



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Is regularization uniform across linguistic levels? Comparing learning and production of unconditioned probabilistic variation in morphology and word order

Citation for published version:

Saldana, C, Smith, K, Kirby, S & Culbertson, J 2021, 'Is regularization uniform across linguistic levels? Comparing learning and production of unconditioned probabilistic variation in morphology and word order', *Language Learning and Development*. <https://doi.org/10.1080/15475441.2021.1876697>

Digital Object Identifier (DOI):

[10.1080/15475441.2021.1876697](https://doi.org/10.1080/15475441.2021.1876697)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Language Learning and Development

Publisher Rights Statement:

This is an Accepted Manuscript of an article published by Taylor & Francis in Language Learning and Development on 19 February 2021, available online:
<https://www.tandfonline.com/doi/full/10.1080/15475441.2021.1876697>

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Is regularisation uniform across linguistic levels? Comparing learning and production of unconditioned probabilistic variation in morphology and word order.

Carmen Saldana*, Kenny Smith, Simon Kirby and Jennifer Culbertson
Centre for Language Evolution, School of Philosophy, Psychology and Language Sciences, The
University of Edinburgh, UK

Abstract

Languages exhibit variation at all linguistic levels, from phonology, to the lexicon, to syntax. Importantly, that variation tends to be (at least partially) conditioned on some aspect of the social or linguistic context. When variation is *unconditioned*, language learners regularise it—removing some or all variants, or conditioning variant use on context. Previous studies using artificial language learning experiments have documented regularising behaviour in learning of lexical, morphological, and syntactic variation. These studies implicitly assume that regularisation reflects uniform mechanisms and processes across linguistic levels. However, studies on natural language learning and pidginisation suggest that morphological and syntactic variation may be treated differently. In particular, there is evidence that morphological variation may be more susceptible to regularisation (Good 2015; Siegel 2006; Slobin 1986). Here we provide the first systematic comparison of the strength of regularisation across these two linguistic levels. In line with previous studies, we find that the presence of a favoured variant *can* induce different degrees of regularisation. However, when input languages are carefully matched—with comparable initial variability, and no variant-specific biases—regularisation can be comparable across morphology and word order. This is the case regardless of whether the task is explicitly communicative. Overall, our findings suggest an overarching regularising mechanism at work, with apparent differences among levels likely due to differences in inherent complexity or variant-specific biases. Differences between production and encoding in our tasks further suggests this overarching mechanism is driven by production.

Keywords: artificial language learning; interaction; word order; morphology; regularisation; variation

1 Introduction

Variation is ubiquitous in language; it can be found at all linguistic levels, from phonology (allophones), through morphology (allomorphs), to the lexicon (synonyms) and syntax (paraphrases)

(Labov 1972). However, linguistic variation *tends* to be at least partially conditioned: the choice of variant is predictable from some aspect of the social or linguistic context. For instance, English exhibits two different realisations of comparative and superlative adjectives: a synthetic realisation via the use of the suffixes *-er* and *-est* respectively (e.g., ‘faster/fastest’), and an analytic realisation via the independent forms *more* and *most* preceding the adjective (e.g., ‘more/most intelligent’). However, this type of morphological variation is highly predictable: monosyllabic and bisyllabic adjectives are very likely to be derived synthetically, whereas longer adjectives are derived analytically. In other cases, variation may be less deterministic, and yet still probabilistically conditioned. For instance, English exhibits probabilistic variation in progressive tense allomorphy, e.g., between [m] or [ɪŋ], but speakers’ choice of variant is partially predictable from social context (Labov 1972; Shuy, Wolfram, & Riley 1967; Trudgill 1974).

Despite the strong tendency for variation to be conditioned, there are some circumstances in which *unconditioned* variation is nevertheless found. For instance, unconditioned variation can occur when new variants are first introduced into an established system, or when input is limited and conventions are still not established, as in the context of language formation (Good 2012; Kouwenberg & Singler 2009; Siegel 1997; Velupillai 2015). Under these circumstances, there is evidence that learners tend to reduce or remove such variation, that is, they regularise the system (Newport 1999; Siegel 1997). This can be achieved either by reducing or removing competing variants, or conditioning variant choice on the context (Ferdinand, Kirby, & Smith 2019; Hudson Kam & Chang 2009; Smith & Wonnacott 2010). Furthermore, regularisation can be seen both in individuals and over time as languages change or become more conventionalised. (Hare & Elman 1995; Kouwenberg & Singler 2009).

Indeed, regularisation (i.e., the reduction, elimination or conditioning of variation) has been documented extensively in natural languages across linguistic levels like phonology, morphology, syntax and the lexicon, including in language acquisition (Fraser, Bellugi, & Brown 1963; Newport 1999; Ross & Newport 1996; Singleton & Newport 2004), in language change (Hare & Elman 1995; Schilling-Estes & Wolfram 1994; van Trijp 2013), and in language formation (Bickerton 2015; DeGraff 1999; McWhorter 2005; Senghas & Coppola 2001; Senghas, Newport, & Supalla 1997; Siegel 2004; Spears 2008; Winford 2003). In addition, experimental studies involving artificial language learning experiments report regularisation behaviour during learning and production of probabilistic variation of diverse linguistic elements, across different linguistic levels: morphology (e.g., Fedzechkina, Newport, & Jaeger 2017; Hudson Kam & Newport 2005, 2009; Perfors 2016; Schumacher, Pierrehumbert, & LaShell 2014; Smith & Wonnacott 2010), word order (e.g., Culbertson, Smolensky, & Legendre 2012; Feher, Ritt, & Smith 2019), and the lexicon (e.g., Ferdinand et al. 2019; Reali & Griffiths 2009). These studies provide evidence for how biases in language learning and use might interact to shape the kind of variation found in natural language.

In a now classic study, Hudson Kam and Newport (2009) used artificial language learning to investigate developmental differences in regularisation behaviour. The artificial language taught to participants contained unconditioned variation in determiner usage. A “main” determiner was used

*Correspondence concerning this article should be addressed to the first author. Carmen Saldana is currently affiliated to the Department of Comparative Language Science at the University of Zurich, Plattenstrasse 54, 8032 Zurich, Switzerland. E-mail: <carmen.saldanagascon@uzh.ch>.

The Version of Record of this manuscript has been published and is available in Language Learning and Development since 19 Feb 2021 at <http://www.tandfonline.com/10.1080/15475441.2021.1876697>.

60% of the time, but what appeared the remainder of the time was manipulated across conditions; either no determiner occurred, or a number of “noise” determiners were used (either 2, 4, 8 or 16). While adult learners were found to roughly match the level of probabilistic variation in the input, children’s productions were more regular and governed by idiosyncratic but invariant patterns. However, adults’ regularisation behaviour was affected by the complexity of the input language: the more noise determiners were present, the more adults regularised by over-producing the main determiner. These results suggest that adult learners are more likely to regularise complex systems of probabilistic variation.

Building on these earlier studies, a number of other researchers have investigated regularisation behaviour in adults, showing that weak biases for regularisation can be amplified over generations of learners in an iterated learning paradigm (Real & Griffiths 2009; Smith & Wonnacott 2010) and that generalisation can increase regularisation (Wonnacott & Newport 2005). Further, recent studies have suggested that regularisation can also be modulated by the nature of specific structures to be learned. For example, Culbertson et al. (2012) and Culbertson and Newport (2015) show that learners regularise harmonic word order patterns (i.e. which are either consistently head-initial or head-final) more than non-harmonic patterns. Taken together, these experiments suggest a link between biases active during learning and the scarcity of unconditioned variation in language: regularisation at the population level is assumed to reflect biases in language learning and/or use at the individual level. While we will return to the precise nature of the bias in section 6.1.3, one possibility is that regularisation is driven by a bias for simplicity. Regularisation via reduction of variants or conditioning results in simpler, more predictable systems (Culbertson & Kirby 2016). This is in line with findings suggesting that regularisation is modulated by the type of input learners receive. A learner’s response to linguistic variation depends on the nature of the structures in question (Culbertson & Newport 2015; Culbertson et al. 2012) and the complexity of the input learners receive (Hudson Kam & Newport 2009).

1.1 Level-specific effects on regularisation

Despite the extensive literature investigating regularisation, relatively little is known about whether and how regularisation might *differ* across linguistic levels and units. Laboratory experiments on regularisation have investigated distinct linguistic phenomena in a number of different domains. Most of these studies implicitly assume uniform mechanisms and processes behind regularisation; no study has directly compared how learners respond to variation at different levels. However, Fedzechkina et al. (2017) indirectly show that given unsystematic input with variable word order and variable case marking, participants tend to regularise case marking systems (by conditioning object marking on word order) but maintain the variability in word order. Although these results do not provide a direct comparison across levels, they suggest nonetheless that morphological systems might be more prone to regularisation than word order when variability at both levels coexists, maybe because participants consider morphology a more informative source than word order and thus tolerate less uncertainty. Further, research on natural language learning and formation suggests that morphological and syntactic variation may not be treated in the same way by learners (Good 2015; Siegel 2006; Slobin 1986).

Research on both first and second language acquisition points to differences in the ease of learning of syntax and morphology, and in some cases there is evidence that variation in morphology is levelled by learners, while variation in syntax is maintained. In the L1 and L2 acquisition literature, it is widely reported that target-like word order precedes target-like morphology. Chil-

dren produce more morphological than word order errors in their native language (Bichakjian 1988; Slobin 1966). Similarly, second language learners make more (and more persistent) errors in inflectional morphology compared to word order, which are often argued to result from early L1 transfer that is quickly corrected (Dietrich, Klein, & Noyau 1995; Montrul 2004; Siegel 2006). The presence of errors alone does not necessarily show that learners regularise morphology more than word order, however it is notable that morphological errors are often reported to involve over-regularisation (Marcus et al. 1992), or in the case of second language learners, omissions of morphological marking altogether (e.g., Prévost & White 2000). Importantly, studies looking specifically at the acquisition of variation (typically conditioned variation) indicate that children's production of variable morphology (e.g., a variably used plural marker) is less likely to be target-like and/or is more prone to regularisation (Miller & Schmitt 2012; Raymond, Healy, McDonnell, & Healy 2009) than variable syntactic features (e.g., object shift or topicalisation, Anderssen, Bentzen, & Westergaard 2010). Similar findings have also been reported in cases of language attrition; Håkansson (1995) observed that whereas expatriate bilingual Swedes produce native-like variation in word order patterns (i.e., without V2 errors), they failed to reproduce NP-internal inflectional morphology (i.e., determiner and adjective inflection for gender, number and definitiveness).

Interesting parallels can be seen in research on pidgin/creole formation. For example, compared to source languages, it is common for early pidgins to initially show more impoverished, less variable inflectional and derivational morphology; by contrast, several studies indicate that word orders tend to be comparably variable (Drechsel 1981; Nichols 1995; Stefansson 1909; Thomason 1980; van der Voort 2013). Some of the clearest cases are the Eskimo pidgins. These languages show variability in word order (van der Voort 2013): although SOV (the default order in the lexifier West Greenlandic) is prevalent, SVO (the default order in the other source languages, i.e., English, Danish, French and Russian) and VSO order are also used. Similarly, they use both adjective-noun and noun-adjective orders, a mixture of those found in their source languages (van der Voort 1997). Conversely, we find that NP-internal inflectional morphology in source languages does not survive in Eskimo pidgins. For example, possessor inflection found in West Greenlandic (and often in other source languages, e.g., English) is omitted, leaving the possessor and the thing possessed in bare juxtaposition (van der Voort 1997, 2013). To summarise, it has been claimed that morphological complexity is generally reduced in pidgins/creoles relative to their source languages, while syntactic complexity is more stable (Bakker 2008; Good 2015). Under the assumption that simplification processes in language formation—as well as in language change more generally (Bichakjian 1988; Slobin 1986)—reflect individual biases during learning (Lefebvre, White, & Jourdan 2006; Siegel 2006) and/or use (McWhorter 2001; Parkvall 2008), the early loss of inflectional morphology and maintenance of word order variation in pidginisation suggest the possibility of a level-specific regularisation bias.

