Eliminating unpredictable linguistic variation through interaction

Kenny Smith (kenny.smith@ed.ac.uk) & Olga Fehér
Language Evolution and Computation Research Unit, School of Philosophy, Psychology & Language Sciences, University of Edinburgh, Dugald Stewart Building, 3 Charles Street, Edinburgh, EH8 9AD, UK

Nikolaus Ritt
Department of English, Universität Wien, Campus d. Universität Wien, Spitalgasse 2-4/Hof 8.3, 1090 Wien, Austria

Abstract

Languages tend not to exhibit unpredictable variation. We explore alignment/accommodation during interaction as a mechanism to explain this cross-linguistic tendency. Specifically, we test the hypothesis (derived from historical linguistics) that interactions between categorical and variable users are inherently asymmetric: while variable users (of e.g. a grammatical marker) can accommodate to their partner by increasing their usage, categorical users should be reluctant to accommodate to variable partners, since this requires them to violate the rules of their grammar. We ran an experiment in which pairs of participants learnt a miniature language (featuring a potentially variable grammatical marker) and then used it to communicate. Our results support the hypothesis: variably-trained participants accommodate to their categorically-trained partners, who do not change their behaviour during interaction. More generally, interaction results in the elimination of variation: accommodation/alignment is a viable mechanism for explaining the absence of unpredictable variation in language.

Keywords: language; regularity; unpredictable variation; communication; alignment; accommodation

Introduction

To what extent are human behaviours a straightforward reflection of the underlying psychological characteristics of the individual? This is a key question in the cognitive sciences, and is central to the debate in linguistics over the relationship between the typological distribution of languages and psychological constraints on language acquisition (see e.g. Chomsky, 1965; Christiansen & Chater, 2008; Evans & Levinson, 2009). Many prominent accounts assume that language universals reflect processes of language acquisition in the individual: for instance, under one account, language acquisition involves statistical computations, and limits on the computations that human learners can carry out necessarily constrain the form of language (e.g. Christiansen & Chater, 2008). However, in such accounts the precise relationship between biases in individuals and constraints on linguistic structure remains an open question. There is evidence that strong learning biases directly shape language: individual learners transform unnatural languages to conform to cross-linguistic universals (e.g. Singleton & Newport, 2004; Hudson Kam & Newport, 2009; Culbertson, Smolensky, & Legendre, 2012). Weaker biases, which appear ineffectual at the individual level and which may even be hard to identify experimentally, might also shape languages: individual learners might not appear to change unnatural input, yet change will occur due to repeated transmission in a population of learners (see e.g. Reali & Griffiths, 2009; Smith & Wonnacott, 2010). This potentially obscures the link between language universals and properties of individual learners: strong or absolute tendencies in language design may not reflect or require strong or absolute constraints in learners, but may instead be a consequence of far weaker biases in language learning, amplified as a result of the transmission of language in populations.

Much of the foregoing work (Singleton & Newport, 2004; Hudson Kam & Newport, 2009; Reali & Griffiths, 2009; Smith & Wonnacott, 2010) uses the learning and transmission of unpredictable or ‘free’ variation (where competing linguistic forms alternate unpredictably) as a test case for studying the link between language learning and language universals. Widespread unpredictable variation is unattested in natural languages: no two linguistic forms will occur in precisely the same environments and perform precisely the same functions; rather, usage is conditioned in accordance with phonological, semantic, pragmatic or sociolinguistic criteria (Givón, 1985). While previous work has focussed on the role of learning and/or inter-generational transmission in eliminating unpredictable variation from language, in this paper we consider a second mechanism by which variation might be systematically removed: alignment during interaction. Interlocutors modify their linguistic behaviour towards that of their partners (a process known as accommodation or alignment, see e.g. Coupland, 2010; Pickering & Garrod, 2004). These low-level adjustments to linguistic behaviour potentially impact on the structure of language, both by changing the long-term behaviour of the individuals involved, and by skewing the linguistic data from which other individuals learn.

