Title
A Naturalness Gradient Shapes the Learnability and Cross-Linguistic Distribution of Morphological Paradigms

Permalink
https://escholarship.org/uc/item/8wv9f99x

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors
Saldana, Carmen
Herce, Borja
Bickel, Balthasar

Publication Date
2022

Peer reviewed

A Naturalness Gradient Shapes the Learnability and Cross-Linguistic Distribution of Morphological Paradigms

Carmen Saldana (carmen.saldanagascon@uzh.ch)
Borja Herce (borja.hercecalleja@uzh.ch)
Balthasar Bickel (balthasar.bickel@uzh.ch)
Department of Comparative Language Science, University of Zurich, Switzerland
Center for the Interdisciplinary Study of Language Evolution, University of Zurich, Switzerland

Abstract

As efficient systems of communication, languages are usually expected to map meanings to forms in a one-to-one way, using for example the same affix form (e.g., *-s* in English) every time a particular meaning is intended (e.g., plural number), and placing affixes with the same meaning consistently in the same position (e.g., always suffixal). Forms and positional rules extending over contexts with a common meaning (e.g., plural in 1PL, 2PL, 3PL) are thus considered *natural*, and those extending over contexts with no consistent common meaning (e.g., 1PL and 3SG) are considered *unnatural*. Natural patterns are most common cross-linguistically, and most learnable in experiments; however, little is yet know about differences between unnatural classes. In this study we explore syncretism (i.e., use of the same form in different functions) and affix position in the domain of person and number agreement in verbs, both cross-linguistically and in artificial language learning experiments. Results from the two approaches and both phenomena converge in finding a gradient of (un)naturalness. Rather than a dichotomous natural/unnatural distinction, we found that both cross-linguistic frequency and learnability are proportional to the amount of shared feature values among the contexts requiring the same form or position. We argue that a cognitive bias towards *similarity-based structure* explains our experimental results and could be driving the patterns observed in natural languages.

Keywords: artificial language learning; linguistic typology; natural class; semantic similarity; morphology; paradigm split

Table 1: Different types of patterns of syncretism in person-number verbal paradigms. Natural, L-type and X-type patterns are illustrated with grey cells in these examples from Dutch (Glottocode: dutc1256), Hindi (hind1269) and Kapau (kapa1251) respectively. In natural patterns, all syncretic forms share a feature value; in L-patterns, syncretic forms share a feature between all but one pair of cells; in X-patterns, more than one pair lacks shared feature values.

	NATURAL PATTERN		L-TYPE PATTERN		X-TYPE PATTERN	
	Dutch *come* PRS		Hindi *be* FUT.F		Kapau *ford water* PST	
	SG	PL	SG	PL	SG	PL
1	kom	komen	hūṃgī	hoṃgī	qäkamanga	qäkamango
2	komt	komen	hogī	hogī	qäkamangn	qäkamanga
3	komt	komen	hogī	hoṃgī	qäkama	qäkamanga

Introduction

Syncretism is a widespread phenomenon in the morphology of languages whereby distinct inflectional values are expressed by a shared form (Baerman, Brown, & Corbett, 2005). It manifests itself in highly diverse ways cross-linguistically. In Dutch, for example, verbal paradigms (see Table 1) normally take one form for all plural person values, another for first singular (1SG) and another one for second and third singular (2SG = 3SG). In this case, all syncretic forms have at least one value in common, either singular (SG) or plural (PL). However, syncretism is not always so orderly. It is not rare to find syncretic forms that lack any common value. For example, Table 1 shows that the paradigm of the verb *to be* in Hindi (McGregor, 1995) contains a shared form across second person (both singular and plural) and third person singular (i.e., 2 = 3SG); and in Kapau (Oates & Oates, 1968), the first person singular shares a form with the second and third person plural forms (i.e., 1SG = 2PL = 3PL). Syncretisms such as those described for Dutch are often referred to as *natural* because all the cells involved share at least one value (Jakobson, 1936), and those of Hindi and Kapau are referred to as *unnatural* because they do not.

Despite this variation in the possibilities of syncretism, certain cross-linguistic tendencies are apparent: The most recurrent types are the natural ones (Cysouw, 2003; Pertsova, 2007). There is a growing body of literature suggesting that this cross-linguistic preference for natural patterns replicates a cognitive bias in language learning (e.g., Pertsova, 2014; Nevins, Rodrigues, & Tang, 2015), that is, a bias towards patterns of syncretism with shared feature values during language learning. This learning bias is in turn taken to shape how languages evolve over time and space, yielding the cross-linguistic preferences we now see (Culbertson & Smolensky, 2012; Kirby, Griffiths, & Smith, 2014; Bickel, 2015; Blythe & Croft, 2021). The idea is that during linguistic evolution, the spread of new variants is subject to the same biases as those manifest in adult language learning (Blythe & Croft, 2021). The bias towards natural patterns resonates with a more general bias favouring categories that comprise closer and more similar meanings in word learning (e.g., Xu & Tenenbaum, 2007) and in concept learning more broadly (e.g., Goodman, Tenenbaum, Feldman, & Griffiths, 2008). We will refer to this learning preference as a bias towards *similarity-based structure*, where similarity is defined in proportion to the amount of shared feature values within inflectional patterns.