While this evidence is suggestive, it remains an open question whether regularisation applies with uniform strength across linguistic levels or not, and to what extent this is driven by level-specific regularisation biases in language learning and use. There are alternative explanations for the suggested differences in regularisation across levels which do not necessarily require level-specific biases. First, as discussed above, there is reason to believe that the complexity of a system (e.g., how much variation it encodes) influences regularisation behaviour (Hudson Kam & Chang 2009; Hudson Kam & Newport 2009). Because we do not have any formal measure of complexity of the morphological and syntactic phenomena cited above, we cannot rule out the possibility that morphological variation simply tends to be more complex than variation in word

order. If that is the case, then differences in regularisation or simplification could be attributed to differences in initial complexity. Secondly, differences between linguistic levels might be driven by specific morphological or syntactic features of the input languages rather than general factors which differentiate morphology and syntax; language acquisition and formation are influenced by the particular variants available to the user (Mufwene 2008; Siegel 2006). Early in L2 acquisition, for example, language learners tend to favour variants which match their prior linguistic knowledge more closely. Consequently, the amount of overlap between features across source languages could also determine the preservation of a given grammatical feature during language formation (Ansaldò, Matthews, & Lim 2007). If morphological paradigms are more complex than word order, then full morphological paradigms will necessarily overlap less than word order systems. It is also less probable that *all* variants in more complex systems will be present in the data available to the learner or that they will all be salient enough to be reproduced (Mufwene 2008). Thus, while evidence from natural language acquisition and language formation point to the intriguing possibility that learners might selectively target morphological variation for regularisation, there are a number of alternative explanations which we cannot rule out on the basis of this data alone.¹

Here, we use a series of artificial language learning experiments to systematically compare the strength of regularisation across the two linguistic levels discussed—morphology and word order—in order to test the extent to which level-specific biases in regularisation exist. Coherent with the literature in language acquisition and formation summarised above, we focus on regularisation defined as the reduction or elimination of variants. Because previous literature suggests that complexity affects regularisation behaviour, we carefully control the relative complexity of the input systems (in terms of variability). To preview: in Experiment 1, we find that the presence of a favoured variant at one level (here, word order)—because that variant is already available to the learner from their L1—can lead to stronger regularisation and make it look as though a level-specific bias exists, suggesting that the pool of variants and its overlap with the learner’s prior linguistic knowledge plays a crucial role in determining the strength of regularisation. However, in Experiment 1b, we show that when input languages have comparable initial variability, and no specific biases targeting particular variants, regularisation behaviour is comparable across morphology and word order. Finally, in Experiment 2, we introduce communication into our experimental task in order to test whether regularisation through alignment (Feher et al. 2019; Fehér, Wonnacott, & Smith 2016) targets linguistic levels differently. We find that it does not. Overall, our results suggest a single regularising mechanism at work, with apparent differences among levels likely due to differences in inherent complexity or variant-specific biases.

2 Experiment 1: regularisation behaviour across linguistic levels in individual production

In Experiment 1, we explore regularisation across linguistic levels using the methodology developed in Culbertson et al. (2012) (following Hudson Kam & Newport 2005, 2009). Adult learners are exposed to a miniature artificial language featuring an inconsistent mixture of synonymous variants, whose use is determined probabilistically, with no conditioning factors. Crucially, we construct input languages which differ *only* in whether variation is in word order or in morphology. The languages are comparable in their inherent complexity—they have the same initial amount of variability. This allows us to ask whether regularisation behaviour is similar across linguistic levels

¹It is also worth mentioning that morphological and syntactic phenomena are not always clearly differentiated, particularly in the context of language acquisition and change.

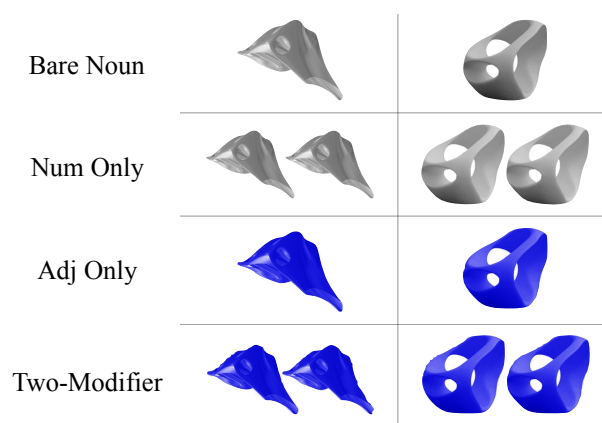


Figure 1. Visual stimuli in Experiment 1.

when complexity is held constant.

2.1 Materials and Methods

2.1.1 Participants. Fifty-two native English speakers (age 19 to 41, mean 23.3) were recruited from the University of Edinburgh’s Careers Services database. Participants were paid £6 for a 30-min long experimental session. Twenty-six participants were assigned to the Morphology condition and 26 to the Word Order condition. The data from a further four participants was excluded on the basis of failure to adequately learn the language. This was determined based on a pre-set limit on incorrect responses in the production testing phase.²

2.1.2 Input languages. The input languages were used to describe simple pictures featuring one of two objects.³ Objects appeared either singly or in a pair and could appear either in grey-scale or coloured in blue. All eight pictures are shown in Figure 1 (stims from Culbertson et al. 2012). When the objects appeared singly and in grey-scale, they were considered bare (un-modified) objects, and were accordingly described by a bare Noun. The lexicons for the two languages are shown in Table 1; they were the same for all participants within a given condition. Descriptions for objects in pairs and/or in colour were phrases composed of a N(oun) plus a Num(eral) and/or Adj(ective) modifier (see Table 2).

The two experimental conditions differed in the type of unconditioned probabilistic variation they involved. The respective probabilistic grammars for these two conditions are shown in Table 2 in more detail. In the Word Order condition, the two modifiers (Adj and Num) could appear either pre- or postnominally. Note that the majority variant is harmonic (i.e., both modifiers tend to appear postnominally) and all variants are *isomorphic* (i.e., linearisations which are compatible with the underlying structure [Num [Adj [N]]]) since these orders are cross-linguistically common and were found to be preferred by learners in Culbertson and Adger (2014). In the Morphology

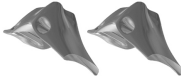

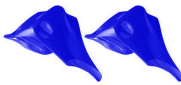
²Similarity between a correct answer in the input language and the response provided on each trial was calculated using normalised Damerau-Levenshtein edit distance (Damerau 1964; Levenshtein 1966). We excluded participants’ data with an average distance of more than two edits (i.e., typos) per response (excluding bare nouns), or with greater than 20% of descriptions in which a word was omitted entirely or inserted (i.e., descriptions which had less or more words than required).

³The languages were designed to be small so as to minimise the effort of learning the lexicon and keep the experiment to under an hour when extended to include dyadic communication in Experiment 2.

Table 1
Lexicon for the two conditions in Experiment 1.

	MORPHOLOGY CONDITION	WORD ORDER CONDITION
Noun	‘mokte’, ‘jelpa’	‘mokte’, ‘jelpa’
Num	‘nefri’, ‘nezno’	‘nefri’
Adj	‘kogla’, ‘kospu’	‘kogla’

Table 2
Probabilistic input languages in the Morphology and Word order conditions of Experiment 1. Probabilities of occurrence for each variant of one- and two-modifier phrases are given to the left of the variant. Majority variants are highlighted in bold.

PHRASE TYPE	MORPHOLOGY CONDITION	WORD ORDER CONDITION
NUM ONLY 	0.6 $NP \rightarrow N \mathbf{Num}_1$	0.6 $NP \rightarrow N \mathbf{Num}$
	0.4 $NP \rightarrow N \text{Num}_2$	0.4 $NP \rightarrow \text{Num} N$
ADJ ONLY 	0.6 $NP \rightarrow N \mathbf{Adj}_1$	0.6 $NP \rightarrow N \mathbf{Adj}$
	0.4 $NP \rightarrow N \text{Adj}_2$	0.4 $NP \rightarrow \text{Adj} N$
TWO-MODIFIER 	0.6 $NP \rightarrow N \mathbf{Adj}_1 \mathbf{Num}_1$	0.6 $NP \rightarrow N \mathbf{Adj} \mathbf{Num}$
	0.1 $\bar{3}$ $NP \rightarrow N \text{Adj}_1 \text{Num}_2$	0.1 $\bar{3}$ $NP \rightarrow \text{Num} \text{Adj} N$
	0.1 $\bar{3}$ $NP \rightarrow N \text{Adj}_2 \text{Num}_1$	0.1 $\bar{3}$ $NP \rightarrow \text{Num} N \text{Adj}$
	0.1 $\bar{3}$ $NP \rightarrow N \text{Adj}_2 \text{Num}_2$	0.1 $\bar{3}$ $NP \rightarrow \text{Adj} N \text{Num}$

condition, Adj and Num modifiers were consistently postnominal, but each modifier had two distinct but related variants. These variants differed only in their second syllable⁴, again conforming to a cross-linguistic preference, in this case for suffixes, which has been found to correlate with the biases of English-speaking participants in the lab (e.g., Hupp, Sloutsky, & Culicover 2009; but cf. Martin & Culbertson in press).⁵

Importantly, although the type of unconditioned variation in languages differs across experimental conditions (morphology vs. word order), system complexity is comparable. We assess a

⁴This type of morphological variation could result from languages with rich inflectional morphology within noun phrases, such as Swedish where adjectives and determiners agree in gender and number with the head noun: e.g. *en grön stol* (‘a green chair’) and *min gröna stol* (‘my green chair’), or *mitt gröna bord* (‘my green table’) and *mina gröna bord* (‘my green tables’).

⁵Hupp et al. (2009) find that speakers of a suffixal language such as English are more likely to treat two words as referring to the same referent if they differ in their endings, rather than in their beginnings. Further, Martin and Culbertson (in press) shows that this preference is not present in speakers of prefixal languages such as Kitharaka.

language’s system complexity by its overall variability, computed using Shannon entropy; higher entropy corresponds to more variability, while lower entropy corresponds to more regularity. Entropy (H) of variant use in a given set of productions is given by

$$H(X) = - \sum_{i=1}^n P(x_i) \log_2 P(x_i), \quad (1)$$

where the sum is over the different variants, and $P(x_i)$ is the relative frequency of variant x_i in the set of productions, X . Take for example a hypothetical language which is completely regular, with only one variant per phrase type (e.g., a single order used for Num Only, Adj Only, and two-modifier phrases). Each phrase type would have an entropy of 0 bits. The overall language entropy will depend on the total number of phrases comprised: if the language contains 30 instances of two-modifier phrases, 10 instances of Adj Only and 10 instances of Num Only, the overall entropy of this completely regular language would be 1.37 bits. For the probabilistic grammars described in Table 2, the entropy is higher. The specific input language in the Word Order condition will also contain 30 instances of two-modifier phrases (i.e., 18 *N Adj Num*, 4 *Num Adj N*, 4 *Num N Adj* and 4 *Adj N Num*), 10 instances of Num Only (i.e., 6 *N Num* and 4 *Num N*) and 10 Adj Only (i.e., 6 *N Adj* and 4 *Adj N*). However, in this case, Num Only and Adj Only phrases will have an entropy of 0.97 each, and two-modifier phrases will have an entropy of 1.67 bits. The input language has an overall entropy of 2.72 bits. Given the same number of instances of each phrase type, the input language in the Morphology condition results in the same exact entropy. This measure of a system’s complexity only assesses the amount of different ways of describing the same thing and allows us to provide a (rough but) theory-neutral comparison—not only of complexity but of regularisation as the reduction of complexity—without the need to take into account specific features of morphological and word order systems which would hinder any cross-level direct comparison.

2.1.3 Procedure. Participants were taught a miniature artificial language through a custom-written program developed in Python 2.7 using the PsychoPy (Peirce 2007, 2009) and Pygame libraries (Pygame Community 2009). Each participant was trained and tested over a single 30-min-long experimental session. Participants sat in a sound-attenuated booth in front of a computer display and wore a headset through which the audio was played. All responses were entered using the keyboard. Before participants started the experiment, they were given detailed instructions by the experimenter on how to proceed during the experiment. Throughout the experiment, they were also given instructions by a “native speaker” of the artificial language, whose speech was synthetically generated using Apple’s text-to-speech software (OS version 10.10.3, speaker “Victoria”). After participants finished the experiment, they were asked to complete a grammaticality judgement task and to fill in an additional questionnaire (described below), which was displayed on the monitor via a web browser. The results from the latter tasks will be analysed separately in section 6.

Phase 1, noun familiarisation. Participants were first trained on the two nouns and their associated objects during a block of 10 trials in which each noun appeared five times (in random order). The first six were simple exposure trials, and the remaining four were picture-selection comprehension trials. In exposure trials (here and throughout), participants were presented with a picture accompanied by its description, displayed both visually and auditorily. In noun familiarisation, the picture was always a single object in grey-scale, and its description was a bare noun. The picture appeared first by itself (1 second), then the text and audio were presented. After the audio finished, the text and the picture remained on display for 2 seconds. Participants were instructed

to repeat the descriptions out loud. In picture-selection trials, participants were prompted with a description (again presented visually and auditorily) and asked to select the corresponding picture (out of an array of four). The foils for the discrimination array in noun familiarisation consisted of the other object they had been trained on and two distractor objects (all in grey-scale)⁶. If the correct picture was selected, a correct-answer sound effect was played along with a display of the correct picture and its text description; if an incorrect picture was selected, a wrong-answer sound effect was played along with a display of the correct picture and its description (in this case presented visually and auditorily).

Phase 2, one-modifier training. In the second phase, participants were trained on noun phrases with a single modifier (either Adj or Num), here referred to as *one-modifier phrases*—which comprise *Adj Only* and *Num Only* phrases. Pictures corresponding to these phrases contained either a single object coloured blue or a pair of two grey-scale objects (see Figure 1). For each picture, a description was selected probabilistically from the grammar corresponding to the participant’s condition. Recall that both grammars contained a majority variant occurring with an empirical probability $P = 0.6$ and a minority variant with an empirical probability $P = 0.4$ for each modifier type (see Table 2).

This phase included 40 trials in total, divided into two blocks of 20 trials; each block consisted of 15 exposure trials followed by five picture-selection trials. Participants saw each of the four one-modifier pictures five times per block (order randomised). The discrimination arrays for the picture-selection trials depended on the modifier category. If the target phrase included an Adj, the array contained each of the two objects in blue and grey-scale, all singly; if the target phrase included a Num, the array contained each of the two objects singly and in pairs, all in grey-scale.

Phase 3, one-modifier testing. In the third phase, participants were tested on their knowledge of one-modifier phrases in the language. They saw pictures as in the previous phase without accompanying text or audio and were asked to type in a description in the language. Participants described 20 pictures in total (each of the four different pictures five times, in random order).