In particular, we explore what happens during interaction between individuals who have different experiences of the variability of a grammatical marker. Such interactions must occur frequently as new grammatical markers develop and spread through languages. For instance, the English language developed from a situation where definiteness on nouns could be marked optionally to one in which definiteness (Sommerer, 2012) and indefiniteness (Rissanen, 1967) of noun phrase reference had to be explicitly marked (with the articles *the* and *a(n)* as default markers). Fehér, Ritt, Smith, and ten Wolde (2014) hypothesise that, in such scenarios, speakers with optional (in)definiteness marking would have found it relatively easy to accommodate to speakers with a categorical rule, simply by increasing their use of a grammatically viable option; in contrast, speakers with categorical rules would have had to violate their grammars to accommodate to their variable partners. The prediction is therefore that, through interaction, categorical and variable speakers would tend to converge on patterns of (in)definiteness marking that favoured the inference of obligatory and categorical rules, so that grammars which incorporated these rules would inevitably spread at the cost of the variable type. More generally, we seek to test the hypothesis that accommodation/alignment during interaction leads to the elimination of unpredictable variation, with categorical patterns being inherently more resistant to change than variable usage.

In order to test these hypotheses, we ran an experiment, completed in pairs, in which participants were asked to learn a miniature language and then use it to communicate with their partner. The miniature language could be used to describe scenes involving animals performing motions: scenes differed in the animal(s) involved, the motion undertaken by the animal(s), and the number of animals in the scene (one or two). Pairs were assigned at random to one of three conditions, which differed in the proportion of training trials on which singular number was signalled with an overt linguistic marker. We were interested in how participants changed their use of the singular marker during interaction, and in particular whether categorically-trained and variably-trained participants systematically differed in the way in which they accommodated to their partner during interaction.

Method

Participants

72 participants (53 female, mean age 21 years 10 months) were recruited from the University of Edinburgh's Student and Graduate Employment service, to take part in a miniature language communication experiment. Participants were paid £8 for their participation, and were briefed that the pair who obtained the highest score on the communicative task would receive an additional £20 in online shopping vouchers.

Procedure

Participants were seated in isolation in sound-proof booths. Participants worked through a computer program which presented and tested them on a semi-artificial language, and then allowed them to use that language to communicate remotely with their partner. The language was text-based: participants observed pictures and text displayed on the screen and entered their responses using the keyboard.

Language Training and Testing Procedures Participants progressed through a three-stage training and testing regime:
1) Noun training: Participants viewed pictures of six cartoon animals (bird, elephant, frog, insect, pig, shark) along with nonsense nouns which were intended to be memorable and transparently related to their associated referent animal (*beeko*, *trunko*, *hoppo*, *bugo*, *oinko* and *fino*). Each presentation lasted 3 seconds, after which the text (but not the picture) disappeared and participants were instructed to retype that text. Participants received 4 blocks of training, each consisting of one presentation of each noun in random order.[1]

2) Vocabulary testing: Participants were presented with a picture of an animal, without accompanying text, and were asked to provide the appropriate label. Participants were tested on each animal once, in random order.
3) Sentence training: Participants were exposed to sentences paired with visual scenes. Scenes showed either single animals or pairs of animals (of the same type) performing one of two possible actions, depicted graphically using arrows: either a straight left-to-right movement, or a bouncing left-to-right movement. Sentences were presented in the same manner as nouns (participants viewed a scene plus text, then retyped the text). The description accompanying each scene consisted of a nonsense verb (*wooshla* for straight movement, *boingla* for bouncing movement), a noun (as above) and a number marker (with the singular marker appearing variably in some conditions, see below). Each pair of participants was assigned two number words, one for singular and one for plural, selected randomly without replacement from the set {*bup*, *dak*, *jeb*, *kem*, *pag*, *tid*, *wib*, *yav*}: for instance, if the randomly-selected markers were *bup* and *yav*, then one bird moving straight would be labelled *wooshla beeko bup* or *wooshla beeko* (depending on whether the singular was marked), and two sharks bouncing would be labelled *boingla fino yav*. Each of the 24 possible scenes (6 animals x 2 motions x 2 numbers) was presented six times (in six blocks, order randomised within blocks).
4) Individual testing: Participants viewed the same 24 scenes without accompanying text and were asked to enter the appropriate sentence. Each of the 24 scenes was presented three times (in three blocks, order randomised within blocks).
5) Interactive testing: Participants played a director-matcher game in which they alternated describing a scene for their partner, and selecting a scene based on their partner's description. When directing, participants were presented with a scene (drawn from the set of 24 possible scenes) and prompted to type the description so their partner could identify it. This description was then passed to their partner[2], who had to identify the correct scene (by button-press) from an array of 8 possibilities: these 8 possibilities contained two animal types (the animal in the director's scene plus one other randomly-selected animal type), both motions (straight and bounce) and both numbers (singular and plural), and thus were guaranteed to contain the target but in themselves provide no information as to the correct target. After each trial both participants then received feedback (either success or failure) and an updated score ("Score so far: X out of Y").

