Beyond Plain and Extra-Grammatical Morphology: Echo-Pairs in Hungarian

Márton Sóskuthy
University of British Columbia, Canada

Péter Rácz
Cognitive Development Center, Central European University, Hungary

Abstract
This paper presents an investigation of echo-pairs in Hungarian. Echo-pairs are formed by duplicating a base with an altered initial consonant and have diminutive, playful, or intimate connotations (e.g., *cica* [tsitʃ] “cat” → *cica-mica* [tsitʃ-mitʃ] “cat.dim”). Echo-pairs are commonly seen as an example of extra-grammatical morphology in the literature. Our goal in looking at this phenomenon is to gain a better understanding of the morphological mechanisms underlying extra-grammatical phenomena and shed new light on the distinction between plain and extra-grammatical morphology. We analyze data from (a) a collection of echo-pairs extracted from a large corpus of online texts and (b) a large-scale online nonce-word experiment with close to 1,500 participants. Our results reveal two key phonological patterns in the data and some additional systematic variation across words and experimental stimuli. We compare two different models of morphology, the Minimal Generalization Learner and the Generalized Context Model, in terms of their ability to capture this variation. We find that echo-pair formation is best captured by lexicon-oriented models such as the Generalized Context Model, but only when they rely on a structured similarity metric that encodes broader generalizations about the data. Our results do not support a clear-cut distinction between extra-grammatical and plain morphological processes, and we suggest that some of the peculiar characteristics of extra-grammatical phenomena such as echo-pair formation may simply follow from their special function and the limited set of contexts in which they appear.

Keywords
Echo-pairs, analogy, extra-grammatical morphology, morphological modelling
Introduction

Two broad categories of word-formation processes are distinguished in the literature: \textit{plain} and \textit{extra-grammatical} morphology (Zwicky & Pullum, 1987; Dressler & Barbaresi, 1994; Mattiello, 2013). The former category contains well-known processes such as affixation, reduplication, and templatic morphology. These processes share a number of core features (Zwicky & Pullum, 1987): they have a fixed function that varies little by context (e.g., plural, past tense, etc.); they apply to a specific input category and produce a specific output category (e.g., noun $\rightarrow$ verb); they show relatively little variation within and across speakers; and they rarely diverge from typical processes of word-formation in a given language (e.g., regular processes of reduplication or templatic morphology are rarely observed in agglutinating languages). Extra-grammatical morphology comprises a medley of processes that do not fit the above criteria in one way or another:\textsuperscript{1} blends (e.g., \textit{breakfast} $+ \textit{lunch} \rightarrow \textit{brunch}$), acronyms (e.g., \textit{President of the United States} $\rightarrow \textit{POTUS}$), echo-pairs (e.g., \textit{hodgepodge}), \textit{shm}-reduplication (e.g., \textit{fancy-shmancy}), ideophones (e.g., \textit{bling bling}), and truncations and clippings (e.g., \textit{Andrew} $\rightarrow \textit{Andy}$; \textit{laboratory} $\rightarrow \textit{lab}$).

Extra-grammatical morphology sits uneasily within the broader context of linguistic inquiry. On the one hand, it is a ubiquitous feature of human languages that often shows a great deal of systematicity (see e.g., Plag, 2003, Vihman, 2019). This alone warrants more detailed inquiry into its implementation and sources. On the other hand, most such processes are not easily captured by models of morphological competence based on core phenomena such as affixation. This is the source of the label “extra-grammatical,” which reflects a widespread conviction that such phenomena fall “outside the realm of morphology” (Haselmath & Sims, 2010, p. 40) and should not be treated alongside regular morphological rules.

The above description suggests a strict binary division between extra-grammatical versus plain morphology, which, however, may not be tenable. First, the extent of regularity in plain morphological processes is often overstated, with much recent work pointing to subtle and often lexically specific patterns of variation that are strongly reminiscent of extra-grammatical phenomena (see Section 1.2 for an overview). In fact, many current models of morphology take such patterns of variation as their key explanandum and draw on a rich lexicon coupled with analogical processes or intricate sets of highly specific rules and constraints to account for speakers’ variable morphological intuitions (e.g., Albright & Hayes, 2003; Dawdy-Hesterberg & Pierrehumbert, 2014; Gouskova et al., 2015). Second, not all extra-grammatical word-formation processes are created equal. They form a continuum from truly idiosyncratic and sporadic processes such as \textit{word manufacture} (e.g., brand names such as \textit{Kodak}; Bauer, 1983) through somewhat more regular processes such as blending (Plag, 2003) all the way to processes that are almost fully amenable to a rule-based analysis such as \textit{shm}-reduplication (Nevins & Vaux, 2003). Although processes from this far end of the continuum may still be distinct from plain morphological phenomena, it is plausible that they share at least some of their underlying mechanics.

In this paper, we examine one such phenomenon: echo-pair formation in Hungarian. Echo-pair formation is the creation of a new word through joining two forms that typically only differ in their initial consonant or their vowels (e.g., \textit{curly-wurly} and \textit{chit-chat} in English). In Hungarian, a particularly productive variant of this process involves duplicating a base form with an altered initial consonant to create a new form with diminutive, playful, or intimate connotations. We refer to the original form as the \textit{base} and the altered duplicate as the \textit{echo}. Examples include \textit{cica} [t͡ʃisɒ] “cat” $\rightarrow \textit{cica-mica}$ [t͡ʃisɒ-mɪsɒ] “cat. dim.” and \textit{csiga} [ʃiɡɒ] “snail” $\rightarrow \textit{csiga-biga}$ [ʃiɡɒ-biɡɒ] “snail. dim.” As these examples demonstrate, the identity of the echo onset (henceforth the \textit{echo behavior}) is variable: in this case, [m] for \textit{cica} and [b] for \textit{csiga}. 

Our main focus in this paper is on this variation in echo behavior and what it can tell us about the distinction between extra-grammatical versus plain morphology and the processes underlying echo-pair formation. We present a detailed analysis of data from two sources. The first of these is a collection of echo-pairs extracted semi-automatically from a large corpus of online texts (the Hungarian Webcorpus; Halácsy et al., 2004), previously described in Sóskuthy (2012). The second dataset comes from a large-scale online nonce-word experiment with close to 1,500 participants, which focuses on the initial consonant of the echo. We use these data to test for broad patterns of phonological conditioning in echo-pair formation of a type that is often associated with plain morphology, and to compare different models of morphology in terms of their ability to go beyond such broad patterns and capture variation of a finer grain in echo-pair formation.

The paper is structured as follows. The rest of this section gives an overview of echo-pair formation, provides further background to the theoretical discussion in the paper, and outlines our research questions. Sections 2 and 3 present the methods and main results of the corpus investigation and the experiment. Section 4 focuses on variation across words in both the corpus and experimental data. This across-word variation also serves as the focus of Section 5, which compares a range of different models to see which can best capture these patterns. Section 6 links the results of the preceding sections to our main research questions, and Section 7 concludes the paper with a brief summary.

The data and code for this paper are available online at the following address: https://doi.org/10.5281/zenodo.3833374.

1.1 Echo-pair formation in Hungarian and elsewhere

Echo-pair formation occurs in a variety of languages and is associated with a range of different meanings (Inkelas & Downing, 2015; Downing & Inkelas, 2015; Southern, 2005). We have already provided examples from English and Hungarian. Other languages with echo-pairs include Turkish (et-met “meat or something;” Southern, 2005, p. 60), Tamil (paampu-kiimpu “snakes and other such creatures;” Keane, 2001), Georgian (xili-mili “fruit and such;” Southern, 2005, p. 83), and Russian (katuška-matuška “my little bobbin;” Southern, 2005, p. 111). Southern (2005) lists examples from a wide range of languages and argues that echo-pairs with labial echo behavior (especially [m]) have their ultimate origin in Turkic languages and have spread across a wide geographical area through contact. Although the evidence presented in Southern (2005) is strongly suggestive of contact effects, it is almost certain that echo-pair formation arose as an independent innovation in at least some languages (e.g., English, which has had very little influence from Turkic and Slavic languages; Vihman, 2019).

The two component forms of echo-pairs typically differ either in their initial consonant or their vowels. Sometimes they differ in both. Hungarian offers examples of all of these combinations: cica-mica [t͡sits̩-mǐsdo] “cat.dim,” csireg-csőrög [t͡ʃi̯e̞reg-t͡ʃo̞rø̞ɡ] “jingle (playful),” zene-bona [zɛ̞n-boɲn] “loud and lively music” (Patay, 2015). The most widely applicable and productive type of echo-pair formation in Hungarian involves replacing the initial consonant of the base with a different consonant, which is typically labial (Sóskuthy, 2012; Patay, 2015). This preference for labials is noted for a range of Turkic and Slavic languages by Southern (2005) and for English by Vihman (2019).

There are a number of different types of echo-pairs in Hungarian:

1. forms where the base exists independently, whereas the echo does not, and the base precedes the echo (e.g., cica [t͡sits̩] “cat” → cica-mica [t͡sits̩-mǐsdo] “cat.dim”);
2. forms where the base exists independently, whereas the echo does not, and the base comes after the echo (e.g., \textit{pici} [pɨsi] “small” \text{\textrarr} \textit{ici-pici} [ɨsi-piʃi] “very small”);
3. forms where both parts of the pair exist independently (e.g., \textit{csillog} [ʃilːog] “glistens” + \textit{villog} [vilːog] “glints” \text{\textrarr} \textit{csillog-villog} [ʃilːog- vilːog] “is very clean”);
4. forms where neither component of the pair exists independently (e.g., \textit{ejnye-bejnye} [eʃɲe-bɛɲe] “tut-tut”).

In this paper, we focus on point 1: forward echo-pair formation based on an existing word. Our intuition is that this type of pattern is much more productive than the others listed above.

