BRILL

LANGUAGE DYNAMICS AND CHANGE 15 (2025) 1–39

brill.com/ldc

Phonotactic mechanisms behind contact-induced morphological simplification in Alorese

An agent-based model of intergenerational transmission

Peter Dekker | ORCID: 0000-0003-4734-2668
AI Lab, Vrije Universiteit Brussel, Brussels, Belgium
research@peterdekker.eu

Marian Klamer | ORCID: 0000-0002-2905-7161
Leiden University Centre for Linguistics, Leiden, The Netherlands
m.a.f.klamer@hum.leidenuniv.nl

Bart de Boer | ORCID: 0000-0001-9429-6129
AI Lab, Vrije Universiteit Brussel, Brussels, Belgium
bart@ai.vub.ac.be

Received 26 March 2024 | Accepted 26 February 2025 |
Published online [UNKNOWN] 2025 ■ to be inserted by typesetter

Keywords

agent-based modeling – language contact – morphology – phonotactics – consonant sequences – phonology-morphology interface

Abstract

In this paper, we study how phonotactic constraints can play a role in contact-induced morphological simplification. We hypothesize that, in languages with constraints on certain consonant sequences, L2 speakers acquiring the language overgeneralize these constraints and avoid consonant sequences formed by morphological affixing. The resulting process of reduction of such sequences could then lead to loss of morphological affixes. We evaluate this hypothesis using data from Alorese, an Austronesian language spoken in eastern Indonesia, which underwent morphological simplification and has a large proportion of L2 speakers. Results from an agent-based model, which

Published with license by Koninklijke Brill BV | DOI:10.1163/22105832-bja10036

allows for mechanisms to be tested in isolation, indicate that phonotactic mechanisms indeed play a role in contact-induced morphological simplification. Phonotactic mechanisms lead to the strongest simplification when combined with a generalization mechanism, which spreads the simplified forms across different verbs. The model results underline the role of multiple interacting mechanisms in language change.

1 Introduction

It is generally known that phonology and morphology interact. On the one hand, phonology can influence morphology, such as when stress determines the choice or ordering of certain morphemes (McCarthy and Prince, 1995; Inkelas, 2011). On the other hand, morphology can influence phonology, such as when certain morphological environments cause phonemes to be realized differently (Inkelas, 2014: 242–280) or when the attachment of certain morphemes causes stress shift. When languages are in contact, phonology and morphology also interact, for example when L1 phonology plays a role in the acquisition of L2 morphology (Bayley, 1996; Bliss, 2006; Hu, 2010; Cabrelli Amaro et al., 2018).

Statistical studies over large sets of languages find a correlation between morphological simplification on the one hand, and large societies with many speakers acquiring the language as adults on the other (Lupyan and Dale, 2010; Bentz and Winter, 2013; Sinnemäki and Di Garbo, 2018). This correlation has been explained by the greater difficulty that adults have in acquiring morphology compared to children (Wray and Grace, 2007; Trudgill, 2011; Kusters, 2003). However, morphological simplification takes place in some contact situations, while in other situations which involve adult language learners it does not occur, or occurs less profoundly. So the loss of morphology cannot be explained by imperfect adult language acquisition alone. Given that phonology and morphology are known to interact, here we explore the role of phonology in contact-induced morphological simplification.

In particular, we try to identify the role of phonotactic factors (relating to the co-occurrence of sounds within words and syllables) in contact-induced morphological simplification. We focus on contact-induced morphological simplification in Alorese, an Austronesian language in eastern Indonesia. Specifically, we propose that L2 speakers of Alorese avoid consonant sequences, due to overgeneralization of the phonotactic constraints in the language they are acquiring. The consonant sequences are created by morphological affixing (Dressler et al., 2019: 93–97) and their reduction eventually leads to the loss of affixes.

We try to identify the role of phonotactic factors using an agent-based model (Smith, 2014), which allows specific mechanisms to be tested in isolation. We study how differences in phonotactic processing in L1 and L2 speakers, through interaction over generations, eventually leads to language change on historical timescales (cf. Enfield, 2014; Brochhagen et al., 2023; Greenhill, 2023).

The paper is structured as follows. In Section 2, we first introduce the case study of Alorese, a language that lost almost all of its morphology through contact. We argue that there are certain phonotactic conditions that apply in processes of contact-induced simplification as part of imperfect L2 acquisition which could have played a role in the observed patterns. In Section 3, we test this hypothesis using an agent-based model, in which we simulate interaction between speakers. The results of this model are presented in Section 4. In Section 5, we discuss the implications and limitations of this study, and the role phonotactic factors may play in other language contact situations.

2 Case study: Alorese (Austronesian)

2.1 *Alorese*

Alorese is an Austronesian (Flores-Lembata; Fricke, 2019; Sulistyono, 2022) language spoken on the coasts of the Alor and Pantar islands in eastern Indonesia. In the interior of these islands, Papuan languages belonging to the Alor-Pantar group are spoken (see Fig. 1). Over the last centuries, Alorese has been used as the language of trade and interethnic communication by Papuan speakers in the region, as well as in marriages of Papuan speakers with speakers of Alorese (Klamer, 2011, 2012, 2020). As a lingua franca in the region, Alorese has enjoyed a high prestige, which has led to asymmetric bilingualism: many L1 speakers of Alor-Pantar languages learn Alorese, but fewer L1 Alorese speakers learn Alor-Pantar languages (Moro, 2021: 15). It is estimated that 50 percent of the speakers of Alorese today are L2 speakers (Moro, 2019; see Section 3.7).

Alorese developed from Proto-Western Lamaholot (Fricke, 2019; Sulistyono, 2022) and it is suggested that the precursors of today's Alorese speakers were originally Western Lamaholot speakers who moved to Pantar in the fourteenth century, and to Alor in the sixteenth century (Klamer, 2012, 2020).

One of the closest relatives of Alorese today is Western Lamaholot (the variety described in Nishiyama and Kelen, 2007). Western Lamaholot is spoken on the islands located to the west of Pantar and Alor (Fricke, 2019: 156–159); see Fig. 2. Western Lamaholot has a significant amount of inflectional and derivational morphology which it inherited from Proto-Malayo Polynesian. Strikingly, in Alorese, almost all of this morphology has been lost. This suggests that

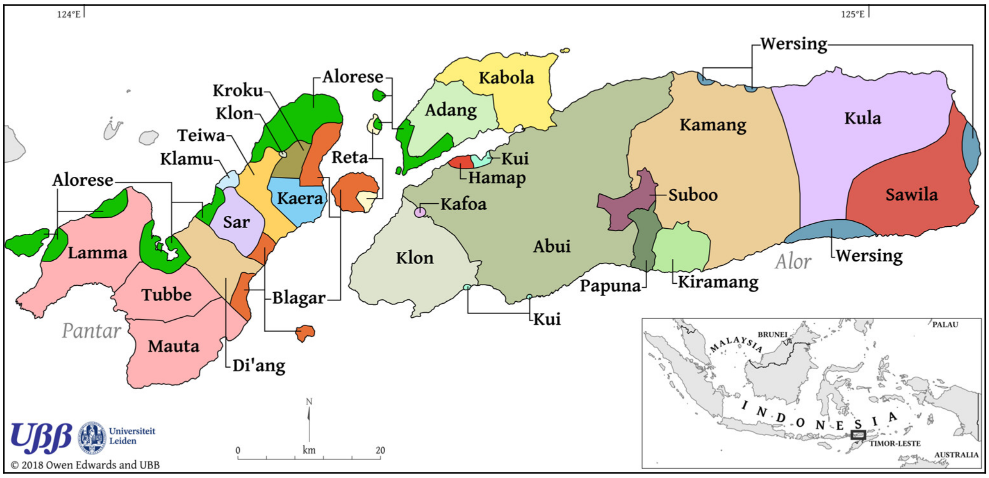


FIGURE 1 The Alor and Pantar islands (main map) and their location within Indonesia (inset). Alorese (Austronesian) is spoken in the bright green areas on Alor and Pantar; in the other areas, Papuan Alor-Pantar languages are spoken
MAP BY OWEN EDWARDS AND UBB, USED WITH PERMISSION

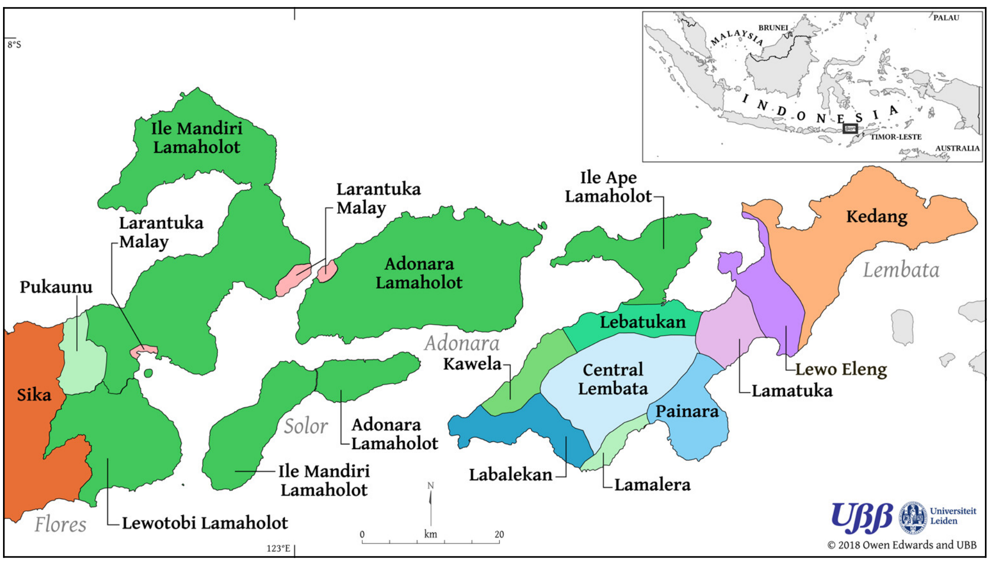


FIGURE 2 Western Lamaholot (in green), sister language of Alorese, is spoken on islands to the west of Pantar (edges in gray on the right)
MAP BY OWEN EDWARDS AND UBB, USED WITH PERMISSION

the loss must have occurred after the speaker group split off from Western Lamaholot and settled on Pantar and Alor, where they came into contact with Papuan-speaking groups.

Western Lamaholot as it is spoken today does not show any evidence of contact with the Papuan languages of Alor and Pantar. Neither is there evidence of other cultural connections between speakers of the Alor-Pantar languages and Western Lamaholot. Due to the strong sea currents, the strait separating Pantar from the islands to the west is hard to cross, which has likely prohibited contacts (Sulistyono, 2021: 342). We thus assume that the history of Alorese involves contact with Alor-Pantar languages, while its close relative Western Lamaholot did not have such contacts.

Much of the large-scale loss of morphology in Alorese is likely caused by imperfect L2 acquisition by adult speakers of a Papuan language, as earlier studies have suggested (Klamer, 2012, 2020). However, in this paper we hypothesize that besides this factor, the phonology of Alorese also played a role in the erosion of certain inflectional morphemes; see Section 2.4.

Alorese is an interesting case of contact-induced morphological simplification to study because there are clear focus, neighbor, and benchmark languages available (see Di Garbo and Napoleão de Souza, 2023; Sinnemäki et al., 2024 for the framework introducing these terms), with limited contact with other languages due to the island geography. Alorese is the *focus* language, which underwent morphological simplification. Alor-Pantar languages are the *neighbor* languages, with which Alorese is in contact. There is a *benchmark* language available, a current language we can view to represent a prior state of Alorese, and with which we can contrast Alorese: Western Lamaholot.

2.2 *Verbal subject marking in Alorese*

This paper focuses on the simplification of verbal inflectional morphology. Inflectional morphology is known to be a vulnerable domain in language contact, and a bottleneck in adult language acquisition (Lardiere, 1998; Prévost and White, 2000; Prévost, 2003; Blom et al., 2007: 315–317). We focus in particular on subject inflection. Subject inflection on verbs relies on syntactic cues: the syntactic relation between the verb and the subject which it encodes. Adult L2 learners have a decreased capacity to acquire this system compared to L1-learning children (Prévost and White, 2000; Prévost, 2003). Encoding syntactic relations in the clause, subject inflection is a case of "contextual" inflection, and differs from "inherent" inflection (like a nominal plural affix), which occurs independently of the syntactic context of a word. Contextual inflection is acquired later by children than inherent inflection (Booij, 1996). Its interpretation requires more knowledge and processing effort than the interpretation of inherent word inflections, and poses problems for adult learners. As a result, it is transferred less easily during contact (Meakins, 2011: 64–65; Gardani, 2008), in particular when adult L2 speakers are involved.