[1] Presentation order for the two members of a pair was randomised independently throughout training and individual testing. In order to keep the participants roughly synchronised, participants were only allowed to progress to the next block of training/testing when their partner was also ready to begin the corresponding block.

[2] In fact the closest legal description was passed to their partner, to prevent participants communicating using English or any system other than the language they were trained on: the string produced by the director was checked against all 36 legal strings in the language the participants were trained on (2 verbs x 6 nouns x three possible markers [null, two possible markers]), and the closest legal string (by Levenstein string-edit distance) was transferred to the matcher.

Participants played 96 such communication games, organised into two blocks of 48 trials, such that each participant directed once for each possible scene within each block (order randomised within blocks, a randomly-selected member of the pair directing first in each block and the participants alternating roles for the remainder of the block).

Variable marking of the singular The training language provided post-nominal particles to mark singular and plural. The plural was consistently marked for all participants throughout training: every sentence labelling a scene featuring two animals included the appropriate post-nominal marker. We manipulated the extent to which participants saw overt marking of the singular during training: participants either saw consistent categorical marking of the singular (100% marking), singular marking on two of every three singulars (for convenience, 66% marking), with the remainder unmarked (i.e. the sentence contained only the verb and the noun), or singular marking on one of every three singulars (33% marking). For variably-trained participants, the training data was constructed such that singular marking was unpredictable: every noun was marked for singular an equal number of times, and every verb appeared with a marked singular an equal number of times.

Participants within a pair differed in the language they were trained on. We ran 12 pairs in each of three combinations: we will refer to the participant trained on the higher frequency of singular marking as P1 and the participant trained on the lower frequency as P2. In the 100-66 condition, P1 is trained on 100% (categorical) marking, P2 is trained on 66% (variable) marking; in the 100-33 condition, P1 is trained on categorical marking, P2 on 33% variable marking; in the 66-33 condition both are trained on variable marking (66% and 33% respectively). These three conditions allows us to test our specific hypothesis regarding differential accommodation by categorical- and variable-trained participants, and also whether interaction in general leads to the reduction or elimination of unpredictable variation in singular marking.

Results

Communicative success

Performance during the communicative portion of the task was extremely high throughout, and varied little across conditions: the mean number of successful trials (in which the matcher selected the picture presented to the director) was 45.04 out of 48 in the 100-66 condition (44.33 in the first block of interaction, 45.75 in the second), 45.62 in the 100-33 condition (45.33 in block 1; 45.92 in block 2), and 47.46 in the 66-33 condition (47.42 in block 1; 47.50 in block 2).

Use of the singular marker

Fig. 1 shows the full data for use of the singular marker across training, individual testing and two blocks of interaction. Fig. 2 provides means for the various phases, plus the mean within-pair difference in marker usage (i.e. proportion of marked singulars produced by P1 - proportion produced by P2; difference is inversely proportional to alignment).

In the 100-66 condition, categorically-trained participants remain categorical throughout. Variably-trained participants mark the singular with a range of frequencies during individual testing, approximately centred round the training proportion (2/3); however, during interaction these participants rapidly align with their categorical partners: 11 of 12 variably-trained participants in the 100-66 condition mark the singular in all trials during the second block of interaction, leading to very low within-pair difference scores.

In the 100-33 condition, we see a slightly different picture: while the majority of categorically-trained participants remain categorical throughout, some do become variable during interaction, and two participants become categorical non-users in the second block of interaction. Variable users in the 100-33 condition exhibit a spread of responses during individual testing; during interaction, three participants accommodate upwards to become categorical, some become categorical non-users (in two cases, leading to dropping of the marker by their partners), and several remain variable users throughout. The level of alignment between categorical and variable users is generally low, as can be seen in from the high within-pair difference scores.