Forward echo-pair formation carries a diminutive meaning in Hungarian. The distribution of such echo-pairs is dominated by a few frequent lexicalized forms (see all the examples given so far). These lexicalized forms show consistent echo behavior: for instance, \textit{cica-mica} is always realized with an [m]-initial echo, and \textit{csiga-biga} with a [b]-initial echo. However, forward echo-pair formation is also used productively, especially in intimate contexts (e.g., as terms of endearment) and in child-directed speech. In such cases, there is substantial variation in echo behavior, and our experience is that speakers’ intuitions vary widely.

We refer to Zwicky and Pullum (1987) to tie the discussion of echo-pair formation in Hungarian to the general discussion of differences between plain and extra-grammatical morphology. They outline a list of characteristics that distinguish extra-grammatical (or, in their terminology, expressive) morphological processes from plain ones. Extra-grammatical morphology (a) has expressive, playful, and poetic associations; (b) is very flexible with regard to input category (i.e., applies to several different word classes); (c) applies to inflected forms; (d) is inaccessible for some otherwise fully competent native speakers; (e) allows several alternative outputs for the same input form; (f) varies across speakers; and (g) has special syntactic properties (e.g., words formed through \textit{shm}-reduplication are typically not incorporated into full sentences).

Forward echo-pair formation in Hungarian fits some but not all of these criteria. It has strong stylistic connotations, can apply to nouns as well as verbs or adjectives, and it can also apply to inflected forms, which means it fits criteria (a)–(c). The situation is less clear with respect to criteria that relate to variation and speakers’ control of the process. Adult speakers are generally aware of this pattern and are able to use it productively, contra (d). Speakers are remarkably consistent in applying forward echo-pair formation to existing forms, and they almost always produce a labial echo behavior (cf. the corpus data in section 2; contra (e) and (f)). However, they do show variation between different labial consonants (especially [b] and [m], e.g., \textit{izé-bizé} [ize-ː-bize] \textasciitilde \textit{izé-mizé} [ize-ː-mize] “thingy”) and there is likely variation within and across speakers, which means the phenomenon does at least partly fit criteria (e) and (f). Finally, words created through echo-pair formation are syntactically regular and can be incorporated into full sentences, contra (g).

Sóskuthy (2012) identifies a number of phonological patterns in forward echo-pair formation. In this paper, we focus on two of these. First, if the base starts with a vowel (i.e., when the initial onset is not filled), the echo behavior tends to be [b] (e.g., \textit{undi-bundi} [undi-bundi] “yucky.dim”). For convenience, we refer to this pattern as \textit{filled O1}. Second, if the consonant immediately following the first vowel is voiceless, the echo behavior tends to be [m] (e.g., \textit{csiga-biga} [ʃiŋ-biŋ] “snail.dim”). This pattern will be referred to as \textit{voiceless V1C}.

In sum, Hungarian forward echo-pair formation sits somewhere between plain and extra-grammatical morphology. As with extra-grammatical patterns, it is relatively rare, playful, morphologically forgiving, shows an unusual distribution in the ambient language and—as a reduplication process—is unusual in the broader context of Hungarian morphology. However, as with plain morphological patterns, most speakers are aware of it, are consistent in using labial echo behaviors, and are able to use echo-pairs in a syntactically regular way. In the following section, we argue that
these attributes of echo-pair formation carry relevance for the broader discourse on morphological processes.

1.2 Models of morphology and the plain versus extra-grammatical distinction

As discussed in the previous section, “plain” morphological processes are traditionally seen as deterministic, regular, and productive, showing little or no variation. Models developed to capture such processes account for this regularity by using rules (Chomsky & Halle, 1968; Halle & Mohanan, 1985). Rules are conceptualized as symbol manipulations applying to abstract underlying representations. In classical generative models, rules are seen as deterministic and exceptionless. Apparent exceptions (such as irregular past tense forms in English) result from the differential application of categorical rules to forms with distinct underlying representations (thus, irregular past tense forms might be marked by diacritics or underlying segments that are not visible on the surface). Although the existence of variability is not denied in such accounts, it is argued to be outside the scope of linguistic competence and is relegated to the fringes.

Zwicky and Pullum’s (1987) approach to the plain versus extra-grammatical distinction derives directly from this view of morphological competence. They show that extra-grammatical morphological phenomena are not only highly variable and of limited productivity, but also often restricted in style and register, associated with playful and/or child-directed language use, and unusually lenient in terms of eligible input forms (see Section 1). They argue that such processes cannot be captured in traditional rule-based accounts, and are therefore outside the purview of grammar. This sets up a categorical distinction between plain and extra-grammatical processes, and implies that models of one have little insight to offer to the other. However, as we show in the rest of this section, models of plain morphology have come to abandon many of the assumptions detailed in the previous paragraph. Our paper makes an important contribution by showing that most aspects of Hungarian echo-pair formation can be captured by these newer models of plain morphology.

There is increasing recognition that characterizing plain morphological processes as fundamentally categorical is flawed (see e.g., Hay & Baayen, 2005). Even basic inflectional processes are often variable. In German, a number of schemata compete for low-frequency words and neologisms in plural formation (see Köpcke, 1988). Variation in Hungarian verbal inflection is rampant and shaped by both lexical similarity and usage frequency (Rácz et al., 2018). Effects of lexical “gangs” that compete with broader generalizations can be observed in both inflectional (Bybee & Moder, 1983) and derivational (Alegre & Gordon, 1999) morphology. Even regular rules do not apply uniformly to new input. In particular, lexical consolidation has been shown to play into the application of the regular English past tense to nonce-word input (Lindsay et al., 2012). The strength and structure of lexical representations has a major role in morphological processes, as evidenced by work on languages ranging from Arabic (Dawdy-Hesterberg & Pierrehumbert, 2014) to Zapotec (Stemberger & Chávez-Peón, 2014).

In response to these observations, models of morphology have gradually shifted away from the idea that morphological competence must be encoded by deterministic rules. There have been numerous attempts to capture variability and irregularity in the context of plain morphology. Some of the earliest modelling efforts were developed in the broader framework of connectionism (Rumelhart & McClelland, 1986, 1987). These models eschew rules altogether and assume the availability of a rich mental lexicon, a network of relationships among lexical items, and associative/analogical mechanisms that are capable of using these relationships to generate novel outputs. The early success of connectionist modelling has inspired a range of similar approaches (Skousen, 1989; Bybee, 1995; Dawdy-Hesterberg & Pierrehumbert, 2014), which (although different in their
implementational details) share these fundamental assumptions about the nature of the lexicon and morphological productivity.

Connectionist models and their cousins have been successfully applied to a wide range of phenomena. The most popular of these is past-tense formation in English, which includes a complex set of irregular processes (e.g., *sing* → *sang*, *think* → *thought*) alongside a simple and regular process of affixation (+ *ed* [t/d/ld], e.g., *lift* → *lifted*). One of the main strengths of these approaches lies in their ability to discover and model pockets of regularity in highly variable data. For instance, there is a cluster of English verbs that form their past tense by changing an [i] in the present tense form to [æ] in the past tense form (e.g., *swim* → *swam*, *sing* → *sang*, *stink* → *stank*). These verbs share a number of phonological characteristics: they are all monosyllabic, their initial consonant is often [s] (but see e.g., *ring* → *rang*), and they typically end in a nasal or a nasal + stop cluster. Connectionist and analogical models have been shown to successfully capture such sub-patterns with minimal supervision (Rumelhart & McClelland, 1986).

Although connectionist and analogical approaches perform well with respect to variation and irregular patterns, their early implementations suffer from a number of flaws (e.g., odd outputs for certain forms, failure to capture systematic regularization in denominal verbs such as *grand-stanced*; Pinker & Ullman, 2002). Dual-route models such as Prasada and Pinker (1993), Pinker and Ullman (2002), and Baayen et al. (1997) address these issues by assuming that morphologically complex forms can be stored wholesale but can also be derived through the application of rules. In Prasada and Pinker’s (1993) model, storage is reserved for irregular forms (such as *sang*, *thought*), whereas regulars are (typically) formed through the application of rules. In Baayen et al.’s (1997) model, the two access mechanisms compete and are mediated by lexical frequency, resulting in a situation where high-frequency regular forms are often accessed wholesale. By positing two separate access mechanisms, dual-route models capture important patterns of variation in the speed of access (Baayen et al., 1997), neurophysiological dissociations and systematic regularization (Pinker & Ullman, 2002). In addition, some versions of dual-route models also assume that the “lexicon has superpositional properties similar to a pattern associator” (Pinker & Ullman, 2002, p. 458). Such models retain the advantages of connectionist and analogical approaches, such as the ability to account for pockets of regularity.

Another set of models attempts to account for variable morphological behavior relying entirely on rules (Albright & Hayes, 2003; Albright, 2009) or constraints (Gouskova et al., 2015; Becker & Gouskova, 2016, building on work by Hayes & Wilson, 2008). Although these models share the characteristic formalisms of rule/constraint-based approaches, they depart from them significantly in the way they use these tools. The Minimal Generalization Learner (MGL; Albright & Hayes, 2003; Albright, 2009; described in more detail in Section 5.2) constructs a large number of potentially overlapping rules that vary widely in terms of their generality and often embody lexically specific patterns. These rules differ from those of traditional rule-based accounts not only in terms of their potential specificity, but also in the fact that they are allowed to apply stochastically.