However, unnatural patterns come in large variety and little is known about any preferences within these. There is no a priori reason to believe that all unnatural patterns must be equal in their cross-linguistic recurrence or in terms of

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*.

Table 2: Different types of patterns of positional splits in person-number verbal agreement paradigms. A positional arrangement is defined by a specific type and number of positions that are occupied by a specific set of agreement affixes (bolded). Natural, L-type and X-type patterns are illustrated with grey cells in these examples from Gumer (Glottocode: gume1239), Koasati (koas1236) and Basque (basq1248) respectively. In natural patterns, all forms with the same positional arrangement share a feature value (e.g., prefix for SG and circumfix for PL in Gumer); in L-patterns, one pair of cells with the same positional arrangement lacks shared feature values; in X-patterns, more than pairs of cells lack shared feature values.

	NATURAL PATTERN		L-TYPE PATTERN		X-TYPE PATTERN	
	GUMER		KOASATI		BASQUE	
	open IPFV		*hear* ACT		*walk* PRS	
	SG	PL	SG	PL	SG	PL
1	**ə-**kəft	**nɨ-**kəft**-ɨnə**	há:lo**-l**	**il-**há:l	**na-**bil	**ga-**bil**-tza**
2	**tɨ-**kəft	**tɨ-**kəft**-o**	**is-**há:l	**has-**há:l	**za-**bil**-tza**	**za-**bil**-tza-te**
3	**yɨ-**kəft	**tɨ-**kəft**-o**	há:l	há:l	**da-**bil	**da-**bil**-tza**

learnability. In a cross-linguistic survey (Baerman, Brown, & Corbett, 2002) we find that unnatural paradigms in which syncretism occurs between cells in an L-shape like the one in Hindi (L-type patterns hereafter) are more frequent (32/36 patterns) than those containing X-shaped syncretism (X-type patterns hereafter) like the one in Kapau (4/36 patterns). Furthermore, we find this cross-linguistic asymmetry between the two unnatural patterns with shared morphology more generally (62 L-type vs 12 X-type patterns; Herce, 2020), that is, not only with cases of whole-word syncretism (where both stem and affixes are shared, as in the examples so far) but also with partial syncretism where only sub-parts of the word (e.g., stem, affixes, etc.) are shared.

In this study we aim to test whether this gradient of cross-linguistic recurrence is mirrored in a gradient of learnability. We hypothesise that the learnability of unnatural patterns is proportional to their similarity-based structure; that is, patterns of syncretism like those of Hindi in Table 1 are easier to learn than those of Kapau because features values are shared across more forms. Although both contain cells that differ in all feature values, there are more of these pairs of cells in the X-type pattern in Kapau (1SG=2PL, and 1SG=3PL) than in the L-type pattern in Hindi (3SG=2PL).

This hypothesis is borne out in two artificial language learning experiments. In a first experiment we contrast the learnability of (un)natural patterns in terms of forms that are shared vs not shared across the cells of an artificial language (i.e., syncretism). Consistent with our hypothesis, we find that L-type unnatural patterns are easier to learn than X-type unnatural patterns, and that natural patterns are the easiest to acquire. In a second experiment, we test the generalisability of this learnability gradient to another morphological phenomenon: affix position (Bickel et al., 2007; Crysmann & Bonami, 2016; Mansfield, Stoll, & Bickel, 2020). We contrast the same (un)natural patterns but now in terms of positional arrangements that share vs do not share feature values (see Table 2). Results are again consistent with the hypothesised gradient *natural* $\gg$ *L-type* $>$ *X-type*, and they are also consistent with frequency trends in a cross-linguistic survey

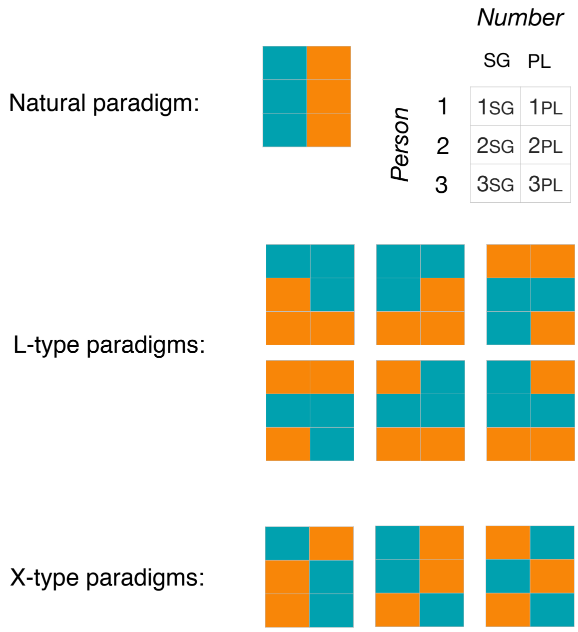

Figure 1: Paradigm splits in the three critical experimental conditions. The grids represent person-number paradigms with a binary number feature (singular and plural) and a ternary person feature (1st, 2nd and 3rd), and each cell is thus a different person-number bundle. Each colour (blue and orange) represents a different agreement affix *form* in Experiment 1, and a different *position* of the agreement affix (prefix or suffix) in Experiment 2.

that we perform. This confirms the notion of a general bias towards similarity-based structure in morphological learning, beyond word and concept learning.