Finally, in the 66-33 condition we see variable responses during individual testing, and rapid alignment during interaction, leading to low within-pair difference scores. Pairs tend to align on either systematic use (4 pairs) or non-use (6 pairs), with an overall preference for non-use reflected in the low mean marking of singulars during interaction.

The statistical analyses below seek to answer three questions. Firstly, did participants *probability match* during individual testing, i.e. reproduce the marker frequency they were trained on? Secondly, did participants align during interaction and if so, did they align more closely in some conditions than others? Finally, and in relation to the specific hypothesis outlined in the Introduction, did categorically- and variably-trained participants differ in the extent or manner in which they accommodated to their partner? Given non-normality of data, non-parametric statistics are used throughout.

Probability matching during individual testing

Participants were grouped according to the frequency with which they saw the singular marked during training, yielding three groups of 24 participants (100%, 66% or 33% marking). The proportion of marked singulars produced during individual testing by each participant was calculated, and converted to a difference score by subtracting this proportion from the proportion the participant was trained on (see Fig. 3a). Participants trained on 100% marking reliably marked all singulars during individual testing, in accordance with their training data. 66%-trained participants produced a distribution of marker usage statistically indistinguishable from probability matching (assessed by one-sample test, $V=139$, $p=.988$); however, 33%-trained participants under-produce marked singulars ($V=68.5$, $p=.035$). While this suggests that

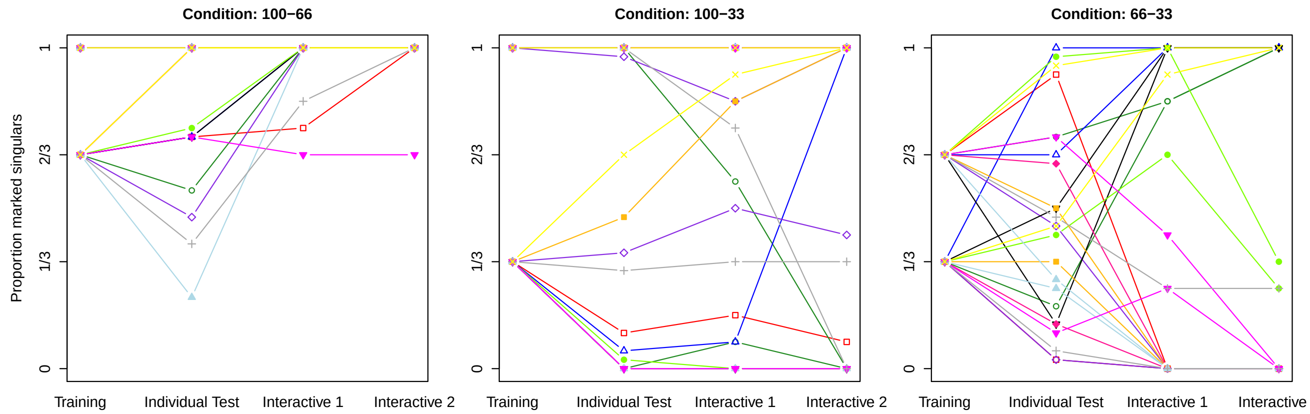

Figure 1: Proportion of trials in which the singular was marked, in training (determined by condition), individual testing, and across two blocks of interaction. Each pair is represented by two lines, one per participant, sharing the same colour and shape: alignment between participants is therefore reflected in lines of matching colour/shape converging.