Gouskova et al. (2015) present a model based on a constraint-based maximum entropy grammar (Hayes & Wilson, 2008). They assume that (a) roots are indexed in memory for the allomorphs that they typically combine with (e.g., *lift* is indexed for -*ed*, whereas *cut* is indexed for zero); (b) words with different indices are organized into sub-lexicons; and (c) for each of these sub-lexicons, speakers construct distinct sets of phonological constraints and assign weights to them, representing their knowledge of the phonotactic patterns associated with different allomorphs. Under normal circumstances, allomorph choices (e.g., regular vs. irregular past tense allomorphs) are based purely on the arbitrary indices carried by roots. This mechanism is reminiscent of the wholesale access route in dual-route models in that the phonological shape of the root plays no role in choosing an allomorph, which is simply retrieved from memory. However, this access route is sometimes
unavailable (e.g., when dealing with novel or infrequent forms). In such cases, speakers decide between different allomorphs for a given root based on the likelihood of that root under the constraint rankings for each allomorph-specific sub-lexicon. Gouskova et al. (2015) show that this model can account for fine-grained and probabilistic patterns of phonological conditioning in Russian speakers’ choices of diminutive allomorphs—a phenomenon that is, in many ways, similar to Hungarian echo-pair formation.

Although extra-grammatical phenomena seem truly outside the realm of grammar when viewed through the lens of categorical rules, the distinction between plain and extra-grammatical processes seems much less clear cut under the approaches reviewed above. As we have seen, flexibility in the input and variability in the output is often seen in plain morphological processes as well (e.g., past-tense formation, plurals, etc.). Plain morphology can and often does entail a number of competing patterns, effects of formal similarity, and sometimes quite striking variation within and across speakers. The models discussed above can all capture these phenomena to some extent.

This naturally leads to the following question: can models that have cut their teeth on plain morphology capture variation in extra-grammatical processes? This question is important for two reasons. First, if extra-grammatical phenomena are found to be amenable to analysis through models developed for plain morphology, it would suggest they emerge from a shared grammatical substrate. Second, comparing the performance of different models in accounting for variation in extra-grammatical phenomena may reveal crucial cues about the nature of their underlying mechanics. Echo-pair formation in Hungarian, then, becomes an especially interesting test case as it straddles the line between plain morphology and extra-grammatical morphology.

1.3 Research questions

Our main focus in this paper is on the nature and sources of variation in echo behavior. We focus on the following research questions.

1. **To what extent do speakers conform to established patterns of phonological conditioning in echo-pair formation?**

   Plain morphology often shows crisp and conventionalized patterns of phonological conditioning, whereas much of extra-grammatical morphology is characterized by gradient and variable conditioning. It is therefore important to map the extent to which speakers of Hungarian conform to the broad patterns of phonological conditioning outlined above (filled $O_1$ and voiceless $V_1C$). We use corpus data and experimental data to address this question in Sections 2 and 3.

2. **Do words show variation in echo behavior above and beyond the two main phonological patterns?**

   The primary interest of across-word variation comes from the fact that it speaks directly to the grammatical machinery underpinning echo-pair formation. If most variation across words is simply due to established patterns of phonological conditioning, a simple rule-based mechanism might be sufficient. However, substantial across-word variation beyond these patterns would call for a different approach such as the models outlined in Section 1.2. We use corpus data and experimental data to map across-word variation in echo-pair formation in Section 4.

3. **What type of model can best account for the variation in Hungarian echo-pair formation?**
As suggested above, comparing different models in terms of their performance can inform us about the grammatical basis of echo-pair formation. In Section 5, we test four different fully implemented models: a simple variable rule-based model, the MGL (Albright & Hayes, 2003) and two different versions of the Generalized Context Model (GCM; Nosofsky, 1986; Dawdy-Hesterberg & Pierrehumbert, 2014). These models were chosen as they vary in terms of their degree of reliance on lexical information: a simple rule-based model uses no lexical information; the MGL can capture lexically specific behavior within certain limits; and the GCM is an analogical model that excels at handling lexical exceptions. Comparing these models therefore allows us to estimate the depth of the lexical basis of echo-pair formation in Hungarian.

2 Corpus analysis

2.1 Methods

We analyze patterns of echo-pair formation in a subset of the echo-pair corpus from Sóskuthy (2012), focusing on forward echo-pair formation. A more detailed account of the creation of our corpus is provided in Sóskuthy (2012). Here, we focus on the essentials.

The corpus was extracted from the 500 million-word Hungarian Webcorpus (Halácsy et al., 2004), which offers written text produced by a wide range of speakers in a variety of registers. Many of these texts are relatively informal, which would favor the occurrence of echo-pairs. Using regular expressions, we automatically extracted a set of roughly 4,000 word types that could potentially result from echo-pair formation, and manually removed false positives. We then collapsed inflected forms into a single entry for each lexeme, and manually identified echo-pairs that were created through forward echo-pair formation. This left 262 echo-pair types.

2.2 Results

As noted in Sóskuthy (2012), labial-initial echos are far more common than would be expected by chance, and this trend is especially strong among echo-pairs created through forward echo-pair formation: 87.4% of all types in our dataset have an echo that starts with [b, m, p, f, v] (in order of frequency). When looking at token frequencies, this trend becomes even stronger: 98.7% of all tokens have a labial-initial echo. This suggests that non-labial echo behaviors are limited to forms that only occur sporadically. The group of forms with labial-initial echos is dominated by two echo behaviors of roughly similar type frequency, [b] (111 types) and [m] (82 types), which make up 73.6% of all types (and 89.1% of all tokens). Because the phonological patterns that we focus on relate specifically to these two echo behaviors, and for reasons of tractability, the rest of this paper considers only the 193 forms with an echo starting in [b] or [m] (for more information on other echo behaviors, Sóskuthy 2012).

Let us now turn to the two main phonological patterns that our paper focuses on. Figure 1 illustrates the distribution of [b] versus [m] behaviors as a function of filled \( O_1 \): whether the initial onset of the base is filled and voiceless \( V_1C \): whether the consonant after the first vowel is voiceless.\(^2\) Forms without an initial onset (\( \text{édi-bédi} \) [edi-bedi] “cute” \( \text{dim} \)) are shown in the left-hand panel, whereas forms with an initial onset (\( \text{cica-mica} \) [t͡ʃiːts-miːts] “cat” \( \text{dim} \)) are shown in the right-hand panel. Within each panel, forms with a voiced postvocalic consonant (\( \text{csönka-bonka} \) [ʃoŋk-boŋk] “truncated” \( \text{dim} \)) are shown on the left, and forms with a voiceless postvocalic consonant (\( \text{pičur-micúr} \) [piʃur-miʃur] “tiny” \( \text{dim} \)) on the right. Both patterns receive ample support from the corpus data: 83.7% of forms with an empty initial onset take an echo with [b] compared to 51.2% for forms with a filled initial onset; and 70.1% of forms with a voiced postvocalic
Consonants take an echo with [b] compared to 26.8% for forms with a voiceless postvocalic consonant. Figure 1 shows little, if any, interaction between the two patterns.

Voiceless $V_1C$ can also be expressed as a restriction on the echo only: if the second consonant of the echo is voiceless, the first consonant is [m] (as opposed to [b]). This raises the interesting possibility that this pattern simply reflects a general preference in the lexicon of Hungarian for postvocalic voiceless consonants to be associated with an initial [m]. We tested this hypothesis by looking at [b] and [m]-initial words among the 10,000 most frequent disyllabic noun lemmata in the Hungarian Webcorpus. The proportion of initial [b] (vs. [m]) among forms with a voiced postvocalic consonant is 59.2%; the same proportion is 49.8% among forms with a voiceless postvocalic consonant. In other words, there is a preference for initial [m] in the presence of a voiceless postvocalic consonant in the lexicon of Hungarian at large, but this preference is much weaker than the trend observed in our echo-pair corpus (where the same proportions are 70.1% and 26.8%). Therefore, general phonotactic preferences cannot account for this pattern on their own.

To test the robustness of the two phonological patterns statistically, we fit a logistic regression model to the 193 types in our dataset. The outcome variable is echo behavior. The predictors are whether $O_1$ is filled (filled $O_1$), whether the postvocalic consonant is voiceless (voiceless $V_1C$), and their interaction. Our model shows there is no significant interaction between filled $O_1$ and voiceless $V_1C$ ($\chi^2 = 0.82$, $df = 1$, $p = 0.36$). However, both filled $O_1$ ($\chi^2 = 10.3$, $df = 2$, $p < 0.01$) and voiceless $V_1C$ ($\chi^2 = 26.97$, $df = 2$, $p < 0.001$) exert a significant overall influence on echo behavior on their own. Figure 1 shows the predictions from this model overlaid on the raw data, with 95% confidence intervals.

In summary, the corpus shows statistical support for the two phonological patterns discussed above. At the same time, most forms in this corpus are exceedingly rare, and, as a consequence, are likely unknown to most Hungarian speakers. Indeed, there are many forms that the two authors of this paper (both native speakers of Hungarian) have never encountered before. This would mean that broad generalizations cannot be made on the basis of these forms. To avoid this issue, we also tested the two patterns in a subset of the 35 most frequent types in our corpus. We used 35 as our

![Figure 1. The colored bars show proportions of echo behavior as a function of filled $O_1$ and voiceless $V_1C$. The filled circles and whiskers indicate predictions and 95% confidence intervals from a logistic regression model.](image-url)
cut-off as it was after this number that forms unknown to both authors started appearing in the list. This test yielded similar results to the ones shown above. However, the results of these tests must be considered inconclusive: a dataset with only 35 types is simply too small to calculate reliable proportions for each combination of filled O₁ and voiceless V₁C.

3 Experiment

3.1 Methods

We ran an online binary forced-choice experiment via Qualtrics asking participants to choose between two possible echo variants for disyllabic nonce stems: one with an initial [m] and one with an initial [b]. Our goal was (a) to see whether participants’ behavior is also governed by the patterns uncovered in the corpus data, and (b) to get a sense of the extent and precise nature of variation across items. In this section, we focus on the first goal. Sections 4 and 5 take up the issue of across-item variation. The experiment was approved by the Ethics Committee of the Department of Language and Linguistic Science at the University of York (where this experiment initiated).