Experiment 1: Syncretism

Materials and Methods

We use an ease-of-learning paradigm where we train and test participants on person-number verb subject agreement paradigms containing different patterns of (whole-word) syncretism and compare how accurately they learn them within 60 trials. Person is defined as a ternary feature (i.e., containing 1st, 2nd, and 3rd person) and number is defined as a binary feature (i.e., containing SG and PL). We ran three experimental conditions with varying degrees of naturalness within the syncretic patterns in the verbal agreement paradigms: natural, L-type or X-type patterns (see Figure 1). Person-number agreement is marked via suffixation, a single suffix that marks both person and number cumualtively. Each agreement paradigm contains only two different suffixes, each present in half of the cells (i.e., three cells will be inflected with one suffix, and the other three with the other suffix). These two agreement suffixes thus constitute two patterns of syncretism within the paradigm and will partition the person-number space according to the experimental condition as illustrated in Figure 1, where each cell colour represents a different verbal agreement affix. Natural paradigms have only one configuration: they contain an agreement suffix for singular and another for plural cells. L-type paradigms can have six different configurations. An example of an L-type paradigm could contain one suffix for 1SG, 1PL and 2PL (i.e., 1=2PL) and another for 2SG, 3SG and 3PL (i.e., 2SG=3). The X-type paradigm has three different configurations; for example, it could have one suffix for 1SG, 2PL and 3PL and another for 1PL, 2SG and 3SG (see Figure 1).

We ran an additional condition without syncretism where agreement with each of the six bundles of person-number feature values in the paradigm is marked by a different affix. This condition is the least ambiguous as each cell is marked via a unique affix and does not require the learner to induce any further category based on specific morphological features; however they require the learner to acquire six different affixes instead of the two that need to be learned elsewhere. The inclusion of this condition allows us to test under which circumstances paradigms containing patterns of syncretism can be easier to acquire.

Transparency All experimental materials and data reported for Experiment 1 are available at osf.io/jpum6/ (Saldana, Herce, & Bickel, 2022, February 9), and the preregistered design and analysis plan is accessible also at osf.io/pwqjg (Saldana, Herce, & Bickel, 2021, August 17).

The Artificial Lexicon The artificial lexicon in the experiment comprises six subject pronouns, three lexical verbs and two verbal agreement suffixes (or six in the non-syncretic condition). The semi-nonce subject pronouns (inspired by Tok Pisin; Glottocode tokp1240) are composed of the person stems *mi* (1st person), *yu* (2nd person), *le* (3rd person), followed by the number suffixes -∅ (SG) or *-pela* (PL). The semi-nonce lexical verbs (based on Basque) are *gidatu*, *igeri* and *oineza* which correspond to 'to cycle', 'to swim' and 'to walk' respectively. In the conditions with syncretic patterns, the two verbal agreement suffixes are randomly selected from an array of four CV syllables {*-na* , *-gu*,*-te*, *-po*}. The suffixes are randomly assigned to person-number agreement bundles according to the condition and specific configuration(see Figure 1). In the non-syncretic condition, each of the six person-number agreement suffixes is different and is randomly mapped to a CV syllable from the array {*-na* , *-gu*, *-te*, *-po*, *-ki*, *-soo*}. The artificial language drops subject pronouns and thus a full sentence meaning 'they walk' is realised only as a verbal form, for example, *oinezagu*, where *oineza* is the verbal stem and *-gu* the 3PL agreement suffix.

Experimental procedure The experimental procedure is divided into two phases. In the first phase, we train and test participants on the artificial lexicon without verbal agreement, that is, only on the pronominal forms and the uninflected lexical verbs (i.e., in isolation without agreement suffixes). In each training trial, participants see an image of an action or a pronoun, and their corresponding forms in the artificial language. In each testing trial, participants see an image and are asked to select the corresponding form in the artificial language out of an array of two, that is, the target, and a randomly selected form of the same lexical category (pronoun or verb) as the target. They will receive feedback after each selection, and see each form-image mapping three times during training and twice during the vocabulary testing.

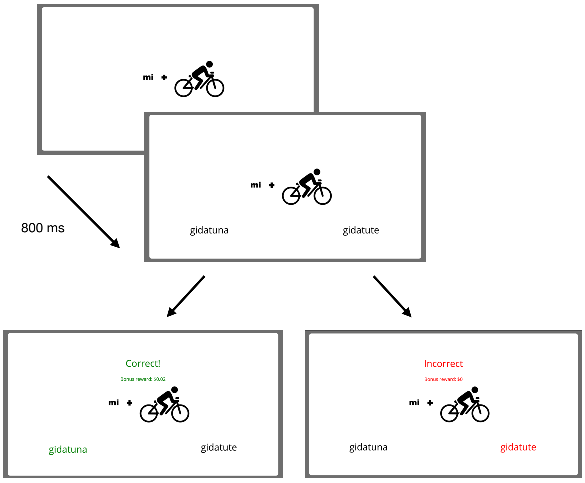

Figure 2: Example test trial in the critical phase in Experiment 1. Participants are shown an image of a pronoun+action combination and are asked to select the corresponding inflected (verb-affix) form of the verb in the artificial language out of an array of two. They receive feedback on whether their choice is correct as well as the bonus amount accumulated. In Experiment 1, the two alternative forms are the only two verbal forms contained in the paradigm and they only differ in the affix form (same stem, different ending). In Experiment 2, however, the alternative forms differ in the affix position (same affix form, either prefix or affix).