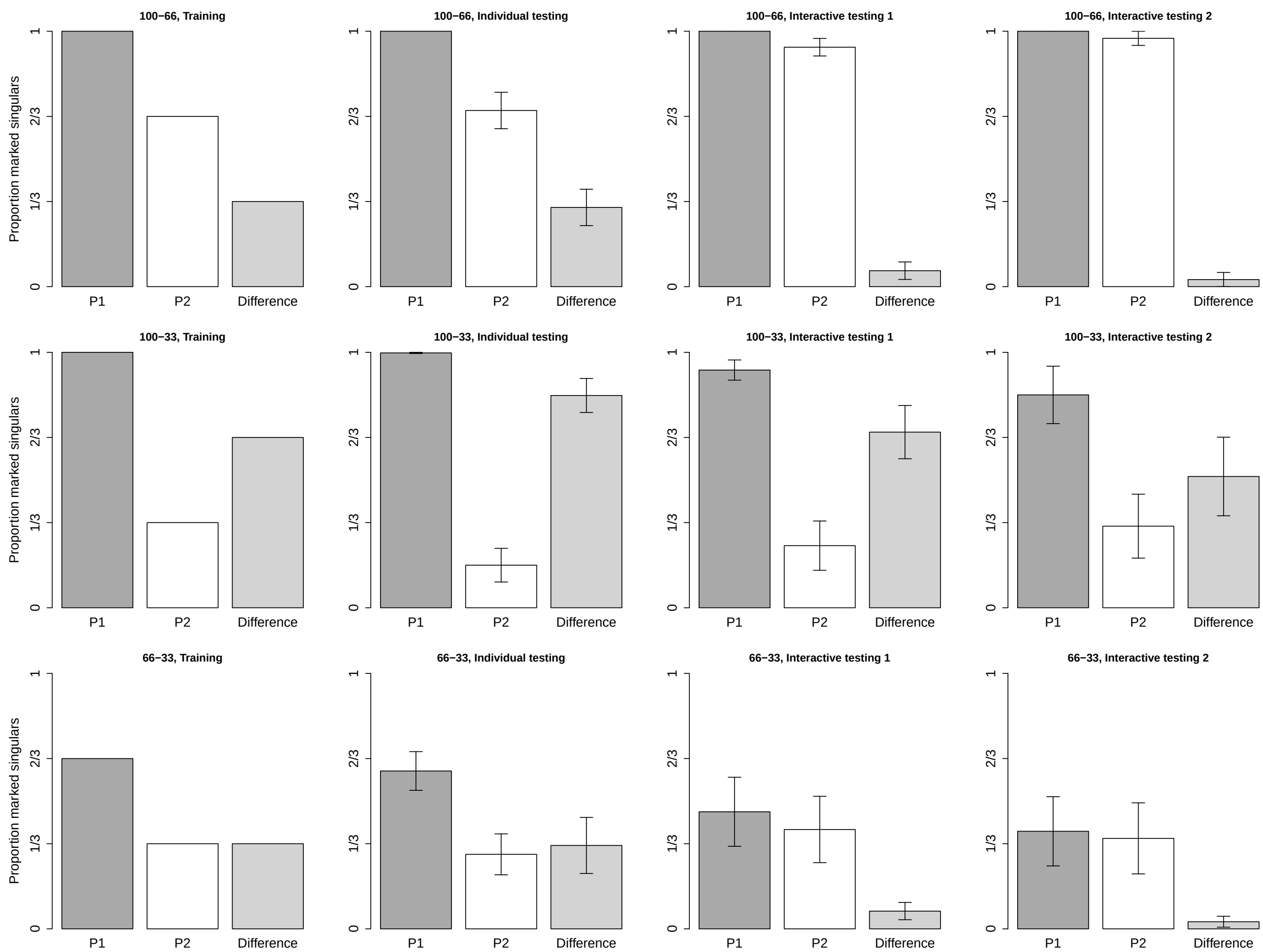

Figure 2: Mean proportion of trials in which the singular was marked (+/- SEM) in training, individual testing, and across two blocks of interaction. Difference bars indicate the mean of the within-pair difference (i.e. the difference between P1's proportion of marked singulars and that produced by P2).

low-frequency variable markers might be more prone to elimination than high-frequency markers, the direct comparison between the difference scores from the 66% and 33% conditions yields no significant difference (W=366, p=.109), suggesting that caution is required in this interpretation.