3.1.1 Stimuli. We created 48 disyllabic nonce forms to test participants’ echo behavior preferences. A large initial pool of forms was generated automatically based on existing stems that are associated with an [m] or [b] echo behavior (i.e., the 193 forms that were entered into the logistic regression models described in the previous section). The forms were all of the following general shape:

\[(O₁) N₁ O₂ N₂ (C₂)\]
\[(O = \text{onset}, N = \text{nucleus}, C = \text{coda})\]

That is, we created disyllabic forms with an optional word-initial onset and an optional word-final coda, without a coda in the first syllable and with an obligatory onset in the second syllable (e.g., [aːtsɔː], [uːsoʊ], [goːci], [vikɛr]). O₁ and O₂ were sampled from initial and word-medial onsets in the corpus; N₁ and N₂ were sampled together from pairs of nuclei in the corpus (to account for patterns of vowel harmony in Hungarian); and C₂ was sampled from word-final codas in the corpus as a function of the previous vowel (using a bigram model). To ensure that the generated forms do not already exist in Hungarian, we discarded all forms that were listed among the 50,000 most frequent disyllabic wordforms in the Hungarian WebCorpus. We also excluded forms that would yield echos that are listed among the same set of wordforms. Finally, we discarded all forms that included a labial stop ([m], [b] or [p]) in any position.

We then selected 48 stimuli from this pool of candidates in a $2 \times 2$ balanced crossed design: 12 forms with an empty O₁ and a voiceless consonant in O₂ (e.g., [aːtsɔː]); 12 forms with a filled O₁ and a voiceless consonant in O₂ (e.g., [goːci]); 12 forms with an empty O₁ and a voiced consonant in O₂ (e.g., [odoː]); and 12 forms with a filled O₁ and a voiced consonant in O₂ (e.g., [zɛdaː]). The selection procedure was partly automatic and partly based on manual inspection. Some of the selected forms were slightly modified to yield more natural-sounding nonce forms or to avoid existing forms that were not successfully filtered out at the previous stage (e.g., because they did not appear among the 50,000 most frequent forms in Hungarian). The full list of stimuli is included in the Appendix.

3.1.2 Procedure. Participants were first presented with a series of instructions that briefly explained echo-pair formation without drawing attention to the fact that echo behavior is influenced by the
phonological makeup of the stem. They were given examples that are frequent but that do not form an echo-pair with labial echo behavior. Participants were then told that they would be given two potential echo behaviors for a set of non-existent Hungarian words, and they would have to choose the one that sounded the most natural to them.

Each of the 48 stimuli was presented on a separate page. The nonce stem was shown in standard Hungarian orthography at the top. The two echo pair alternatives with [m] and [b] were shown next to each other below (e.g., stem gótyi [goːci] and echo pairs gótyi-mótyi [goːci-moːci] and gótyi-bótyi [goːci-boːci]). Both the order of the stimuli and the order of the echo pairs on the screen were randomized separately within each participant.

We recorded participants’ choices ([m] or [b]), which side their choice appeared on (left or right) and the amount of time they took to answer each question. We also asked participants to specify their sex, age, level of education (i.e., how many years they have spent in education starting from primary school and potentially leading up to the end of their postgraduate studies), and current place of residence.

3.1.3 Participants. We recruited participants using a snowball sampling method relying on social media. The experiment was completed by a total 1,496 participants. Due to the nature of our sampling method, we could not control factors such as sex, age, education, or location. Women are strongly overrepresented in our sample with an overall proportion of 72% (n = 1084). Our participants cover a reasonable age range with a median age of 38, a lower quartile of 30 and an upper quartile of 45. We also asked participants to specify how many years they have spent in education (starting from primary school and potentially leading up to the end of their postgraduate education). The median for level of education is 17 years, which typically corresponds to 12 years in primary and secondary education, and 5 years at university or college. There is minimal variation in level of education, with a lower quartile of 15 and an upper quartile of 18. This suggests that the majority of our participants are university educated. Participants were also asked to indicate their current place of residence. The majority of our participants (n = 822, 55%) reside in Budapest, the capital, which holds roughly 20% of the population of Hungary. Other locations are represented by much smaller numbers of participants.

In summary, the majority of our participants are female, between 30–45 years old, highly educated and from the capital. Although this is certainly a biased sample, it is still substantially more diverse than typical experimental pools in psychological and linguistic experiments, which tend to be restricted to undergraduate students from a single location. It should also be noted that a preliminary analysis of the results (not reported here) found no appreciable effects of sex, age, or education (place of residence was not tested due to the diversity of locations in the sample). Therefore, it is unlikely that sampling issues have a substantial influence on the generalizability of our findings.

3.1.4 Data processing. We filtered our dataset in a number of ways to exclude responses that were unreliable and participants who were clearly not following the experimental instructions. First, some participants consistently chose answers that appeared on a specific side of the screen. We excluded 51 participants who chose the left-hand side option more than 38 times (~80% of all trials), and four participants who chose the right-hand side more than 38 times. We excluded a further 72 participants whose median response time fell in the bottom 2.5% of all participants (lower than 2.9 seconds) or in the top 2.5% (higher than 8.9 seconds). Responses to individual stimuli with a recorded response time higher than a minute were also dropped from the data (n = 182), and we excluded a further four participants who had more than three such responses.

Our final dataset includes 65,089 responses from a total of 1,359 participants.
3.2 Experimental results

Figure 2 shows the raw results from the experiment in a format analogous to Figure 1. Three trends are immediately apparent from this visualization of the results. First, there is a very slight preference for [m] as the echo behavior (it is chosen in 55% of all the trials), but the results are fairly close to chance (50%) across most contexts. Second, the rate of [b] responses is generally higher when O1 is empty (which is consistent with filled O1) and lower when the consonant in O2 is voiceless (which is consistent with voiceless V1C). In other words, participants appear to follow the subtle patterns of phonological conditioning observed in the corpus data. Third, the differences across conditioning contexts are not nearly as strong as in the corpus results: the overall difference in echo behavior between empty versus filled O1 is 8% (compared to 33% in the corpus), whereas the overall difference between voiced versus voiceless O2 is 7% (compared to 43% in the corpus).

We fit a mixed-effects logistic regression model to the results to test whether the observed differences are statistically significant. As before, the outcome variable is echo behavior. The predictors are filled O1, voiceless V1C and whether the [b] response appeared on the left-hand side of the screen or not (left [b]; included as a control variable). Because the corpus results did not provide support for an interaction between filled O1 and voiceless V1C, the model fitted to the experimental data does not include any interaction terms. The model also included the following random effects: random intercepts by item, random intercepts by speaker, and random slopes over filled O1 and voiceless V1C by speaker (along with all corresponding correlation parameters). The predictions from the model are shown in Figure 2 alongside the raw data. Model comparisons reveal significant effects of filled O1 ($\beta = -0.4$, $\chi^2 = 6.14$, $df = 1$, $p = 0.014$) and voiceless V1C ($\beta = -0.34$, $\chi^2 = 4.64$, $df = 1$, $p = 0.031$). Although these effects are significant, they are dwarfed by the effect of left [b]: when the [b] response appeared on the left-hand side, participants were 15% more likely to choose it (based on the raw data; significance according to model comparison: $\chi^2 = 1763.5$, $df = 1$, $p < 0.001$)\(^5\).

In summary, the results of the experiment suggest that participants’ choices about echo behavior are subtly influenced by the phonological patterns uncovered in the previous section, providing a
partial answer to RQ1 (see Section 1.3). However, these effects are weak: their absolute sizes are much smaller than those of the corpus results. Moreover, participants appear to latch on to random factors such as the location of the response on the screen to guide their responses. This would be much less likely to happen if their choices were strongly grounded in phonological patterns. Based on this and informal comments from some participants, it appears that the task of deciding between the two echo behaviors was not a straightforward one.

4 Variation across items

So far, we have focused on general trends in the data without discussing how individual items deviate from them. The standard deviation of the by-item random intercepts in the model from the previous section is 0.53, which is high compared to the main predictors, whose coefficients are \(-0.4\) (\(\text{filled } O_1\)) and \(-0.34\) (\(\text{voiceless } V_1C\)). To translate this into probabilities, although the effects of the main predictors on echo behavior are on a scale of 8–10%, individual words can deviate from the predicted probabilities by up to 25%. This is clearly visible in Figure 3, which shows raw proportions of echo behavior separately for each item (grouped by the main predictors). For instance, the item [eːdɛɡ] (rightmost bar in the leftmost panel) takes an echo with an initial [b] for nearly 71% of our participants, whereas the same figure is only 23% for the item [sɒtʃoɡ] (first bar in the rightmost panel). The filled black circles and 95% confidence intervals indicate roughly where the proportions for individual items should lie if there was no systematic variation across items apart from the effects of \(\text{filled } O_1\) and \(\text{voiceless } V_1C\) (based on posterior predictive simulations from the model in Section 3.2 with by-item random intercepts suppressed). The overwhelming majority of items are well outside the predicted ranges, suggesting that (a) there is a great deal of variation across items that is not due to the phonological patterns we have identified, and (b) participants had relatively strong intuitions about certain items.
Although a lot of the variation in Figure 3 seems idiosyncratic, there are some more general patterns relating to \(O_2\). Focusing first on the panels corresponding to items with a voiced \(O_2\) (which are predicted to have more [b] responses based on the corpus data), it seems that items with sonorants in \(O_2\) (e.g., [uɲog], [oːli], [kunor]) attract a lower proportion of [b] responses than items with obstruents. This trend seems fairly reliable: most of the items that deviate from the predicted pattern in a negative direction (i.e., more [m] responses) have a sonorant in \(O_2\) (three out of six in the first panel and four out of five in the third panel). One possible interpretation of this trend is that participants made a distinction between active voiced consonants versus sonorants/voiceless obstruents instead of a simple distinction between voiced versus voiceless consonants. Second, items with fricatives in \(O_2\) have generally lower proportions of [b] responses. However, both of these observations ought to be treated with caution: the experiment was not set up to test patterns beyond the ones we identified in the corpus data, and the range of sonorants and fricatives exemplified in the data is severely limited.