The subject inflection in Western Lamaholot (Lewoingu variety; Nishiyama and Kelen, 2007) comprises two sets of affixes—a set of prefixes marking transitive subjects, and a set of suffixes marking intransitive subjects—as well as agreement on nonverbal predicates (Klamer, 2012, 2020).[1]

In Western Lamaholot, the subject prefixes are obligatory. The condition for the use of the prefixes is phonological: only vowel-initial verb stems can take these prefixes, and 20 such verbs are mentioned (Nishiyama and Kelen, 2007: 32). As the prefixes consist of a single consonant (*k*- '1SG,' *m*- '2SG,' *n*- '3SG,' and so on; see Table 1 below), they form the onset of the initial syllable of the inflected verb, as in example (1).[2]

(1) *go* *k-oiro* Western Lamaholot
1SG 1SG-know
'I know'

In contrast to the mandatory subject prefixes, Western Lamaholot subject suffixes are optional, but occur on a wider range of verbs. Their use is not conditioned by the shape of the (final) syllable of the verb. The suffixes are all consonant-initial syllables (*-kən, -ko, -na, -te, -ke, -ka*), and may attach to verbs ending in a vowel or a consonant, as in (2) and (3). Suffixed to the verbal base, they remain a separate CV(C) syllable.

(2) *go* *simo-kən* Western Lamaholot
1SG smile-1SG
'I smiled'

(3) *go* *gəkər-kən*
1SG be.surprised-1SG
'I'm surprised'

In Alorese, most of the morphology that is still attested in Western Lamaholot has been lost, including all the subject suffixes. For example, the third person

1 Western Lamaholot also has possessor suffixes on nouns and seven derivational affixes; these morphemes are not of concern here.

2 The following abbreviations are used in the glosses in this paper: 1/2/3 first/second/third person; EXCL exclusive; INCL inclusive; PL plural; SG singular.

singular of the suffixing verb *lodo* 'to go down' has the 3SG marker *-na* in Western Lamaholot (4) (cf. Nishiyama and Kelen, 2007: 32), while it lost the person marker in Alorese (5) (cf. Klamer, 2011: 65):

(4) *lodo-na* Western Lamaholot
go.down-3SG
'he goes down'

(5) *lodo-∅* Alorese
go.down-3SG
'he goes down'

What has been retained in Alorese, however, are the subject prefixes, albeit on just a small set of eight high-frequency, vowel-initial verbs, including *-ala* 'to pass,' *-ang* 'to eat,' *-ang* 'to use, to make,' *-ate* 'to carry,' *-ei* 'to go,' *-enung* 'to drink,' *-oing* 'to know,' *-ong* 'to be with.' The first person plural inclusive form for *enung* 'to drink' is the same in Alorese (Klamer, 2011: 61, 128) as in Lamaholot (Nishiyama and Kelen, 2007: 31):

(6) *t-enung* Western Lamaholot
1PL.INCL-drink
'we drink'

(7) *t-enung* Alorese
1PL.INCL-drink
'we drink'

Some of the verbs that are prefixing in Western Lamaholot, have been replaced in Alorese by different consonant-initial verb forms. For instance, the verb 'to hunt' is *-iu* (inflected) in Lamaholot, but *kori* (not inflected) in Alorese, the verb 'to wait' is *-awan* (inflected) in Lamaholot, but *baing* (not inflected) in Alorese, and the verb 'to stay' is *-awa* (inflected) in Lamaholot, but *tobo* (not inflected) in Alorese. For those Alorese verbs we cannot attest either omission or fossilization of the subject prefix (Moro, 2019).

2.3 *Phonological constraints on simplification*

While it is clear that L2 acquisition during contact is an important factor mediating the morphological simplification in Alorese, we explore the possibility that the phonology of Alorese also played a role. As mentioned above, all the Alorese inflectional *suffixes* were lost, while the inflectional *prefixes* were still retained in a number of verbs. That is, of the two inflection types, the verbal suffixes were simplified more, and/or faster than the verbal prefixes. We suggest that the explanation for this asymmetry may be found in the phonotactics of the inflected verbs.

It is generally known that phonology may play a role in morphological simplification; both independent of language contact, as well as in situations where language contact has occurred. For example, Sims-Williams and Baerman (2021: 24) show how internal sound change leads to the loss of an inflectional marker in Geʾez, and the merger of markers in Classical Greek. Kusters (2003) shows that Quechua underwent less profound morphological change than Scandinavian, even though both were societies with many L2 speakers. He explains this difference by the phonological changes that Scandinavian had already undergone: because of the stress shifting to the beginning of the word in Germanic languages, the unstressed end of the word underwent phonological processes in Scandinavian that made it less morphologically transparent, and more prone to morphological simplification, compared to Quechua. Kusters also describes contact-induced morphological simplification in Arabic, which was phonologically conditioned. And in Molineaux (2012), it is argued that the different stress patterns of Romance borrowings were a factor in the loss of certain affixes in early English.

Phonotactic constraints may be an important factor in contact-induced morphological simplification for at least two reasons. Firstly, because the conditions on the co-occurrence of sounds within words and syllables are sensitive to contact. For example, simplification of consonant clusters in creole languages is well attested (Blevins, 1995; Wurzel, 2001: 379).[3] And crosslinguistically, the complexity of the syllable structure can increase through contact with a donor language that has a more complex syllable structure than the target language (Napoleão de Souza and Sinnemäki, 2023). Because phonotactic structure is sensitive to contact in general, it is plausible that it also plays a role in those contact processes that simplify morphology.

Secondly, crosslinguistically, there seems to be a relationship between phonotactic and morphological complexity. The maximal phonotactic complexity of syllables in a language and the degree of morphological synthesis (number of morphemes per word) have been shown to correlate (Easterday et al., 2021).

3 This is not a universal process however. Schramm (2014) stresses there is large variation in creole syllable complexity and there are many instances in which consonant clusters are not simplified.

The phonotactic phenomenon we focus on here is that of consonant sequences. In the remainder of this article, we will use the term *consonant sequences* to describe any sequence of consonants, including sequences which cross syllable boundaries, which is the object of study of this paper. In much of the literature, the term *consonant clusters* is used for sequences of consonants within a syllable, and we will use this term when referring specifically to these sequences (which are not the main focus of this work).

In both Western Lamaholot and Alorese, words generally consist of two syllables, with CV and CVC as its main types of syllable structures (Klamer, 2011: 32–35; Nagaya, 2011: 71–79; Nishiyama and Kelen, 2007: 10). Verb stems are mostly disyllabic forms, usually taking the shape (C)V.(C)V or (C)V.CVC.[4] Consonant sequences within a morpheme only occur in the combination [nasal.voiced obstruent], where the nasal is the coda of the first syllable and the stop is the onset of the second syllable (with the full stop indicating the syllable boundary). Other consonant sequences are not allowed in Alorese. Morpheme-internal consonants share their place features, so that a sequence like [m.b] is attested but one like *[m.d] is not.

Coda consonants occurring at the end of verb stems in Alorese and Western Lamaholot are *n*, *k* or *ʔ*, and *t*. All the subject inflection suffixes start with *n*, *k* and *t*; see Table 1. So, when subject suffixes attach to verb stems with a coda consonant, this would create consonant sequences over the morpheme boundaries, like *n.k, n.t, k.n, k.t, t.n, t.k*.[5] Such sequences of non-homorganic stop-nasal or nasal-stop sequences do not occur word-internally, and they are highly marked across morpheme boundaries. These consonant sequences over morpheme boundaries are possible in Western Lamaholot (which can be regarded as a hypothetical earlier state of Alorese), but not in Alorese.

We suggest that in Alorese, the asymmetry in the subject inflections shown in the partial retention of inflectional *prefixes* versus the complete loss of inflectional *suffixes* may be explained by a constraint on consonant sequences, which

4 About 66 percent of the 522 Alorese words in Klamer's (2011) limited corpus end in an open syllable, while 34 percent end in a syllable with a final consonant.

5 When the final consonant of the verb stem and the initial consonant of the suffix are identical, the ensuing consonant sequence might potentially be realized as a long or geminate consonant (*n.n, k.k, t.t*). Alorese lacks phonemic geminates, while consonant lengthening only occurs in word-medial position as an effect of stress on the penultimate syllable, and when preceded by a vowel that etymologically relates to Proto-Alorese schwa *ə (Klamer, 2011: 30; Sulistyono, 2022: 75). The hetero-morphemic sequences of identical consonants that arise by adding a stem-final subject suffix to an unstressed syllable occur in a completely different environment than the word-internal phonetic geminates. We assume that they are subject to the same constraints as the other consonant sequences.

is particularly strong at morpheme boundaries. Attaching a consonantal prefix to a verb that is vowel-initial will always result in a simple CV syllable at the beginning of the verb, and never in a sequence of consonants. However, attaching a consonant-initial suffix to a consonant-final verb creates (heteromorphemic) consonant sequences, such as the unsimplified forms *gəkə**r**-**k**ən* (example (3)) and *bali**k**-**n**a* 'return-3SG' (Nishiyama and Kelen, 2007: 32, 95).

We propose that the conditions on consonant sequences in Alorese, combined with a tendency of L2 speakers to overgeneralize their strategies to avoid such constructions, contributes to contact-induced morphological simplification. We will now discuss our hypothesis in more detail.

2.4 *Avoidance of consonant sequences by L2 speakers*

In this paper, we study what happens with consonant sequences created through affixing (see Dressler et al., 2019: 93–97) and that are part of separate syllables. The sequences cross the morpheme boundary between the verbal base and the pronominal suffix, and are part of the derived (phonological, morphological) word. For consonant sequences in other contexts, such as consonant clusters within the syllable or at the word end, it has been shown how the clusters may be reduced in L1 and L2 acquisition (Dressler and Dziubalska-Kołaczyk, 2006; Zydorowicz, 2010; Dziubalska-Kołaczyk and Zielińska, 2011; Lardiere, 2003). We think some of the underlying principles also support the reduction of consonant sequences over morpheme and syllable boundaries, the focus of this work. Studies on both L1 and L2 speakers suggest that word-medial consonant sequences are somewhat more stable than word-final sequences (Dziubalska-Kołaczyk and Zielińska, 2010, 2011; Dziubalska-Kołaczyk and Zydorowicz, 2014),[6] but the studies show that in all positions consonant sequences undergo reduction. The word-medial consonant sequences in these studies partially overlap with the consonant sequences over syllable boundaries and morpheme boundaries which are the focus of this paper.

Our hypothesis is that, when L2 learners learn a language with phonotactic constraints on consonant sequences in specific contexts, the L2 speakers will *overgeneralize* this avoidance of consonant sequences. Overgeneralization lets L2 speakers avoid consonant sequences even in contexts where consonant sequences would normally occur, such as when combining stems and affixes. This overgeneralization account is consistent with evidence that L2 speakers are less sensitive to the phonological environment than L1 speakers when

6 The results in Dziubalska-Kołaczyk and Zielińska (2010) additionally indicate that word-medial consonant sequences are more stable than word-initial consonant sequences.

reducing consonant clusters (Bayley, 1996). Simplification through overgeneralization in this paper is assumed to be a universal effect of L2 acquisition, with an influence of the typological distance between the L1 and L2.

