Second-Language Acquisition and Morphological Irregularity in Simulations of Natural Language Evolution

Weldon Scott

I. Foreword

THIS paper is an adaptation of my work to the College Board's AP Research course from the 2020–2021 academic year. I am still proud of what I managed to do at the age of sixteen and my paper and accompanying presentation ended up receiving a perfect score from the College Board, which was only achieved by 1.56% of students that year. There is a cool article about this here.

However, this being said, there are some significant shortcomings of the project that I aim to address here. Firstly, some of my original discussion of the theoretical basis of my methodology is not as rigorous as it could be. Secondly, I would like to provide more detailed commentary on my results as I am no longer constrained by a word limit. Thirdly, I wanted to rewrite the paper in IEEE style for both practical and aesthetic reasons as it initially used APA formatting guidelines. I may also rewrite the code I used to run the simulations, but while it is somewhat clunky, it does work, so this is not a major priority for me at the moment.

To whomever is reading this paper, thank you for taking the time to engage with it! Whether this be in the form of a brief skim or a thorough dissection, I hope that you are able to glean something from it.

II. Introduction

In linguistics, morphological compositionality and morphological irregularity are terms that refer to the manner in which a word is inflected. Compositionality is the representation of complex meanings via a combination of representations of their parts. For instance, in the English phrase "three red dogs," each of the four morphemes, which are "three", "red", "dog", and "-s," can be understood in isolation and each contributes to the overall meaning of the phrase. Compositional structure is prevalent in natural language because speakers can easily combine components of meaning into a larger phrase, even if the phrase itself is new. In constrast, morphological irregularity involves representations that have a nontrival relationship to their constituent parts. As an example, the simple past tense conjugation of the English verb "to go" is "went." This cannot be decomposed simply as the verb stem with the "-ed" suffix, which marks the simple past tense for regular verbs. This is contrasted with the regular verb "to walk," which does indeed have the simple past form "walked." Irregular forms must be individually learned as compositional patterns do not apply to them, which means that irregular inflection is harder for second-language (L2) speakers to acquire than compositional inflection [1].

It is understood that languages adapt to the settings in which they are used, following patterns that may be reminiscent of those observed in biological evolution [2]. Consequently, one may wonder how such adaptation presents itself when examining compositionality and irregularity. More specifically, one point of curiosity surrounds how the presence of second-language speakers affects the degree of irregularity in natural languages. This is difficult to examine through cross-linguistic analyses because of logistical problems, in addition to a lack of relevant data. Given that simulated languages do not have these issues, they are employed to study this relationship in this paper.

III. Literature Review

A. Lingustic Adaptation

The eminence of linguistic adaptation research is rather new. In the past, culture and linguistic structure were not often correlated [3], although efforts to investigate such a correlation did exist during this time period (e.g. [4]). In the years since, there has been much more interest in the interplay of various factors that influence linguistic evolution [3]. This has led to an abundance of research in this field (e.g. [2], [5], [6] [7]), especially with respect to social structure.

In this regard, an important piece of work was produced by Lupyan and Dale in 2010 [8], in which the authors performed a cross-linguistic analysis of numerous grammatical structures. They found that the social context in which a language is spoken influences its grammar. This theory is known as the linguistic niche hypothesis. In subsequent research, the hypothesis was expanded to account for stressors including ecology, communication technology, and genetic influences [9]. One proposed explanation for the impact of social determinants on linguistic structure involves mechanisms of language acquisition. In particular, it has been hypothesized that languages with mature learners are less grammatically complex because features that are difficult to acquire are comparatively less likely to be imparted upon subsequent generations of speakers. On the other hand, languages with higher grammatical complexity generally encode more redundancy, which means they are easier for infants and young children to acquire [8]. These notions are supported by existing research in the field of L2 acquisition.