The substantial amount of across-word variation alongside the presence of possible additional regularities suggests that a simple rule-based account cannot fully capture the behavior of echo-pairs in Hungarian. The question, then, is whether models developed to capture finer-grained patterns are predictive of echo behavior in a way that goes beyond the two main phonological patterns. The next section explores this possibility by comparing the performance of different probabilistic models.

5 Modelling echo-pair formation

The previous section suggested that across-word variation in echo behavior goes beyond what can straightforwardly be captured using broad phonological generalizations. In this section, we ask (a) whether models allowing for finer-grained generalizations can account for some of this additional variation; (b) whether broad generalizations turn out to be epiphenomenal under these models; and (c) what specific type of grammatical architecture best captures the observed patterns. We compare four different models. The first of these is a simple rule-based architecture that relies on the two main phonological generalizations identified above. This model will be used as a baseline in assessing the others. The second model is the MGL (Albright & Hayes, 2003), which also relies on a rule-based formalism but constructs rules automatically from the data, allows them to apply stochastically, and is capable of identifying rules of vastly different specificity (ranging from rules capturing lexical idiosyncrasies to broad rules of the type discussed above). The third and fourth models eschew rules altogether and attempt to capture echo behavior relying on whole-word similarity instead. Both of these models are implemented using the formalism of the GCM (Nosofsky, 1986; Dawdy-Hesterberg & Pierrehumbert, 2014; Rácz et al., 2018). They differ in that the third model (which we term the "informed GCM") can learn to weight different aspects of similarity differently (e.g., by putting a heavier weight on consonantal than vocalic material), whereas the fourth model (the "naive GCM") weights all aspects of similarity to the same extent.

As noted in Section 1.3, these four models can be arranged along a continuum that represents their degree of reliance on lexical information. While a simple rule-based architecture uses no lexical information, the MGL can capture lexically idiosyncratic behavior to some extent—though rules representing such behavior will receive lower weights due to their limited generality (see below). Closer to the other end of the continuum, the informed GCM relies heavily on lexical information in the form of whole-word similarity, but complements it with generalizations about what types of similarity it should attend to. Finally, the naive GCM relies solely on lexical information and does not use any broader generalizations. By assessing these models, we not only gain information about the plausibility of their specific implementational details, but we also learn more
generally about the depth of the lexical basis of echo-pair formation in Hungarian. Although it is important to sample models that represent the full range of variation along this continuum, we do not mean to suggest that all of these models are equally plausible. In fact, most contemporary models lie closer to the middle of this spectrum (such as the MGL and the informed GCM). We also do not mean to imply these are the only existing models that could be tested. The models that we focus on suffice for our purposes in that they represent a broad range of variation in terms of lexical specificity—but there are other alternatives such as Gouskova et al.’s (2015) maxent-based model that are not examined here.

Our general strategy for evaluating these models is as follows. First, we train the models on the corpus data using supervised learning. This has already been done for the simple rule-based model in Section 2 by fitting a logistic regression model to the corpus data. This logistic model allows us to make predictions about the probabilities of [b]/[m]-initial echos based on filled O₁ and voiceless V₁C. This is equivalent to using variable phonological rules (cf. Cedergren & Sanko, 1974) to make predictions about echo behavior. For the MGL and the GCM, training consists of extracting rules and parameter values from the corpus data. The trained models can then be used to make predictions about the likelihood of a specific form taking an echo with [b]. These predictions are compared to the patterns observed in the corpus and the participants’ behavior in the experiment.

In what follows, we provide a brief outline of the MGL (5.1) and the GCM (5.2). This is followed by an evaluation of the different models on the basis of their performance on the corpus data (5.3) and the experimental data (5.4).

5.1 The MGL

The MGL (Albright & Hayes, 2003; Albright, 2009) is based on the assumptions that (a) morphological behavior is rule based, (b) there may be many different rules of different degrees of generality implementing the same morphological process, and (c) rules apply stochastically. The MGL discovers these rules directly from the data through a learning algorithm that receives morphologically related pairs (in our case, the base and the echo) as its input and generates a large set of weighted rules as its output. For instance, an MGL trained on our data may construct several different rules whose structural descriptions are met by the base cica [tʃiʃsɒ]. Some possible examples are shown below:

1. \{C, \emptyset\} → m / # __V[-voi]
2. \{C, \emptyset\} → m / # __Vtʃs
3. \{C, \emptyset\} → b / # __i

These rules will be weighted differently as a function of their generality (i.e., the number of forms that match their structural description) and their accuracy (i.e., the proportion of matches for which they make correct predictions). Their probability of applying to a given form is calculated as a function of these weights, such that more general and more accurate rules are more likely to apply. Thus, 1 and 2 above would likely receive heavier weights than 3, and therefore would be more likely to generate the echo corresponding to cica [tʃiʃsɒ].

At the heart of the MGL is a learning algorithm that constructs rules in an iterated fashion. It starts with a set of lexically specific rules, which are then compared pairwise to construct more rules with broader coverage if possible. This process is iterated many times, resulting in gradually more general rules. For instance, after comparing the lexically specific rules for cica-mica [tʃiʃsɒ-miʃsɒ] and kócimóci [koʃsɪ-moʃsɪ] “dishevelled.” the MGL will construct a rule along the lines of 2 above. If, at a later iteration, it compares this new rule to the lexically specific rule for
husi-musi [huʃi-muʃi] “meatDIM” (also used as a term of endearment), it will come up with a general rule closer to 1. Both general and specific rules are retained by the system.

The version of the MGL trained on our data used the default settings of the MGL as implemented by Albright and Hayes (2003), available from the following website: http://www.mit.edu/albright/mgl/. The model did not use the token frequency of training words. To allow feature-based generalization, we adapted the feature matrix in Albright and Hayes (2003) to Hungarian. The supplementary materials provide more information about the details of the model-fitting procedure. This included a reshuffling of the initial segments of the input forms to allow the MGL to make feature-based generalizations on the basis of $O_1$ and $C_1 / O_2$. Otherwise, these consonants would not have been available for feature-based generalization due to locality constraints and further specifics of the MGL algorithm. We note that the version of the model reported in the paper provided a better fit to both the corpus and the experimental data than a model fitted to the unmodified inputs.

The relative probabilities of [m]/[b]-initial echo forms were calculated following a method proposed by Albright and Hayes (2003). This involves finding the highest-rated rules with [m] and [b] as their outputs (this is done automatically by the MGL), and then setting the probabilities of these outcomes to be proportional to their weights. For instance, if the two highest-rated rules for a nonce-form pice [piʃse] are 1 and 3, and their weights are 0.7 and 0.1, the probability of an [m]-initial echo will be $0.7 / (0.7 + 0.1) = 0.875$, whereas the probability of a [b]-initial echo will be $0.1 / (0.7 + 0.1) = 0.125$.

5.2 The GCM

The specific implementation of the GCM used in this paper is based closely on Nosofsky (1986) and can be downloaded from the following website: https://github.com/soskuthy/rgcm (Sóskuthy, 2015). This section provides an outline of the GCM at a conceptual level. The supplementary materials provide more information about its computational details.

The GCM operates on instances that are specified as a combination of (a) a set of features (in this case, its segmental makeup) and (b) the outcome category it belongs to (in this case, its echo behavior). We use six features for describing individual forms, each corresponding to a segmental position in a specific syllable: $O_1$, $N_1$, $C_1$ (the onset, nucleus and coda of the first syllable), and $O_2$, $N_2$, $C_2$ (the same positions for the second syllable). The values of these features are simply the segments in the relevant positions. Thus, the form nyulam [ɲulɒm] “my rabbit” is specified as $O_1=[ɲ]$, $N_1=[u]$, $C_1=∅$, $O_2=[l]$, $N_2=[ɒ]$, $C_2=[m]$. The GCM calculates the distance between forms by counting the number of featural mismatches: for instance, the distance between nyulam and cica [ʦiʃɑ] “kitty” $O_1=[ʦ]$, $N_1=[i]$, $C_1=∅$, $O_2=[ʦ]$, $N_2=[ɒ]$, $C_2=∅$ is 4, because they differ in $O_1$, $N_1$, $O_2$ and $C_2$. Distances are then transformed into similarities using an exponential decay function (see the supplementary materials for more detail). The probability of a given outcome category ([m]/[b]) for an input form is proportional to its summed similarity to all existing forms with that outcome. If the input is highly similar to forms with a given outcome category and/or that outcome category is much more frequent than other categories, the predicted probability of the outcome category will be high.

The behavior of the GCM is controlled by a number of different parameters. These parameters determine the weight that individual features should be given in calculating similarities (e.g., $O_1$ could have a heavier weight than $N_1$); the prior probability of choosing a given outcome category (e.g., all things being equal, [b] could be assigned a higher probability as the outcome category based on the corpus data); and how featural mismatches should be combined into a distance measure, then converted to a similarity metric and used to calculate categorization probabilities. These parameters are described in detail in the supplementary materials.
An important aspect of the corpus data is the extreme skew in frequency. In total, 5% of all types cover 90% of the tokens. Although the GCM can take token frequency into account, we did not use it as part of our training method.