From one perspective, the effect is universal, rather than a transfer effect. The Alor-Pantar languages with which Alorese is in contact, such as Adang, Blagar, Kaera, and Western Pantar, all have CV and CVC as their basic syllable structure types, like Alorese. They only allow consonant sequences in one specific context: word-initial onsets with a liquid as second consonant (e.g., Western Pantar *bro* 'dust'; Holton and Klamer, 2018: 587; Holton and Robinson, 2017: 68). Adang, the Alor-Pantar language which has a long established history of contact with Alorese, does not allow consonant sequences within the syllable at all (Robinson and Haan, 2014: 231). As mentioned, in Alorese, consonant sequences within the morpheme are restricted to the combination of homorganic nasal and stop across two syllables. Thus, both the Alor-Pantar languages and Alorese have either very strong conditions or do not allow consonant sequences within the morpheme. As the phonotactics of Alorese and the Alor-Pantar languages are largely similar in this respect, we assume the simplification does not happen through transfer effects of the first language of the L2 speakers, in which linguistic material from this first language (e.g., a morphological pattern) is imported into the L2. Instead, we analyze the simplification as being based on universal principles of language processing in L2 acquisition (see Muysken, 2013).

Additionally, a universal effect of the typological distance between the L1 and L2 of L2 speakers may play a role. In Western Lamaholot, consonant sequences are not allowed in most contexts, but certain consonant sequences (of base-final coda, and suffix-initial onset) are allowed across morpheme boundaries due to suffixing processes. If we assume that Western Lamaholot is close to the pre-Alorese that served as the target for earlier generations of L2 speakers acquiring Alorese, L2 speakers did thus observe consonant sequences in specific contexts in pre-Alorese, but still avoided them and overgeneralized this avoidance. This may be due to the typological distance between Alorese and the Alor-Pantar languages. The Alor-Pantar languages that are in contact with Alorese (i.e., the L1 of the Alorese L2 speakers), lack subject affixes altogether, while verbal prefixes express objects (Robinson and Haan, 2014: 266–267). This leads to a mismatch, where L2 speakers fail to parse the meaning of the subject suffixes on the verb, hence neither recognize the morpheme boundaries nor the existence of consonant sequences that occur across those boundaries. We thus assume that a universal effect of the typological distance between languages (Thomason, 2001: 71, 76–77; see also Klamer, 2002: 377–380), that have different morpho-syntactic alignment patterns, leads to overgeneralization of constraints on consonant sequences. In the Alorese case, the constraint is that a word may not have /C.C/ sequences, except a nasal-obstruent sequence in word-medial position.

Avoidance of consonant sequences or clusters can result in deletion, sound change, or vowel epenthesis to break up the consonant sequence (Zydorowicz, 2010; McCarthy, 2008), or avoidance and replacement of words with non-permitted segment combinations (Dressler and Dziubalska-Kołaczyk, 2006: 60–61). Even though these are relatively small changes, there is broader evidence that small phonological changes, combined with ensuing processes, can lead to loss of affixes (see Sims-Williams and Baerman, 2021, and Kusters, 2003, discussed above). It seems therefore not unreasonable to assume that the phonological changes resulting from the reduction of consonant sequences could, under the influence of subsequent processes, eventually lead to loss of affixes.

2.5 *From local to holistic changes: Generalization*

The avoidance of consonant sequences only applies to forms in the language where consonant sequences appear during affixing. This makes it a relatively local process. We hypothesize that this leads to a holistic process of morphological simplification for all verbs in the language. The whole verbal inflectional system loses the expression of the categories of person and number.

In Western Lamaholot, which could be regarded as a predecessor state of Alorese, about one-third of the attested suffixing verbs end in a consonant and thus form a consonant sequence in combination with the suffix. How does the loss of affixes spread from the verb forms with consonant sequences to the verb forms without consonant sequences? We think that processes of generalization between lexical concepts are needed to explain this extension of forms from one verb to the other (Prasada and Pinker, 1993). For example, after a simplified form is used for the third person singular of a specific verb, it will then via analogy also be used for the third person singular of another verb. Generalization through analogy has been shown to be a powerful force in morphology, such as for the choice of linking morphemes within groups of words in Dutch (Krott et al., 2001) and in the change of verbal morphology in Greek (Sims-Williams, 2016). Court et al. (2023) show, using a computational model, how analogy plays a role in contact-induced morphological change in Maltese.

Our hypothesis thus becomes as follows: avoidance of consonant sequences by L2 speakers, combined with mechanisms of generalization between lexical concepts, could play a role in contact-induced morphological simplification in Alorese and possibly in other languages. To test this hypothesis more clearly, we will investigate it using an agent-based model.

3 Agent-based model

3.1 *Motivation*

We want to study how phonotactic mechanisms (introduced in Section 2.4) influence contact-induced morphological simplification in Alorese. As we believe generalization plays an important role in spreading of changes through the language system of an individual (Section 2.5), we would also like to test how generalization conditions the influence of phonotactics on simplification. Please note that we do not claim that phonotactic mechanisms are the only mechanisms behind morphological simplification, but we would like to study them in isolation to establish to what extent they could be responsible for processes of contact-induced simplification.

Agent-based computer models (simulations of interactions between language users) are one method to study these types of questions (Smith, 2014). Agent-based models allow isolation and evaluation of conditions which are difficult to manipulate in the real world. These models can be used to explore hypotheses, which can be subsequently further validated using empirical methods (de Boer, 2012; Gong et al., 2014). Agent-based models have been used before to study the influence of social factors (like population size and language contact) on linguistic structure, usually in more abstract settings (Dale and Lupyan, 2012; Lou-Magnuson and Onnis, 2018; Reali et al., 2018; de Bie and de Boer, 2007). The models have also been applied to more concrete case studies of contact-induced morphological and phonological change, such as German case syncretism (van Trijp, 2013), convergence of the phonological inventory in the Tibeto-Burman language Duoxu (Chirkova and Gong, 2019), contact-induced morphological change in Mozambican Portuguese (Jon-And and Aguilar, 2019), and propagation of lexical items in the creole Sranan in Suriname (Cheung et al., 2022).[7] To our knowledge, our study is the first agent-based model investigating whether phonotactic factors play a role in contact-induced morphological simplification.

We would like to construct an agent-based model of the verbal morphology of Alorese that stays close to the real language, to be able to model the language-specific phonotactics. The model is initialized with linguistic data from Lewoingu Lamaholot, a variety of Western Lamaholot, which we consider as a predecessor state of Alorese before it underwent morphological simplification. After running the model, the model outcome is compared to the linguistic

7 An additional line of work are mathematical models of language change, which study the population as a whole, and have been used to investigate contact-induced simplification, evaluated on Afrikaans and Afro-Peruvian Spanish data (Kauhanen, 2022).

situation in Alorese, where morphological simplification has taken place, in order to evaluate which mechanisms and parameter settings lead to an outcome close to the situation in Alorese. The attested language history of Alorese developing from Western Lamaholot and the geographical separation of Western Lamaholot from contact with Alor-Pantar languages (see Section 2.1) make it possible to consider Western Lamaholot as a predecessor state of Alorese.

Because the difference between L1 and L2 learning has been characterized as perfect versus imperfect intergenerational transmission (Kusters, 2003; Dubois and Noetzel, 2005), we create a model of intergenerational transmission of verbal affixes, in which the next generation has to learn from the previous generation. This "vertical transmission" approach (see also iterated learning approaches in Kirby and Hurford, 2002; Kirby et al., 2014) contrasts with "horizontal transmission" approaches to agent-based modeling, where interaction within one generation of speakers is modeled. In our model, the difference between L1 and L2 speakers is defined by the fact that L1 speakers receive a perfect copy of the language from the previous generation, while L2 speakers have to learn it through interaction, which can lead to imperfect acquisition. In reality, L1 speakers also acquire the language through interaction, but we abstract away from this in order to focus on L2 transmission of the language. When L2 speakers produce an utterance, phonotactic reduction and generalization mechanisms are applied.

In our model, we study how different factors interact: which combinations of levels of phonotactic reduction, proportion of L2 speakers, and generalization lead to morphological simplification? Although other factors could possibly influence morphological simplification, such as community size, network structure, and other cognitive mechanisms, these are not included in this study, in order to fully focus on the effects of the mechanisms under study. Also, the possibility of the innovation of new affixes is not part of this model.

3.2 *Setup*

Every simulation consists of a number of generations (default: 50), where every next generation of agents has to learn the affixes for different verb-person combinations from the previous generation (see Fig. 3).[8] The first generation g_0 consists of only L1 speakers: all agents in this generation are initialized with the full language (see Section 3.3). In every next generation g_i, a new set of

8 As the model abstracts reality in several aspects, a generation cannot be exactly mapped to a certain time period. However, the whole run of the model should approximately correspond to the time it took to develop from the full morphology in Western Lamaholot to no morphology in Alorese.

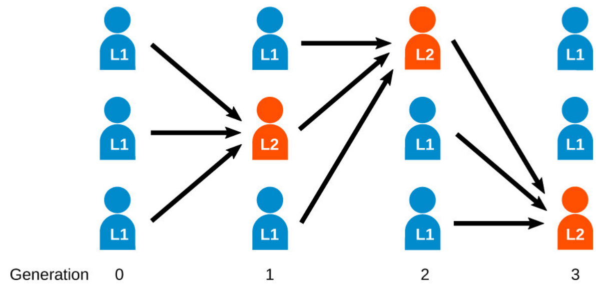


FIGURE 3 Overview of the intergenerational nature of our agent-based model: L2 speakers in every new generation learn from all other speakers in the previous generation, while L1 speakers get the language transmitted faithfully

agents is created, with a proportion of L2 agents which is set by a parameter (see Section 3.7) and the rest of the agents being L1 agents. L1 agents in the new generation do not get spoken to, they receive the full language faithfully. The L1 agents thus serve as a stabilizing factor. L2 agents in the next generation learn the language from all agents (L1 and L2) in the previous generation. When the new generation g_i is created, every agent from the previous generation g_{i-1} engages in n_{ipg} interactions, speaking to the L2 agents in the new generation.[9] In every interaction, this agent in previous generation g_{i-1} randomly chooses an L2 agent in the new generation g_i and speaks to it. If in an interaction, communication is successful, the L2 agent in the new generation updates its internal representation, thus learning the language in this way from the previous generation (see Section 3.4).

3.3 *Initialization*

The model is exemplar-based (Pierrehumbert, 2001; Gahl and Yu, 2006; Wedel, 2006): every agent has a memory of previously perceived utterances (exemplars; see Bloem et al., 2015, for another example). In general, in an exemplar-based model, the more instances of a certain utterance there are in the list, the higher the probability of using that utterance. Because an agent stores multiple different possible utterances for a concept, there will be intra-individual variation in usage of utterances to express a concept, just as in the real world.

9 The network structure between agents in different generations is thus fully connected. Other types of network structures could lead to different patterns of spread of innovations (Fagyal et al., 2010; Pierrehumbert et al., 2014; Josserand et al., 2021, 2024; Buzato and Cunha, 2024).

TABLE 1 Excerpt of the internal table of an agent, for five prefixing and five suffixing verbs (out of the total of 56 verbs)

Concept	Form	Type	Affixes						
			1SG	2SG	3SG	1PL.INCL	1PL.EXCL	2PL	3PL
to sleep with	*itə*	prefix	*k-*	*m-*	*n-*	*m-*	*t-*	*m-*	*r-*
to hunt	*iu*	prefix	*k-*	*m-*	*n-*	*m-*	*t-*	*m-*	*r-*
will	*odi*	prefix	*k-*	*m-*	*n-*	*m-*	*t-*	*m-*	*r-*
to improve	*olin*	prefix	*k-*	*m-*	*n-*	*m-*	*t-*	*m-*	*r-*
to dream	*urən*	prefix	*k-*	*m-*	*n-*	*m-*	*t-*	*m-*	*r-*
to return	*balik*	suffix	*-kən*	*-ko, -no*	*-na*	*-kən*	*-te*	*-ke, -ne*	*-ka*
to eat	*bu'a*	suffix	*-kən*	*-ko, -no*	*-na*	*-kən*	*-te*	*-ke, -ne*	*-ka*
to open	*buka*	suffix	*-kən*	*-ko, -no*	*-na*	*-kən*	*-te*	*-ke, -ne*	*-ka*
to be born	*bəsuk*	suffix	*-kən*	*-ko, -no*	*-na*	*-kən*	*-te*	*-ke, -ne*	*-ka*
to stand	*de'in*	suffix	*-kən*	*-ko, -no*	*-na*	*-kən*	*-te*	*-ke, -ne*	*-ka*

The left part (lexical concept, form) is the same for every agent and thus common knowledge. The right part ■ additional information from caption moved here, OK? (affixes) is different per agent, and changes throughout the simulation. Every agent keeps a list of affixes per concept. For L1 agents, it is initialized like in this table. For L2 agents, the affix lists are initialized empty: the affix columns in this table would be empty.

