Islands of reliability in Portuguese nominalization

This study explores the question of how speakers decide among derivational patterns in situations of morphological rivalry, taking as its empirical domain the competition between two patterns of deverbal nominalization in Portuguese. This type of derivation is predominantly carried out in Portuguese by means of two affixes, -mento and -ção, whose distributions overlap to a large extent, giving rise to doublets such as alagamento and alagação (both meaning ‘flood’), derived from the verb alagar.

Different surveys show that -ção has formed a significantly larger number of words than -mento in recent centuries and in present-day language (Basilio, 1996; Quadros, 2015; Sandmann, 1988) – Quadros (2015) reports a rate of vocabulary growth twice as great for -ção. This situation is intriguing since both suffixes have been part of the language since its first records, which leads to the question of how the least productive of them has been able to remain productive despite the generality of its rival.

Exploring the hypothesis that the stability of -mento can be explained by the existence of lexical niches in which its occurrence is favored, we employed the Minimal Generalization Learner (MGL), a rule induction procedure proposed by Albright & Hayes (2003). We fed the MGL a list of 1,919 pairs of deverbal nouns containing -mento or -ção, extracted from Dicionário Houaiss da Língua Portuguesa, a general dictionary of Portuguese.

We found a number of reliable phonotactic contexts for -mento affixation (islands of reliability in the terminology of Albright & Hayes (2003)), including some that had not been previously described in the literature. Among these are contexts strongly favoring the attachment of -mento to first conjugation verbs with roots ending in alveolar consonants, e.g. engess-, in palatal consonants, e.g. embaralh-, and in vowels, e.g. zoa-. The first conjugation is particularly important in this regard, since nearly all new verbs in the language fall into this class. Hence we suggest that the existence of these contexts has helped support the stability of -mento in the language.

The grammar output of the MGL was then applied to a set of test data comprising bases of 185 new derived words extracted from a 13 million word corpus of Brazilian Portuguese. The goal was to check whether the rules learned from dictionary data would match the choices made by real speakers in the coinages found in the corpus, particularly whether the islands of reliability found by the MGL have been used by speakers as synchronic generalizations.
We found a general agreement of 87% between the predictions of the model and the choices made for new deverbal nouns in the corpus. When considering only the 50 cases where the model predicts affixation of -mento, we see 88% of agreement with corpus attestations, \( p < 0.001, 95\% \text{ CI} [0.76, 0.95] \). This suggests that the islands of reliability we found for this suffix are generalizations currently exploited by Brazilian Portuguese speakers.

On the other hand, while we find a similar rate of agreement for the 135 cases where the model predicts affixation of -ção (87%, \( p < 0.001, 95\% \text{ CI} [0.80, 0.92] \)), we note that most of them involve verbs formed by the suffix -izar, which is known to potentiate -ção affixation. When these verbs are removed from the test data, we are left with only 36 bases for which the model predicts affixation of -ção, and the agreement for these cases drops to 53%, not significantly better than chance, \( p = 0.87, 95\% \text{ CI} [0.35, 0.69] \). Some limitations of our test are discussed in this regard, particularly the use of dictionary data, including words that are not in current usage, in our training set, which seems to have caused the underestimation of some islands of reliability for -mento and the overestimation of the reliability of phonotactic contexts for -ção affixation, resulting in the poor performance observed in the prediction of -ção nouns not involving -izar.

Our results give support to our initial hypothesis that the suffix -mento has remained a viable means of forming deverbal nouns in Portuguese because speakers have been able to find productive contexts where its affixation is more predictable than the affixation of its rival. On a general level, these results show that derivational morphology is sensitive to fine-grained subregularities which must be closely investigated for a complete understanding of morphological productivity. This poses the question of how to integrate this aspect of speakers’ knowledge into our theories of morphology.

References