Phase 4, full training. In the fourth phase, participants were trained on a mix of one-modifier and two-modifier phrases (i.e., Noun plus both Num and Adj). Two-modifier phrases were used to describe pairs of blue objects. For each picture, a description was selected according to the grammar corresponding to the participant’s condition. Recall that grammars contained a majority two-modifier phrase variant, and the three equally-frequent minority two-modifier phrase variants (see Table 2).

This stage comprised 100 trials (20 Num Only, 20 Adj Only and 60 two-modifier phrases), divided into four blocks of 25 (15 exposure trials followed by 10 picture-selection trials). Participants saw each one-modifier picture 10 times and each two-modifier picture 30 times (order randomised). In this phase, the discrimination arrays in the picture-selection trials were of two types. In the first type, the array contained the four different possible pictures featuring the target object (i.e., grey-scale single object, blue single object, pair of grey-scale objects, pair of blue objects). In the second type, the array contained the four different possible pictures with the target modifier (if Adj, the bottom two rows in Figure 1; if Num, the top two rows in Figure 1).

Phase 5, full testing. In the final phase participants were tested on their ability to produce all phrase types in the language. They saw all eight different pictures they had been trained on, and were asked to type in corresponding descriptions. They had to describe 52 pictures in total (in

⁶These are not included in Figure 1.

random order): 20 one-modifier pictures (10 Adj Only and 10 Num Only; five per object in each), 30 two-modifier pictures (15 per object) and the two bare object pictures (one each).

Grammaticality judgement. After participants completed all training and testing phases, they were asked to provide grammaticality judgements for phrases in the language. Participants saw a picture and a description and were asked to accept or reject the description as part of the language. They saw 24 picture-phrasal pairs, 12 one-modifier trials (six Num Only and six Adj Only) and 12 two-modifier trials. One third of the trials featured violations (i.e., variants absent from the input language), the remainder featured grammatical variants in the language.

Participants in each condition saw the 16 variants they were trained on (eight variants for each noun, see Table 2), plus eight ungrammatical phrases containing word order or morphological violations. In the Morphology condition, all ungrammatical one-modifier phrases consisted of word order violations: they featured prenominal modification (e.g., *nezno jelpa*). In the Word Order condition, ungrammatical one-modifier phrases consisted of morphological violations: modifiers featured different suffixes (e.g., *nezno* instead of *nefri*). Across conditions, half of the ungrammatical two-modifier phrases contained word order violations and the other half, morphological violations. For example, word order violations could consist of unseen prenominal xmodifier orders (e.g., *jelpa nefri kogla*, N Num Adj, or *kogla nefri mokte*, Adj Num N); morphological violations again consisted of unseen modifier variants (in grammatical word order).

Post-experimental questionnaire. After participants completed the experiment and the grammaticality judgement task, they were asked to fill in a questionnaire (displayed on the monitor in a web browser). Participants were asked to estimate the relative frequencies of the majority variants in the input languages. They did so by selecting from a scale of 0% to 100% with 10% intervals.

2.2 Analyses and results

2.2.1 Confirmatory analysis: Does the strength of regularisation vary between linguistic levels? We assess regularity within a set of productions using Shannon entropy as described in section 2.1.2, equation 1. Entropy measures the variability of a given participant's productions; higher entropy corresponds to more variability, while lower entropy corresponds to more regularity. We then assess each participant's regularisation in production by computing the drop in entropy between the input and the output systems. Using entropy drop instead of simply comparing the proportion of productions using the majority input variant allows us to capture regularisation behaviour more robustly (Ferdinand et al. 2019). If half of the participants produced one variant categorically and the other half produced another variant for the same meaning, an average-proportion analysis would not pick up that participants are indeed regularising their input languages because the average proportion of one variant would be 0.5. In the same situation, entropy drop would be maximal, correctly capturing regularisation behaviour.

Figure 2 shows the entropy of participants' productions for both the Morphology and Word Order conditions. Analyses are run exclusively on participants' last set of production responses (phase 5, see section 2.1.3).⁷ Lexical items were corrected for typos before analysis.⁸ N1 ('mokte')

⁷Individual trials in which words were inserted or omitted were also excluded from analysis: the mean proportion of excluded trials (out of 50) per participant was 0.014 ($SD = 0.042$) and 0.01 ($SD = 0.029$) for the Morphology and Word Order conditions respectively.

⁸ Typos were generally corrected to the closest vocabulary item, that is, the vocabulary item with the lower Damerau-Levenshtein's distance (Damerau 1964; Levenshtein 1966); if there was not a single closest vocabulary item, they were

and N2 ('jelpa') were treated as the same Noun when the entropy of the phrases was calculated, thus no variability was introduced by the correct use of the different nouns.⁹ An entropy of 0 corresponds to a set of productions that only contains one variant and therefore no variability whatsoever. Entropy lower and upper bounds vary according to the number of required and possible variants as well as to the number of production trials (see Figure 2).¹⁰ Note that whereas the input entropy and the output entropy lower bound is held constant, output entropy upper bounds will increase if participants introduce additional variants.

Figure 2 shows entropy for the different phrase types separately: one-modifier Num (Num Only), one-modifier Adj (Adj Only), and two-modifier phrases. A visual inspection of the output entropy across the Morphology and Word Order conditions suggests that many participants did not reproduce the full variability of the input languages; their productions are generally more regular than the input (i.e., entropy is lower). Moreover, the mean output entropies in Figure 2 hint at higher regularity in the Word Order condition.

We used the *stats* and *lme4* packages developed in R (Bates, Mächler, Bolker, & Walker 2015; R Core Team 2015) to run a linear mixed-effects model predicting regularisation behaviour. We used entropy drop (difference between input and output entropy) as our dependent variable; for each participant, we have three data points, that is, the entropy drops for Num Only, Adj Only and two-modifier NPs. As fixed effects, we included Phrase Type (reverse Helmert coded with 3 ordered levels: Num Only, Adj Only and two-modifier)¹¹, Condition (reverse Helmert coded, Word Order vs Morphology)¹², and the interaction between them. In addition to these fixed effects, random

not corrected. More specifically, we corrected one-off misspellings with the correct initial syllable and final vowel (e.g., *nepli* instead of *nefri* or *kolpra* instead of *kogla*) or the systematic misspelling of lexical items (e.g., consistent use of *kolga* or *korpa* instead of *kogla*). Note that if the participants produced two variants (more than once each) that could be corrected to the same vocabulary item (e.g., *konga* and *kolga*), only one variant was corrected (i.e., the variant with the lowest edit distance, or the most frequent one otherwise); in cases where one of these variants was the target one, the others were not corrected (e.g., *konga* would not be corrected if *kogla* was also used more than once). However, we also allowed for innovations by participants. For example, in the Morphology condition, we retained additional variants introduced by a given participant that did not fall within the aforementioned categories of typo. Similarly, in the Word Order condition, production of two-modifier phrase word orders which were not present in the input were not corrected. These additional variations introduced by participants could therefore increase their entropy scores beyond the maximal entropy of the distribution of variants in the input language. After correction, the mean proportion of trials with innovative lexical items and word orders per participant in the Morphology and Word Order conditions respectively were 0.042 ($SD = 0.104$) and 0.049 ($SD = 0.108$), which amount to an average of approximately two innovative trials (out of 50) per participant; note that we only count those innovations that increase the number of variants participants produce beyond the input number because these are the only relevant cases for the entropy measure.

⁹We previously ruled out the possibility that participants were producing conditioned variation: the mutual information between the different variants and different nouns was 0, which means that there exists no relation between the variant that was used and the noun it was used with.

¹⁰The most regular language which is still expressive (i.e., contains a unique description for each picture) would consist of three different variants, one Num Only (e.g. *N nefri*), one Adj Only (e.g. *N kogla*) and one two-modifier (e.g. *N kogla nefri*). As the most regular system would contain a single variant per phrase type, the minimum entropy for the set of productions for a given phrase type considered individually is 0. The final production phase consisted of 50 trials (excluding the 2 Noun trials), divided up into 20 one-modifier trials (half Num and half Adj) and 30 two-modifier trials: the entropy lower bound for the *overall* language (i.e. not treating each phrase type separately) is thus 1.37 bits (represented as a solid vertical line in Figure 2). The *overall* input entropy for the same number of trials would be 2.72 bits (represented as a dotted vertical line in Figure 2).

¹¹Reverse Helmert coding compares levels of a variable with the mean of the previous levels of the variable, intercept being the grand mean. It allows us compare the means of the one-modifier phrases to each other and the mean of the two-modifier phrases against the mean of those.

¹²Using reverse Helmert coding with two levels is equivalent to using simple contrast coding, which compares each

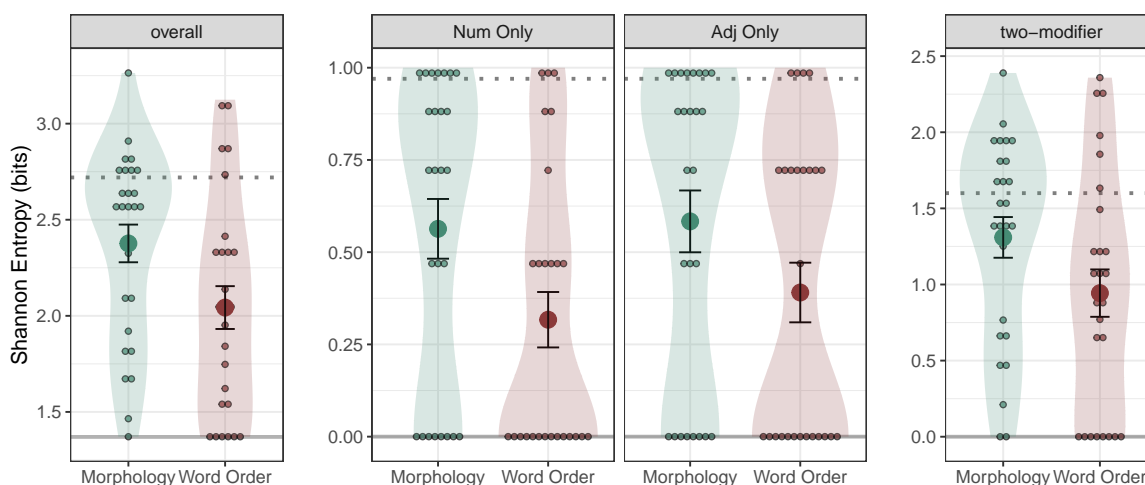


Figure 2. Entropy scores of participants' output productions in Experiment 1 for the Morphology (green) and Word Order (red) conditions. From left to right, entropies of participants' complete production sets as well as entropies for one modifier Num (Num Only), one-modifier Adj (Adj Only) and two modifier (two-Mod) phrases. The small dots represent the individual entropy scores and the big dots represent the mean. The bars around the mean represent standard errors. The dashed lines indicate input entropy; solid lines are minimum entropy scores (0 for each individual phrase type and 1.37 for the overall system with three phrase types).

intercepts for Subject were also included.¹³ P-values were obtained through the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen 2015), using Satterthwaite approximation (Satterthwaite 1946). Results reveal a significant intercept ($\beta = 0.488, SE = 0.056, p < 0.001$), showing that the difference between input and output is significantly different from 0 and thus that learners are regularising input variability in their productions. We also found a significant effect of Condition ($\beta = 0.142, SE = 0.056, p = 0.015$), showing that participants regularise their input significantly more in the Word Order condition. We did not find any significant effect of Phrase Type or its interaction with Condition (all $|p| > 0.413$). In sum, these results show that participants regularised input languages across conditions and phrase types, and that participants in the Word Order condition regularised their input *slightly* more than those in the Morphology condition on average.

2.2.2 Exploratory analysis: How are learners regularising input variation? The previous analysis shows that participants regularised input variability in both conditions, but more so in the Word Order condition. However, this reduction in variability could be due to over-production of majority input variants or it could reflect over-production of minority variants. Figure 3 shows the proportion of majority input variants produced by participants across conditions. Unlike in the Morphology condition, all distributions in the Word Order condition are bimodal, with modes of the distributions of majority variant use at $P \leq 0.1$ and $P > 0.9$ across phrase types (and more so in two-modifier phrases), pointing to opposite trends amongst participants: one towards the over-production of the majority input word order variants and another towards under-production of ma-

level to the reference level but unlike in treatment coding, the intercept is the grand mean.

¹³A model including by-Subject random slopes for Phrase Type given the parameters and power we have is over-fitted and thus we report the simpler model.