As agents convey combinations of lexical concepts and grammatical persons using affixes, every agent keeps a separate list of affixes they have heard for every combination of lexical concept and person; see Table 1. We initialize L1 agents with verb forms and affixes from a grammar of Lewoingu Lamaholot (Nishiyama and Kelen, 2007: 31–32), which can be seen as the starting state of Alorese, before it underwent morphological simplification. We use 56 verb concepts, which are the only verbs in the grammar explicitly classified as prefixing or suffixing: 17 prefixing and 39 suffixing verbs. We do not take into account frequency:[10] the list of affixes is initialized with one exemplar if there is one possible affix, or two exemplars if there are two possible affixes. The list gets extended with more exemplars during the course of the simulation. For L2 agents, this list is initialized empty: the lexical concepts (verbs) and persons are known, but no affixes are in the list. Initializing all agents with the same

10 In 2022, a corpus of Alorese became available (Sulistyono, 2022), with the dataset available at https://dataverse.nl/dataset.xhtml?persistentId=doi:10.34894/APQDTX. Due to the small size (one hour of recordings), only very rough estimates of concept frequencies could be extracted.

table of concepts and forms (with affixes for L1 and without affixes for L2) is a simplifying assumption that introduces a certain common knowledge into the model. The internal representation of the language in the agents stays close to the real word forms in Alorese, which allows for modeling of the language-specific phonotactics.

3.4 *Interaction*

In an interaction between a speaker and a listener, agents play a language game (Steels, 1998): agents exchange signals (in this case affixes) and have to infer the meanings (in this case grammatical persons); communication is successful if the hearer ends up retrieving the meaning that the speaker intended. Table 2 gives an overview of one interaction between a speaker (from the previous generation) and a listener (from the new generation).

When an agent speaks, it randomly selects a concept it will speak about: a combination of a lexical concept (e.g., *itə* 'to sleep with') and a person (e.g., 1SG). Based on the concept (lexical concept + grammatical person), it draws an affix from the exemplar list for this concept. The speaker then sends an utterance, consisting of the verb form (stem) and the affix, to the listener.

We assume the listener can understand the lexical concept from the verb form without further inference: this is considered common knowledge. This is a simplifying assumption, to be able to focus purely on the role of the affixes and their simplification in the model. The affix is used by the speaker, as the only source of information, to determine the grammatical person that was meant by the speaker. To keep the model focused on affix morphology, no information from the pronoun is used to determine meaning by the listener.[11] The listener then infers the grammatical person by choosing the person of which the highest proportion of exemplars match the received affix. If multiple persons have the same proportion of matching exemplars, one of these persons is randomly chosen. Also if there is no matching exemplar, a random person is chosen: this happens for example for just initialized L2 agents with an empty system. The listener then sends the inferred concept, consisting of a lexical concept and a person, to the speaker. The speaker answers with a binary feedback signal, indicating whether the concept is correct or not. If the concept was correct, the interaction was successful. The listener then adds the affix used by the speaker to its affix list for that concept.

11 Often the subject is expressed with a pronoun in Alorese (Klamer, 2011: 60–61) and Western Lamaholot (Nishiyama and Kelen, 2007: 13–15). Using only the affix, and not the pronoun, to determine grammatical person is thus an abstraction of the situation in the language. See Section 5 for a further discussion of the role of pronouns.

TABLE 2 Algorithm of one interaction in the simulation. l: lexical concept, p: grammatical person, a: affix, p_{gen}: generalization probability, p_{red}: phonotactic reduction probability. Subscripts s and l stand for speaker and listener

Speaker	Listener
Sample $l_s, p_s \sim$ *concepts*	
$f \leftarrow$ concept form map(l_s)	
With probability p_{gen} :	
Sample $a \sim$ affix list(concepts affix type(l_s))	
Else	
Sample $a \sim$ affix list(l_s, p_s)	
With probability p_{red}:	
$a \leftarrow$ phonotactic reduction(a,f)	
Send f, a to listener	
	Receive f, a
	$l_l \leftarrow$ form concept map(f)
	$p_l \leftarrow p$ for which affix list(l_l, p) has most occurrences of a
	If tie, randomly choose p_l from p with highest occurrences
	Point to l_l, p_l
Receive concept l_l, p_l	
If ($l_l, p_l = l_s, p_s$)	
Send nonverbal feedback: *Correct*	
Else	
Send nonverbal feedback: *Incorrect*	
	Receive nonverbal feedback
	If *Correct*:
	Append a to affix_list(l_l, p_l)
	Else
	Nothing

3.5 *Phonotactic reduction*

We implement a variant of the model that features phonotactic reduction. As discussed in Section 2.4, reduction of consonant sequences can have several outcomes, including deletion of segments, sound change, or vowel epenthesis, which in the long run and in combination with other processes can lead to full simplification of the affix. As a strong hypothesis to explore the question, we operationalize phonotactic reduction as a full simplification of the affix once a consonant sequence arises during affixing. This full affix reduction serves as a first approximation, which can later be refined by a more subtle mechanism. One other variation we experimented with is dropping only the single phoneme in the affix that borders on the stem (see Appendix A).

In the full phonotactic reduction model, with a probability p_{red}, which allows for intra-individual variation, a producing L2 agent drops the full affix if adding the affix to the stem would lead to the formation of a consonant sequence. The chosen data representation, staying close to the real word forms in Alorese, allows for this modeling of language-specific phonotactics. This phonotactic reduction is only performed by L2 speakers, and happens after an affix has been chosen from the exemplar list and before it is sent to the listener. If the phonemes on the morpheme boundary between stem and affix form a consonant sequence, the affix is dropped.[12] In this simplified representation, a sequence of two different consonants or two of the same consonant is treated in the same way: in both cases the affix is dropped.[13] The stems of the verb forms in the dataset do not contain consonant sequences; this sequence thus always arises due to concatenation with an affix. The implementation of the phonotactic reduction mechanism can be found in Table 3. The *phonotactic reduction probability* is the probability that during one interaction, phonotactic reduction is performed. The default setting is 1.0, so it is always performed by L2 speakers.

3.6 *Lexical concept generalization*

As discussed in Section 2.5, a generalization mechanism is needed to spread the changes through phonotactic reduction of individual lexical items caused by phonotactic reduction across the full morphological system. Generalization is operationalized in the following way: during production, with a probability p_{gen} per interaction, the agent picks an affix from the probability distribution of that grammatical person for all lexical concepts for that affix type (prefix or suffix), instead of just for the current lexical concept. Generalization only takes

12 The phonemes on the boundaries of the verb stem and the affix are converted to a CV pattern, marking each phoneme as either a consonant (C) or a vowel (V), using the `pyclts` library (List et al., 2020).

13 In some languages, reduction of geminates may lead to different outcomes than other consonant sequences: degemination can lead to a chain of other sound changes (Campbell, 2013: 41–42). In Alorese, geminates are not phonemic: lengthening of consonants is only caused by stress in a list context (Klamer, 2011: 30). We assume geminates behave in the same way as other consonant sequences, see also footnote 5.

TABLE 3 Definition of the phonotactic reduction function

```
phonotactic_reduction (affix, form):
    fa ← morpheme boundary phonemes(form, affix)
    cv ← cv pattern(fa)
    If CC ∈ cv
        If drop_boundary:
            affix ← drop boundary phoneme(affix)
        Else
            affix ← ∅
    Else:
        affix unchanged
    Return affix
```

place over lexical concepts, not over grammatical persons, and only takes place over lexical concepts within the same affix type (prefix or suffix).

3.7 *Evaluation*

Separate models with proportions of L2 speakers between 0.2 and 1.0 are evaluated, to analyze the dynamics of smaller and larger proportions of L2 speakers in a society. The exact number of L2 speakers on the Alor-Pantar islands is not known (especially not for historical times), but based on demographic data gathered during fieldwork, it is estimated that more than 50 percent of the speakers of Alorese are L2 speakers (Moro, 2019).[14] Because this estimate is rough and based on contemporary rather than historical data, it seems useful to evaluate a range of values for the proportion of L2 speakers. This also allows for the evaluation of how the morphological complexity changes with an increasing proportion of L2 speakers.

The morphological complexity of prefixes and suffixes, for L2 speakers, is evaluated by the proportion of non-empty affixes in the exemplars of the agents (average over all agents in population). This is a metric of syntagmatic morphological complexity, taking into account the existence or number of morphemes that are bound to a stem. An alternative would be to focus on paradigmatic morphological complexity, based on the predictability of forms within

14 Moro (2019) asked local informants for estimates of the number of current speakers in Papuan Alor-Pantar villages who speak Alorese as L2. There are at least as many L2 Alorese speakers in Papuan villages as there are L1 Alorese speakers in Alorese villages. This leads to an estimate of at least 50 percent L2 speakers.

the paradigm (Stump and Finkel, 2015). In the current model setting, including intra-individual variation, the syntagmatic metric of proportion of non-empty affixes seems most straightforward to compute and understand. For an overview of measures of morphological complexity, see Miestamo et al. (2008) and Sagot (2013).

We evaluate the morphological complexity of L2 speakers, because the language of L1 speakers is by definition stable in this model. This is also the reason we do not evaluate for a proportion of L2 speakers of 0.0: in that case all agents would learn perfectly, and no change would take place.

4 Results

Simulations were run with the default model settings described in Table 4, unless otherwise stated. See Appendix B for a parameter exploration of some of these settings.

4.1 *Model results*

We evaluate the model with full phonotactic reduction (the whole affix is dropped if the morpheme boundary constitutes a consonant sequence) for different proportions of L2 speakers (for results for the boundary phoneme reduction model, see Appendix A). We plot the morphological complexity for different affix types (prefix/suffix, L1/L2), measured by the number of affixes in the language system of the agents that are non-empty. Figure 4a shows the result for the model without lexical concept generalization: here the morphological complexity (proportion of non-empty affixes) of suffixes drops by about 10 percent with increasing proportions of language learners, while the morphological complexity for prefixes does not decrease. The prefix result in this figure could be regarded as a kind of baseline model: in this model, no generalization takes place, and because prefixes do not give rise to any consonant sequences, it can be established from the mechanics of the model that no phonotactic reduction will take place here. Figure 4b shows that with a generalization probability of 0.1, a sharper decrease in morphological complexity of suffixes (about 20 percent) takes place with an increasing proportion of L2 speakers. The complexity of prefixes stays unchanged: this corresponds to the situation in Alorese. This result thus shows that a combination of the effects of incomplete transmission through adult language learning, a phonotactic reduction mechanism, and generalization leads to the largest decrease of morphological simplification. Suffixes simplify, while prefixes do not, which is congruent with the documented changes in Alorese.

TABLE 4 Model settings used in the agent-based model

Setting	Value	Note
# generations	50	
# agents per generation	10	
# interactions per generation (n_{ipg})	10,000	
# model runs	10	results are an average over these runs
generalization probability (p_{gen})	0.1	
phonetic reduction probability (p_{red})	1.0	

Generalization is needed to spread the use of the empty form from lexical concepts which give rise to consonant sequences to lexical concepts which will never lead to the creation of consonant sequences. Figure 4c shows the number of unique affixes per agent. Figure 4d shows the proportion of interactions where the concept is correctly inferred, a measure of communicative success. Both of these metrics stay relatively stable across different proportions of L2 learners.

4.2 *Evaluation of lexical concept generalization probability*

Figure 5a shows, for the model with phonotactic reduction enabled, how different values of the lexical concept generalization probability influence the morphological complexity (here only reporting suffix of L2 speakers). As seen in the model results (Section 4.1), this evaluation graph shows that without generalization (yellow line), the phonotactic reduction mechanism gives only a decrease in morphological complexity of about 10 percent. Surprisingly, a low generalization probability of 0.1 (green line) led to the highest decrease in complexity of 20 percent, while higher values of the generalization probability led to smaller decreases in complexity. This probably happens because for higher generalization probabilities, the communicative success, measured by the proportion of interactions where the concept is correctly inferred, becomes lower (Fig. 5b), because it becomes more difficult to comprehend a concept when the distribution over all concepts is used when producing a form. When there is high communicative success (at low generalization probabilities), it allows the empty affix, which is most common across all lexical concepts, to spread.