In the second and critical phase, we test participants on the verbal paradigms with the agreement suffixes. For this phase, we use feedback learning whereby the same testing trials will serve as training. In each critical testing trial, participants see an image combining a pronoun in the artificial language and an action and after a second, the only two potential verbal forms within a given paradigm (same stem, different affixes) are displayed (see Figure 2). Participants have to select which form they think is the one that corresponds to the specific pronoun+action combination, in other words, they have to select the verbal form they think agrees in person and number with the given pronoun. They receive feedback on their selection so they can learn the correct correspondence as they move along testing. This phase comprises ten blocks of six trials, each containing all six different person-number agreement bundles.

Participants We recruited 405 participants through Amazon Mechanical Turk for a 20-minute long session. Participants were all over 18 years old, based in the US and had approval ratings of $> 95\%$. We excluded the data from participants who failed to provide at least 80% of correct responses in the second block of vocabulary testing during the training phase ($N = 59$), and participants who responded that they had used pen and paper during the study ($N = 0$) in a post-experimental questionnaire. After exclusions, our anal-

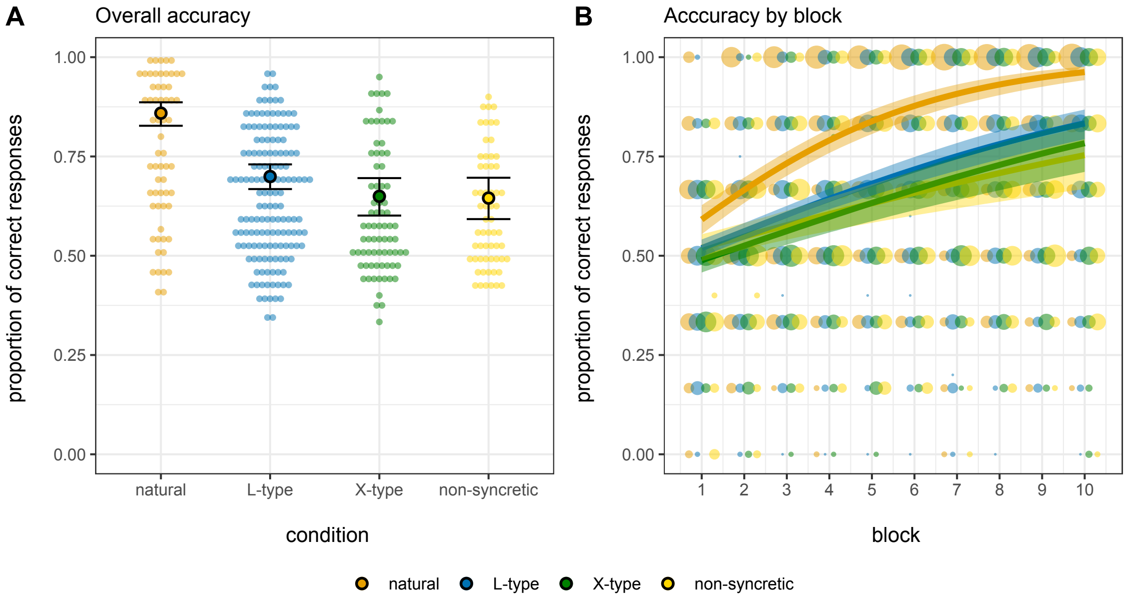

Figure 3: Accuracy scores and model estimates in Experiment 1. (A) Overall accuracy by condition. Shaded dots represent participants' individual scores; black-circled dots represent the model's predicted mean accuracy scores and the error bars represent the model's predicted 90% credible intervals. (B) Accuracy by testing block for each of the four conditions. Shaded dots represent participants' individual scores, and larger dots represent more individuals; thick lines represent the model's predicted accuracy means conditioned on experimental condition and block. The shaded area shows the 90% credible intervals.

ysis contains the data from 60, 61, 150 and 75 participants in the non-syncretic, natural, L-type and X-type conditions respectively. Participants within L-type and X-type conditions were distributed evenly across the different paradigm configurations; we recruited 25 participants for each of the 6 and 3 paradigm configurations within L-type and X-type respectively (see Figure 1). Participants were paid a base rate of $2.5 plus they received a bonus of $0.02 for each correct response (maximum bonus reward = 1.56).