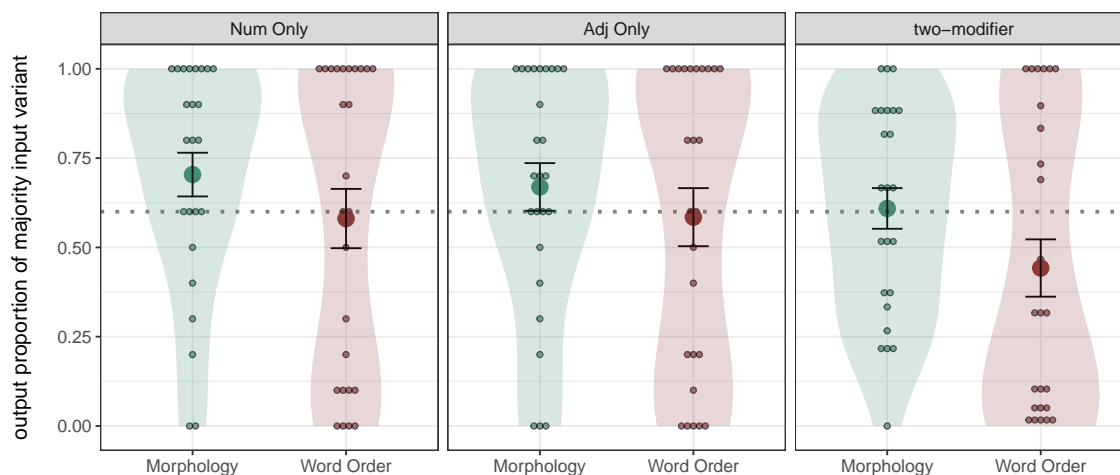


Figure 3. Proportion of the majority input variant used for each of the phrase types in the Morphology (green) and Word Order (red) conditions of Experiment 1. Small dots represent the individual proportions and the big dots their mean. Error bars represent the standard error. The dotted horizontal lines indicate the input proportions.

majority variants.

We ran a logistic mixed-effects model to evaluate the proportion of the majority input variant in the output across conditions. Responses in production were coded according to the presence or absence of the majority input variants in each trial. The fixed effects structure was the same as in the previous model, and we included random intercepts for Subject. We found a significant effect of the two-modifier Phrase Type ($\beta = -0.305, SE = 0.039, p < 0.001$) as well as a significant interaction between two-modifier Phrase Type and Word Order ($\beta = -0.160, SE = 0.039, p < 0.001$), showing that participants across conditions produced majority input variants less frequently for two-modifier phrases than for one-modifier phrases, and that participants in the Word Order condition produced majority variants significantly less frequently than participants in the Morphology condition. We calculated the z-scores of the model's estimates (with the input log-odds as the mean) and their p-values to estimate the deviation from probability matching behaviour for input majority variants: results show that participant behaviour is not significantly different from probability matching on average for input majority variants ($p = 0.315$), but that participants in the Word Order condition significantly under-produce them ($p = 0.001$). Participants under-producing the majority word order variant in *one-modifier* phrases are necessarily over-producing modifiers preminally, but what are participants producing for *two-modifier* phrases? Figure 4 shows us the overall proportions of the variants produced for two-modifier phrases. The word order produced the most is the majority input variant N Adj Num, but it is regularised only by a minority of participants. Although the three remaining word order variants were equally frequent in the input language, the opposite harmonic order Num Adj N is produced more frequently by participants (though again, only by a minority as indicated by the median value 0). Interestingly, despite the fact that these two orders—N Adj Num and Num Adj N—appear to be most over-used, only 30% of participants produced them *both* (and only 19% produced both variants more than once). In other words, most participants tended not to mix these two orders in their productions. These results thus support (or

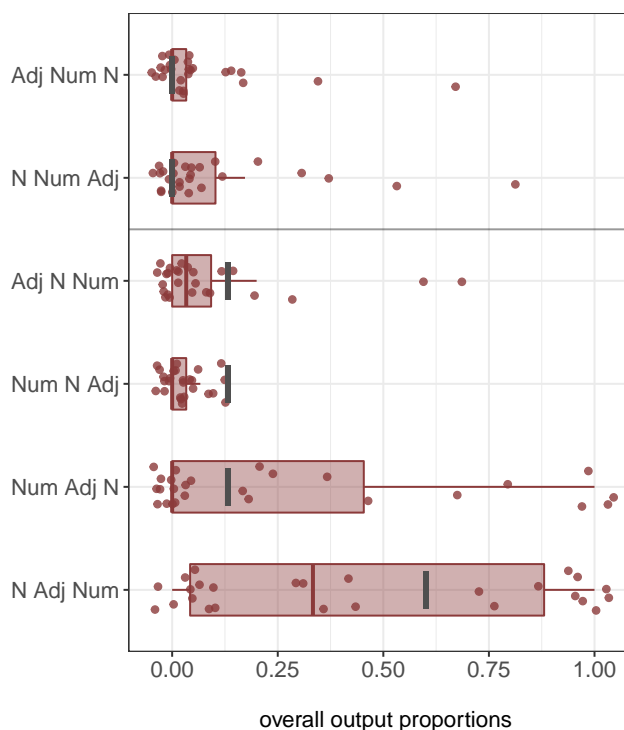


Figure 4. Overall proportion of each two-modifier variant produced in the Word Order condition of Experiment 1. Seen (bottom) and unseen (top) variants during training are divided by a solid grey line. Vertical grey lines within the bars indicate input proportions.

at least do not contradict) a tendency to avoid the coexistence of two opposite N-peripheral variants, that is, N Adj Num and Num Adj N.

2.3 Discussion

Experiment 1 set out to test whether learners preferentially regularise unconditioned variation in morphology relative to word order when the initial complexity of the systems to be learned is matched. This hypothesis was based on evidence from first and second language acquisition and language formation (e.g., pidginisation) which suggests that complex morphological variation may be regularised or simplified to a greater degree than syntactic variation. Our results fail to confirm this; in fact word order was regularised to a greater degree than morphology in this experiment. Our results *are* consistent with previous work showing that adult learners regularise unconditioned variation. Although the majority of participants remained probabilistic users (only a minority used a completely deterministic system), most participants in both conditions regularised their input by eliminating variants and/or by increasing the frequency of one variant and reducing others.

However, an exploration of variant production in the word order condition suggests a possible explanation for the increased entropy drop we observed in this condition: an over-use of Num Adj N order by a subset of participants. This is likely to be reflecting two factors. First, although this order is a minority variant in the input, it matches participants' native language (i.e., English). Second, because this order is the opposite of the majority variant, N Adj Num, participants may have actively avoided using both these orders together; in terms of the linear order of elements,

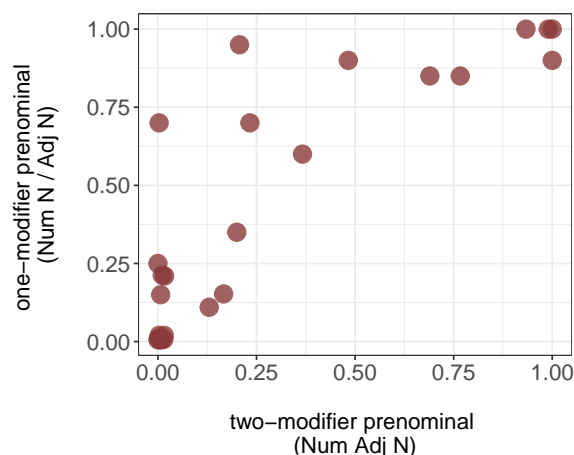


Figure 5. Relationship between each participant’s production of prenominal one-modifier phrases, and use of Num Adj N in two-modifier phrases in Experiment 1. Note that most participants do not use prenominal word orders (lower left corner), but those who use the English-like prenominal two-modifier order tend to use prenominal orders across phrase types.

these two are the most distinct from each other. The presence of Num Adj N in the input might have therefore been enough to have led some participants to use it with a high level of regularity. In fact, these same participants also over-used prenominal modification in one-modifier phrases. This is illustrated in Figure 5, which suggests a positive relationship between the proportion of prenominal orders in one-modifier phrases and the proportion of Num Adj N in two-modifier phrases.¹⁴ The result of this might contribute to the additional reduction in word order variation relative to the Morphology condition. Notably, this is precisely the type of variant-specific bias which could in principle contribute to the appearance of a level-specific regularisation bias in natural language acquisition or emergence. In Experiment 1b, we exclude this variant.

3 Experiment 1b: replication of the word order condition without Num Adj N

Experiment 1b follows the same design as the Word Order condition described in Experiment 1, with a single change: the set of two-modifier input variants. We swapped the two-modifier Num Adj N variant for N Num Adj.¹⁵ These changes should mitigate both of the effects outlined above: it is not the English pattern (reducing the possibility of L1 transfer) and it is not the direct opposite of the majority variant (reducing the dispreference for using both harmonic variants in pro-

¹⁴Additionally, we ran a mixed-effects logistic regression to test whether the proportions of *prenominal* two-modifier phrases differed from those predicted from the product of Num N and Adj N productions within one participant’s use. If a participant produced the variant Num N 40% of the time and the variant Adj N also 40% of the time and they combined them to produce two-modifier variants proportionally, they would produce Num Adj N, N Adj Num, Num N Adj and Adj N Num orders 16%, 36%, 24% and 24% of the time respectively. With the model, we wanted to test whether the proportion of two-modifier prenominal variants was proportional to the proportion of prenominal productions of one-modifier phrases. Results show that there is no difference between the actual output proportions of two-modifier phrase variants and the predicted proportional productions from one-modifier variants ($\beta = -0.345, SE = 0.682, p = 0.613$). These results at least do not contradict a non-trivial relationship between prenominal modification in one-modifier and two-modifier variants.

¹⁵Note that although N Num Adj (non-isomorphic) is dispreferred relative to Num Adj N by English participants in the lab, recent work shows that this dispreference is not strong (Martin, Holtz, Abels, Adger, & Culbertson under review).

duction). If these effects were indeed the cause of the difference between conditions in Experiment 1, then we expect decreased regularisation of word order. We will refer to this as the *Word Order 2* condition, and compare results to the two conditions from Experiment 1.

3.1 Participants

Twenty-six native-English speakers (age 18 to 35, mean 24.8) were recruited via the University of Edinburgh’s Careers Service database. Participants received £6. The data from a further two participants were excluded on the basis of failure to adequately learn the language. This was determined as in Experiment 1 (see section 2.1.1).

3.2 Analyses and results

3.2.1 Confirmatory analysis: Does the strength of regularisation vary between linguistic levels? Entropy scores obtained in the Word Order 2 condition are shown in Figure 6,¹⁶ contrasted with those in the Morphology condition of Experiment 1. We ran a linear mixed-effects model to explore the effect of condition on regularisation behaviour now including Word Order 2 as well as Morphology and Word Order of Experiment 1. The model structure was the same as in Experiment 1 but this time Condition was reverse Helmert coded so that Word Order 2 was directly compared to the Morphology condition from Experiment 1, and the Word Order condition (predicted to be the odd one out) was compared to the average of those. Results revealed the expected significant effect of Condition Word Order ($\beta = 0.073, SE = 0.035, p = 0.037$) but no significant effect of condition Word Order 2 ($\beta = 0.063, SE = 0.060, p = 0.295$), confirming that even though participants regularise their input significantly more in the Word Order condition, participants in the Word Order 2 condition do not regularise more than those in the Morphology condition. We did not find any significant effect of Phrase Type or its interaction with Condition (all $|p| > 0.272$). To summarise, these results show that excluding the Num Adj N variant from the input language eliminated any difference in the reduction of variation across morphology and word order.¹⁷

4 Discussion: the effect of linguistic level on regularisation behaviour in production

In Experiment 1b, participants were taught a language with variable word order, similar to the Word Order condition in Experiment 1, but without the English-like harmonic pattern, Num Adj N. Across both experiments, we found evidence for regularisation of unconditioned variation in morphology and word order, potentially pointing to an overarching simplicity bias (Culbertson & Kirby 2016). Moreover, our results suggest that when input languages have comparable initial complexity, and no variant-specific biases are at play—e.g., variants which match the L1—regularisation rates

¹⁶Individual trials whose phrases did not contain the right number of words were excluded from analysis as in Experiment 1: the mean proportion of excluded trials per participant was 0.006 ($SD = 0.012$). Participants were also allowed to introduce new variants not present in the input: the mean proportion of trials with innovative word orders per participant in Word Order 2 was 0.06 ($SD = 0.11$), which amounts to an average of three variants (out of 50) per participant.

¹⁷We used the *bayestestR* library (Makowski, Ben-Shachar, & Lüdtke 2019) to further explore the strength of the evidence in favour of the null hypothesis—that entropy scores do not differ across conditions—by comparing a regression model as presented in the main text (but only with the data from Morphology and Word Order 2) with an intercept-only model (i.e., not containing condition as a predictor). Models are compared by their BIC measures, allowing a Bayesian comparison of non-nested frequentist models (Wagenmakers 2007). The results reveal a Bayes factor in favour of the full model of < 0.01 . This indicates very strong evidence in favour of the null (intercept-only) model over the full model, suggesting that we can have high confidence in the similarity in entropy scores across conditions.