The generalization mechanism thus functions through the spread of an innovation that is most common across lexical concepts. In the current model, the source of the innovative form is the phonotactic reduction mechanism, which (when full phonotactic reduction is performed), reduces every affix to

2024493 [LDC-2025-15.1] 10036-Dekker-et-al-proof-01 [version 20250423 date 20250423 16:21] page 23

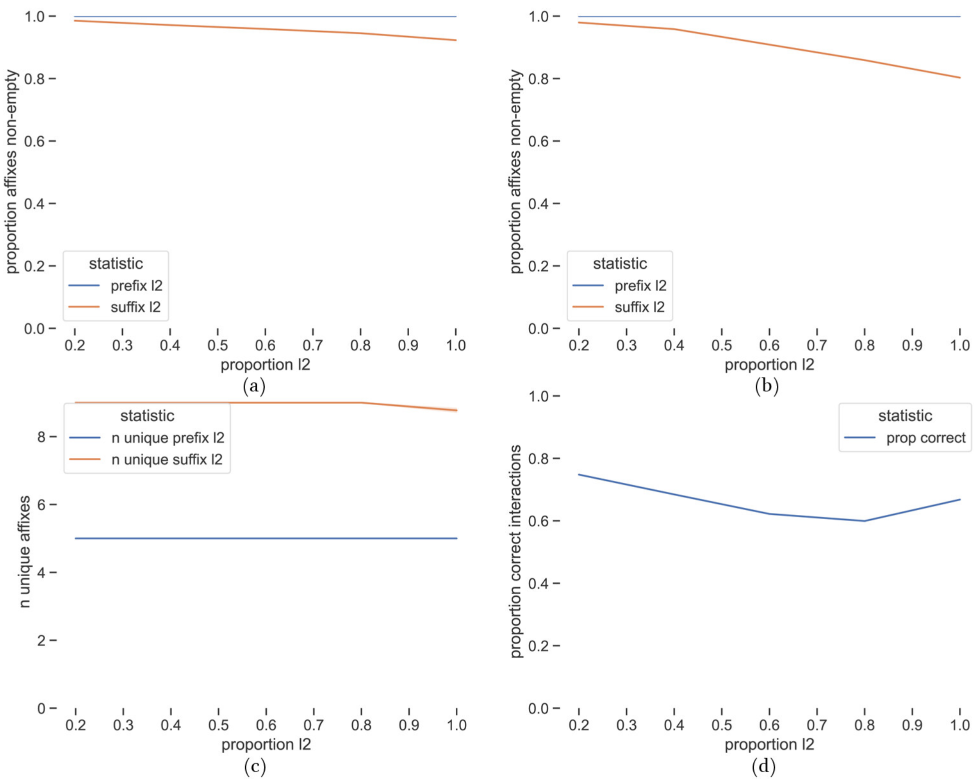


FIGURE 4 Results for model with phonotactic reduction, for different proportions of L2 speakers and different affix types (prefix/suffix, L1/L2). Phonotactic reduction probability 1.0; lexical concept generalization probability of 0.0 in (a) and 0.1 in (b)

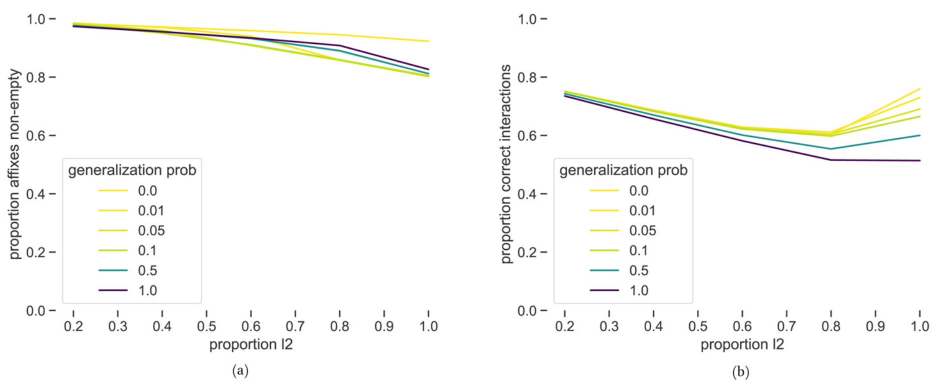


FIGURE 5 Evaluation of different lexical concept generalization probabilities. Phonotactic reduction probability fixed to 1.0

the empty affix, leading to a situation with one clear most common innovation. However, if a similar generalization mechanism were used in a model studying a different linguistic question, the most common innovative form could take a completely different shape (e.g., an innovative lexical item), but the generalization mechanism would spread this innovation in the same way.

For evaluation of other parameters (phonotactic reduction probability, number of agents, number of interactions per iteration), see Appendix B.

5 Conclusion and discussion

We studied the role that phonotactic mechanisms could play in contact-induced morphological simplification using an agent-based model of Alorese, an Austronesian language in contact with Papuan languages. We hypothesized that the phonotactic structure of the language subject to change is a factor in contact-induced morphological simplification. We proposed that L2 speakers overgeneralize the phonotactic restrictions in the language they acquire and avoid consonant sequences caused by affixing. This process then leads to the loss of affixes. Our results show that, under the assumptions of our model, this phonotactic reduction mechanism in combination with a generalization mechanism and a high proportion of L2 speakers, leads to a similar situation as that observed in Alorese: simplification of the suffixing morphology, while keeping the prefixing morphology intact. Generalization is needed to spread the empty form from verbs which give rise to consonant sequences to verbs that do not give rise to consonant sequences. The generalization mechanism fosters the use of an innovation that is frequently used over many lexical concepts, in this case the empty affix, which could be a more general mechanism that plays a role in other instances of language change. Our results highlight the role of multiple interacting mechanisms in morphological change.

The implemented phonotactic reduction mechanism, which deletes the entire affix when a consonant sequence arises when concatenating stem and affix, is a strong mechanism, and an abstraction of the processes of reduction of consonant sequences that happen in reality. A criticism of computer models is that sometimes the output of the model can easily be predicted from the mechanisms in the model (see de Boer, 2012). However, in our model, a strong mechanism has a surprisingly small effect: it causes only a moderate amount of morphological simplification with increasing proportions of L2 speakers. This can be explained by the pressure for communicative success in the model: forms which do not disambiguate meaning well, such as the empty affix, have a low chance of being retained by the listener and reproduced. A further exploration of the model (see Appendix A) suggests that a less far-reaching operationalization of phonotactic reduction, namely, deleting only the boundary phoneme of the affix, leads to a possible precursor of morphological simplification, with qualitatively similar outcomes to the model where the full affix is dropped.

We have given modeling evidence for the role of phonotactic mechanisms, specifically avoidance of consonant sequences, in contact-induced simplification in only one language, Alorese, which has a specific contact history, and linguistic profile. However, as interactions between other domains of phonology and contact-induced morphological simplification have been reported for other languages (see Section 2.3), we expect that the phonotactic mechanism in morphological simplification discussed in this paper is a more general one, that plays a role in more languages. We specifically expect this mechanism to be active in language contact situations where the acquired language has strong constraints on its phonotactics and where there is a large typological distance between the L1 and L2 of the L2 speakers.

The subtle effect that phonotactic reduction has in the model indicates that phonotactic mechanisms form just one part of the explanation of contact-induced morphological simplification, alongside multiple other processes. The case study of Alorese already shows a few of these other processes, with yet other mechanisms possibly acting in other languages. Firstly, stress may be an additional phonological factor that could play a role in contact-induced morphological simplification, as exemplified by Kusters (2003) and Molineaux (2012). Stress in Alorese is on the penultimate syllable of the stem (Klamer, 2011: 32–33). Because the suffix is unstressed, and attaches to the ultimate, unstressed, stem syllable, the suffix may be more prone to loss. Secondly, the prefixing verbs that retained their prefix in Alorese are in general frequently used verb concepts ('to go,' 'to eat,' 'to drink'): high frequency of use may have helped to protect irregular forms (Bybee and Thompson, 1997; Diessel, 2007). Finally, in Alorese, the number of L2 speakers has been remarkably high (50 percent or more of the population) for a prolonged period of time (600–700 years); no language shift of Papuan L1 speaking communities to Alorese has occurred; and only 5 percent of the Alorese lexicon consists of Papuan loanwords. In Sika, a related language from the Flores-Lembata group, a situation of massive morphological simplification also occurred, but this case likely involved a rapid shift to Sika, possibly within just one generation, and the transfer of a large number of loanwords (making up 38 percent of the lexicon) from the L1 languages into Sika (Klamer, to appear). This shows that contact situations resulting in similar morphological simplification may have different outcomes in other linguistic respects. In addition, the socio-demographics of the contact situation, including the proportion of L2 speakers, the age of L2 acquisition (adults/children), and the duration and intensity of the contact, can fundamentally change the linguistic outcome. For example, Moro (2021) shows that in Alorese, an earlier phase of children (with L1 Alorese) acquiring Alor-Pantar languages led to additive transfer of constructions from Alor-Pantar languages to Alorese.

In the current model, the listener infers meaning only from the affix. It could be interesting to study a model where, once the morphology has been lost to such an extent that affixes fail to disambiguate, pronouns and syntactic phenomena like word order become used as means to convey meaning. This could possibly lead to a dynamic where the rise of pronouns and syntactic constructions further speeds up the morphological simplification because the affixes become even less needed to carry meaning, leading to a self-reinforcing process of morphological simplification. These complexity trade-offs between linguistic domains are a topic of active discussion (Plank, 1998; Fenk-Oczlon and Fenk, 2008; Sinnemäki, 2008; Bentz et al., 2023; Benítez-Burraco et al., 2024).

In this paper, we hope to have shown that a language's phonotactics, and specifically the avoidance by L2 speakers of using consonant sequences during morphological affixing, is one piece of the puzzle to explain how morphological simplification under adult language contact occurs.

Supplementary materials

All code can be found in Zenodo repository https://zenodo.org/records/15085932 and Github repository https://github.com/peterdekker/agents-intergenerational.

Acknowledgments

Thanks to Owen Edwards for permission to use the maps in this article. Thanks to Andreas Baumann, Francesca Moro, Katie Mudd, Limor Raviv, Kaius Sinnemäki, Ricardo Napoleão de Souza, Ljiljana Progovac, Antonio Benítez-Burraco, David Gil, and Charles Redmon for insightful comments and discussions that improved this work. All errors remain our own. This work was supported by funding from the Flemish Government under Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen. PD was supported by a PhD Fellowship fundamental research (11A2821N) of the Research Foundation—

Flanders (FWO). MK was supported by funding from OCSEAN project, under the Marie Skłodowska-Curie grant agreement No. 873207.

References

Bayley, Robert. 1996. Competing constraints on variation in the speech of adult Chinese learners of English. In Robert Bayley and Dennis R. Preston (eds.), *Second Language Acquisition and Linguistic Variation*, Studies in Bilingualism, 97–120. Amsterdam: John Benjamins. doi:10.1075/sibil.10.05bay.

Benítez-Burraco, Antonio, Sihan Chen, and David Gil. 2024. The absence of a trade-off between morphological and syntactic complexity. *Frontiers in Language Sciences* 3: 1340493. doi:10.3389/flang.2024.1340493.

Bentz, Christian, Ximena Gutierrez-Vasques, Olga Sozinova, and Tanja Samardžić. 2023. Complexity trade-offs and equi-complexity in natural languages: A meta-analysis. *Linguistics Vanguard* 9(s1): 9–25. doi:10.1515/lingvan-2021-0054.

Bentz, Christian, and Bodo Winter. 2013. Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change* 3(1): 1–27. doi:10.1163/22105832-13030105.

Blevins, Juliette. 1995. The syllable in phonological theory. In J. Goldsmith (ed.), *Handbook of Phonological Theory*, 206–244. Oxford, United Kingdom: Blackwell.

Bliss, Heather. 2006. L2 acquisition of inflectional morphology: Phonological and morphological transfer effects. In Mary Grantham O'Brien, Christine Shea and John Archibald (eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference* (*GASLA 2006*), Sommerville: Cascadilla Proceedings Project. https://www.lingref.com/cpp/gasla/8/paper1481.pdf (accessed 2025-03-24).

