

## Stem shortening in Romance verbs: the 'S morpheme' at the intersection of token frequency and paradigmatic structure

Although some stem alternation patterns (aka. 'morphemes') in Romance (N, L, PYTA, see Maiden 2018) have been quite substantially described and analysed, others remain underexplored. Here we focus on a pattern of alternation that involves irregular stem shortening (S) in parts of the paradigm (e.g. It. *d-ire* 'say.INF' [vs *dic-iamo*], or *f-a* 'do.3SG.PRS.IND' [vs *fac-iamo*]). Because these short stems have never been subject to a systematic pan-Romance investigation (but see Malkiel [1977], Maiden [2004:237], Mariño Paz [2019], and Dubert García [2021]), we explore their paradigmatic and lexical domain across Romance, and their likely historical origin and motivation.

An initial qualitative inspection (in Maiden et al.'s 2010 ODRVM database) identified the reflexes of *faciō*, *dīcō*, *habeō*, *sapiō*, *possum*, and *volō* as the verbs that show these alternations most frequently. For these we coded, across the 70+ varieties in the database, the paradigmatic distribution of short stems in the paradigm, which yielded 2773 short-stem forms (18.67%), and 12082 long-stem forms (81.33%). Short stems were found to be most frequent in the cells 3SG.PRS.IND, 2SG.PRS.IND, 3PL.PRS.IND, and 2SG.IMP. A quantitative phylogenetic reconstruction of the ancestral states (i.e. presence of a short or long stem in a particular cell in a particular verb) at different points in time along the family tree found that the likelihood of short stems increases through time quite early in the history of the family. Results suggests that various short stems are likely ancestral to Proto-Western-Romance (i.e. Romance minus Sardinian and Balkan).

We propose an explanation for the timing of the emergence of short stems and their lexical and paradigmatic domain. The first part of the explanation relies on the well-known relation between length of expression and frequency of use (e.g. Zipf 1935, Bybee 2006, Gahl et al. 2012). The cells and verbs in which short stems are most common are all extremely frequent: among the 10 most frequent cells and among the 20 most frequent verbs respectively in Latin (Delatte et al. 1981). Having shorter forms for the expression of very common lexical and morphosyntactic meanings is an adaptive property for the efficient transfer of information, which would provide a motivation to prefer these in situations of competition (note that short stems would have appeared accidentally in the paradigm from regular sound change in some forms like *fa*<*fac* 'do.2SG.IMP' and *di*<*dīc* 'say.2SG.IMP').

The exact domain for the spread of short stems, however, must have been influenced, in addition, by the paradigmatic domains of extant stem alternations. The domain of short stems corresponds closely to those cells that partake in N alternations (i.e. stem-vowel differences related to stress), but not in L alternations (i.e. stem-final consonant alterations resulting from palatalization). N-L is an area of the paradigm, hence, within which a single stem would have been expected, whereas stem differences could easily exist with other parts of the paradigm. This links with the timeline of the better-known Romance morphemes. While PYTA is ancestral to all of Romance (i.e. it was present already in Classical Latin), those known as L and N emerged later. The sound changes that generated them all took place in Western-Romance but not always in the varieties that split before. The morphomic niche for the short-stem allomorphy, thus, is the direct result of the cross-classification of the domains of L and N, which means that it must have followed (and not preceded) the emergence in Romance of the L and N morphemes.

## References

Bybee, Joan. 2006. *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Delatte, Louis, Étienne Evrard, Suzanne Govaerts & Joseph Denooz. 1981. *Dictionnaire fréquentiel et index inverse de la langue latine*. L.A.S.L.A, Liege.

Dubert García, Francisco. 2021. Las formas reducidas de «facēre» en el noroeste de la península ibérica. Convergencia y divergencia geolingüísticas." *Studia linguistica romanica* 6: 163-191.

Gahl, Susanne, Yao Yao, and Keith Johnson. 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of memory and language* 66, 4: 789-806.

Herce, Borja. 2019. Morpheme interactions. *Morphology* 29, 1: 109-132.

Maiden, Martin. 2004. When lexemes become allomorphs: on the genesis of suppletion. *Folia Linguistica* 38: 227-256.

Maiden, Martin. 2018. *The Romance verb: Morphomic structure and diachrony*. Oxford: Oxford University Press.

Maiden, Martin, John Charles Smith, Silvio Cruschina, Marc-Olivier Hinzelin & Maria Goldbach. 2010. Oxford online database of Romance verb morphology. Available at: <http://romverbmorph.clp.ox.ac.uk/>.

Malkiel, Yakov. 1977. Editorial Post-Script: Old Spanish far, fer, fazer. *Romance Philology* 31, no. 2: 257-262.

Mariño Paz, Ramón 2019. O cambio *dizes > dis, diz > di, dizen > din* nas fontes do galego medio. Xosé A. Fernández Salgado, Aquilino S. Alonso Núñez (eds.). *Non haberá illa, pro hai o nome. Homenaxe a Antón Palacio*. Vigo: Universidade de Vigo, 97-113. <https://secretaria.u.vigo.gal/uv/web/publicaciones/public/show/342>.

Zipf, George Kingsley. 1935. *The Psycho-Biology of Language: An Introduction to Dynamic Philology*. Boston, MA: Houghton Mifflin Company.