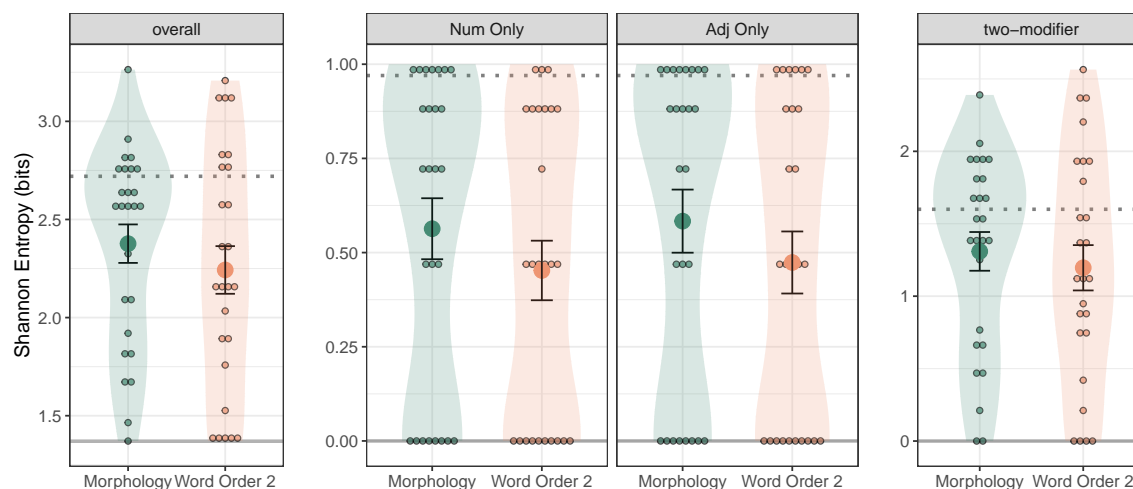


Figure 6. Entropy scores of participants’ output productions in the Morphology condition in Experiments 1 (green) and the Word Order 2 condition in Experiment 1b (orange). From left to right, entropies of participants’ complete production sets as well as entropies for one modifier Num (Num Only), one-modifier Adj (Adj Only) and two modifier (two-Mod) phrases. The dashed lines indicate input entropy; solid lines are minimum entropy scores (0 for each individual phrase types and 1.37 for the overall system with three phrase types).

are similar across these two linguistic levels. Specifically, in Experiment 1 we found higher levels of regularisation in word order than in morphology, but this was driven by the specific set of variants in the input language: when both harmonic prenominal and postnominal two-modifier variants were included, results suggest that some participants over-used the English-like Num Adj N variant. Further, the coexistence of these two quite distinct variants in a single participant’s productions was rare, leading to decreased variability in this condition overall. In Experiment 1b, we showed that eliminating the English-like order from the set of two-modifier variants (by swapping Num Adj N for N Num Adj) generally eliminated the difference in regularisation between levels. Overall, our findings do not support level-specific regularisation biases as an explanation for the apparent asymmetry between simplification of morphological and word order variation hinted at in the literature on language learning and pidgin formation (Bichakjian 1988; Drechsel 1981; Good 2015). Instead, this study suggests that variation is reduced to a similar degree across linguistic levels, with any asymmetry likely due to differing complexity of linguistic paradigms or other features of the contact (or second) languages (Ansaldo et al. 2007; Mufwene 2008).

While our results so far suggest similar strengths of regularisation behaviour in production across linguistic levels, Experiments 1 and 1b do not necessarily reflect the context in which languages—including newly formed pidgins—are learned and used. Language production typically takes place in communicative interaction and so far we have tested production of variants without any explicit communicative task. Some studies on language emergence (i.e., pidgin/creole formation) have argued that grammatical simplification is driven by the reduction of features in source languages which are “incidental to basic communication” (McWhorter 2001) (for similar views in L2 acquisition more broadly, see Klein & Perdue 1997). These studies assume that language users

will generally produce the minimum grammatical information required to convey meanings, which will of course be dependent on the communicative context. On this view, language users will thus fail to reproduce uninformative variation in their productions as a direct consequence of communication. Relatedly, previous experimental studies have shown that an additional mechanism driving regularisation comes from alignment during active communication. For example, Fehér et al. (2016) report results from an artificial language learning task in which participants exposed to unconditioned variation must interact with other learners after training. In their study, pairs of participants were trained on a linguistic system with variable word order and were asked to recall it in isolation, then use it to communicate with an interlocutor, and finally recall it in isolation again. They report increased regularisation during the communication phase, suggesting that reciprocal priming or alignment might contribute to explaining the tendency for avoiding unconditioned variation in natural language. If alignment and/or reduction of uninformative features during communication are effectively independent mechanisms for regularisation, differences across linguistic levels may be revealed by introducing a communicative task.

5 Experiment 2: regularisation behaviour across linguistic levels in communicative interaction

In Experiment 2 we utilise the methodology developed by Fehér et al. (2016) to incorporate communicative interaction into the experimental paradigm. As in the previous experiments, adult learners are exposed to a miniature artificial language featuring an inconsistent mixture of synonymous variants in either morphology or word order. However, this time all production takes place in the context of dyadic communication.

5.1 Materials and methods

5.1.1 Participants. Forty-eight English-native speakers (age 18 to 26, mean 21.8) were recruited from the University of Edinburgh's Careers Service database. Each was compensated £6. Twenty-four participants (12 pairs) were assigned to the Morphology Dyads condition, and the other 24 (12 pairs) to the Word Order Dyads condition. No participants were excluded in this experiment as they all successfully learned the lexicon in the input languages.

5.1.2 Input languages. The input languages used for the Morphology Dyads and the Word Order Dyads conditions were the same ones used for the Morphology condition in Experiment 1 and the Word Order 2 condition in Experiment 1b respectively.

5.1.3 Communicative interaction. The experimental procedure was the same as in Experiments 1 and 1b (see section 2.1.3) with the only modification being the introduction of dyadic communicative interaction during all testing stages. The instructions provided were changed accordingly: participants were told that they were going to be taught an artificial language which they had to learn in order to later communicate with a partner. In Phase 3 (one-modifier testing), participants in a dyad took turns describing one-modifier pictures to their partner, who then had to choose the corresponding picture from a set. They completed 20 trials in total (10 trials per one-modifier type, five per object); this means that each participant completed 10 trials as a director (the one providing descriptions) and 10 trials as a matcher (the one matching the provided descriptions to a meaning). In Phase 5 (full testing), participants in a dyad took turns describing and interpreting descriptions for 52 trials. These 52 trials comprised the following set of pictures (shown in random

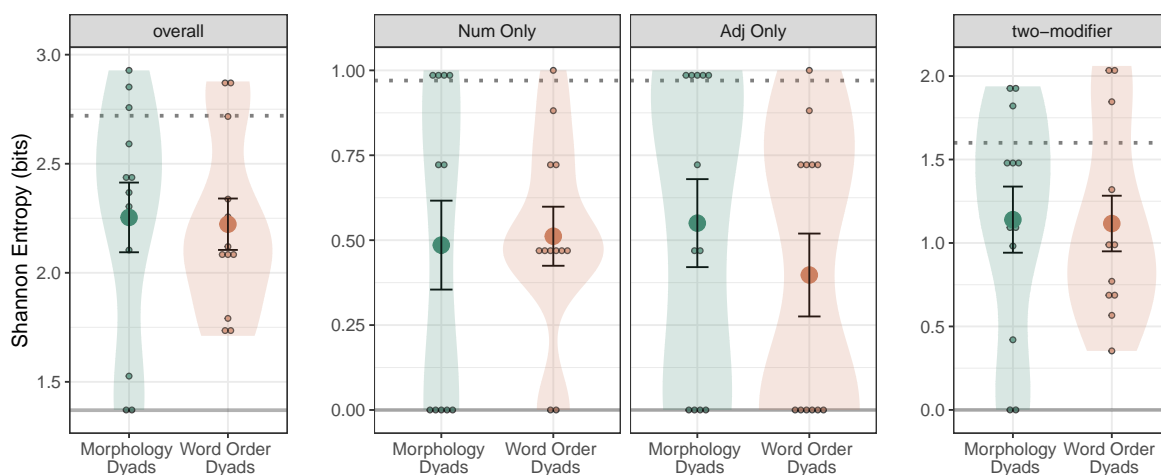


Figure 7. Entropy scores of the output productions of dyads in Experiment 2. Scores are shown for the Morphology Dyads (green) and Word Order Dyads (orange) conditions. From left to right, entropies of participants’ complete production sets as well as entropies for one modifier Num (Num Only), one-modifier Adj (Adj Only) and two modifier (two-Mod) phrases. The dashed lines indicate input entropy; solid lines are minimum entropy scores.

order): 10 Adj Only (five per object), 10 Num Only (five per object), 30 two-Mod (15 per object) and two noun pictures (one per object, singly and in grey-scale).

At each testing trial, one participant was assigned the role of director and the other that of the matcher. The director typed in a description for the picture appearing in their screen. The description was sent to the matcher, who saw the description in the middle of the screen surrounded by an array of four different pictures (chosen as described in section 2.1.3 for picture selection trials). The matcher had to select the picture they thought their partner wanted to convey. After the matcher selected a picture, both participants were given feedback on their communicative success: a green screen was displayed after the matcher correctly selected the picture the director described, and a red screen when the matcher selected a different picture. Green screens were accompanied by a correct-answer sound effect and red screens by a wrong-answer sound effect. The roles of director and matcher were swapped after each trial.

5.2 Analyses and results

5.2.1 Confirmatory analysis: Does the strength of regularisation vary between linguistic levels during communicative interaction?

Figure 7 shows the Shannon entropy of participants’ output productions in Morphology Dyads and Word Order Dyads (computed using each dyad’s set of productions; $N = 12$ for each condition). Most dyads failed to reproduce the full variability of the input languages—they regularised. This was confirmed by a linear mixed-effects model, testing the effects of Condition (i.e., word order vs. morphology) and communicative interaction on regularisation (entropy as dependent variable). We compared the two isolate production conditions—Morphology in Experiment 1 and Word Order 2 in Experiment 1b—to the two communicative conditions—Morphology Dyads and Word Order Dyads.. The model thus included fixed effects for Condition (word order vs. morphology), Experiment (isolate production vs. communica-

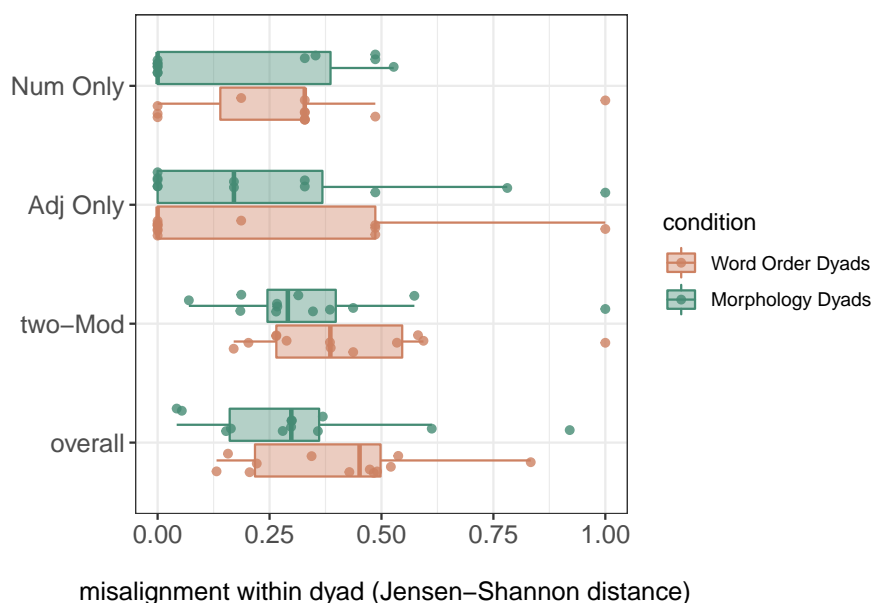


Figure 8. Misalignment between participants in each dyad in Experiment 2. From bottom to top, the Jensen-Shannon distances for the complete output production sets, the two-Mod phrases alone, and each one-modifier phrase type. Individual data points from each dyad as well as a boxplot summarising the data are provided for Morphology Dyads (green) and Word Order Dyads (orange). Dyads are not as closely aligned as expected if participants were matching each other’s behaviour, specially in two-Mod productions, where the distances between variant production are highest.

tive dyads), Phrase Type (Num Only, Adj Only and two-Mod), and their interactions. As random effect, we included subject intercepts (a dyad in Experiment 2 or an individual participant otherwise). As in the previous models, we used reverse Helmert contrast coding across fixed effects. Results revealed a significant intercept ($\beta = 0.444, SE = 0.052, p < 0.001$), showing that the difference between input and output is significantly higher than 0 and thus that across conditions learners are regularising input variability in their productions overall. We did not find any other significant effects (all $|p| > 0.368$), suggesting that the level of regularisation did not differ between linguistic levels nor between production in isolation and during communicative interaction.¹⁸

5.2.2 Exploratory analysis: Do participants within dyads align in their productions?

The results above contrast with those reported in Fehér et al. (2016); we found similar degrees of regularisation behaviour across production in isolation and during communicative interaction. There are several possible reasons for this. First, input languages in this study are more complex in terms of the variants available: eight different variants across three phrase types. Indeed, participants in Experiments 1 and 1b already regularise these systems when they produce phrases in isolation. If the baseline level of regularisation by participants producing the language in isolation is already relatively high, it may not increase further once communication is added.