B. Second-Language Acquisition

For example, it has been observed that L2 speakers of English overgeneralize low-frequency irregular forms [1]. In other words, such speakers tend to use compositional inflection paradigms for uncommon words when irregular inflection patterns would be grammatically correct, while common words with irregular morphology are correctly inflected [10]. Moreover, the native language of the speaker is correlated with the details of their performance; native Korean speakers were found to conjugate irregular verbs with higher accuracy than native Spanish speakers, for instance [1]. Considering that Korean verb phrases are particularly rich [11], a case could be made for a speaker's native language shaping the nature of their abilities in English. The above results corroborate the mechanism that Lupyan and Dale propose to explain their findings: the presence of L2 speakers that have difficulty acquiring certain features of a language could apply an evolutionary pressure onto the language. Regardless of this, such data cannot establish a causal relationship between non-native speakers learning a language and morphological shifts in the grammar of that language over time.

C. Knowledge Gaps

Although research that discusses morphological irregularity relative to L2 acquisition exists (e.g. [1], [12]), it only examines this for individual speakers and their cognitive processes. There is much less research about the correlation between L2 speakers and linguistic variation. Some investigation into this matter has been undertaken (e.g. [5], [13]), but this work does not specifically examine irregularity. However, the results of such research do indicate that significant quantities of L2 speakers are correlated with noticeable changes in grammar. It is not clear whether or not L2 speakers are the primary factor driving linguistic evolution in this manner as other metrics, such as the total number of speakers, may determine linguistic structure more directly [14]. In spite of this, the notable changes that L2 speakers often accompany further the notion that research focused on examining morphological irregularity and L2 acquisition is likely to yield valuable results.

The lack of research into this subject is likely due to the difficulty of acquiring suitable data. Namely, it is difficult to account for external factors. For instance, languages with a higher speaker count tend to have greater proportions of L2 speakers [8]. Since languages spoken by large populations undergo different evolutionary trends from those spoken by small populations [14], Bromham, it is difficult to determine which factors are likely to influence linguistic evolution, or if another confounding variable is driving such adaptation. In addition, contact between languages is difficult to avoid or control, posing immense logistical difficulties.

This gap in the body of knowledge within linguistics is significant as its presence limits the validity of proposed mechanisms for linguistic change within the framework of the linguistic niche hypothesis. In particular, if the presence of L2 speakers does not lead to an increase in the degree of morphological compositionality in a language, it is difficult to argue that the process of second-language acquisition is a vehicle for this type of linguistic evolution. Consequently, the goal of this paper is to analyze the nature of the causal relationship between the prominence of L2 speakers and morphological compositionality. This is accomplished through the examination of computationally generated languages in order to avoid external influences.

D. Computational Approaches

Simulations of natural language development have become increasingly common in recent years [15] since this practice has any advantages over its traditional analogs. As an example, simulated languages can be employed to study language as an interconnected system because all relevant aspects of a simulated language can be directly recorded or observed from the examination of samples [16]. The same cannot be said for human languages; the availability of appropriate samples is often a limiting factor in linguistic research [17]. Likewise, simulated language samples can be generated in large supply and tailored to the requirements of the study [16].

Simulated languages are also useful because they can be altered as much or as little as one desires. On one hand, they can arise from disorder without outside intervention (e.g. [18]). Although some degree of internal logic on the part of the speakers is required to accomplish this, no intervention is required once the generation process commences. On the other hand, the researcher is able to alter the grammar of these languages at any point, which is impossible to do with natural language. Moreover, interactions between speakers of any two simulated languages can interact, which enables control in the proportion of L2 speakers present.

References

[1] D. Birdsong and J. E. Flege, "Regular-irregular dissociations in L2 acquisition of English morphology" in *BUCLD 25: Proceedings of the 25th Annual Boston University Conference on Language Development*. Somerville, MA, USA: Cascadilla Press, 2001, pp. 123–132.

[2] W. Croft, "The Darwinization of linguistics," *Selection*, vol. 3, no. 1, pp. 75–91, Nov. 2002, doi: 10.1556/select.3.2002.1.7.

