



Variables and similarity in linguistic generalization: Evidence from inflectional classes in Portuguese



João Veríssimo^{a,*}, Harald Clahsen^b

^a Faculty of Psychology, University of Lisbon, Alameda da Universidade, 1649-013 Lisboa, Portugal

^b Potsdam Research Institute for Multilingualism, University of Potsdam, Haus 2, Campus Golm, Karl-Liebknecht-Straße 24–25, 14476 Potsdam, Germany

ARTICLE INFO

Article history:

Received 11 November 2013

Received in revised form 19 May 2014

Available online 9 July 2014

Keywords:

Variables

Similarity

Rules

Morphological generalization

Productivity

Computational modeling

ABSTRACT

Two opposing viewpoints have been advanced to account for morphological productivity, one according to which some knowledge is couched in the form of operations over variables, and another in which morphological generalization is primarily determined by similarity. We investigated this controversy by examining the generalization of Portuguese verb stems, which fall into one of three conjugation classes. In Study 1, an elicited production task revealed that the generalization of 2nd and 3rd conjugation stems is influenced by the degree of phonological similarity between novel roots and existing verbs, whereas the 1st conjugation generalizes beyond similarity. In Study 2, we directly contrasted two distinct computational implementations of conjugation class assignment in how well they matched the human data: a similarity-driven model that captures phonological similarities, and a dual-mechanism model that implements an explicit distinction between context-free and similarity-based generalizations. The similarity-driven model consistently underestimated 1st conjugation responses and overestimated proportions of 2nd and 3rd conjugation responses, especially for novel verbs that are highly similar to existing verbs of those classes. In contrast, the expected proportions produced by the dual-mechanism model were statistically indistinguishable from human responses. We conclude that both context-free and context-sensitive processes determine the generalization of conjugations in Portuguese, and that similarity-based algorithms of morphological acquisition are insufficient to exhibit default-like generalization.

© 2014 Elsevier Inc. All rights reserved.

Introduction

One of the striking features of human language is its productivity, that is, the fact that speakers are able to produce and comprehend linguistic expressions that they have not encountered before (Chomsky, 1965). At the heart of this ability is the generalization of linguistic patterns and constraints to novel items. For example, if a novel verb such

as *ploamph* were to enter the English language, speakers would be readily able to form its different variants (e.g., *ploamphed*, *ploamphing*; Prasada & Pinker, 1993), as well as incorporate them into acceptable sentences (e.g., *Why do you think I should have ploamphed it?*). Given that knowledge of language is finite, but the number of complex forms and sentences that can be produced and understood is infinite, one of the central goals of the language sciences is to characterize the representational substrate that accounts for linguistic generalizations.

Broadly speaking, two opposing viewpoints have been advanced. On one side, proponents of symbol-manipulation approaches hold that linguistic knowledge is primarily

* Corresponding author. Present address: Potsdam Research Institute for Multilingualism, University of Potsdam, Haus 2, Campus Golm, Karl-Liebknecht-Straße 24–25, 14476 Potsdam, Germany. Fax: +49 (0) 331/977 2687.

E-mail address: joao.verissimo@uni-potsdam.de (J. Veríssimo).

couched in the form of operations over *variables*, that is, placeholders that stand for every instance of a category (e.g., Chomsky, 1980; Fodor & Pylyshyn, 1988; Marcus, 2001). Variables are insensitive to the idiosyncratic properties of the tokens they instantiate, and as such, allow free and unbounded generalization to novel instances. For example, if the rules or constraints of sentence formation make reference to a variable '(V)erb', then every lexical item that satisfies this condition can be used in a well-formed sentence. Likewise, if producing a progressive form involves concatenating an instance of a variable with the appropriate suffix (i.e., V + *-ing*), then this operation can be productively extended to any novel verb.

A radically different approach to linguistic generalization is espoused by proponents of similarity-based approaches, which here we will take to encompass a broad class of models, including (amongst others) connectionist and exemplar-based architectures (e.g., Daelemans, 2002; Elman, 1993; Rumelhart & McClelland, 1986; Skousen, Lonsdale, & Parkinson, 2002).

A distinctive feature of such models is the notion that generalization is primarily determined by *similarity*. More specifically, the higher the representational overlap between a novel item and a set of learned instances, the higher the probability that it will be responded to in the same way (e.g., Hahn & Nakisa, 2000). This aspect stands in sharp contrast to how generalization is treated in variable-based approaches. Rather than the same operation being applied equally to all members of a category, analogical models typically produce graded and probabilistic outcomes which reflect overlap at different levels of representation (e.g., phonological, semantic) and are influenced by the statistical properties of previously learned input-output pairs.