Complementarily, another possible reason why interaction may not have led to additional

¹⁸Further, we found a Bayes Factor in favour of the full model (over an alternative not including condition, phrase type, or experiment as fixed effects) of < 0.01 . This again points to very strong evidence for the null (intercept-only) model, and allows us to conclude with confidence that entropy scores are similar across conditions and phrase types, as well as across isolate production and dyadic communication.

regularisation is the misalignment between two interlocutors' productions within a dyad. If participants within a dyad regularise the input in different ways and therefore the distance between their production systems is substantial, the admixture of the two systems will result in higher variability. In Fehér et al. (2016), there were only two word order variants which appeared equally often, agent-first or patient-first. Consistent with an agent-first bias (and likely strengthened by participants' L1), regularisation was almost exclusively driven by the over-production of agent-first orders. This bias towards agent-first orders explains the strong alignment the authors observed during interaction; participants' biases were reinforced during communication leading to stronger regularisation in communicative interaction than in isolate production. In our study, there is no such strong bias towards a specific pattern (particularly after the removal of the English-like two-modifier variant), and therefore we would not expect the same degree of alignment within dyads. We use Jensen-Shannon distance¹⁹ (Endres & Schindelin 2003) to measure the distance between the frequency distributions over variants produced by each participant in a dyad. Figure 8 shows this misalignment for all dyads—across all phrase types as well as for each phrase type individually. This shows that dyads are not perfectly aligned in most cases, particularly for productions involving both modifiers, where the use of variants is often quite distinct. We ran a linear mixed-effects model to explore misalignment across conditions and across the different phrase types. The fixed effects were Condition (Morphology Dyads and Word Order Dyads) and Phrase Type (Num Only, Adj Only and two-Mod) along with their interaction. We used the same contrasts as in previous models. Random intercepts for Subject were also included. The model intercept shows that the Jensen-Shannon distance is significantly different from 0 (where 0 would indicate perfect alignment) on average across conditions and Phrase Types ($\beta = 0.294, SE = 0.043, p < 0.001$). Moreover, we found a significant difference between one-modifier and two-modifier phrases ($\beta = 0.049, SE = 0.016, p = 0.004$), confirming that misalignment between participants' productions is indeed higher for these more complex phrase types. No significant differences between conditions were found (largest: $\beta = -0.043, SE = 0.028, p = 0.128$).

5.3 Discussion: the effect of linguistic level on regularisation behaviour during communicative interaction

Previous work has suggested that communicative interaction provides a potential context for regularisation of probabilistic unconditioned variation, either by encouraging speakers to drop uninformative features of the language, or through alignment (Fehér et al. 2019; Fehér et al. 2016; Klein & Perdue 1997; McWhorter 2001). In Experiment 2, we therefore incorporated a communicative task into the experiment in order to determine whether differences in regularisation across linguistic levels might emerge in this context. In particular, participants had to use phrases in the

¹⁹The Jensen-Shannon distance is the square root of the Jensen-Shannon divergence, which is a symmetrised version of the more general Kullback-Leibler divergence metric. Let P and Q be two probability distributions. Kullback-Leibler divergence is defined as:

$$D_{KL}(P||Q) = \sum_i P(i) \log_2 \frac{P(i)}{Q_i}. \quad (2)$$

we then symmetrise this expression and take the square root to obtain the Jensen-Shannon distance, given by

$$JSD(P||Q) = \sqrt{\frac{D_{KL}(P||M) + D_{KL}(Q||M)}{2}}, \quad (3)$$

where $M = (P + Q)/2$.

language to describe pictures to a partner, and had to interpret phrases produced by that partner. However, results revealed no significant additional regularisation compared to production in isolation in Experiments 1 and 1b, and more importantly, no differences across linguistic levels emerged. As discussed above, the lack of additional (across-the-board) regularisation may be explained by a relatively high baseline level of reduction of variation in production in isolation compounded by the lack of alignment in complex two-modifier phrases (which appear in four different variants in the input). While our communicative task is obviously not completely naturalistic, the fact that communication failed to reveal differences in regularisation across conditions suggests again that differences between morphology and word order reported in the literature on language acquisition and formation are likely to reflect specific features of the relevant systems and/or regularisation of other types of complexity.

6 Encoding

Before we discuss the overall conclusions we draw from the experiments reported here, we will first return to the two remaining experimental tasks which we have not yet touched on: grammaticality judgements and variant frequency estimation. We included these tasks because previous studies suggest the possibility that regularisation behaviour may be driven by production rather than learning *per se*. The majority of experiments on regularisation illustrate its effects using production under the assumption that what learners produce reflects what they have encoded during learning (e.g., Culbertson et al. 2012; Hudson Kam & Chang 2009; Hudson Kam & Newport 2009). For example, it could be that a bottleneck in memory affects the encoding of variants and their relative frequencies (potentially explaining why children’s regularise more robustly than adults). However, recent studies have suggested the possibility that variation is actually encoded with relatively high fidelity, even when regularisation during production occurs. For example, Ferdinand et al. (2019) find that when trained on lexical variation, adult learners regularise this variation across a number of conditions, from a very simple task involving a single object with two alternating labels, to a much more difficult task with many objects and alternating labels (i.e., one versus many frequencies to track). Importantly, Ferdinand et al. (2019) show that even when they regularise in production, participants are able to accurately report input proportions. Further evidence to reject a memory bottleneck as the primary driver of regularisation behaviour comes from Perfors (2012a, 2012b, 2016). These studies show that although increasing working memory load during training does not lead to regularisation, *perception of variation does*. The more random the system is perceived to be by participants, the more they regularised input variation. This suggests that it is not learning conditions, but explicit expectations about the nature of the task that might drive regularisation. To fully understand the cognitive mechanisms driving regularisation (rather than only the resulting impact on variation in the language), it is therefore crucial to look both at production and encoding. Above we have reported whether participants *produce* regularised output across conditions. Below we explore whether regularisation is found in encoding as well, based on grammaticality judgements and post-experimental questionnaires in which participants were asked to estimate the relative frequencies of variants in the input languages (as described in section 2.1.3).

6.1 Results

6.1.1 Grammaticality judgement.

Morphology. Participants in morphology conditions generally achieved near-perfect accuracy in distinguishing between grammatical and ungrammatical phrases: the proportion of correct judgements was very high for both Morphology in Experiment 1 (median = 1.00, MAD (median absolute deviation) = 0.00) and Morphology Dyads in Experiment 2 (median = 0.958, MAD = 0.042).

Word order. Scores in the word order conditions were also high across experiments (median = 0.917, MAD = 0.00). However, they were not as high as in the morphology conditions. The pattern of errors suggests that participants in the word order conditions correctly rejected morphologically ungrammatical variants, but were less likely to reject ungrammatical word orders.²⁰ One way to interpret this is that participants may have perceived order to be free in the input grammars. Alternatively, it could be that different orders of the same words (at least in complex phrases) are generally more difficult for participants to tell apart than differences in morphology. In any case, this provides some clue as to why two-modifier phrases are not as regularised by over-production of the majority variant (in comparison to the Morphology Condition). If specific word orders are more difficult to tell apart, it might be that participants in the word order conditions encode pre- and post-modification altogether as the majority input feature rather than a specific variant leading to the maintenance of (and even a slight preference for) minority variants which share the noun position of majority input variants.

6.1.2 Frequency reports. To explore whether regularisation behaviour was driven by production or encoding, we had participants fill out a post-experimental questionnaire where they were asked to estimate how frequently they had heard majority input variants during training. By comparing participants' reports of input frequencies with the frequencies they used in their productions, we can assess whether the regularisation bias is in encoding during learning, or whether it is driven primarily by production.

Figure 9 (isolates) and Figure 10 (communicative interaction) show the reported proportions of majority input variants in the input languages compared to the actual proportions in the input and in participants output productions. We ran a logistic mixed-effects regression model to test whether participants' production of majority variants differs from their estimates of variant frequency.²¹ For ease of comparison across experiments, we excluded the Word Order condition of Experiment 1 in the analyses. Four fixed effects were included in the model along with their interactions: Response Type (Input Report as reference, and Production), Condition (word order vs. morphology), Experiment (communicative vs. isolate production) and Phrase Type (Num Only, Adj Only and two-modifier). We used reverse Helmert contrast coding across fixed effects except for Response Type, which was treatment coded (as we wanted to interpret the intercept for Input Report alone). As random effects, we included intercepts for Subject as well as by-Subject random slopes for the effect of Response Type and Phrase Type and their interaction. The coefficients from the model's intercept show that participants' estimates of major-

²⁰Note that the ungrammatical variants judged by participants in the word order conditions differ slightly across experiments, due to our exclusion of prenominal two-modifier variants in Experiments 1b and 2. In Experiment 1, word order violations were two-modifier phrases with either Adj Num N or N Num Adj order. In Experiments 1b and 2, the ungrammatical word orders were Adj Num N and Num Adj N. In both cases, however, failure to reject ungrammatical word order variants was highly systematic: in the task, there were two relevant trials out of a total of 24, resulting in a score of 22/24 (0.917) for all participants.

²¹We converted the input and reported relative frequencies into dummy coded absolute frequencies weighted according to the number of input and output trials respectively. For example, if a participant reported that the two-modifier majority variant N Adj Num occurred 70% of the time, and they produced that variant 30 times during production, we would code 0.7×30 presences and 0.3×30 absences of the majority input variant respectively.

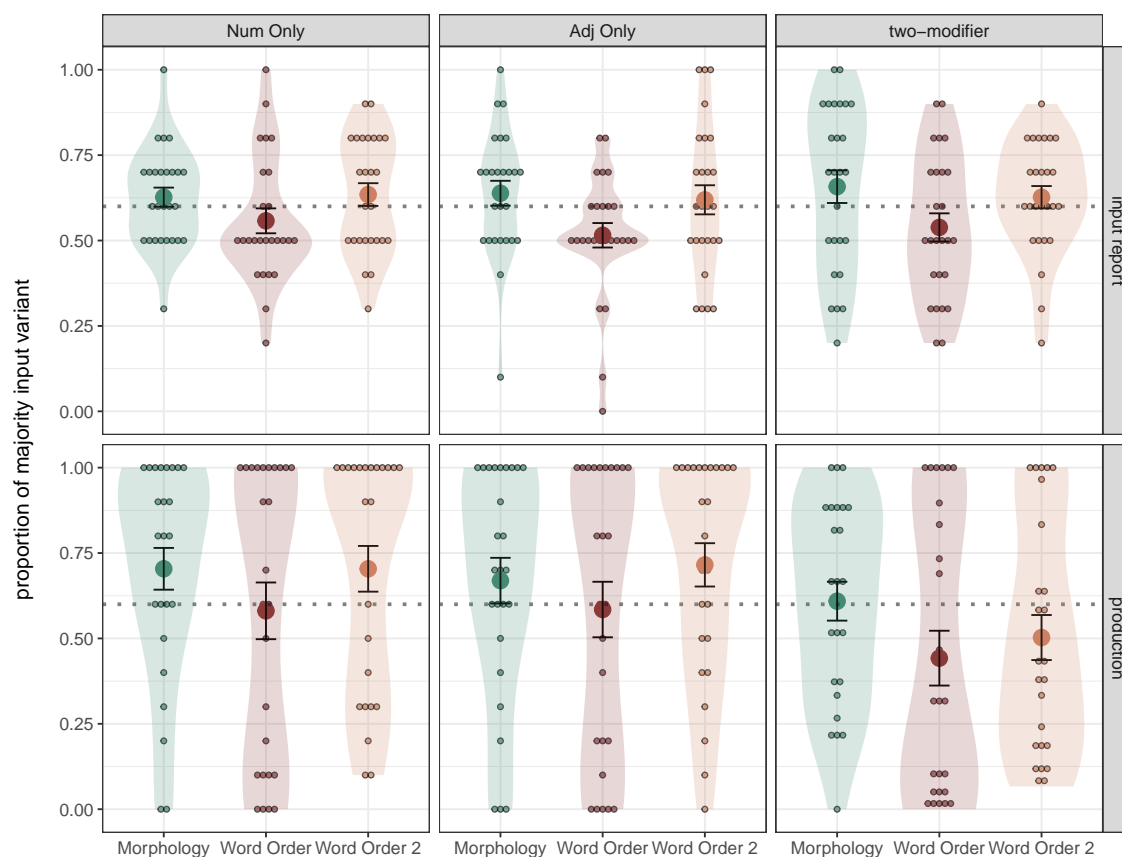


Figure 9. Experiment 1 and 1b: Participants’ estimates of the proportion of majority variant observed during training versus the proportions in production during testing. Results are shown for Morphology (green), Word Order (red) and Word Order 2 (orange) conditions of Experiments 1 and 1b, for each phrase type. Dash lines indicate input proportions of the majority variants ($P_{input} = 0.6$).

ity variants are not significantly different from the input proportion $P = 0.6$ ($p = 0.378$). Results from the model further suggest that participants’ output proportions are overall significantly higher than their input estimates ($\beta = 0.793, SE = 0.261, p = 0.002$) but this difference is smaller in two-modifier phrases ($\beta = -0.580, SE = 0.096, p < 0.001$). We did not find a main effect of Linguistic Level ($\beta = -0.029, SE = 0.072, p = 0.684$) or of its interaction with Response Type ($\beta = -0.063, SE = 0.256, p = 0.805$).