[3] R. D. Perkins, *Deixis, Grammar, and Culture*. Amsterdam, Netherlands: John Benjamins Publishing Company, 1992.

[4] E. Sapir, "Language and environment," in *Amer. Anthropologist*, vol. 14, no. 2, pp. 226–242, Jun. 1912.

[5] C. Bentz and B. Winter, "The impact of L2 speakers on the evolution of case marking," in *The Evolution of Language*, T. C. Scott-Phillips, M. Tamariz, E. A. Cartmill, and J. R. Hurford, Eds., Singapore, Singapore: World Scientific, 2012, pp. 58–63.

[6] D. Nettle, "Coevolution of phonology and the lexicon in twelve languages of West Africa," *J. of Quantitative Linguistics*, vol. 5, no. 3, pp. 240–245, Jul. 2008, doi: 10.1080/09296179808590132.

[7] A. Wray and G. W. Grace, "The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form," *Lingua*, vol. 117, no. 3, pp. 543–578, Mar. 2007, doi: 10.1016/j.lingua.2005.05.005.

[8] G. Lupyan and R. Dale, "Language structure is partly determined by social structure," *PLOS One*, vol. 5, no. 1, Jan. 2010, doi: 10.1371/journal.pone.0008559.

[9] G. Lupyan and R. Dale, "Why are there different languages? The role of adaptation in linguistic diversity," *Trends in Cogn. Sci.*, vol. 20, no. 9, pp. 649–660, Sep. 2016, doi: 10.1016/j.tics.2016.07.005.

[10] G. Matiini, "Overgeneralization in singular/plural nouns and suffixed nouns of IELTS course students," *J. of Lang. and Literature Educ.*, vol. 16, no. 2, pp. 144–159, Oct. 2016, doi: 10.17509/bs_jpbsp.v16i2.4478.

[11] P. Sells, "Korean and Japanese morphology from a lexical perspective," *Linguistic Inquiry*, vol. 26, no. 2, pp. 277–325, May 1995.

[12] A. Housen, "Verb semantics and the acquisition of tense-aspect in L2 English," *Studia Linguistica*, vol. 54, no. 2, pp. 249–259, Dec. 2002, doi: 10.1111/1467-9582.00064.

[13] C. Bentz, A. Verkerk, D. Kiela, F. Hill, and P. Buttery. "Adaptive communication: Languages with more non-native speakers tend to have fewer word forms," *PLOS One*, vol. 10, no. 6, Jun. 2015, doi: 10.1371/journal.pone.0128254.

[14] A. Koplenig. "Language structure is influenced by the number of speakers but seemingly not by the proportion of non-native speakers," *Roy. Soc. Open Sci.*, vol. 6, no. 2, doi: 10.1098/rsos.181274.

[15] L. Becerra-Bonache, G. Bel-Enguix, M. D. Jiménez-López, and C. Martín-Vide, "Mathematical foundations: Formal grammars and languages," in *The Oxford Handbook of Computational Linguistics (2nd edn)*, R. Mitkov, Ed., London, United Kingdom: Oxford Univ. Press, 2018, pp. 207–229.

[16] A. Cangelosi and D. Parisi, "Computer simulation: A new scientific approach to the study of language evolution," in *Simulating the Evolution of Language*, A. Cangelosi and D. Parisi, Eds., London, United Kingdom: Springer-Verlag, 2002, pp. 3–28.

[17] H. Mannila, T. Nevalainen, and H. Raumolin-Brunberg, "Quantifying variation and estimating the effects of sample size on the frequencies of linguistic variables," in *Research Methods in Language Variation and Change*, M. Krug and J. Schlüter, Eds., Cambridge, United Kingdom, Cambridge Univ. Press, 2013, pp. 337–360.

[18] S. Kirby, "Spontaneous evolution of linguistic structure—An iterated learning model of the emergence of regularity and irregularity," *IEEE Trans. on Evol. Computation*, vol. 5, no. 2, pp. 102–110, Apr. 2001, doi: 10.1109/4235.918430.