The study of morphological generalization and processing has played an important role in this debate, particularly in what concerns the contrast between regular (e.g., *walked*) and irregular (e.g., *sang*) past tense forms in English: whereas the regular *-ed* pattern is productively extended to new roots, in a way that appears to be insensitive to their phonological characteristics (e.g., *plamphed*; see, e.g., Berko, 1958; Prasada & Pinker, 1993; Ullman, 1999; but see Albright & Hayes, 2003), irregular patterns are seldom generalized and are applied only to novel items that phonologically resemble clusters of irregular verbs (e.g., *spling*, which in analogy with verbs such as *sing*, can be inflected as *splang*; Bybee & Moder, 1983).

The case of the English past tense clearly illustrates a tension that is also visible in the inflectional and derivational systems of many other languages: that between context-independence and context-sensitivity (Keuleers et al., 2007). More specifically, because many inflectional contrasts and word-formation processes are not applied in the same way for each and every member of a grammatical class (such as all verbs), then morphological operations must, at least for some items, be conditioned by lexical information. At the same time, because some patterns can be productively extended in an unbounded fashion (i.e., even to novel items that are very dissimilar to existing forms in the language), then it would appear that at least some morphological operations can behave as a default,

applying when “all else fails” (Bybee, 1995, p. 452) and in a way that is independent of the idiosyncratic properties of individual tokens (see, e.g., Berent, Pinker, & Shimron, 1999, for Hebrew; Marcus, Brinkmann, Clahsen, Wiese, & Pinker, 1995, for German; Prasada & Pinker, 1993, for English).

The balance between lexically conditioned and productive generalizations is most explicitly incorporated in a class of *dual-mechanism* models of morphology (e.g., Clahsen, 1999; Marslen-Wilson & Tyler, 1997, 2007; Pinker, 1999; Pinker & Ullman, 2002). According to such models, morphological operations can either be instantiated: (1) by the application of a grammatical rule, which operates over a variable and generates a structured representation (e.g., adding the English regular *-ed* affix to any verbal root); or (2) through direct retrieval of an exceptional form (e.g., an irregular past tense, such as *brought*), and in the case of generalization to novel words, via analogy from the associations between lexically specified representations (e.g., *splought* as a possible inflection for *spling*).

In contrast, according to the class of *similarity-based* models mentioned above, a single context-sensitive mechanism based on the overlap between lexical representations is purported to be sufficient to capture both the generalizations that are similarity-based and those that are made outside of restricted areas of the similarity space. In such models, approximation to default-like behavior is thought to emerge naturally for those morphological patterns that are most frequent or that display significant heterogeneity in their phonological distributions (e.g. Hahn & Nakisa, 2000; Hare & Elman, 1995; Hare, Elman, & Daugherty, 1995). In other words, in the single-mechanism view, what appears to be the result of an operation over a variable is in fact the product of the same frequency- and similarity-based mechanisms that are responsible for restricted and lexically conditioned generalizations.

In the present paper, we set out to investigate the generalization properties of conjugation classes in Portuguese, an example of pure morphology, which we believe provides a better test case for assessing the mechanisms involved in morphological generalization than the familiar contrast between regular and irregular inflection. In order to assess the role of phonological similarity, we have used a metric derived from a computational implementation of a similarity-based model, the Minimal Generalization Learner (MGL) proposed by Albright (Albright, 2002a; Albright & Hayes, 2003). In addition, by minimally changing the MGL model to embed a more explicit dual-mechanism distinction, we were able to directly compare two specific computational implementations in how well they matched elicited production data obtained from native speakers of Portuguese.

Linguistic background

In linguistic treatments of Portuguese morphology, the structure of Portuguese verbs has been proposed to display three hierarchical levels: a root constituent, a stem constituent, and a word constituent (Villalva, 2000, 2003). The root (e.g., *cant-*, in the infinitive form *cantar*, ‘to sing’) is taken to be the locus of all semantic, syntactic and morphological information, and transmits this information to the stem

Download English Version:

<https://daneshyari.com/en/article/931814>

Download Persian Version:

<https://daneshyari.com/article/931814>

[Daneshyari.com](https://daneshyari.com)