Data Analysis We use *R*'s *brms* (Bürkner, 2018) as an interface to *RStan* (Stan Development Team, 2021) to fit a Bayesian logistic regression model predicting participants' performance by condition and test block. Our dependent variable is participants' responses for each of the 60 critical test trials (coded as 1 if correct, and 0 if incorrect). As fixed effects, we include Condition (natural, L-type, X-type and non-syncretic) and Block as well as their interaction. The categorical predictor Condition is Helmert contrast-coded comparing X-type to non-syncretic, L-type to the average of the two, and natural to the average of all the other conditions; Block is coded as a centered continuous variable. As random effects, we included intercepts for participants as well as by-participant slopes for the effect of Block. We set the same Student-*t* prior on all fixed effects and the intercept ($DF = 6, \mu = 0, \sigma = 1.5$); for the random effects, we set a half-Cauchy prior with a scale parameter 10.

Results

Based on our hypothesis, we predict natural patterns of syncretism to be the most learnable, and within unnatural patterns, we expect L-type patterns to be easier to acquire than X-type. We further expect that non-syncretic paradigms will be harder to learn than (at least) natural patterns given that they contain more forms. Our experimental results are consistent with our hypothesis (see Figure 3). We find that participants in the natural condition have higher accuracy scores than the average of all other conditions ($\hat{\beta} = 0.280$, 90%CI = $[0.214, 0.347]$, $SE = 0.040$, $P(\hat{\beta} > 0) = 1$). We also find that participants in the L-type condition score higher than those in the X-type and non-syncretic conditions ($\hat{\beta} = 0.079$, 90%CI = $[0.005, 0.150]$, $SE = 0.044$, $P(\hat{\beta} > 0) = 0.962$); and we find no difference between accuracy scores in X-type and non-syncretic conditions ($\hat{\beta} = 0.010$, 90%CI = $[-0.145, 0.161]$, $SE = 0.093$, $P(\hat{\beta} > 0) = 0.542$). The model's results also inform us about the change in accuracy over time. We find very strong evidence in favour of an increase in accuracy by block of testing ($\hat{\beta} = 0.188$, 90%CI = $[0.166, 0.212]$, $SE = 0.014$, $P(\hat{\beta} > 0) = 1$). This increase is higher in the natural condition than in the other conditions on average ($\hat{\beta} = 0.044$, 90%CI = $[0.030, 0.059]$, $SE = 0.009$, $P(\hat{\beta} > 0) = 1$). Participants in the L-type condition also seem to improve their accuracy by block slightly more than participants in the X-type or non-syncretic condition ($\hat{\beta} = 0.013$, 90%CI = $[-0.002, 0.028]$, $SE = 0.009$, $P(\hat{\beta} > 0) = 0.926$). Finally, we do not find a strong difference between X-type and non-syncretic conditions in regards of their increase in accuracy by block ($\hat{\beta} = 0.016$, 90%CI = $[-0.015, 0.048]$, $SE = 0.019$, $P(\hat{\beta} > 0) = 0.809$).

Experiment 2: Positional splits

In Experiment 1 we provide evidence for a learnability gradient *natural* $\gg$ *L-type* $>$ *X-type* of paradigmatic splits based on syncretic forms, which mirrors the cross-linguistic trend. This gradient, however, need not be specific to syncretism and might be a general property of how forms are distributed over feature values. In order to explore this possibility, we looked at where agreement morphemes are positioned in verb forms. Specifically we explore cases where different person-number bundles are arranged in different positions (e.g., prefix or suffix) across forms within the paradigm (as illustrated above in Table 2): We refer to these paradigmatic splits by positional arrangment as *positional splits*.

Data from 227 languages from 97 different families—based on AUTOTYP (Bickel et al., 2017) plus additional data collected from the WALS 100-language sample (Dryer & Haspelmath, 2013)—show that the majority of agreement paradigms (subject and object) only require reference to a single position (e.g., only suffixation), thus obeying the principle of *category clustering* (Mansfield et al., 2020). A sizeable minority (128 paradigms, 39.38%), however, require reference to two or more positions and show splits whereby different person-number markers appear in different positions. Within positional splits, we found 44 natural, 73 L-type, and 24 X-type patterns. These raw numbers need to be interpreted relative to baseline expectations since each type has a different probability of occurring by chance (e.g., there are less logically possible configurations of natural patterns than L-type patterns). In response to this we fitted Bayesian mixed-effects models comparing the natural occurrences of each pattern type (natural, L-type and X-type) to the occurrences we would expect by chance from all logically possible configurations in person-number 3×2 paradigms (with language and family as random intercepts). We find that the natural patterns (illustrated by Gumer in Table 2; Völlmin, 2017) are over-represented in natural languages (baseline vs natural: $\hat{\beta} = -0.679$, 90%CI = [-1.051,-0.304], $SE = 0.230$, $P(\hat{\beta} < 0) = 1$). The most unnatural X-patterns (e.g., Basque in Table 2; Hualde & De Urbina, 2011) in turn, are under-represented in natural languages (baseline vs natural: $\hat{\beta} = 1.250$, 90%CI = [0.791,1.735], $SE = 0.290$, $P(\hat{\beta} > 0) = 1$). Intermediate-naturalness L-patterns (e.g., Koasati in Table 2; Kimball, 1985) occur with a similar frequency as expected by chance (baseline vs natural: $\hat{\beta} = 0.035$, 90%CI = [-0.320, 0.390], $SE = 0.217$, $P(\hat{\beta} < 0) = 0.56$). These results are consistent with the *natural* $>$ *L-type* $>$ *X-type* gradient that we find in the worldwide distribution of syncretism and that we confirmed in Experiment 1. In order to test whether the gradient in positional splits also appears in artificial learning, we ran a replication of Experiment 1 on them.