Alignment during interaction

We use the within-pair difference scores provided in Fig. 2 as our measure of alignment, low difference scores indicating high alignment. Taking each condition separately and comparing difference scores at individual testing (i.e. before interaction) to those in the second block of interaction, differences significantly reduce (i.e. participants align to a statistically significant extent due to interaction) in all conditions (100-66: V=44, p=.013; 66-33: V=69, p=.016; 100-33: V=33, p=.04; all comparisons remain significant after Holm correction for multiple comparisons). Looking across conditions, the three conditions differ in their levels of alignment at individual testing (Kruskall-Wallis χ^2=16.692, p<.001) and in the second block of interaction (Kruskall-Wallis χ^2=7.482, p=.024). These differences are driven by significant differences between the 100-33 condition and the other conditions (100-66 against 100-33 at individual testing: W=9, p<.001; 66-33 against 100-33 at individual testing: W = 129.5, p<.001; 100-66 against 100-33 at interactive testing: W = 38, p = .025; 66-33 against 100-33 at interactive testing: W = 104, p = .041). This accords with the impression from Figs. 1 and 2 that there is less alignment in the 100-33 condition, possibly due to the larger difference in frequency of marked singulars in training.

Accommodation

Accommodation is characterised, in this experiment, by systematically changing from one's personal estimate of the frequency of singular marking during training (as evidenced during individual testing) towards the frequency used by one's partner. We quantify this by calculating, for each participant, the difference between the proportion of marked singulars produced during individual testing and the proportion produced during the second block of interaction (subtracting the former from the latter: positive change indicates increased marker use). These values are shown in Fig. 3b. The hypothesis outlined in the Introduction is that categorically-trained participants will not accommodate downwards to variable-trained participants, but that variable-trained participants will accommodate upwards to categorically-trained participants.

A statistical analysis supports this hypothesis, most strongly in the 100-66 condition. Collapsing across the 100-66 and 100-33 conditions, variably-trained participants change significantly more than categorically-trained participants (W = 159, p = .003); furthermore, this is a reliable *increase* in marker usage (change significantly greater than 0, V=126.5, p=.003). Comparison with the 66-33 condition shows that it is not the case that variably-trained participants tend to increase their marker usage (collapsing across all participants in the 66-33 condition, change in marker usage between individual testing and interaction is not significantly different from 0, V=103, p=.294), nor do participants who are paired with a partner who was trained on a higher frequency of marker usage automatically increase their usage (considering only the 33%-trained individuals from the 66-33 condition, change not significantly different from 0, V=30, p=.824). A three-way comparison of P2s across all three conditions reveals a difference in amount of change (Kruskal-Wallis χ^2 = 8.586, p=.014); post-hoc tests show that the difference between P2s in the 100-66 and 100-33 conditions is not statistically significant (W = 92, p=.251), whereas the other contrasts are at least marginally significant (P2s in 100-66 vs P2s in 66-33: W = 107.5, p=.043; P2s in 100-33 vs P2s in 66-33: W = 105.5, p=.054). Finally, while the lower-frequency-trained P2 participants in the 100-66 and 100-33 conditions do not change by different amounts according to this comparison, and collectively they increase their frequency of marking, this effect seems smaller in the 100-33 condition: testing the two groups separately, P2s in the 100-66 condition change by an amount significantly greater than 0 (V = 44, p=.013), whereas P2s in the 100-33 condition do not (V = 23, p=.15). Again, this suggests that the large difference in training frequencies in the 100-33 condition modulates the tendency for variably-trained participants to accommodate upwards to their categorically-trained partners.

Discussion

Our experiment provides support for both our specific hypothesis that accommodation between categorically- and variably-trained participants is asymmetric (variably-trained participants tend to accommodate upwards to their partners, while categorically-trained participants tend not to accommodate downwards) and that, more generally, alignment during interaction leads to the loss of linguistic variation (there is a striking loss of variable marker usage even in the 66-33 condition, where both participants were trained on variable systems). Additionally, our data suggests that the extent to which participants accommodate may be modulated by the extent to which they differ linguistically: while participants in the 100-33 condition do align somewhat through interaction, the level of alignment is far less than in the 100-66 and 66-33 conditions, where participants are trained on more similar languages. A useful avenue for future work would be to identify the critical distance threshold for alignment.