Bloem, Jelke, Arjen Versloot, and Fred Weerman. 2015. An agent-based model of a historical word order change. In *Proceedings of the Sixth Workshop on Cognitive Aspects of Computational Language Learning*, 22–27. Lisbon, Portugal: Association for Computational Linguistics. doi:10.18653/v1/W15-2404.

Blom, Elma, Daniela Polišenská, and Fred Weerman. 2007. Effects of age on the acquisition of agreement inflection. *Morphology* 16(2): 313–336. doi:10.1007/s11525-007-9110-1.

Booij, Geert. 1996. Inherent versus contextual inflection and the split morphology hypothesis. In Geert Booij, Jaap Van Marle, Stephen Anderson, Mark Aronoff, Laurie Bauer, Mark Baker, Rudie Botha, Joan Bybee, Andrew Carstairs-McCarthy, Wolfgang Dressler, Jack Hoeksema, Rochelle Lieber, Peter Matthews, Franz Rainer, Sergio Scalise, Henk Schultink, Arnold Zwicky, Geert Booij and Jaap Van Marle (eds.), *Yearbook of Morphology 1995*, 1–16. Dordrecht: Springer Netherlands. doi:10.1007/978-94-017-3716-6_1.

Brochhagen, Thomas, Gemma Boleda, Eleonora Gualdoni, and Yang Xu. 2023. From language development to language evolution: A unified view of human lexical creativity. *Science* 381(6656): 431–436. doi:10.1126/science.ade7981.

Buzato, Dalmo, and Evandro Cunha. 2024. Agent-based modeling of language change in a small-world network. In Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti and Nianwen Xue (eds.), *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation* (*LREC-COLING 2024*), 594–599. Torino, Italy: ELRA and ICCL. https://aclanthology.org/volumes/2024.lrec-main/ (accessed 2025-03-24).

Bybee, Joan, and Sandra Thompson. 1997. Three frequency effects in syntax. *Annual Meeting of the Berkeley Linguistics Society* 23(1): 378–388.

Cabrelli Amaro, Jennifer, Gonzalo Campos-Dintrans, and Jason Rothman. 2018. The role of L1 phonology in L2 morphological production: L2 English past tense production by L1 Spanish, Mandarin, and Japanese speakers. *Studies in Second Language Acquisition* 40(3): 503–527. doi:10.1017/S0272263117000122.

Campbell, Lyle. 2013. *Historical Linguistics*. Edinburgh, United Kingdom: Edinburgh University Press.

Cheung, Ching-Yat, Kofi Yakpo, and Christophe Coupé. 2022. A computational simulation of the genesis and spread of lexical items in situations of abrupt language contact. In Andrea Ravignani, Rie Asano, Daria Valente, Francesco Ferretti, Stefan Hartmann, Misato Hayashi, Yannick Jadoul, Mauricio Martins, Yohei Oseki, Evelina Daniela Rodrigues, Olga Vasileva and Sławomir Wacewicz (eds.), *Proceedings of the Joint Conference on Language Evolution* (*JCoLE*), 115–122. Kanazawa, Japan. doi:10.17617/2.3398549.

Chirkova, Katia, and Tao Gong. 2019. Modeling change in contact settings: A case study of phonological convergence. *Language Dynamics and Change* 9(1): 1–32. doi:10.1163/22105832-00802006.

Court, Sara, Andrea Sims, and Micha Elsner. 2023. Analogy in contact: Modeling Maltese plural inflection. *Proceedings of the Society for Computation in Linguistics* 6(1): 35–46. doi:10.7275/y45e-k264.

Dale, Rick, and Gary Lupyan. 2012. Understanding the origins of morphological diversity: The Linguistic Niche Hypothesis. *Advances in Complex Systems* 15(03n04): 1150017. doi:10.1142/S0219525911500172.

de Bie, Pieter, and Bart de Boer. 2007. An agent-based model of linguistic diversity. In Anton Benz, Christian Ebert and Robert van Rooij (eds.), *Proceedings of the ESSLLI 2007 Workshop on Language, Games and Evolution*, 1–8. Dublin, Ireland.

de Boer, Bart. 2012. Modelling and language evolution: Beyond fact-free science. In L. McCrohon, T. Fujimura, K. Fujita, R. Martin, K. Okanoya, R. Suzuki and N. Yusa (eds.), *Five Approaches to Language Evolution: Proceedings of the Workshops of the 9th International Conference on the Evolution of Language*, Kyoto, Japan: Evolang9

Organizing Committee. https://kyoto.evolang.org/sites/default/files/EvolangWorkshopProceedings.pdf (accessed 2025-03-24).

Di Garbo, Francesca, and Ricardo Napoleão de Souza. 2023. A sampling technique for worldwide comparisons of language contact scenarios. *Linguistic Typology* 27(3): 553–589. doi:10.1515/lingty-2022-0005.

Diessel, Holger. 2007. Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology* 25(2): 108–127. doi:10.1016/j.newideapsych.2007.02.002.

Dressler, Wolfgang U., and Katarzyna Dziubalska-Kołaczyk. 2006. Proposing morphonotactics. *Rivista di Linguistica* 18(2): 249–266.

Dressler, Wolfgang U., Alona Kononenko, Sabine Sommer-Lolei, Katharina Korecky-Kröll, Paulina Zydorowicz, and Laura Kamandulytė-Merfeldienė. 2019. Morphological richness, transparency and the evolution of morphonotactic patterns. *Folia Linguistica* 53(s40-s1): 85–106. doi:10.1515/flih-2019-0005.

Dubois, Sylvie, and Sibylle Noetzel. 2005. Intergenerational pattern of interference and internally-motivated changes in Cajun French. *Bilingualism: Language and Cognition* 8(2): 131–143. doi:10.1017/S136672890500218X.

Dziubalska-Kołaczyk, Katarzyna, and Daria Zielińska. 2010. Predicting phonotactic difficulty in second language acquisition. In Andréia S. Rauber, Michael A. Watkins, Rosane Silveira and Rosana D. Koerich (eds.), *The Acquisition of Second Language Speech: Studies in Honor of Professor Barbara O. Baptista*, Florianópolis, Brazil: Editora Insular. http://hdl.handle.net/10593/9785 (accessed 2025-03-24).

Dziubalska-Kołaczyk, Katarzyna, and Daria Zielińska. 2011. Universal phonotactic and morphonotactic preferences in second language acquisition. In Magdalena Wrembel, Malgorzata Kul and Katarzyna Dziubalska-Kolaczyk (eds.), *Achievements and Perspectives in SLA of Speech: New Sounds 2010*, 53–64. Lausanne, Switzerland: Peter Lang.

Dziubalska-Kołaczyk, Katarzyna, and Paulina Zydorowicz. 2014. The production of high-frequency clusters by native and non-native users of Polish. In Walcir Cardoso and Pavel Trofimovich (eds.), *Proceedings of the International Symposium on the Acquisition of Second Language Speech. Concordia Working Papers in Applied Linguistics*, vol. 5, 130–144. Montreal, Canada: Concordia University.

Easterday, Shelece, Matthew Stave, Marc Allassonnière-Tang, and Frank Seifart. 2021. Syllable complexity and morphological synthesis: A well-motivated positive complexity correlation across subdomains. *Frontiers in Psychology* 12: 638659. doi:10.3389/fpsyg.2021.638659.

Enfield, N.J. 2014. *Natural Causes of Language: Frames, Biases, and Cultural Transmission*. Language Science Press. doi:10.26530/OAPEN_533873.

Fagyal, Zsuzsanna, Samarth Swarup, Anna María Escobar, Les Gasser, and Kiran Lakkaraju. 2010. Centers and peripheries: Network roles in language change. *Lingua* 120(8): 2061–2079. doi:10.1016/j.lingua.2010.02.001.

Fenk-Oczlon, Gertraud, and August Fenk. 2008. Complexity trade-offs between the subsystems of language. In Matti Miestamo, Kaius Sinnemäki and Fred Karlsson (eds.), *Language Complexity: Typology, Contact, Change*, (Studies in Language Companion Series 94), 43–65. Amsterdam: John Benjamins. doi:10.1075/slcs.94.05fen.

Fricke, Hanna Lotte Anneliese. 2019. *Traces of Language Contact: The Flores-Lembata Languages in Eastern Indonesia*. PhD dissertation, Leiden University.

Gahl, Susanne, and Alan C.L. Yu. 2006. Introduction to the special issue on exemplar-based models in linguistics. *The Linguistic Review* 23(3). doi:10.1515/TLR.2006.007.

Gardani, Francesco. 2008. *Borrowing of Inflectional Morphemes in Language Contact*. Frankfurt: Peter Lang.

Gong, Tao, Lan Shuai, and Menghan Zhang. 2014. Modelling language evolution: Examples and predictions. *Physics of Life Reviews* 11(2): 280–302. doi:10.1016/j.plrev.2013.11.009.

Greenhill, Simon J. 2023. A shared foundation of language change. *Science* 381(6656): 374–375. doi:10.1126/science.adj2154.

Holton, Gary, and Marian Klamer. 2018. 5. The Papuan languages of East Nusantara and the Bird's Head. In Bill Palmer (ed.), *The Languages and Linguistics of the New Guinea Area*, 569–640. De Gruyter. doi:10.1515/9783110295252-005.

Holton, Gary, and Laura C. Robinson. 2017. The internal history of the Alor-Pantar language family. In Marian Klamer (ed.), *The Alor-Pantar Languages: History and Typology*, 2nd edn., 49–91. Berlin: Language Science Press. doi:10.5281/zenodo.569388.

Hu, Chieh-Fang. 2010. Phonological bases for L2 morphological learning. *Journal of Psycholinguistic Research* 39(4): 305–322. doi:10.1007/s10936-009-9143-1.

Inkelas, Sharon. 2011. The interaction between morphology and phonology. In John Goldsmith, Jason Riggle and Alan C.L. Yu (eds.), *The Handbook of Phonological Theory*, 2nd edn., 68–102. Chichester, United Kingdom: John Wiley. doi:10.1002/9781444343069.ch3.

Inkelas, Sharon. 2014. *The Interplay of Morphology and Phonology*. Oxford, United Kingdom: Oxford University Press.

Jon-And, Anna, and Elliot Aguilar. 2019. A model of contact-induced language change: Testing the role of second language speakers in the evolution of Mozambican Portuguese. *PLOS ONE* 14(4): e0212303. doi:10.1371/journal.pone.0212303.

Josserand, Mathilde, Marc Allassonnière-Tang, François Pellegrino, and Dan Dediu. 2021. Interindividual variation refuses to go away: A bayesian computer model of language change in communicative networks. *Frontiers in Psychology* 12: 626118. doi:10.3389/fpsyg.2021.626118.

Josserand, Mathilde, Marc Allassonnière-Tang, François Pellegrino, Dan Dediu, and Bart de Boer. 2024. How network structure shapes languages: Disentangling the factors driving variation in communicative agents. *Cognitive Science* 48(4): e13439. doi:10.1111/cogs.13439.

Kauhanen, Henri. 2022. A bifurcation threshold for contact-induced language change. *Glossa: A Journal of General Linguistics* 7(1). doi:10.16995/glossa.8211.

Kirby, Simon, Tom Griffiths, and Kenny Smith. 2014. Iterated learning and the evolution of language. *Current Opinion in Neurobiology* 28: 108–114. doi:10.1016/j.conb.2014.07.014.

Kirby, Simon, and James R. Hurford. 2002. The emergence of linguistic structure: An overview of the iterated learning model. In Angelo Cangelosi and Domenico Parisi (eds.), *Simulating the Evolution of Language*, 121–147. London: Springer.

Klamer, Marian. 2002. Typical features of Austronesian languages in Central/Eastern Indonesia. *Oceanic Linguistics* 41(2): 363–383. doi:10.1353/ol.2002.0007.

Klamer, Marian. 2011. *A Short Grammar of Alorese* (*Austronesian*). München, Germany: Lincom Europa.

Klamer, Marian. 2012. Papuan-Austronesian language contact: Alorese from an areal perspective. In Evans, Nicholas and Klamer, Marian (eds.), *Language Documentation & Conservation. Melanesian Languages on the Edge of Asia: Challenges for the 21st Century*, 72–108. Honolulu, Hawaii: University of Hawaii Press.