Materials and Methods

We use the same ease-of-learning paradigm as in Experiment 1. We train and test participants on person-number verb subject agreement paradigms containing different patterns of positional splits and compare how accurately they learn them. We ran the same critical experimental conditions with varying degrees of naturalness within paradigmatic splits: natural, L-type or X-type patterns (see Figure 1). Person-number verbal agreement is marked together in a single affix and can appear in a different position (either as suffix or prefix). Each agreement paradigm contains only two different positional arrangements of person-number affixes, each present in half of the cells (i.e., three cells will be inflected via suffixation, and the other three via prefixation). These two positional arrangements will split the person-number space according to the experimental condition as illustrated in Figure 1, where each cell colour would here represent a different positional arrangement (either as suffix or prefix). We ran a further condition where we taught participants a system of person-number agreement where all markers were either suffixes or prefixes, that is, without a positional split: We expect this condition to be the most learnable as all that needs to be learned is whether agreement is prefixal or suffixal for the whole paradigm. The experimental procedure was identical to that of Experiment 1. The preregistered design and analysis plan is accessible at osf.io/yzcxp (Saldana, Herce, & Bickel, 2022, January 21).

The Artificial Lexicon The artificial lexicon in Experiment 2 comprises six pronouns, three lexical verbs and six person-number agreement markers. The pronouns and verbs are the same as in Experiment 1 (except we use the artificial verbs *figeri* and *moineza* instead of *igeri* and *oineza* so all stems start with a consonant and end with a vowel). The agreement markers are selected from an array of six CV syllables {*na*, *gu*, *te*, *po*, *ki*, *so*}, and randomly assigned to each of the six person-number bundles. These markers can be either suffixes (in three cells) or prefixes (in the other three cells); these correspond to the two different colours in the experimental conditions shown in Figure 1. In the no-split condition, however, all the agreement markers are in the same position (i.e., either suffixes or prefixes).

Participants We recruited 247 participants as per Experiment 1, each randomly assigned to an experimental condition: 65, 60, 62 and 62 participants in the no-split, natural, L-type, and X-type conditions respectively.

Data Analysis We fit a Bayesian logistic regression model as per Experiment 1. Our dependent variable is participants' responses for each of the critical test trials (coded as 1 if correct, and 0 if incorrect). As fixed effects, we include Condition, Block, and their interaction. We apply Helmert contrast coding to the categorical predictor of Condition: We compare L-type to X-type, natural to the average of the two, and no-split to the average of all the rest. Priors are set like in Experiment 1.

Results

Our experimental results confirm the predicted gradient of learnability natural $\gg$ L-type $>$ X-type (see Figure 4), thus replicating the gradient obtained in Experiment 1 for a dif-

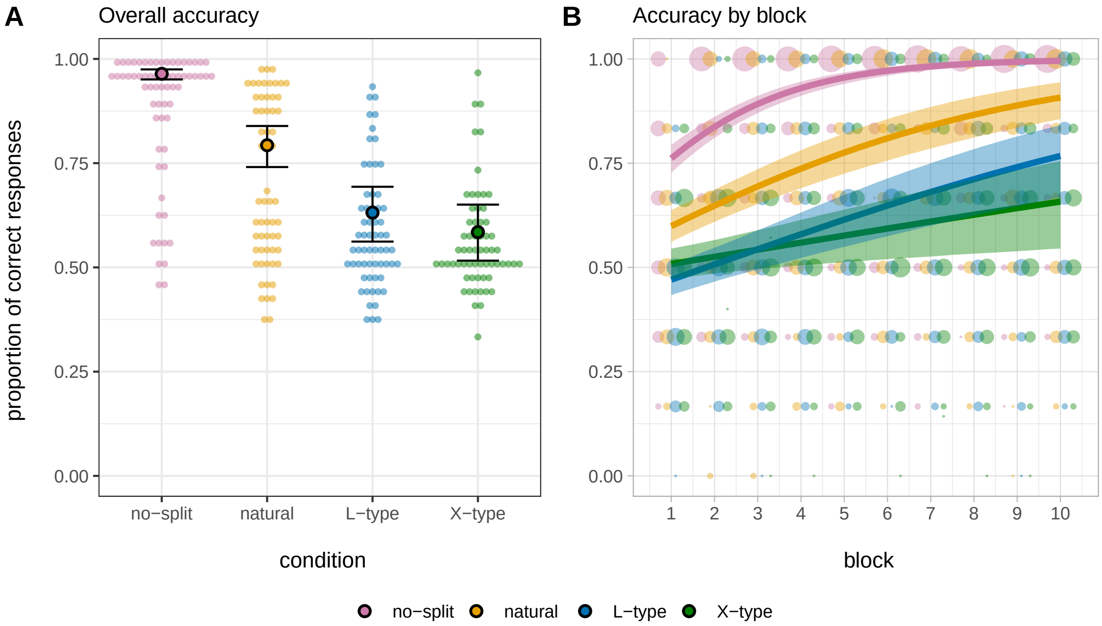

Figure 4: Accuracy scores and model estimates in Experiment 2. (A) Overall accuracy by condition. Shaded dots represent participants' individual scores; black-circled dots represent the model's predicted mean accuracy scores conditioned on experimental condition, and the error bars represent the model's predicted 90% credible intervals. (B) Accuracy by testing block for each of the four conditions. Shaded dots represent participants' individual scores, and larger dots represent more individuals as per the legend; thick lines represent the model's predicted accuracy means conditioned on experimental condition and block. The shaded area shows the 90% credible intervals.