A number of other questions remain to be addressed. Firstly, we have only considered presence/absence variation: other paradigms (Hudson Kam & Newport, 2009; Smith & Wonnacott, 2010) look at variation where there are two or more overt markers for a single function, and it may be that alignment during interaction proceeds differently in such cases. Secondly, we look only at alignment within pairs who undergo a relatively short period of training and a relatively long, intense period of interaction with a single partner: since the real-world case involves longer learning (perhaps entailing greater commitment to the trained system) and interaction with a wider range of partners, this seems like a worth-

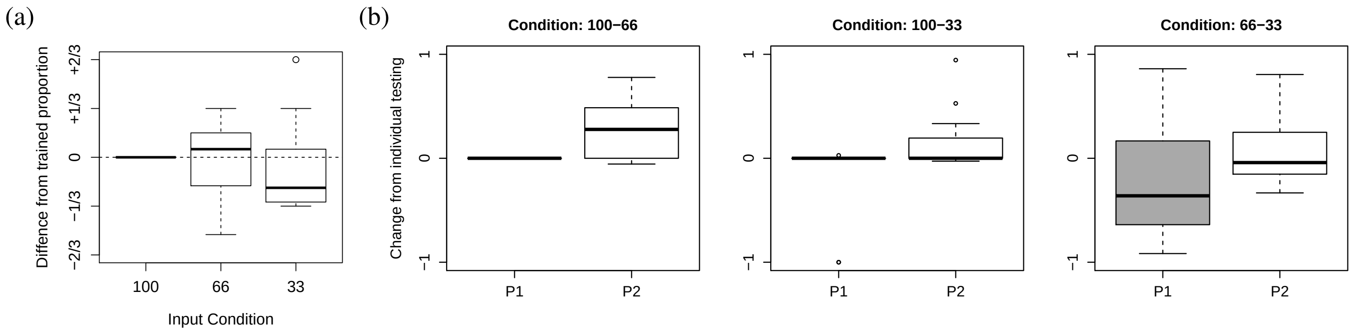

Figure 3: (a) Difference between the proportion of marked singulars participants encountered during training and the proportion produced during individual testing. (b) Difference between the proportion of marked singulars produced during individual testing and the proportion produced during the second block of interaction, giving a measure of accommodation.

while scenario to explore experimentally. Finally, accommodation is surprisingly rapid in our study, with a great deal of alignment taking place in the first few trials of interaction: it would be intriguing to investigate the lower-level processes by which participants make judgements about how to use marking on the basis of one or two exposures to the marking behaviour of their partner, and how these might be influenced by modifying a participant's knowledge of the language used by their partner (e.g. by manipulating whether participants were trained entirely in isolation, as here, or together).

Conclusions

Alignment during interaction leads to the elimination of unpredictable variation, and consequently provides a second (complimentary) mechanism by which the absence of unpredictable variation in natural language might be explained. Furthermore, as suggested by the historical literature, accommodation during interaction is inherently asymmetric: while variable users can accommodate to their partner by increasing their frequency of usage, categorical users tend not to accommodate to their variable partners by becoming variable. As such, once a grammatical marker reaches a critical threshold in a population such that many individuals are variable and some are categorical, alignment during interaction should drive the population towards uniform categorical marker use.

References

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*, 489–509.

Coupland, N. (2010). Accomodation theory. In J. Jaspers, J. Verschueren, & J. Östman (Eds.), *Society and language use* (pp. 21–27). Amsterdam: Benjamins.

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, *122*, 306-329.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*, 429-492.

Fehér, O., Ritt, N., Smith, K., & ten Wolde, E. (2014). *The spread of (in-)definiteness marking in Early English: reconstructing category emergence in the lab.* (Paper presented at the 18th International Conference on English Historical Linguistics, Leuven, Belgium)

Givón, T. (1985). Function, structure, and language acquisition. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition* (Vol. 2, pp. 1005–1028). Hillsdale, NJ: Lawrence Erlbaum.

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, *59*, 30-66.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169-225.

Reali, F., & Griffiths, T. L. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, *111*, 317–328.

Rissanen, M. (1967). *The uses of "one" in old and early middle english*. Helsinki: Societé Néophilologique.

Singleton, J. L., & Newport, E. L. (2004). When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive Psychology*, *49*, 370–407.

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*, 444-449.

Sommerer, L. (2012). Investigating the emergence of the definite article in Old English: About categorization, gradualness and constructions. *Folia Linguistica Historica*, *33*, 175–213.