 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.

Longer 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 and SBJ:

(2)

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

While 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, however, raises what Bachrach and Nevins (2008) call the 'inclusion question': why is the 2sg not included – in other words, *why is the pattern 'L-shaped', rather than 'sideways T' shaped*?

2. Naturalness Biases in Learning

Certainly, the very existence of the term *morphome* is predicated on the fact that (2) is 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); Becker et al. (2011) finds that *unnatural patterns are memorized for the existing items, but underlearned*: not generalized to novel items.

3. Implicational Nonce Testing

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, using three novel alternations, none of which are extant in Portuguese: *p*∼*f*, *t*∼*s*, *k*∼*x*.

Participants were divided into two groups, each of which saw 15 items. The methodology thus followed Wilson's (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 first* group (I → S) was exposed to frames such as (3), 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.

(3) 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.'

The *subjunctive first* (S → I) group was exposed to frames such as (4), 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.

(4) 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.'

Importantly, participants saw frames in which the linear order of 1SG.IND and 2SG.IND was the reverse in the I → S, 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.

3.1. Method

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

There were 15 target items (5 from each place of articulation), and 15 fillers (with nasal or liquid stem-final consonants, non-alternating). The 30 items were randomly presented with one of 10 different frames, 5 of which had the fricative form first and 5 the stop form first.

3.2. Results

250 participants completed the experiment. In the I→ S group, participants massively preferred forming the 2.SBJ on the basis of the 2.IND, rather than the 1.IND, with 67% of the responses choosing the fricative form, as shown in Figure 1. The overwhelming choice of base, therefore, was chosen on the basis of morphosyntactically identical PERSON.

In the S → I group, by contrast, participants showed little preference between forming the 1.IND on the basis of the 2.IND or the 2.SBJ, with a slight preference (53%) towards choosing the base on the basis of morphosyntactically identical MOOD, as shown in Figure 2.

The difference between the I → S and S → I groups was highly significant ($p < .0001$) in a mixed-effects logistic regression in which participant, item, and frame were modeled as random effects.

3.3. Discussion

Neither group displayed results compatible with the predictions of the L-morphome theory. Speakers given a chance to base the 2.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 slight preference for uniformity within MOOD.

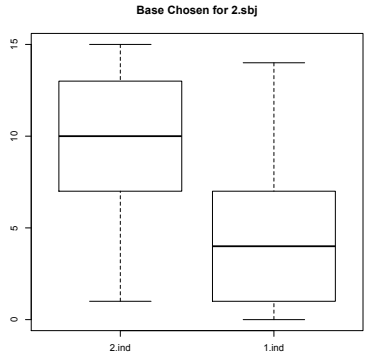

Figure 1: Results for I $\rightarrow$ S condition; Portuguese.

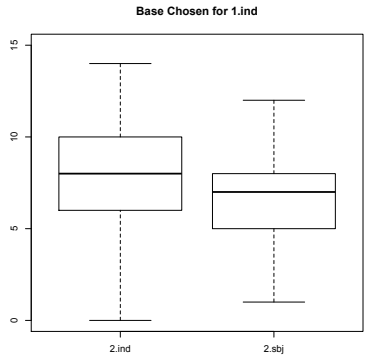

Figure 2: Results for S $\rightarrow$ I condition; Portuguese.

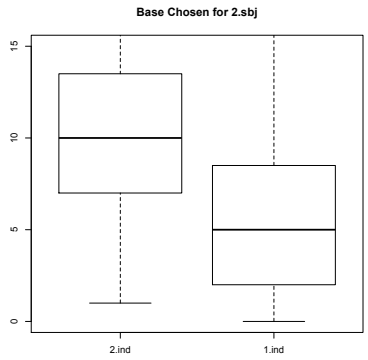

Figure 3: Results for I → S condition; Italian.

The L-morphome pattern, when it was 'thriving' in Old Portuguese, would have involved the cumbersome fractionation UNIFORM-1ST >> UNIFORM-MOOD >> UNIFORM-2ND. Contemporary learners lack sufficient evidence to overcome the biases against forming this grammar, instead bundling UNIFORM-PERSON >> UNIFORM-MOOD.

Our results are a first step towards experimentally testing whether learners 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.

4. Replication in Italian

We wished to replicate the finding in Italian, another L-shaped language. We used two new alternations: a stop∼ nasal change (5) and a high∼mid vowel change (6), neither of which are extant in Italian.