ferent morphological phenomenon. They further comfirm that paradigms obeying to the principle of *category clustering* (i.e., no-split) are thes easiest to acquire. We find that accuracy scores for L-type and X-type are similar at the intercept ($\hat{\beta} = 0.097$, 90%CI = [-0.100, 0.298], $SE = 0.122$, $P(\hat{\beta} > 0) = 0.786$) but they increase more by block in L-type than in X-type ($\hat{\beta} = 0.038$, 90%CI = [0.005, 0.072], $SE = 0.021$, $P(\hat{\beta} > 0) = 0.970$). We also find that natural paradigms show both higher accuracy scores ($\hat{\beta} = 0.302$, 90%CI = [0.185, 0.425], $SE = 0.073$, $P(\hat{\beta} > 0) = 1$) and faster learning rates ($\hat{\beta} = 0.034$, 90%CI = [0.014, 0.056], $SE = 0.013$, $P(\hat{\beta} > 0) = 0.998$) than L and X-type paradigms. Further, accuracy scores for the no-split condition are also overwhelmingly higher ($\hat{\beta} = 0.640$, 90%CI = [0.548, 0.738], $SE = 0.011$, $P(\hat{\beta} > 0) = 1$), and its learning rates faster ($\hat{\beta} = 0.084$, 90%CI = [0.065, 0.103], $SE = 0.073$, $P(\hat{\beta} > 0) = 1$), than the average of all other conditions.

Discussion

A growing body of work uses experimental methods to investigate how language learning correlates with cross-linguistic preferences in the partitions of semantic space (e.g., Silvey, Kirby, & Smith, 2015; Maldonado & Culbertson, 2021; Carr, Smith, Culbertson, & Kirby, 2020; Pertsova, 2014; Lee, 2020). These studies suggest a bias towards partitions where all the members share some set of feature values (i.e., natural classes) over partitions where that is not the case (i.e., unnatural classes). For example, lexical words do not tend to mean both *chair* and *rain* because the meanings do not share any obvious properties. In a similar way, morphological affixes should not tend to mean both 1st person singular and 3rd person plural, as these do not share any feature values. Unnatural patterns are common in morphological paradigms, yet little is hitherto known about their cross-linguistic recurrence and learnability. In this paper, we investigated a naturalness gradient in morphological paradigms. We surveyed the possible cross-linguistic asymmetries between different types of (un)natural paradigm splits based on two different linguistic phenomena (shared forms or syncretism, and shared ordering rules or positional splits), and tested their learnability in two artificial language learning experiments. We found cross-linguistic evidence consistent with the recurrence hierarchy natural > L-type unnatural > X-type unnatural patterns, and our experimental results provide evidence for a learnability gradient consistent with it: Natural patterns are by far easier to learn than unnatural patterns, and L-type unnatural patterns are easier to learn than X-type unnatural patterns. We propose that this gradient in learnability reflects a general bias towards similarity-based structure in morphological learning, which can also be found in word learning as well as in category and concept learning more generally. Our results thus support a more nuanced view of the natural-unnatural distinction in morphological paradigms—which ought to be conceptualised as a gradient rather than a dichotomous property—and suggest a causal link between differences in learnability and the frequency of different patterns of syncretism and positional identity.

Acknowledgements This research has been partially funded by the NCCR Evolving Language, Swiss National Science Foundation (Agreement Nr. 51NF40_180888).

We would like to thank the developers of the open-source softwares used for this study: Josh de Leeuw and `collaborators` for the development of *jsPsych* (De Leeuw, 2015), Vanessa Sochat for the curation of *The Experiment Factory* (which provided us with valuable ideas for the coding of our software; Sochat, 2018).

Ethics The study was approved by the Ethics Committee of the School of Philosophy at the University of Zurich (Authorisation Nr. 20.4.16 and 21.9.15). Research practices follow the ethical guidelines for psychologists of the Swiss Society for Psychology (SGP), and the Ethical Principles of Psychologists and Code of Conduct of the American Psychological Association (APA).

References

Baerman, M., Brown, D., & Corbett, G. (2002). *Surrey person syncretism database.* Retrieved from `http://www.smg.surrey.ac.uk/syncretism/index.aspx`

Baerman, M., Brown, D., & Corbett, G. G. (2005). *The syntax-morphology interface: A study of syncretism* (Vol. 109). Cambridge: Cambridge University Press.

Bickel, B. (2015). Distributional typology: Statistical inquiries into the dynamics of linguistic diversity. In B. Heine & H. Narrog (Eds.), *The Oxford Handbook of Linguistic Analysis* (p. 901-923). Oxford: Oxford University Press.