Klamer, Marian. 2020. From Lamaholot to Alorese: Morphological loss in adult language contact. In David Gil and Antoinette Schapper (eds.), *Austronesian Undressed: How and Why Languages Become Isolating*, 339–367. Amsterdam: John Benjamins.

Klamer, Marian. to appear. Papuan-Austronesian contact in pre-modern eastern Indonesia. In Nick Evans and Sebastian Fedden (eds.), *The Oxford Guide to the Papuan Languages*, Oxford: Oxford University Press.

Krott, Andrea, R. Harald Baayen, and Robert Schreuder. 2001. Analogy in morphology: Modeling the choice of linking morphemes in Dutch. *Linguistics* 39(1). doi:10.1515/ling.2001.008.

Kusters, Christiaan Wouter. 2003. *Linguistic Complexity: The Influence of Social Change on Verbal Inflection*. PhD dissertation, LOT Utrecht.

Lardiere, Donna. 1998. Dissociating syntax from morphology in a divergent L2 end-state grammar. *Second Language Research* 14(4): 359–375. doi:10.1191/026765898672500216.

Lardiere, Donna. 2003. Second language knowledge of [±Past] vs. [±Finite]. In Juana Liceras, Helmut Zobl and Helen Goodluck (eds.), *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference* (*GASLA 2002*), 176–189. Sommerville: Cascadilla Press.

List, Johann-Mattis, Cormac Anderson, Tiago Tresoldi, and Robert Forkel. 2020. PYCLTS. A Python library for the handling of phonetic transcription systems. Zenodo. doi:10.5281/zenodo.4321268.

Lou-Magnuson, Matthew, and Luca Onnis. 2018. Social network limits language complexity. *Cognitive Science* 42(8): 2790–2817. doi:10.1111/cogs.12683.

Lupyan, Gary, and Rick Dale. 2010. Language structure is partly determined by social structure. *PLoS ONE* 5(1): e8559. doi:10.1371/journal.pone.0008559.

McCarthy, John J. 2008. The gradual path to cluster simplification. *Phonology* 25(2): 271–319. doi:10.1017/S0952675708001486.

McCarthy, John J., and Alan Prince. 1995. Prosodic morphology. In J. Goldsmith (ed.), *Handbook of Phonological Theory*, 318–366. Oxford, United Kingdom: Blackwell.

Meakins, Felicity. 2011. Borrowing contextual inflection: Evidence from northern Australia. *Morphology* 21(1): 57–87. doi:10.1007/s11525-010-9163-4.

Miestamo, Matti, Kaius Sinnemäki, and Fred Karlsson (eds.). 2008. *Language Complexity: Typology, Contact, Change*. Studies in Language Companion Series. Amsterdam: John Benjamins.

Molineaux, Benjamin J. 2012. Prosodically conditioned morphological change: Preservation vs loss in Early English prefixes. *English Language and Linguistics* 16(3): 427–458. doi:10.1017/S1360674312000184.

Moro, Francesca R. 2019. Loss of morphology in Alorese (Austronesian): Simplification in adult language contact. *Journal of Language Contact* 12(2): 378–403. doi:10.1163/19552629-01202005.

Moro, Francesca Romana. 2021. Multilingualism in eastern Indonesia: Linguistic evidence of a shift from symmetric to asymmetric multilingualism. *International Journal of Bilingualism* 25(4): 1102–1119. doi:10.1177/13670069211023134.

Muysken, Pieter. 2013. Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism: Language and Cognition* 16(4): 709–730. doi:10.1017/S1366728912000727.

Nagaya, Naonori. 2011. *The Lamaholot Language of Eastern Indonesia*. PhD dissertation, Rice University.

Napoleão de Souza, Ricardo, and Kaius Sinnemäki. 2023. Beyond segment inventories: Phonological complexity measures and suprasegmental variables in contact situations. *Journal of Language Contact* 15(3–4): 439–480. doi:10.1163/19552629-15030001.

Nishiyama, Kunio, and Herman Kelen. 2007. *A Grammar of Lamaholot, Eastern Indonesia*. München, Germany: Lincom Europa.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Joan Bybee and Paul J. Hopper (eds.), *Frequency Effects and Emergent Grammar*, Typological Studies in Language, 137–158. Amsterdam: John Benjamins. doi:10.1075/tsl.45.08pie.

Pierrehumbert, Janet B., Forrest Stonedahl, and Robert Daland. 2014. A model of grassroots changes in linguistic systems. *arXiv:1408.1985* doi:10.48550/arXiv.1408.1985.

Plank, Frans. 1998. The co-variation of phonology with morphology and syntax: A hopeful history. *Linguistic Typology* 2: 195–230. https://www.degruyter.com/document/doi/10.1515/lingty-2017-1007/html (accessed 2023-12-21).

Prasada, Sandeep, and Steven Pinker. 1993. Generalisation of regular and irregular morphological patterns. *Language and Cognitive Processes* 8(1): 1–56. doi:10.1080/016909 69308406948.

Prévost, Philippe. 2003. Truncation and missing inflection in initial child L2 German. *Studies in Second Language Acquisition* 25(1): 65–97. doi:10.1017/S0272263103000032.

Prévost, Philippe, and Lydia White. 2000. Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research* 16(2): 103–133. doi:10.1191/026765800677556046.

Reali, Florencia, Nick Chater, and Morten H. Christiansen. 2018. Simpler grammar, larger vocabulary: How population size affects language. *Proceedings of the Royal Society B: Biological Sciences* 285(1871): 20172586. doi:10.1098/rspb.2017.2586.

Robinson, Laura C, and John W. Haan. 2014. Adang. In Antoinette Schapper (ed.), *The Papuan Languages of Timor, Alor and Pantar: Volume 1*, 221. Boston/Berlin: De Gruyter Mouton.

Sagot, Benoît. 2013. Comparing complexity measures. In *Workshop "Computational Approaches to Morphological Complexity"*, Paris, France: Surrey Morphology Group.

Schramm, Mareile. 2014. *The Emergence of Creole Syllable Structure: A Cross-linguistic Study*. Berlin/Boston: Walter de Gruyter.

Sims-Williams, Helen. 2016. *Analogy in Morphological Change*. PhD dissertation, University of Oxford.

Sims-Williams, Helen, and Matthew Baerman. 2021. A typological perspective on the loss of inflection. In Svenja Kranich and Tine Breban (eds.), *Lost in Change: Causes and Processes in the Loss of Grammatical Elements and Constructions*, (Studies in Language Companion Series 218), 21–49. Amsterdam: John Benjamins. https://benjamins.com/catalog/slcs.218.01sim.

Sinnemäki, Kaius. 2008. Complexity trade-offs in core argument marking. In Matti Miestamo, Kaius Sinnemäki and Fred Karlsson (eds.), *Language Complexity: Typology, Contact, Change*, 67–88. Amsterdam/Philadelphia: John Benjamins.

Sinnemäki, Kaius, and Francesca Di Garbo. 2018. Language structures may adapt to the sociolinguistic environment, but It matters what and how you count: A typological study of verbal and nominal complexity. *Frontiers in Psychology* 9: 1141. doi:10.3389/fpsyg.2018.01141.

Sinnemäki, Kaius, Francesca Di Garbo, Ricardo Napoleão De Souza, and T. Mark Ellison. 2024. A typological approach to language change in contact situations. *Diachronica* 41(3): 379–413. doi:10.1075/dia.23029.sin.

Smith, Andrew D.M. 2014. Models of language evolution and change. *Wiley Interdisciplinary Reviews: Cognitive Science* 5(3): 281–293.

Steels, Luc. 1998. Synthesising the origins of language and meaning using co-evolution, self-organisation and level formation. In James Hurford, Michael Studdert-Kennedy and Chris Knight (eds.), *Approaches to the Evolution of Language*, 384–404. Cambridge, United Kingdom: Cambridge University Press.

Stump, Gregory, and Raphael A. Finkel. 2015. The complexity of inflectional systems. *Linguistics Vanguard* 1(1): 101–117. doi:10.1515/lingvan-2014-1007.

Sulistyono, Y. 2022. *A History of Alorese (Austronesian): Combining Linguistic and Oral History*. PhD dissertation, Leiden University. https://hdl.handle.net/1887/3275052.

Sulistyono, Yunus. 2021. Interpreting oral history from the Alorese people in east Indonesia. *Revista Universidad y Sociedad* 13(4): 339–350.

Thomason, Sarah Grey. 2001. *Language Contact: An Introduction*. Repr. edn. Edinburgh: Edinburgh University Press.

Trudgill, Peter. 2011. *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford, United Kingdom: Oxford University Press.

van Trijp, Remi. 2013. Linguistic assessment criteria for explaining language change: A case study on syncretism in German definite articles. *Language Dynamics and Change* 3(1): 105–132. doi:10.1163/22105832-13030106.

Wedel, Andrew B. 2006. Exemplar models, evolution and language change. *The Linguistic Review* 23(3). doi:10.1515/TLR.2006.010.

Wray, Alison, and George W. Grace. 2007. The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua* 117(3): 543–578. doi:10.1016/j.lingua.2005.05.005.

Wurzel, Wolfgang Ullrich. 2001. Creoles, complexity, and linguistic change. *Linguistic Typology* 5(2/3): 377–387.

Zydorowicz, Paulina. 2010. Consonant clusters across morpheme boundaries: Polish morphonotactic inventory and its acquisition. *Poznań Studies in Contemporary Linguistics* 46(4). doi:10.2478/v10010-010-0028-0.

Appendix A: Boundary phoneme reduction model

A.1 *Model results*

In the boundary phoneme reduction model, instead of dropping the full affix (reported in the main article), only the phoneme of the affix bordering on the stem is removed in case of a consonant sequence. The boundary phoneme mechanism leads to a more gradual reduction, shortening affixes but not fully deleting them, so we now measure the morphological complexity by the average affix length (number of phonemes). Again, this is evaluated only for the L2 speakers.

Fig. 6 shows the results for the boundary phoneme reduction model. In this model, average affix length is used as a proxy for morphological complexity, which cannot be fully compared to the proportion of non-empty affixes used for the full affix reduction model. However, the results show some similarities to the full affix reduction model: without lexical concept generalization, the morphological simplification is small (Fig. 6a). With a generalization probability of 0.1, which gave a larger decrease in morphological complexity in the full affix reduction model, no large decrease is reached in this boundary phoneme model (Fig. 6b). However, in contrast, in this boundary phoneme reduction model, higher generalization values lead to more morphological simplification, so a generalization probability of 0.5 leads to a stronger morphological simplification (Fig. 6c). Although this shortening in terms of affix length is not complete morphological simplification, it could be a stage prior to morphological simplification. Removing phonemes from affixes leads to a higher, but still stable, number of unique affixes than in the full affix reduction model (Fig. 6d). Apparently, this larger variation in suffixes does not lead to more ambiguity, as the proportion of interactions where the concept is correctly inferred (Fig. 6e) stays stable with increasing proportions of L2 speakers.

A.2 *Evaluation of lexical concept generalization probability*

For the full affix phonotactic reduction model presented in the main article, the decrease in morphological complexity (measured in proportion of non-empty affixes) was highest for low generalization probabilities (0.1). The morphological complexity in the boundary phoneme reduction model shows a different pattern, although it cannot be directly compared, because it is measured using a different metric, the average affix length. For the boundary phoneme model (Fig. 7a), the average affix length simplifies with increasing generalization probabilities. The proportion of correct interactions decreases with increasing generalization probabilities (Fig. 7b) like in the full affix reduction model.