(5) 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'

(6) 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.'

175 participants completed the experiment. In the I→S condition, 64% chose the 2nd.indic base, as shown in Figure 3; in the S →I condition, 73% chose the 2nd.indic base, as shown in Figure 4. The results were thus even more decisively non-morphomic than the Portuguese version.

The difference between the I → S and S → I groups was significant ($p < .05$) in a mixed-effects logistic regression in which participant, item, and frame were modeled as random effects.

5. Replication and Test of Person in Spanish

If 1sg.IND and 2sg.IND form are provided and 3rd.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 → S_3 condition potentially gives the L-Morphome theory its best chance. We had 15 items and 15 fillers, much like the Portuguese experiment, with the distribution of stop and fricative alternants reversed (stop in 2sg, fric in 1sg), to additionally test whether a blanket preference for fricatives was responsible for those results.

We compared I→S_2 in frames such as (7) and I→S_3 in frames such as (8).

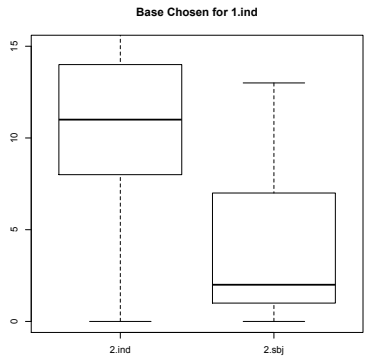

Figure 4: Results for S → I condition; Italian.

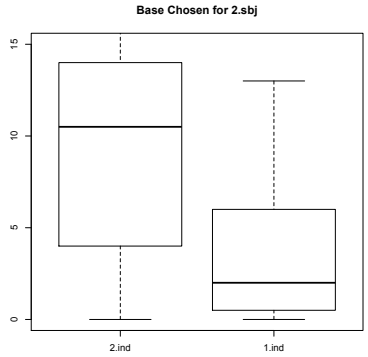

Figure 5: Results for I → S_2 condition; Spanish.

(7) 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 *lluto*1SG.IND with either hand. It's necessary that you ____2SG.SBJ with both hands in order to be more productive.'

(8) 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.'

148 participants completed the experiment. In the I→S_2 condition, 72% chose the 2ND.INDIC BASE, as shown in Figure 5. In the I→S_3 condition, 71% chose the 2ND.INDIC BASE, as shown in Figure 6.

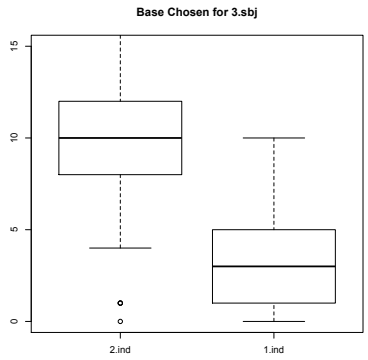

Figure 6: Results for I $\rightarrow$ S_3 condition; Spanish.

6. Discussion

These results in the Spanish I$\rightarrow$$S_3$ condition 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. 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.

The second interpretation is that speakers are implicitly aware of the fact that the 1sg.indic 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.indic constitutes an odd-man out, to be avoided as a base for any 2nd/3rd person forms. Evidence from *defective* verbs such as *abolir* may provide exactly this evidence.

In fact, these two interpretations 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. 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.

References

Aronoff, Mark. 1994. *Morphology by Itself*. Cambridge, MA: MIT Press.

Bachrach, Asaf, and Andrew Nevins. 2008. Introduction: Approaching Inflectional Identity. In *Inflectional Identity*, 1–28. Oxford University Press.

Becker, Michael, Nihan Ketrez, and Andrew Nevins. 2011. The Surfeit of the Stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* 87.1:84–125.

Becker, Michael, and Jonathan Levine. 2010. Experigen - an online experiment platform. Available at https://github.com/tlozoot/experigen.

Maiden, Martin. 2004. Morphological Autonomy and Diachrony. *Yearbook of Morphology 2004* 137–175.

Pertsova, Katya. 2010. Learning Biases in the Acquisition of Russian Genitive Plural Allomorphy. Ms., Univ. North Carolina.

Schwindt, Luiz Carlos. 2007. Paradigmatic correspondences in the brazilian portuguese verbal vowel system. *Acta Linguistica Hungarica* 54.4:393–407.

Wilson, Colin. 2006. Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science* 30:945–982.

Zhang, Jie, Yuwen Lai, and 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.