6.1.3 Discussion: the roots of regularisation behaviour. In this section, we investigated whether the regularisation found in *production* in our experiments was also found in encoding. We did so through grammaticality judgements, and a post-experimental questionnaire in which participants were asked to estimate the relative frequencies of the majority variants in the input. Results of the grammaticality judgement task suggest that participants generally had little problem recognising grammatical variants in their input languages across conditions and experiments. This suggests that participants did indeed encode the variants they heard, even if they regularised in production. Interestingly however, participants in the word order conditions failed to learn the restricted set of word order options, instead judging ungrammatical word orders to be allowed in the language. The

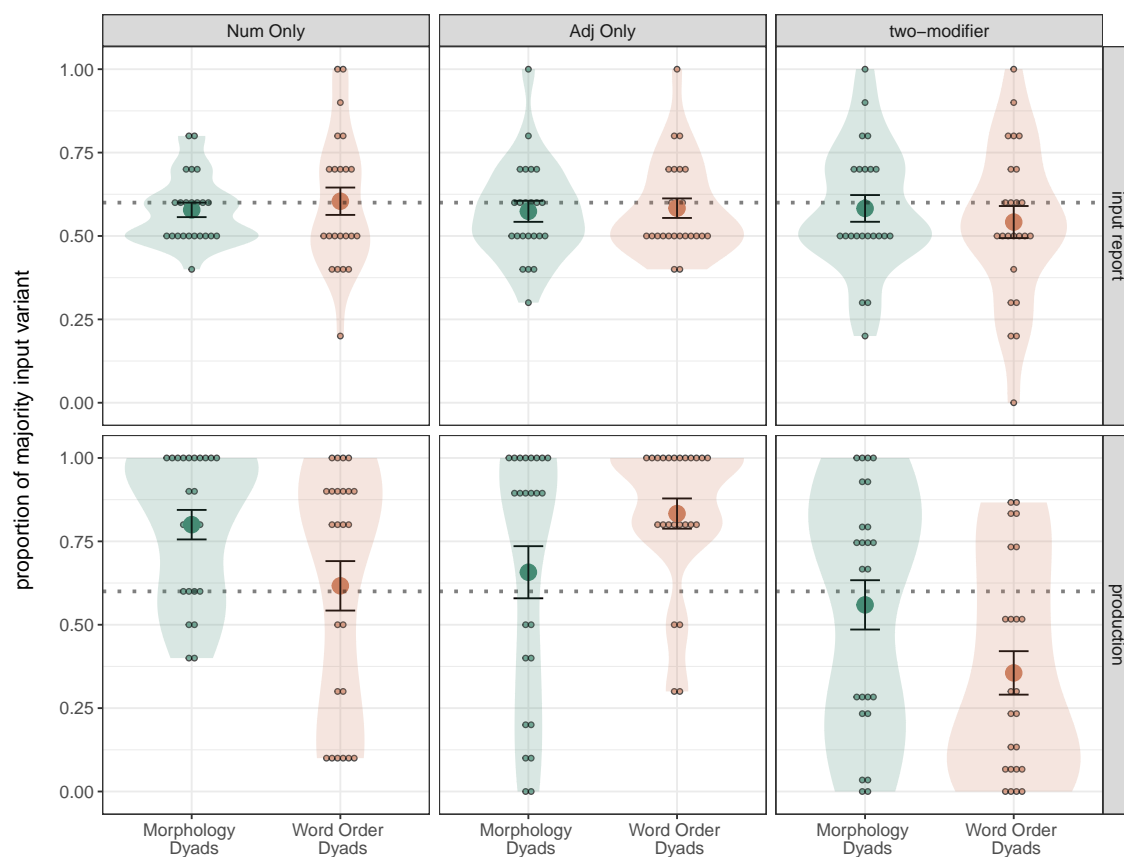


Figure 10. Experiment 2: Participants' estimates of the proportion of majority variant observed during training versus the proportions in production during testing. Results are shown for Morphology Dyads (green) and Word Order Dyads (orange) conditions in Experiment 2, for each phrase type. Dash lines indicate input proportions of the majority variants ($P_{input} = 0.6$).

consistency of this result across participants suggests that, in the presence of many variants, arbitrary restrictions on word order are less likely to be learned. By contrast, participants did not judge ungrammatical morphological variants to be grammatical. The type of unconditioned morphological variation participants are trained on is very close to unconditioned lexical variation: when learners are not exposed to sufficiently rich input to infer the rules behind morphological variation, the only difference between lexical variation and morphological variation might be phonological similarity.²² The results from the grammatical judgement task could be interpreted to suggest that recalling specific lexical items is easier than specific word orders. However, while this difference in the perception of variation suggests that participants in word order conditions perceived a more variable system than participants in the morphology condition, this did *not* lead to significantly stronger regularisation behaviour (as shown in our analyses of entropy reduction above, cf. Hudson Kam & Newport 2009; Perfors 2012a).

²²Although we cannot say whether participants perceive the variation as lexical or morphological, the errors that participants produce in the Morphology conditions indicate that they learn the initial common phonemes between variants (or 'stems') better than the endings (or 'suffixes'); 70% percent of the errors contained the right 'stems'.

Results from the post-experimental questionnaire point to further differences between production and encoding. Although participants regularise in their productions, they correctly estimate the input probabilities—at least for one-modifier phrases. Given that input frequencies of majority variants are 60%, it is somewhat difficult to discern between accuracy in encoding and perception of random variation (i.e., 50%), nevertheless these results suggest that regularisation rates in production do not match participants' frequency encodings for one-modifier phrases. For two-modifier phrases, on the other hand, participants do not over-produce majority variants, leading to an apparently closer match between production and encoding. However, to draw strong conclusions for these phrases we would need estimates of all the different variants.

To summarise, in line with Ferdinand et al. (2019), these findings suggest that regularisation behaviour in production is not necessarily rooted in a memory bottleneck affecting the encoding of less frequent variants, at least for less variable phrase types (i.e., one-modifier phrases). Precisely *why* regularisation occurs more in production is still an open question. It could be related to problems with lexical retrieval; participants can correctly recall input frequencies and the different variants in the input, but they might struggle to retrieve them without being cued during production. For example, Hudson Kam and Chang (2009) show that adults regularise their input less when lexical items are provided to them at test; similar results are reported in Schwab, Lew-Williams, and Goldberg (2018) for regularisation behaviour in children. Memory limitations may therefore interact with mechanisms at play during production such as priming, which might impede the production of alternative variants either during production in isolation or communication. In order for this explanation to extend to regularisation of word order, there would need to be evidence for an effect of structural retrieval or structural priming (Pickering & Branigan 1999).

7 General discussion

7.1 Regularisation behaviour can be uniform across linguistic levels

Previous experimental studies investigating regularisation have typically *assumed* that regularisation is driven by a level-independent mechanism; however, research on first and second language acquisition (Anderssen et al. 2010; Raymond et al. 2009; Slobin 1966) as well as language formation (i.e., pidginisation Bakker 2008; Drechsel 1981; Siegel 2006) suggests intriguing differences in how variation in morphology and word order are learned and transmitted. To explore the extent to which regularisation behaviour in language learning and use differs across linguistic-levels, we conducted a series of experiments in which learners were trained on systems with variable morphology *or* word order. Crucially, the input systems were of comparable initial complexity. In Experiment 1, the presence of a minority word order variant matching English (and the direct opposite of the majority variant in the language) lead to a difference in regularisation between conditions. However in Experiment 1b, when this variant was excluded, we showed that participants regularise unconditioned variation in a similar way across morphology and word order. Thus the present study does not provide supporting evidence for a level-specific bias driving any asymmetries in regularisation of morphology and word order in natural language.

Importantly, comparable regularisation behaviour across levels in this study is dependent on two factors. First, the complexity of the systems of variation—i.e., the amount variants and their frequency of use—must be comparable across levels. Previous work has shown that system complexity matters in regularisation (Ferdinand et al. 2019; Hudson Kam & Newport 2009), and the findings reported in this paper are in line with the idea that apparent level-specific differences may

in fact be driven by asymmetries in complexity. Second, there must not be biases targeting specific alternative variants, for example, an L1-similar word order. If this condition is not met, as we showed in Experiment 1b, over- or under-production of particular variants may lead to overall more regular systems. Taken together, this suggests that in order to fully evaluate why particular natural language systems are regularised more than others, we must have some measure of the inherent complexity of these systems, and the extent to which particular aspects or features of the system might be targeted by independent, variant-specific biases. Inherent complexity and the amount of overlap between L1 and L2 systems still remain to be systematically explored as drivers for asymmetries between levels in language acquisition and formation. It is possible that morphological variation simply tends to be more complex than variation in word order; paradigmatic variation is often greater in morphology than in word order, for example, inflectional systems generally contain more (meaningful) variants than word order in any given phrase (cf. Good 2015). If that is the case, then differences in regularisation or simplification in natural language formation could be attributed to differences in initial complexity, for two reasons. Firstly, the complexity of a system (e.g., how much variation it encodes) is known to influence regularisation behaviour, with more regularisation of more complex input (Hudson Kam & Chang 2009; Hudson Kam & Newport 2009). If morphological paradigms tend to be more complex than word order, it is therefore possible that *not all* variants within the given systems will be present in the data available to the learner, or that they will not all be frequent or salient enough to be reproduced (Mufwene 2008). If the learner does not have adequate access to the full paradigm, this may result in regularisation due to underspecification. Further, if morphological paradigms are more complex than word order, then full morphological paradigms will necessarily vary more between languages. Consequently, the features within them will overlap less cross-linguistically than in word order systems. This difference in the amount of overlap between features across source languages could also determine the preservation of a given grammatical feature during language formation (Ansaldi et al. 2007); the lower the probability of overlap, the lower the probability of being robustly transmitted (however, see Blasi, Michaelis, & Haspelmath 2017).

7.2 Level-specific differences in regularisation do not emerge in communicative interaction

In Experiment 2, we introduced communicative interaction into our task in order to explore whether a level-specific bias in regularisation would emerge in a more naturalistic context of language use. We found that participants regularised to similar degrees when they produced phrases in isolation and during communicative interaction. This suggests that regularisation—at least in the context of L2 production—is not fully driven by communication (cf. Klein & Perdue 1997; McWhorter 2001). This is in line with a number of previous studies which have found evidence of regularisation without communication (e.g., Culbertson et al. 2012; Hudson Kam & Newport 2009) under certain circumstances (e.g., when the systems are highly complex, or certain variants are preferred). More importantly however, we found that communication does not lead to level-specific differences in regularisation (cf. Jansson, Parkvall, & Strimling 2015).

7.3 Regularisation behaviour takes place in production as well as in learning

Across all three experiments, we have also found evidence that regularisation may be at least partially due to production mechanisms rather than learning biases. We observed that participants successfully encoded all variants, not only the most frequent ones. Results from grammaticality

judgements suggested that participants generally learned the specific input variants (although they had some trouble with complex word order variants). Regularisation behaviour was therefore not due to failure to encode minority variants (Ferdinand et al. 2019; Perfors 2012b). Further, the results of our post-experimental questionnaire indicate that participants were able to report the input frequencies of (one-modifier) majority variants quite accurately, even if they over-produce that variant in their own productions. Altogether these results provide additional evidence, in line with Ferdinand et al. (2019), that adult learners regularise unconditioned probabilistic variation in the input mainly through production.

8 Conclusions

The causal link between regularisation and the scarcity of unconditioned variation in natural languages is strongly supported by many experimental studies using artificial language learning experiments (see Smith et al. 2017 for review). However, despite this extensive literature, relatively little is known about whether and how regularisation behaviour might differ across linguistic levels. This is surprising given evidence for apparent differences in regularisation processes across linguistic levels in natural language learning, change and formation: while morphological variation seems to be highly susceptible to regularisation, variation in other syntactic features, like word order, appears more likely to be reproduced (Bakker 2008; Good 2015; Siegel 2006). Our experiments provided the first controlled investigation of level-specific regularisation by comparing regularisation behaviour across morphology and word order systems with matched formal complexity (entropy). We found no evidence for an effect of linguistic level on regularisation behaviour: morphological and syntactic variation were regularised to the same degree so long as no variant-specific biases (e.g. preference for L1-like patterns in only one level) were present. While it remains possible that multiple mechanisms drive regularisation (e.g., production biases, alignment during communication), we conclude these are *not* likely to be independent from a systems' inherent complexity. Instead, our results suggest that understanding asymmetries in regularisation in natural language requires careful analysis of formal complexity, and a better understanding of how specific features of particular linguistic variants are treated in language learning and use.