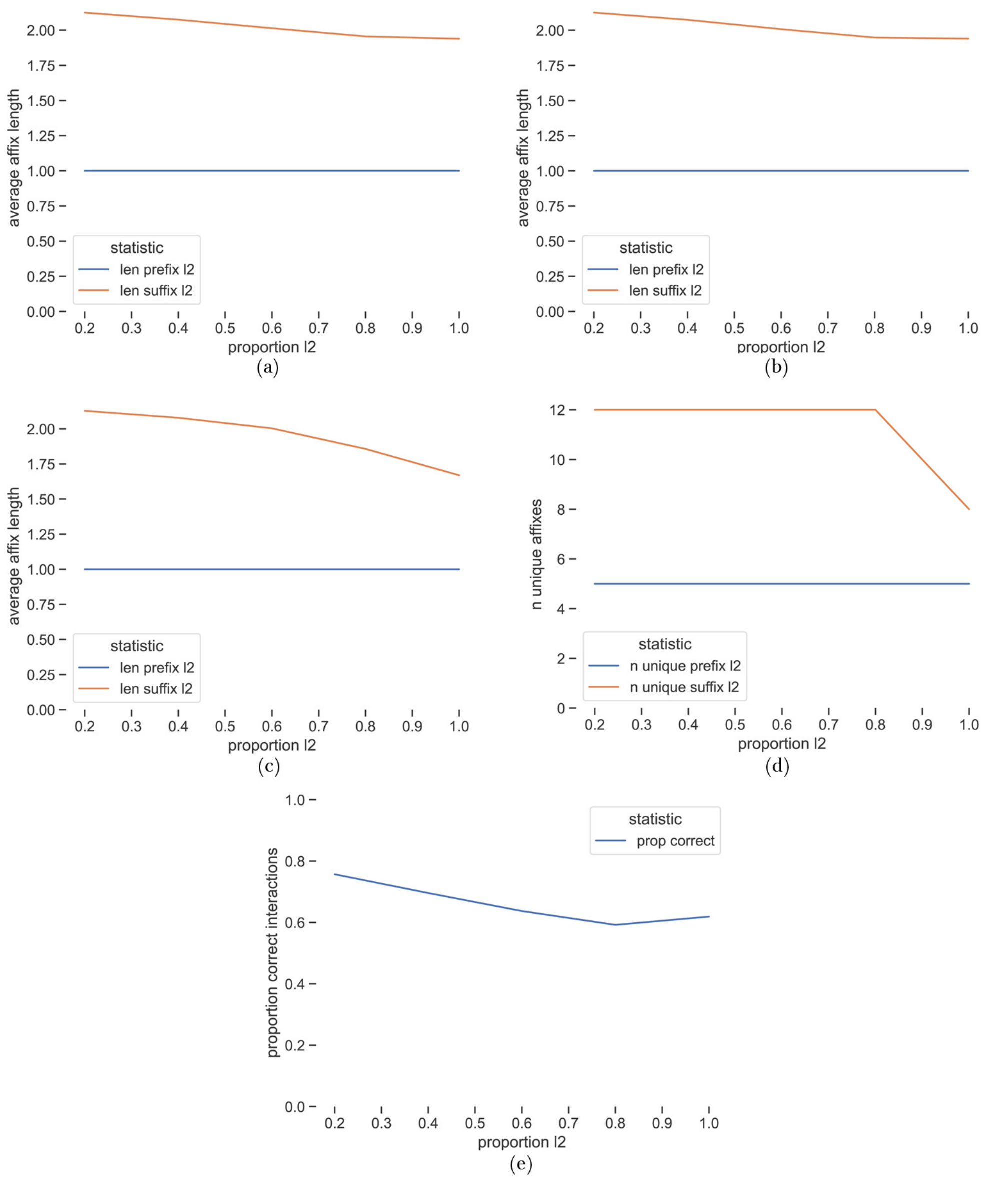
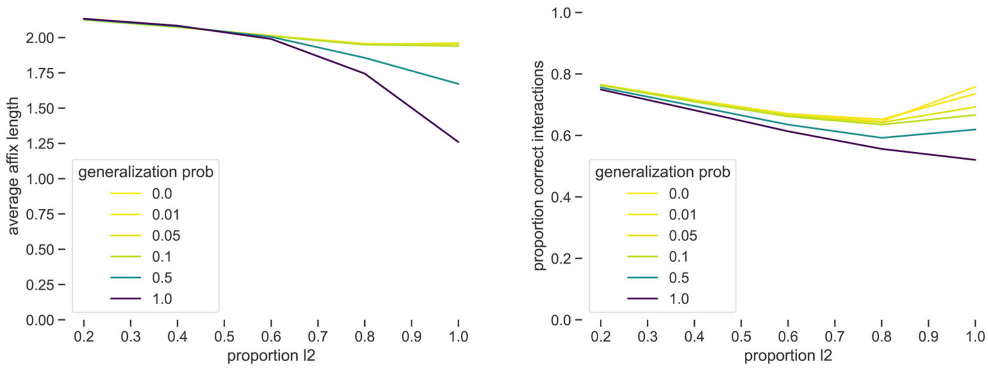


FIGURE 6 Model results with phonotactic reduction of boundary phoneme. Influence of the proportion of L2 speakers on the different affix types (prefix/suffix, L1/L2). Phonotactic reduction probability fixed to 1.0; lexical concept generalization probability of 0.0 in (a), 0.1 in (b) and 0.5 in (c)



(a) (b)

FIGURE 7 Evaluation of different lexical concept generalization probabilities for boundary phoneme reduction model. Phonotactic reduction probability fixed to 1.0

In the full affix reduction model discussed in the main article, the generalization mechanism spreads the empty affix as most common form to all concepts. For the data this model is initialized with (Table 1), the boundary phoneme reduction model does not lead to the formation of empty affixes, so there is no clear candidate for a most common form that is spread by the generalization mechanism.

Appendix B: Evaluations of parameters

B.1 *Phonotactic reduction probability*

In the experiments in the main article, a phonotactic reduction probability of 1.0 is used. Fig. 8 shows how morphological complexity, only for the suffix of L2 speakers, changes with an increasing proportion of L2 speakers, for different values of the phonotactic reduction probability. In these experiments, the generalization probability is fixed to 0.1. The results show that morphological complexity drops faster when the phonotactic reduction probability is higher. This is to be expected, as the phonotactic reduction probability is a force towards simplification.

B.2 *Number of agents*

In the reported experiments, a default of 10 agents per generation are used. Fig. 9 shows how the number of agents per generation influences the morphological complexity in the model. The results show that the model outcomes do not depend crucially on the number of agents

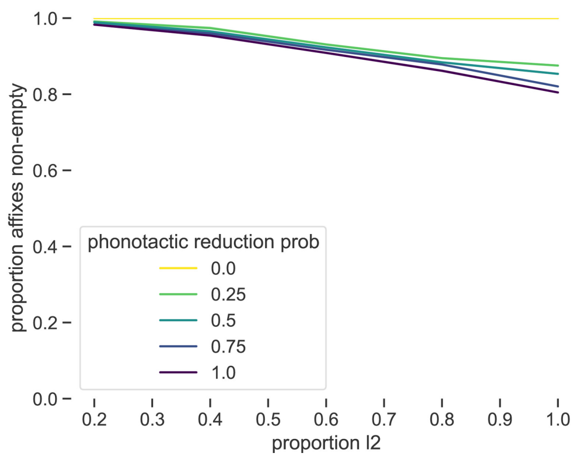


FIGURE 8
Evaluation of morphological complexity (proportion of affixes non-empty, only suffix of L2 speakers) for different phonotactic reduction probabilities and proportions of L2 speakers. Lexical concept generalization probability is fixed to 0.1, full phonotactic reduction

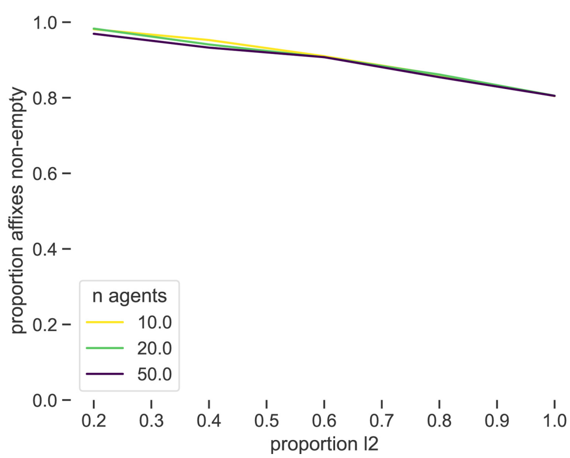


FIGURE 9
Evaluation of morphological complexity (proportion of affixes non-empty) for different numbers of agents. Lexical concept generalization probability is fixed to 0.1 and phonotactic reduction probability 1.0, full phonotactic reduction

B.3 *Number of interactions per generation*

In the main experiments, a default of 10,000 interactions per generation are used. Fig. 10 shows how the number of interactions per generation influences the model outcomes. A number of interactions per generation lower than 10,000 leads to a low communicative success, while from 10,000 and upwards the communicative success is quite comparable (Fig. 10b). Models with fewer interactions give in some cases a larger drop in morphological complexity (Fig. 10a), but this is then caused by the low communicative success. We evaluated the model also over a higher number of generations than in the main experiments: 100 instead of 50. We can see in the graph over generations that the results stabilize after 50 generations (Fig. 10c). 10,000 interactions per generation seems a level at which the communicative success is high and above

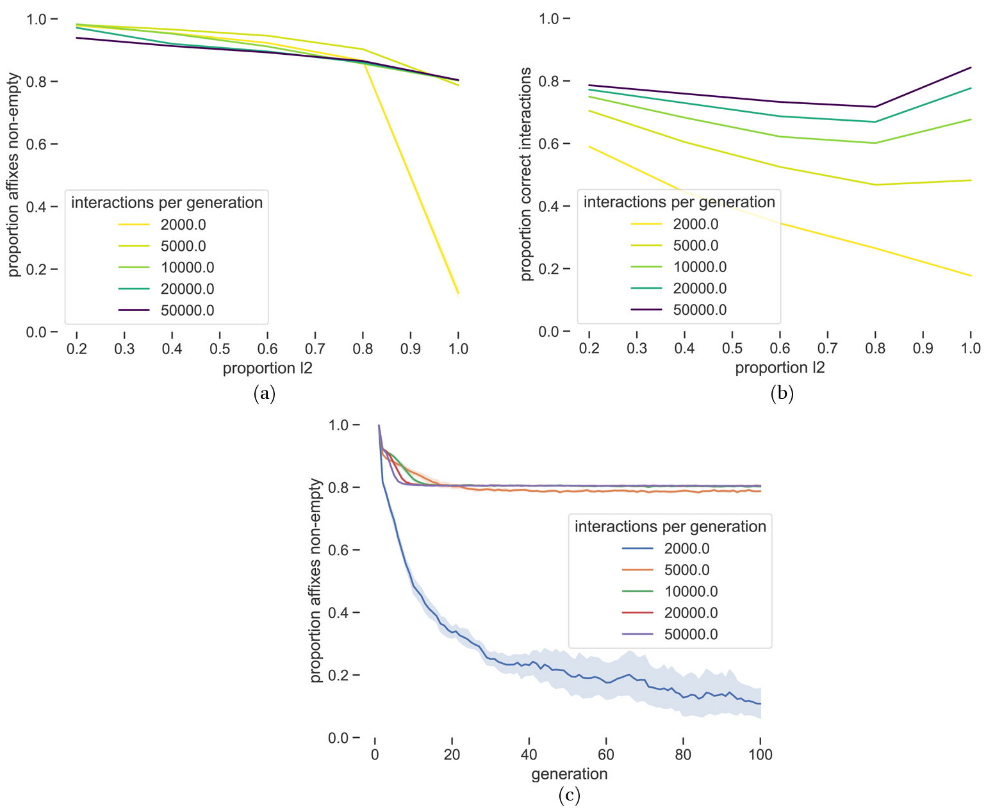


FIGURE 10 Evaluation of suffix L2 for different phonotactic reduction probabilities and proportions of L2 speakers. Lexical concept generalization probability is fixed to 0.1 and phonotactic reduction probability is fixed to 1.0, full phonotactic reduction

which the model results do not qualitatively differ, so we used this for our main simulations.

It is insightful to evaluate the number of interactions per generation in the light of how many interactions take place per paradigm cell during a generation. Every interaction, a random paradigm cell (7 person-number combinations) is chosen for a random verb (56 verbs), which leads to a total of 392 paradigm cells. With 10,000 interactions per generation, on average 10,000 / 392 = 25 interactions take place per cell. For 2,000 interactions per generation, this amounts to only 5 interactions per cell. For these low numbers of interactions per cell, it will be more difficult to obtain stable model behavior, which also shows in the number of successful interactions (Fig. 10a).