As explained above, we fit two versions of the GCM to our corpus data: a naive version that assumes that all features and outcome behaviors have the same weight and an informed version that estimates these weights directly from the corpus data using Bayesian parameter estimation (cf. Vanpaemel, 2009; Bartlema et al., 2014; more information about the parameter estimation is provided in the supplementary materials). The naive model relies purely on lexical information: it assumes that speakers bring no prior generalizations to the task of creating new echo-pairs, deriving such forms entirely through a holistic comparison with existing forms (see Dawdy-Hesterberg & Pierrehumbert, 2014, and Albright & Hayes, 2003 for similar implementations of the GCM that do not use feature weighting). Albright and Hayes (2003) and Albright (2009) refer to the type of similarity metric underlying such comparisons as “variegated” or “unstructured,” as it does not weight any specific aspects of similarity (e.g., shared endings vs. shared initial segments; shared consonants vs. shared vowels, etc.) more heavily than others. The informed GCM also relies heavily on lexical information, but it complements it with broader generalizations about what types of similarity actually matter for echo-pair formation. This is referred to as “structured” similarity by Albright and Hayes (2003) and Albright (2009). Reliance on structured similarity is also a key design feature of the MGL. In the case of the GCM, we would expect a structured similarity metric to weight $O_1$ and $O_2$ more heavily than other positions as a reflection of filled $O_1$ and voiceless $V_1C$ (though these generalizations themselves are not explicitly coded by the model). The actual feature weights in our model are as follows: $w(O_1) = 0.18$, $w(N_1) = 0.07$, $w(C_1) = 0.13$, $w(O_2) = 0.27$, $w(N_2) = 0.05$, $w(C_2) = 0.28$ (see the supplementary materials for more detail on how these weights were estimated from the data).

5.3 Modelling the corpus data

This section evaluates the performance of each of the four models based on the corpus data. To predict echo behavior probabilities in our corpus based on the GCM and the MGL, we used the leave-one-out method: the training corpus for each item is the entire corpus minus that specific item. For the simple rule-based model, the probabilities are taken directly from the logistic model in Section 2. The end result is, for each form, a probability corresponding to the likelihood of that form taking an echo with [b]. Using a simple majority decision rule (i.e., if the probability of [b] is higher than 0.5, choose [b]; otherwise, choose [m]), our models can also propose categories, that is, predict echo behavior for each form. These predictions can be compared with the actual echo behaviors for the forms in our corpus to calculate an accuracy score. The accuracy scores of the four models are all in the same narrow range: 71% for the simple rule-based model; 73.6% for the MGL; 71.5% for the naive GCM; and 77.7% for the informed GCM. The baseline accuracy, that is, the proportion of forms with the dominant behavior [b], is 57.5%. All of the models outperform this baseline by a wide margin. The MGL and the informed GCM produce the highest accuracy scores, which suggests echo-pair formation relies both on fine-grained lexical detail and broader generalizations.

Although the accuracy of these models is of interest in and of itself, it does not directly tell us about the kinds of patterns in the data captured by the MGL and the GCM. Therefore, we now look at the extent to which these models can emulate the simple rule-based account, and the extent to which they can go beyond it in terms of their coverage. To investigate these questions, we fit three expanded versions of the logistic regression model in Section 2 to the corpus data (one for the MGL, one for the naive GCM, and one for the informed GCM). These models all share the predictors filled $O_1$ and voiceless $V_1C$ (corresponding to the broad generalizations in the simple rule-based
account), and an additional term that differs across the three models. This term corresponds to the predictions from the MGL and the two versions of the GCM (naive and informed). The model predictions were logit transformed before being entered in the logistic regression model. We refer to these predictors as model weights. We perform model comparisons for each predictor in each model, which allow us to see (a) whether the MGL and the GCM have predictive value beyond the two main patterns and (b) whether filled O₁ and voiceless V₁C retain any predictive value once the MGL and GCM-based predictors are included in the models.

The comparisons are shown in Table 1. All of the model weights are significant even when included alongside the two broad phonological predictors, indicating they capture unique variance in the data beyond the reach of these predictors. The χ² values suggest that the two versions of the GCM outperform the MGL, with the informed GCM providing the strongest performance. In every case, the χ² values for filled O₁ and voiceless V₁C are substantially reduced compared to the model presented in section 2 (10.3 to 5.78/1.08/0.52 for filled O₁ and 26.97 to 4.51/8.59/6.81 for voiceless V₁C). This suggests that both the MGL and the GCM compete for some of the same variance as the simple-rule based model.

### Table 1. Model comparisons based on likelihood ratio tests for logistic regression models including MGL and GCM-based predictors fitted to the corpus data. The degrees of freedom for all comparisons are df = 1.

<table>
<thead>
<tr>
<th></th>
<th>MGL</th>
<th></th>
<th>GCM</th>
<th></th>
<th>GCM</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>χ²</td>
<td>p</td>
<td>χ²</td>
<td>p</td>
<td>χ²</td>
<td>p</td>
</tr>
<tr>
<td>filled O₁</td>
<td>5.78</td>
<td>0.016</td>
<td>1.08</td>
<td>0.300</td>
<td>0.52</td>
<td>0.469</td>
</tr>
<tr>
<td>voiceless V₁C</td>
<td>4.51</td>
<td>0.034</td>
<td>8.59</td>
<td>0.003</td>
<td>6.81</td>
<td>0.009</td>
</tr>
<tr>
<td>model weight</td>
<td>5.98</td>
<td>0.014</td>
<td>25.37</td>
<td>&lt; 0.001</td>
<td>34.60</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

GCM: Generalized Context Model; MGL: Minimal Generalization Learner.

5.4 Modelling the experimental data

We now turn to the experimental results. Given that the corpus data provide evidence for the involvement of lexically specific patterns in echo behavior, we expect that participants also rely on these in their judgments. Moreover, as before, it will be interesting to see the extent to which such lexically specific patterns overlap with filled O₁ and voiceless V₁C.

As in the previous section, we analyze the results in two separate ways. We first examine the proportion of variance in the experimental results accounted for by the four different models. This analysis is based on a dataset aggregated at the level of individual experimental stimuli, which contains raw echo-behavior proportions (i.e., the same information that is shown in Figure 3) and predictions from the four models. We then look at how the MGL and the GCM hold up in a model that also controls for broad phonological patterns. To do this, we revisit the regression model reported in Section 3. As above, we fit extended versions of this model that also include model weights based on the MGL and the GCM.

Figure 4 shows the fit between the model predictions (plotted along the horizontal axis) and the patterns of echo-behavior in the corpus (plotted along the vertical axis).

The R² values for each model are shown above the scatterplots. To aid interpretability and comparability across the models, all predictions and proportions are logit transformed, centered, and scaled (note that the R² values show exactly the same patterns without these transformations). In addition, the positions of the IPA labels in the plot are very slightly jittered to avoid overlap. This does not affect the overall visual patterns in the plots. All four models show a positive correlation
with participant behavior. However, both the scatterplots and the $R^2$ values clearly indicate that the GCM outperforms the two rule-based models. The informed GCM performs particularly well, accounting for nearly a quarter of all the variance in the experimental data. Somewhat surprisingly, the MGL ranks the lowest in terms of its prediction accuracy.

We now turn to the second part of the analysis. The model weights are created by using the corpus as the training set and predicting echo behavior probabilities for the experimental stimuli using the MGL and the GCM. As before, the predicted probabilities are logit transformed and the contributions of model weight, filled $O_1$, and voiceless $V_1 C$ are tested through model comparison.

The comparisons are shown in Table 2. The results are in many ways similar to those for the corpus data. In general, filled $O_1$ and voiceless $V_1 C$ make less of an improvement to model performance in the presence of model weights from the MGL and the GCM (with the exception of filled $O_1$ for the MGL): compared to the model in Section 3, $\chi^2$ values change from 6.14 to 6.24/1.56/1.58 for filled $O_1$ and from 4.64 to 0.22/3.9/0.24 for voiceless $V_1 C$. As before, the GCM outperforms the MGL, and the informed GCM appears to do especially well. After controlling for the broad phonological patterns, the model weights for the MGL and the naive GCM are no longer
significant. Although it is not possible to draw firm conclusions from non-significant results, this is a strong indication that the MGL and the naive GCM may only have limited predictive power beyond the two broad rule-based generalizations. Both filled $O_1$ and voiceless $V_1C$ lose most of their predictive power when included alongside the model weights from the informed GCM.

5.5 Summary of modelling results

In summary, and tying back to RQ2 and RQ3 (see Section 1.3), both our corpus and experimental results show fine-grained patterns that go beyond those captured by broad phonological rules (and, in fact, for some models these phonological rules become wholly redundant). The informed GCM performs particularly well for both datasets. This suggests that speakers are more likely to rely on the type of structured similarity captured by this model than variegated similarity (represented by the naive GCM) or the structured similarity metric that underlies the MGL. Although voiceless $V_1C$ does seem to make a significant independent contribution in most of the models that we fitted, the model weights generated by the MGL and the GCM act as strong suppressors for filled $O_1$ and voiceless $V_1C$. This suggests the MGL and the GCM capture some of the same variance in the data as filled $O_1$ and voiceless $V_1C$. Perhaps the most cautious interpretation of these findings is that the corpus results show evidence of both broad phonological generalizations and more fine-grained lexically specific patterns, whereas the experimental results are well captured by analogical mechanisms alone—albeit similarity in this case incorporates broader generalizations about what parts of a wordform are relevant to echo behavior.

6 Discussion

We now summarize the main results of the paper and link them to the three main research questions posed in Section 1.3.

1. To what extent do speakers conform to established patterns of phonological conditioning in echo-pair formation?

The corpus data analyzed in Section 2 provide strong support for filled $O_1$ and voiceless $V_1C$: there is a higher likelihood of [b] as the echo behavior in forms whose base starts with a vowel, and forms with a voiced consonant after the vowel in the first syllable. The same section also shows that voiceless $V_1C$ cannot be explained as the extension of a general phonotactic preference that holds over the entire Hungarian lexicon. The experimental data
in Section 3 demonstrate that speakers can extend filled $O_1$ and voiceless $V_1C$ to novel forms, although their usage shows considerably more variation than the corpus data.