Bickel, B., Banjade, G., Gaenszle, M., Lieven, E., Paudyal, N. P., Rai, I. P., ... Stoll, S. (2007). Free prefix ordering in chintang. *Language*, 43–73. https://doi.org/10.1353/lan.2007.0002

Bickel, B., Nichols, J., Zakharko, T., Witzlack-Makarevich, A., Hildebrandt, K., Rießler, M., ... Lowe, J. B. (2017). *The AUTOTYP typological databases*. Retrieved from `https://github.com/autotyp/autotyp-data/tree/0.1.0`

Blythe, R. A., & Croft, W. (2021). How individuals change language. *PLoS ONE*, *16*(6), e0252582. https://doi.org/10.1371/journal.pone.0252582

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395–411. https://doi.org/10.32614/RJ-2018-017

Carr, J. W., Smith, K., Culbertson, J., & Kirby, S. (2020). Simplicity and informativeness in semantic category systems. *Cognition*, *202*, 104289. https://doi.org/10.1016/j.cognition.2020.104289

Crysmann, B., & Bonami, O. (2016). Variable morphotactics in information-based morphology. *Journal of Linguistics*, *52*(2), 311–374.

Culbertson, J., & Smolensky, P. (2012). A Bayesian model of biases in artificial language learning: The case of a word-order universal. *Cognitive Science*, *36*(8), 1468–1498. https://doi.org/10.1111/j.1551-6709.2012.01264.x

Cysouw, M. (2003). *The paradigmatic structure of person marking*. Oxford: Oxford University Press.

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior research methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Dryer, M. S., & Haspelmath, M. (Eds.). (2013). *WALS Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from `https://wals.info/`

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, *32*(1), 108–154. https://doi.org/10.1080/03640210701802071

Herce, B. (2020). *A typological approach to the morphome*. Unpublished doctoral dissertation, University of Surrey & University of the Basque Country.

Hualde, J. I., & De Urbina, J. O. (2011). *A grammar of Basque* (Vol. 26). Berlin: De Gruyter Mouton.

Jakobson, R. (1936). Beitrag zur allgemeinen Kasuslehre: Gesamtbedeutungen der russischen Kasus. *Travaux du Cercle Linguistique de Prague 6*, 240–288.

Kimball, G. D. (1985). *A descriptive Grammar of Koasati (Louisiana)*. Unpublished doctoral dissertation, Tulane University.

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, *28*, 108–114. https://doi.org/10.1016/j.conb.2014.07.014

Lee, N. (2020). Learning (im)possible number syncretisms: Investigating innate featural representations. In *Proceedings of the 50th Annual Meeting of the North East Linguistic Society.* Amherst: University of Massachusetts GLSA.

Maldonado, M., & Culbertson, J. (2021). Person of interest: Experimental investigations into the learnability of person systems. *Linguistic Inquiry*. https://doi.org/10.1162/ling_a_00406

Mansfield, J., Stoll, S., & Bickel, B. (2020). Category clustering: A probabilistic bias in the morphology of verbal agreement marking. *Language*, *96*(2), 255–293.

McGregor, R. S. (1995). *Outline of Hindi grammar*. Oxford: Oxford University Press.

Nevins, A., Rodrigues, C., & Tang, K. (2015). The rise and fall of the L-shaped morphome: Diachronic and experimental studies. *Probus*, *27*(1), 101–155. https://doi.org/10.1515/probus-2015-0002

Oates, W., & Oates, L. (1968). *Kapau pedagogical grammar*. Canberra: Australian National University.

Pertsova, K. (2007). *Learning form-meaning mappings in the presence of homonymy*. Unpublished doctoral dissertation, University of California Los Angeles.

Pertsova, K. (2014). Logical complexity in morphological learning: effects of structure and null/overt affixation on learning paradigms. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 38, pp. 401–413).

Saldana, C., Herce, B., & Bickel, B. (2021, August 17). *Learnability of morphomic patterns of syncretism.* https://doi.org/10.17605/OSF.IO/PWQJG

Saldana, C., Herce, B., & Bickel, B. (2022, February 9). *Learnability of morphomic patterns of syncretism.* Retrieved from https://osf.io/jpum6/

Saldana, C., Herce, B., & Bickel, B. (2022, January 21). *Learnability of morphological patterns of syntagmatic splits.* https://doi.org/10.17605/OSF.IO/YZCXP

Silvey, C., Kirby, S., & Smith, K. (2015). Word meanings evolve to selectively preserve distinctions on salient dimensions. *Cognitive science*, *39*(1), 212–226. https://doi.org/10.1111/cogs.12150

Sochat, V. (2018). The experiment factory: Reproducible experiment containers. *Journal of Open Source Software*, *3*(22), 521. https://doi.org/10.21105/joss.00521

Stan Development Team. (2021). *RStan: the R interface to Stan.* Retrieved from https://mc-stan.org/ (R package version 2.21.3)

Völlmin, S. (2017). *Towards a grammar of gumer-phonology and morphology of a western gurage variety*. Unpublished doctoral dissertation, University of Zurich.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245. https://doi.org/10.1037/0033-295X.114.2.245