References

- Anderssen, M., Bentzen, K., & Westergaard, M. (2010). *Variation in the input: Studies in the acquisition of word order* (Vol. 39). Springer. doi: 10.1007/978-90-481-9207-6
- Ansaldo, U., Matthews, S., & Lim, L. (2007). *Deconstructing creole* (Vol. 73). Amsterdam/Philadelphia: John Benjamins.
- Bakker, P. (2008). Pidgins versus creoles and pidgincreoles. In (pp. 48–73). Oxford: Wiley-Blackwell. doi: 10.1002/9781444305982.ch6
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi: 10.18637/jss.v067.i01
- Bichakjian, B. H. (1988). *Evolution in language* (Vol. 18). Karoma Pub.
- Bickerton, D. (2015). *Roots of language*. Berlin: Language Science Press. doi: 10.17169/langsci.b91.109
- Blasi, D. E., Michaelis, S. M., & Haspelmath, M. (2017). Grammars are robustly transmitted even during the emergence of creole languages. *Nature Human Behaviour*, 1(10), 723. doi: 10.1038/s41562-017-0192-4

- Culbertson, J., & Adger, D. (2014). Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences*, *111*(16), 5842–5847. doi: 10.1073/pnas.1320525111
- Culbertson, J., & Kirby, S. (2016). Simplicity and specificity in language: Domain-general biases have domain-specific effects. *Frontiers in Psychology*, *6*, 1964. doi: 10.3389/fpsyg.2015.01964
- Culbertson, J., & Newport, E. L. (2015). Harmonic biases in child learners: In support of language universals. *Cognition*, *139*, 71–82. doi: 10.1016/j.cognition.2015.02.007
- Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, *122*(3), 306–329. doi: 10.1016/j.cognition.2011.10.017
- Damerau, F. J. (1964). A technique for computer detection and correction of spelling errors. *Communications of the ACM*, *7*(3), 171–176. doi: 10.1145/363958.363994
- DeGraff, M. (1999). *Language creation and language change*. Cambridge, MA: The MIT Press.
- Dietrich, R., Klein, W., & Noyau, C. (1995). *The acquisition of temporality in a second language* (Vol. 7). Amsterdam: John Benjamins.
- Drechsel, E. J. (1981). A preliminary sociolinguistic comparison of four indigenous pidgin languages of north america (with notes towards a sociolinguistic typology in american indian linguistics). *Anthropological Linguistics*, 93–112. Retrieved from <http://www.jstor.org/stable/30027542>
- Endres, D. M., & Schindelin, J. E. (2003). A new metric for probability distributions. *IEEE Transactions on Information theory*, *49*(7), 1858–1860. doi: 10.1109/TIT.2003.813506
- Fedzechkina, M., Newport, E. L., & Jaeger, T. F. (2017). Balancing effort and information transmission during language acquisition: Evidence from word order and case marking. *Cognitive science*, *41*(2), 416–446.
- Feher, O., Ritt, N., & Smith, K. (2019). Asymmetric accommodation during interaction leads to the regularisation of linguistic variants. *Journal of Memory and Language*, *109*, 104036. doi: 10.1016/j.jml.2019.104036
- Fehér, O., Wonnacott, E., & Smith, K. (2016). Structural priming in artificial languages and the regularisation of unpredictable variation. *Journal of Memory and Language*, *91*, 158–180. doi: 10.1016/j.jml.2016.06.002
- Ferdinand, V., Kirby, S., & Smith, K. (2019). The cognitive roots of regularization in language. *Cognition*, *184*, 53–68. doi: 10.1016/j.cognition.2018.12.002
- Fraser, C., Bellugi, U., & Brown, R. (1963). Control of grammar in imitation, comprehension, and production. *Journal of verbal learning and verbal behavior*, *2*(2), 121–135. doi: 10.1016/S0022-5371(63)80076-6
- Good, J. (2012). Typologizing grammatical complexities: or why creoles may be paradigmatically simple but syntagmatically average. *Journal of Pidgin and Creole Languages*, *27*(1), 1–47. doi: 10.1075/jpcl.27.1.01goo
- Good, J. (2015). Paradigmatic complexity in pidgins and creoles. *Word Structure*, *8*(2), 184–227. doi: 10.3366/word.2015.0081
- Håkansson, G. (1995). Syntax and morphology in language attrition: A study of five bilingual expatriate swedes. *International Journal of Applied Linguistics*, *5*(2), 153–169.
- Hare, M., & Elman, J. L. (1995). Learning and morphological change. *Cognition*, *56*(1), 61–98. doi: 10.1016/0010-0277(94)00655-5
- Hudson Kam, C. L., & Chang, A. (2009). Investigating the cause of language regularization

- in adults: Memory constraints or learning effects? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3), 815–21. doi: 10.1037/a0015097
- Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language learning and development*, 1(2), 151–195. doi: 10.1080/15475441.2005.9684215
- Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive psychology*, 59(1), 30–66. doi: 10.1016/j.cogpsych.2009.01.001
- Hupp, J. M., Sloutsky, V. M., & Culicover, P. W. (2009). Evidence for a domain-general mechanism underlying the suffixation preference in language. *Language and Cognitive Processes*, 24(6), 876–909.
- Jansson, F., Parkvall, M., & Strimling, P. (2015). Modeling the evolution of creoles. *Language Dynamics and Change*, 5(1), 1–51. doi: 10.1163/22105832-00501005
- Klein, W., & Perdue, C. (1997). The basic variety (or: Couldn't natural languages be much simpler?). *Second language research*, 13(4), 301–347. doi: 10.1191/026765897666879396
- Kouwenberg, S., & Singler, J. V. (2009). *The handbook of pidgin and creole studies*. Oxford: Wiley-Blackwell.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). lmerTest: Tests in linear mixed effects models [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=lmerTest> (R package version 2.0-29)
- Labov, W. (1972). *Sociolinguistic patterns* (No. 4). University of Pennsylvania Press.
- Lefebvre, C., White, L., & Jourdan, C. (2006). *L2 acquisition and creole genesis: dialogues* (Vol. 42). Amsterdam: John Benjamins.
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady* (Vol. 10, pp. 707–710). Retrieved from <http://adsabs.harvard.edu/abs/1966SPhD...10...707L> (Provided by the SAO/NASA Astrophysics Data System)
- Makowski, D., Ben-Shachar, M., & Lüdtke, D. (2019). bayestestr: Describing effects and their uncertainty, existence and significance within the bayesian framework. *Journal of Open Source Software*, 4(40), 1541. doi: 0.21105/joss.01541
- Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., & Clahsen, H. (1992). Overregularization in language acquisition. *Monographs of the society for research in child development*, 57(4), i–178. doi: 10.2307/1166115
- Martin, A., & Culbertson, J. (in press). Revisiting the suffixing preference: Native language affixation patterns influence perception of sequences. *Psychological Science*.
- Martin, A., Holtz, A., Abels, K., Adger, D., & Culbertson, J. (under review). *Experimental evidence for the influence of structure and meaning on linear order in the noun phrase*.
- McWhorter, J. (2001). The world's simplest grammars are creole grammars. *Linguistic typology*, 5(2/3), 125–166. doi: 10.1515/lity.2001.001
- McWhorter, J. (2005). *Defining creole*. Oxford: Oxford University Press.
- Miller, K. L., & Schmitt, C. (2012). Variable input and the acquisition of plural morphology. *Language Acquisition*, 19(3), 223–261. doi: 10.1080/10489223.2012.685026
- Montrul, S. A. (2004). *The acquisition of spanish: Morphosyntactic development in monolingual and bilingual l1 acquisition and adult l2 acquisition* (Vol. 37). John Benjamins.
- Mufwene, S. S. (2008). *Language evolution: Contact, competition and change*. London: Blooms-

bury.

- Newport, E. L. (1999). Reduced input in the acquisition of signed languages: Contributions to the study of creolization. In M. DeGraff (Ed.), *Language creation and language change: Creolization, diachrony, and development*. Cambridge, MA: The MIT Press.
- Nichols, J. D. (1995). The ojibwe verb in ‘broken oghibbeway’. *Amsterdam Creole Studies*, 12, 1–18.
- Parkvall, M. (2008). The simplicity of creoles in a cross-linguistic perspective. In (pp. 265–285). doi: 10.1075/slcs.94.17par
- Peirce, J. W. (2007). Psychopy–psychophysics software in python. *Journal of neuroscience methods*, 162(1), 8–13. doi: 10.1016/j.jneumeth.2006.11.017
- Peirce, J. W. (2009). Generating stimuli for neuroscience using psychopy. *Frontiers in Neuroinformatics*, 2(10). doi: 10.3389/neuro.11.010.2008
- Perfors, A. (2012a). Probability matching vs over-regularization in language: Participant behavior depends on their interpretation of the task.. Retrieved from <https://mindmodeling.org/cogsci2012/papers/0155/paper0155.pdf>
- Perfors, A. (2012b). When do memory limitations lead to regularization? an experimental and computational investigation. *Journal of Memory and Language*, 67(4), 486–506. doi: 10.1016/j.jml.2012.07.009
- Perfors, A. (2016). Adult regularization of inconsistent input depends on pragmatic factors. *Language Learning and Development*, 12(2), 138–155.
- Pickering, M. J., & Branigan, H. P. (1999). Syntactic priming in language production. *Trends in Cognitive Sciences*, 3(4), 136–141.
- Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? evidence from tense and agreement. *Second Language Research*, 16(2), 103–133.
- Pygame Community. (2009). *Pygame - python game development*. <http://pygame.org/>.
- R Core Team. (2015). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Raymond, W. D., Healy, A. F., McDonnell, S., & Healy, C. A. (2009). Acquisition of morphological variation: The case of the english definite article. *Language and Cognitive Processes*, 24(1), 89–119. doi: 10.1080/01690960802075489
- Reali, F., & Griffiths, T. L. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, 111(3), 317–328. doi: 10.1016/j.cognition.2009.02.012
- Ross, D. S., & Newport, E. L. (1996). The development of language from non-native linguistic input. In *Proceedings of the 20th annual boston university conference on language development* (Vol. 2, pp. 634–645). Retrieved from http://wearables.cc.gatech.edu/paper_of_week/Ross_and_Newport_1996.pdf
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics bulletin*, 2(6), 110–114. doi: 10.2307/3002019
- Schilling-Estes, N., & Wolfram, W. (1994). Convergent explanation and alternative regularization patterns: Were/weren’t leveling in a vernacular english variety. *Language variation and change*, 6(03), 273–302. doi: 10.1017/S0954394500001691
- Schumacher, R. A., Pierrehumbert, J. B., & LaShell, P. (2014). Reconciling inconsistency in encoded morphological distinctions in an artificial language. In *Pro-*

- ceedings of the 36th annual conference of the cognitive science society*. Retrieved from http://www.phon.ox.ac.uk/jpierrehumbert/publications/schumacher_pierrehumbert_lashell_CogSci2014.pdf
- Schwab, J. F., Lew-Williams, C., & Goldberg, A. E. (2018). When regularization gets it wrong: children over-simplify language input only in production. *Journal of child language*, 1–19. doi: 10.1017/S0305000918000041
- Senghas, A., & Coppola, M. (2001). Children creating language: How nicaraguan sign language acquired a spatial grammar. *Psychological science*, 12(4), 323–328. doi: 10.1111/1467-9280.00359
- Senghas, A., Newport, E. L., & Supalla, T. (1997). Argument structure in nicaraguan sign language: The emergence of grammatical devices. In *Proceedings of the 21st annual boston university conference on language development*. Retrieved from https://www.researchgate.net/publication/237295284_Argument_Structure_in_Nicaraguan_Sign_Language_The_Emergence_of_Grammatical_Devices
- Shuy, R. W., Wolfram, W., & Riley, W. K. (1967). *Linguistic correlates of social stratification in detroit speech*. Technical Report U. S. Office of Education Final Report No. 6-1347.
- Siegel, J. (1997). Mixing, leveling, and pidgin/creole development. In A. K. Spears & D. Winford (Eds.), *The structure and status of pidgins and creoles* (pp. 111–149). Amsterdam/Philadelphia: John Benjamins. doi: 10.1075/cll.19.09sie
- Siegel, J. (2004). Morphological simplicity in pidgins and creoles. *Journal of Pidgin and Creole languages*, 19(1), 139–162. doi: 10.1075/jpcl.19.1.06sie
- Siegel, J. (2006). Links between sla and creole studies: Past and present. *Language Acquisition and Language Disorders*, 42, 15.
- Singleton, J. L., & Newport, E. L. (2004). When learners surpass their models: The acquisition of american sign language from inconsistent input. *Cognitive psychology*, 49(4), 370–407. doi: 10.1016/j.cogpsych.2004.05.001
- Slobin, D. I. (1966). The acquisition of russian as a native language. *The genesis of language: A psycholinguistic approach*, 129–148.
- Slobin, D. I. (1986). The acquisition and use of relative clauses in turkic and indo-european languages. *Studies in Turkish linguistics*, 8, 273. doi: 10.1075/tsl.8.16slo
- Smith, K., Perfors, A., Fehér, O., Samara, A., Swoboda, K., & Wonnacott, E. (2017). Language learning, language use and the evolution of linguistic variation. *Phil. Trans. R. Soc. B*, 372(1711), 20160051. doi: 10.1098/rstb.2016.0051
- Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, 116(3), 444–449. doi: 10.1016/j.cognition.2010.06.004
- Spears, A. K. (2008). Pidgins/creoles and african american english. In S. Kouwenberg & J. V. Singler (Eds.), *The handbook of pidgin and creole studies* (pp. 512–42). Oxford: Blackwell. doi: 10.1002/9781444305982.ch21
- Stefansson, V. (1909). The eskimo trade jargon of herschel island. *American Anthropologist*, 11(2), 217–232.
- Thomason, S. G. (1980). On interpreting ‘The Indian Interpreter’. *Language in Society*, 9(2), 167–193.
- Trudgill, P. (1974). *The social differentiation of english in norwich* (Vol. 13). CUP Archive.
- van der Voort, H. (1997). New light on eskimo pidgins. In A. K. Spears & D. Winford (Eds.), *The structure and status of pidgins and creoles: Including selected papers from meetings of the*

- society for pidgin and creole linguistics* (Vol. 19). Amsterdam: John Benjamins.
- van der Voort, H. (2013). Eskimo pidgin structure dataset. In S. M. Michaelis, P. Maurer, M. Haspelmath, & M. Huber (Eds.), *Atlas of pidgin and creole language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://apics-online.info/contributions/76>
- van Trijp, R. (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
- Velupillai, V. (2015). *Pidgins, creoles and mixed languages: An introduction* (Vol. 48). Amsterdam/Philadelphia: John Benjamins.
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic bulletin & review*, 14(5), 779–804. doi: 10.3758/BF03194105
- Winford, D. (2003). *An introduction to contact linguistics*. Oxford: Wiley-Blackwell.
- Wonnacott, E., & Newport, E. L. (2005). Novelty and regularization: The effect of novel instances on rule formation. In *BuclD 29: Proceedings of the 29th annual Boston university conference on language development* (Vol. 29, pp. 663–673). Somerville: Cascadia Press. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.222.7380&rep=rep1&type=pdf>