2. Do words show variation in echo behavior above and beyond the two main phonological patterns?

Section 4 shows that the items in our nonce-word experiment vary substantially in echo behavior, and this variation is not simply due to phonological predictors such as filled $O_1$ and voiceless $V_1C$. The percentage of [b] can be as low as 23% for some words, and as high as 71% for others.

3. What type of model can best account for the variation in Hungarian echo-pair formation?

In Section 5, we demonstrate that the MGL and the GCM—which have both previously been used to account for variation in plain morphological phenomena—can explain a substantial amount of variance in the data. However, their performance differs markedly. The GCM (and especially the informed GCM) is a significant and strong predictor of echo behavior in both the corpus and the experimental data. When included alongside the simple rule-based predictions, these become largely redundant, whereas the GCM predictions remain robust. This indicates the GCM can capture these rule-based generalizations as well as some further unique variance in the data. The GCM performs best when it can also access broader generalizations about which parts of the base are relevant to echo-pair formation.

The MGL is weaker in terms of its predictive power. Although it accounts for some unique variance in the corpus data, it cannot fully replace the simple rule-based predictors. Moreover, it loses most of its predictive power when included alongside the simple rule-based predictors in a model of the experimental data.

In what follows, we tie these results back to the broader questions put forward in Section 1. We first compare different models of morphology in terms of their ability to capture the above observations. We then interpret our results in the broader context of the extra-grammatical versus plain morphology distinction.

Let us first discuss fully rule-based accounts, which do not rely on lexical information at all. As noted above, both the corpus and the experimental data provide support for two phonological patterns, filled $O_1$ and voiceless $V_1C$. Both of these patterns can be formulated as rules that link the shape of the echo to the phonological make-up of the base. When these rules are used to make probabilistic predictions about echo behavior, they yield relatively accurate predictions about the corpus data and account for a substantial amount of variance in the experimental data.

However, the analyses presented above do not support an account based purely on broad phonological generalizations implemented by rules. First, the results in Section 4 show that our experimental participants have strong intuitions about specific stimuli that go far beyond what can be captured using filled $O_1$ and voiceless $V_1C$. Although some of these intuitions could be explained by positing further rules (e.g., one that takes the manner of articulation of $O_2$ into account), the success of the GCM at accounting for this additional variance indicates its lexical basis may be deeper than would be expected under a simple rule-based approach. Some additional mechanism is needed to account for these fine-grained patterns of variation, which argues for a dual-route access model (where rules are complemented by lexically based processes) or a model where both general and lexically specific patterns are captured by the same processes, such as the
MGL or the GCM. The success of the informed GCM at dealing with both types of patterns suggests that rules based on broad generalizations may be entirely unnecessary for explaining echo-pair formation in Hungarian.

There is another—less forceful, but nonetheless important—argument against simple rule-based approaches. Both of the phonological patterns investigated in this paper are highly unusual from a phonological point of view. This is unexpected from the perspective of approaches that rely on rules require them to be “natural” or “unmarked” in some sense (e.g., Donegan & Stampe, 1979; this is also true of approaches that capture rule-like patterns using constraints such as Prince & Smolensky, 2008). Looking first at filled $O_1$, there is absolutely no reason why one might expect a missing onset in the base to attract an initial [b] in the echo. There is no phonetic or phonological motivation for this pattern, nor any similar corresponding phenomena in other languages. As for voiceless $V_1C$, this pattern could be argued to be a case of long-distance voicing harmony, where the first postvocalic consonant of the echo extends its voicing specification to the initial consonant. Such processes are cross-linguistically rare, but they do occur (Hansson, 2004). However, there is only a vague sense in which such an interpretation fits the current pattern. This is because patterns of voicing harmony typically only apply to obstruents and involve the spreading of active voicing. Voiceless $V_1C$ goes against this generalization. If it is formalized as a rule that applies in the context $\_V [+\text{voice}]$ to attract a $[+\text{voice}] [b]$ behavior, the triggering context must involve both passive and active voiced consonants (i.e., sonorants and obstruents). If, in contrast, it is formalized as a rule that applies in the context of $\_V[−\text{voice}]$ to attract an $[m]$ behavior, it must create a passively voiced $[m]$ through the spreading of voicelessness. Both of these analyses are problematic from a phonological point of view.

We now move on to the MGL, an example of a model that relies on the same formal tools as deterministic rule-based and constraint-based approaches, but can capture fine-grained and even lexically specific patterns through a fundamentally stochastic architecture. Although the MGL is designed to discover generalizations at all levels of specificity, it performed rather poorly on our data. Our results suggest that it cannot fully capture the predictions of a simple rule-based account, nor can it go far beyond it by capturing additional variation. There are two possible reasons for these shortcomings of the MGL. First, the MGL sets itself an ambitious task by attempting to discover both morphological operations (e.g., $C \rightarrow b$) and their phonological contexts (e.g., $# \_V_0[+\text{voi}]$) simultaneously in a bottom-up fashion. In other models, such as the GCM or the maxent-based model of Gouskova et al. (2015), the morphological operations are typically treated as given, and the focus is on learning their phonological conditioning. Although the MGL does well at performing both tasks at the same time, it does so at a cost: it can only generalize across classes of sounds (by relying on phonological features) when those sounds are in the local context of the structural change of a posited rule. We have partly circumvented this issue by reordering the sounds in the input to the MGL, so that both $O_1$ and $O_2$ are available for feature-based generalizations, but it appears that this does not fully resolve the issue. Gouskova et al. (2015) point out similar issues with the MGL in their paper on Russian diminutive formation, a phenomenon that closely resembles Hungarian echo-pairs in terms of its (often non-local) phonological conditioning. Therefore, the failure of this approach may stem from limitations that are specific to the MGL, and it is possible that a different model (e.g., based on a maxent algorithm) may perform better.

However, it is also possible that the poor performance of the MGL stems directly from one of its key design features, namely its tendency to weight general rules more heavily than specific ones. Although we have found robust evidence for some broader phonological trends in echo-pair formation, these patterns are somewhat unreliable as shown by the immense variability in the data. An approach that models echo-behavior directly on the basis of whole-word similarity may be inherently better at capturing it. We leave it to future research to determine whether other models
that prioritize broader generalizations (such as Gouskova et al.’s 2015 maxent-based approach) can improve on the MGL’s performance or if a more lexically driven model such as the GCM is needed to account for the data. Our suspicion is that constraint-based models may be harder to distinguish from the informed version of the GCM in terms of their performance, and that this may be a direct result of the fact that both types of models can integrate lexical information and broader generalizations in an effective manner. This is in line with Albright and Hayes (2003), who note that the empirical coverage of MGL-style models and informed analogy may, in fact, be very similar.

This takes us to the GCM. The GCM in its “informed” incarnation can account for a large amount of the variance in both the corpus and the experimental data without the need for additional explicit generalizations in the form of rules (although voiceless V₁C still provides a modest improvement to the corpus-based model even in the presence of predictions from the informed GCM; cf. Table 1). The naive model provides a worse fit for both datasets and has no significant predictive power beyond the two main phonological patterns in the case of the experimental data. In other words, although an analogical approach based on whole-word similarity is clearly superior to the other models that we evaluate in this paper, it must rely on a structured similarity metric that incorporates information about the relative importance of different dimensions of similarity. Variegated similarity on its own is not sufficient, and speakers must bring at least some broader generalizations to the task of generating novel echo-pairs.

An important aspect of the corpus data is the extreme skew in frequency. Of all types (nine forms), 5% cover 90% of the tokens in our corpus. While we can only speculate on this point, it is likely the behavior of some target words in the experiment was largely driven by one or two frequent real words. Although some versions of the GCM can capture such effects by weighting frequent forms in the training set more heavily, our implementation did not include weighting by token frequency (as frequency-weighted estimates led to a worse overall model fit). To give an example where the GCM may have failed to capture the effects of a single analogical model, the forms [ëtsɛ] and [ëki] both show a strong preference for /b/ in the participants’ data, contrary to the predictions of the informed GCM. This could be driven by the frequent, extant word édi-bédi [ëdi-be-di] “endearing dim.” For the MGL, the situation might be worse in that similarity to frequent real words could run counter to the structural descriptions of the rules determining the behavior of a target form.

Let us now turn to the distinction between plain and extra-grammatical morphology. As noted in Section 1.1, echo-pair formation in Hungarian shows many but not all of the traits associated with extra-grammatical processes. In this sense, it can be viewed as an intermediate type of process: not quite extra-grammatical but also not plain. The above conclusions about the mechanisms underlying echo-pair formation certainly do not contradict this view. In discussing the closely related distinction between regular versus irregular morphology, Albright and Hayes (2003) argue that the debate about underlying mechanics is orthogonal to the question of regularity. They suggest that both regular and irregular processes can be captured by models relying on structured similarity. The same argument can be extended to the case of echo-pair formation in Hungarian. Although it is, in some ways, an unusual process, the patterns of phonological variation in this phenomenon are amenable to an analysis using tools associated with plain morphological phenomena. In a broader sense, this suggests that the difference between extra-grammatical and plain morphological phenomena may simply be one of degree, not of kind.

Our account above emphasizes the shared grammatical origins of plain and extra-grammatical processes—but do formal models such as the MGL and the GCM really perform comparably on plain and extra-grammatical phenomena? To explore this question, we focus on Albright and Hayes’ (2003) nonce-form study of past tense formation in English (an example of a plain morphological phenomenon). Similar to our study, Albright and Hayes (2003) also compare predictions
from the MGL and the GCM to participants’ intuitions. The bulk of their paper focuses on goodness ratings, which we did not include in our paper. Based on these data, they conclude that the MGL provides a better account of the data than the GCM, with $R^2$ figures (presented simply as the raw correlation coefficient $R$ in their paper) ranging from 0.23–0.56 for the MGL and from 0.10–0.30 for the GCM (cf. Albright & Hayes, 2003, Tables 21, 22, and 23). However, they also present “production probabilities” for a subset of their data in Table 21 (p. 142), which are directly comparable to the echo-behavior proportions that we analyzed. The MGL is still in the lead when modeling the production probabilities of regular -ed responses, with $R^2 = 0.46$ versus $R^2 = 0.20$ for the GCM. However, for irregular responses this relationship turns around: the GCM provides a better fit ($R^2 = 0.27$) than the MGL ($R^2 = 0.11$). These numbers are very close to those in Figure 4 of our paper ($R^2 = 0.17 \sim 0.24$ for the two different versions of the GCM and $R^2 = 0.11$ for the MGL).

This—admittedly limited—comparison suggests that formally implemented models such as echo-pair formation and (b) plain but irregular morphological processes. To fully explore this observation, it would be instructive to also consider goodness rating data for extra-grammatical phenomena and further examples of both plain and extra-grammatical processes. We leave these tasks for future research, but speculate that work along these lines may reveal further parallels between extra-grammatical and plain but irregular morphology.

If extra-grammatical processes are implemented by the same mechanisms as plain ones, what explains the clear differences between echo-pair formation in Hungarian and plain morphological phenomena? Specifically, why is it stylistically marked, why is it so promiscuous in terms of its input (nouns, verbs, adjectives as well as inflected forms), why is it typologically distinct from plain morphological processes in Hungarian, and why does it show so much variation within and across speakers? In our view, the most likely source of these characteristics lies in the usage of echo-pair formation. Compared to plain morphological markers such as plural or past-tense affixes, echo-pairs occur in a very limited set of contexts: in playful and/or intimate interactions, and often specifically in child-directed speech. Their stylistic markedness follows straightforwardly from the function they play in these contexts, which is simply to reinforce their intimacy and playfulness. Their typological weirdness in the context of the host language (in the current case, reduplication in an agglutinating language) and their syntactic peculiarities are likely a consequence of the speakers’ intention to make such forms distinct and salient, thereby increasing their expressive power. Finally, the degree of variation seen in echo-pair formation likely follows from two facts. First, the playful function of echo-pairs encourages creative language use, which results in a steady influx of forms that deviate from “standard” patterns of echo-pair formation. Second, the fact that echo-pairs are only used in a limited set of contexts means that their token frequency is also rather low. Forms with low frequency are often argued to have weak lexical representations (see e.g., Bybee, 2001; Pierrehumbert, 2001), which means they retain a considerable amount of flexibility instead of becoming crystallized in a specific lexicalized form. This allows variation to survive in such forms. Note that this account of echo-pair formation is very similar to Vihman's (2019) discussion of phenomena that fall under the rubric of “language at play.” She argues that playfulness and creativity are crucial in shaping such extra-grammatical word formation processes.

There is one final interesting aspect of the data that we have not yet addressed: the main phonological patterns are substantially stronger in the corpus results than in the experimental data. If speakers explicitly represent such patterns in the form of rules, constraints or feature weights, why do they apply these much less consistently with novel forms? There are likely two main factors underlying this observation. First, as we noted above, a nonce-word experiment is a somewhat artificial context for language use, and it is clear that our participants found the task quite challenging. It is not surprising, then, that they were less able to apply their knowledge of phonological patterns in this task. Second, the corpus results at least partly represent lexicalized forms that are
shared by most speakers of Hungarian, and show little variation. Such lexicalized forms likely attained their current shape through a large number of interactions among speakers. It is possible that patterns such as filled O₁ and voiceless V₁C are only present as weak biases in individual speakers. However, these biases have a chance to apply every time a given echo-pair is used, and, over many interactions, they may turn into the kind of robust patterns that we now see in the corpus results. This type of explanation is frequently evoked in diachronic accounts of linguistic patterns, and is often referred to as the production-perception feedback loop (Pierrehumbert, 2001; Wedel, 2007; Sóskuthy and Hay, 2017).

7 Conclusion

In this paper, we presented an analysis of echo-pair formation in Hungarian based on corpus data and a large-scale online nonce-word experiment. Consistent with previous results in Sóskuthy (2012), we demonstrated that echo-pairs are primarily formed with a labial-initial echo, and found evidence for two broad phonological patterns: a preference for [b] as the initial consonant of the echo when (a) the base has no initial onset and (b) when the first postvocalic consonant of the base is voiced. The corpus results follow these two patterns fairly consistently, whereas the experimental results show weaker (although still significant) trends. Our analysis revealed a considerable degree of variation across words and experimental stimuli beyond the two main phonological patterns. We used this across-word variation to evaluate two models that differ in terms of the degree to which they rely on lexical information: the MGL and the GCM. The GCM performed particularly well at capturing patterns in the data, though only when allowed to rely on a structured similarity metric that draws on broader generalizations about the relevance of different parts of the base. On the basis of these results, we argued that echo-pair formation is implemented by mechanisms that are also involved in plain morphological processes, making the difference between extra-grammatical and plain morphological phenomena one of degree, not one of kind. We suggested that most of the peculiarities of phenomena such as echo-pair formation follow from the specific contexts and functions in which they are used, and do not require an explanation in terms of morphological models that are distinct from those associated with plain morphology.

Acknowledgements

We would like to thank two anonymous reviewers and the editor for their valuable input, as well as Dániel Szeredi and Bálint Feyér for their help in collecting the data which our corpus analysis is based on.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Márton Sóskuthy https://orcid.org/0000-0002-5074-4767

Supplemental material

Supplemental material for this article is available online.

Notes

1. For convenience, all examples are from English. However, processes similar to the ones listed here are found in a vast array of unrelated languages, and some of them may well be universal.
2. This consonant is usually the onset of the second syllable, but sometimes the coda of the first syllable.
3. Figure 1 does not show these proportions directly. It presents separate bars for each combination of initial onset and postvocalic consonant. In other words, it allows us to look not only at the overall effects of these factors, but also their interaction. However, the same trends can be inferred from this graph.

4. We test these predictors using model comparisons based on likelihood-ratio tests. Following Esarey and Sumner (2017), the comparisons for testing the overall (or marginal) effects of filled O₁ and voiceless V₁C are between a full model (including the interaction term) and nested models that exclude both the interaction and one of the main terms. This is necessary because the interpretation of main terms changes in the presence of interactions: for instance, simply excluding filled O₁ without also excluding the interaction term would only tell us whether filled O₁ has a significant influence when O₂ is not voiceless. Excluding both the main term and the interaction tells us about the overall influence of the predictor. Because the interaction term was found not to be significant, we can be reasonably confident that this overall influence does not vary substantially across contexts.

5. The statistics for this last comparison ought to be taken with caution as the model did not include random slopes over left [b].

6. We have also tested another version of this model that estimates distances based on phonological features instead of segment identity. However, this model was consistently outperformed by the segment-based model, and is therefore not discussed here. The analyses in the accompanying GitHub repository also include this feature-based model.

References


Appendix: Experimental stimuli

Table 3. List of experimental stimuli organized by filled $O_1$ and voiceless $V_1C$.

<table>
<thead>
<tr>
<th>SPELLING</th>
<th>IPA</th>
<th>FILLED $O_1$</th>
<th>VOICELESS $V_1C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>esző</td>
<td>ɛsøː</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>óti</td>
<td>ɔtti</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>ácő</td>
<td>aːtsøː</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>éće</td>
<td>eːtsɛ</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>ükér</td>
<td>yker</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>etyő</td>
<td>eːcoː</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>úszog</td>
<td>uːsoɡ</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>ecsa</td>
<td>eːʃɔː</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>ityő</td>
<td>ici</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>éki</td>
<td>eːki</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>ucser</td>
<td>ucʃɛr</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>ócő</td>
<td>otsøː</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>ége</td>
<td>eːɡɛ</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>ilász</td>
<td>ilas</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>azsu</td>
<td>ɲsu</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>inó</td>
<td>inoː</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>édeg</td>
<td>edeɡ</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>uzsó</td>
<td>uːsoː</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>udok</td>
<td>udok</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>agyőr</td>
<td>ɲɾɔr</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>óli</td>
<td>oːli</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>odó</td>
<td>oːdoː</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>unyog</td>
<td>ɲʊɡə</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>ugó</td>
<td>ʊɡɔː</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>főke</td>
<td>fəke</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>csükő</td>
<td>tʃylkə</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>gótyi</td>
<td>ɡoːci</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>szece</td>
<td>sæʃɛ</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>lászi</td>
<td>lɑʃi</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>vikér</td>
<td>vikɛr</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>hácog</td>
<td>хаːtsɔɡ</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

(Continued)
### Table 3. (Continued)

<table>
<thead>
<tr>
<th>SPELLING</th>
<th>IPA</th>
<th>FILLED O₁</th>
<th>VOICELESS V₁C</th>
</tr>
</thead>
<tbody>
<tr>
<td>kücsi</td>
<td>kytʃi</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>kasó</td>
<td>koʃo:</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>getye</td>
<td>ɡɛɛɛ</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>szacsog</td>
<td>sɒtsʃog</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>láki</td>
<td>ɬaki</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>kunor</td>
<td>kʊnor</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>fázsi</td>
<td>faʃi</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>zinya</td>
<td>zɪnɪ</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>csanyi</td>
<td>tʃɪɲi</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>rudog</td>
<td>rʊdʊɡ</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>gyudó</td>
<td>juːdo:</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>zédál</td>
<td>zɛːdɔl</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>runya</td>
<td>rʊɲa</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>cségi</td>
<td>tʃɛɟi</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>kegya</td>
<td>keɲa</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>cege</td>
<td>ʈʃɛɡə</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>rüli</td>
<td>rʊli</td>
<td>+</td>
<td>–</td>
</tr>
</tbody>
</table>