COGNITION AND THE STABILITY OF EVOLVING COMPLEX MORPHOLOGY: AN AGENT-BASED MODEL

Erich Round[*1,2], Stephen Mann[1,3], Sacha Beniamine[1], Emily Lindsay-Smith[1], Louise Esher[4], and Matt Spike[5]

[*]Corresponding Author: e.round@surrey.ac.uk
[1]Surrey Morphology Group, University of Surrey, Guildford, UK
[2]School of Languages and Cultures, University of Queensland, St Lucia, Australia
[3]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[4]CNRS LLACAN, Paris, France
[5]School of Philosophy Psychology & Language Sciences, University of Edinburgh, Edinburgh, UK

Cultural attractors enable evolving cultural traits to gain the stability that underpins cumulative cultural evolution, yet the conditions that support their existence are poorly understood. We examine conditions affecting the stability of a salient kind of complex cultural attractor in human language, known as inflectional classes. We present a model of the evolution of inflectional classes, as they are reconstructed across generations via a combination of direct transmission and analogical inference. Parameters examined pertain to diversity of the lexicon and the cognitive policies governing inferential reasoning. We discover that persistence of stable inflection classes interacts in complex ways with features which affect how inflection classes are inferred. Thus we contribute to a greater understanding of factors affecting cultural attractors' existence, and to insights into a widespread and complex trait of human language.

1. Introduction

Human languages present an explanatory challenge: to reconcile their seemingly endless structural diversity with an equally striking tendency to exhibit recurrent similarities. Cultural evolutionary perspectives offer a solution. The adaptive flexibility of **cultural** transmission introduces the noise which makes linguistic innovation and change inevitable, while the **cognitive** nature of language influences the origin, direction and uptake of these changes, and their subsequent stability. Cultural Attraction Theory (Claidière & Sperber, 2007) foregrounds the role of cognition in shaping the adaptive landscape over which culture evolves. Cultural attractors are regions of meta-stability in cultural trait space. Their existence enables traits to remain stable in certain states but not others, thus promoting the emergence of shared, cumulative culture (see Griffiths & Kalish, 2007; Kirby, Tamariz, Cornish, & Smith, 2015). Here we examine *inflectional class systems* as cultural attractors. We investigate how parameters governing the inferential reconstruction of inflectional classes affect their evolutionary stability. Inflectional class systems are known to persist across millennia, thus the cognitive strategies

common among language speakers are likely to be those which promote stability. Consequently, we view our approach as a method for identifying parameter settings of cognitive policies that may promote the persistence of a highly complex cultural attractor known to be a salient, recurrent trait of human language.

Table 1. Paradigms illustrating three inflectional classes in Swedish nouns.

		SG.INDEF	SG.DEF	PL.INDEF	PL.DEF
IC1	'school'	*skol**a***	*skol**an***	*skol**or***	*skol**orna***
	'bottle'	*flask**a***	*flask**an***	*flask**or***	*flask**orna***
IC2	'chair'	*stol*	*stol**en***	*stol**ar***	*stol**arna***
	'box'	*ask*	*ask**en***	*ask**ar***	*ask**arna***
IC3	'idol'	*idol*	*idol**en***	*idol**er***	*idol**erna***
	'Basque'	*bask*	*bask**en***	*bask**er***	*bask**erna***

Inflectional Classes as Cultural Attractors In many languages, lexemes can possess a set of multiple, contextually-conditioned wordforms, termed an *inflectional paradigm* as illustrated in Table 1 for Swedish nouns. In paradigms, each wordform occupies a *cell*. The wordforms in a paradigm typically contain formal differences, termed the *exponents* of the cell, and Inflectional Classes (ICs) are distinct *patterns* of exponents found in paradigms. In Table 1, IC 1 has exponents {*-a,-an,-or,-orna*} whereas IC 2 has {*Ø,-en,-ar,-arna*}.

As languages change, ICs are known to serve as attractors in the sense that anomalous paradigms (which may arise as the result of minor disruptions) often undergo alteration, to become more like other ICs in the language (Maiden, 2018). Recently, an explanation for this has been formulated in terms of the reconstruction and transmission of paradigms.

Reconstruction and Transmission Paradigms are reconstructed not *en bloc* but in piecemeal fashion, one cell at a time: a single utterance might contain the wordform '*sang*', but not the whole paradigm '*sing–sings–singing–sang–sung*'. In many languages, paradigms are sufficiently large (often comprised of dozens if not hundreds of wordforms) that it is implausible that speakers would always have heard a specific wordform of a lexeme before they first need to produce it (Ackerman, Blevins, & Malouf, 2009; Blevins, Milin, & Ramscar, 2017). Yet speakers can perform this task and they agree upon the solution. Psycholinguistic results show that speakers have clear inferential intuitions about paradigm forms they have not previously heard. English speakers for instance nominate *splung* but not *splong* as a plausible past tense form of novel *spling* (Albright & Hayes, 2002). Thus, cognitively, ICs are supported by processes of inferential reasoning in addition to mere storage of wordforms.

Conditional Predictability In natural languages, the distribution of exponents in different cells is not random, rather the exponent of one cell is typically relatively predictable from the exponent of another (Carstairs-McCarthy, 1987); conditional entropy is low compared to random covariation (Ackerman & Mal-

ouf, 2013). Linguists have been interested in understanding this interpredictability as a potentially emergent property of IC evolution (Ackerman & Malouf, 2015; Round, Beniamine, & Esher, 2021).

2. Modelling of Attractor Evolution

Inflectional Classes The task of inferring the content of a paradigm cell is known as the *paradigm cell filling problem*, or PCFP (Ackerman et al., 2009). Ackerman and Malouf (2015) modelled an analogical PCFP process, in which one cell of a lexeme's paradigm (which we call the *focal* cell) is inferred from analogical relationships that exist between it and one other cell (which we call the *pivot* cell, and which is sampled at random). To infer the focal cell's exponent, the agent attends to other lexemes that share a similar pivot cell, and samples from their focal cells. In an iterated learning simulation, with a single agent who learns all forms of the language except for one, which is inferred analogically, a system that is seeded initially with random paradigms will gradually self-organize into a limited number of ICs. Ackerman and Malouf (2015) show that across the lexicon, the mean conditional entropy of pairs of cells falls over time, and interpret this as reflecting a rising interpredictability between cells.

However, Round et al. (2021) note that conditional entropy can fall for two reasons: because interpredictability is increasing or because total entropy is falling as a system becomes uniform. They show that in Ackerman and Malouf (2015)'s model, conditional entropy falls for the latter reason: unlike the stable distinctiveness found in real IC systems, ICs in the model are inherently unstable and inevitably collapse together entirely. They advocate an improved measure of interpredictability based on mutual information. Mutual information will rise and fall as interpredictability does, including when total entropy is changing.

Inferential Policies In Ackerman and Malouf (2015) and Round et al. (2021), the PCFP is based solely on the contents of lexemes' inflectional paradigms. However, other information sources, including similarities along semantic and phonological dimensions, are known to affect the probability with which one lexeme influences another, both psycholinguistically and in language change (Hayes, Zuraw, Siptár, & Londe, 2009; Maiden, 2020). Here we investigate the emergence and stability of inflectional classes when analogical inference in the PCFP makes reference to *extra-paradigmatic similarity*: features other than morphological exponents that lexemes can share.

3. Model Description

We examine conditions for the (in)stability of cultural attractors, namely inflectional classes, in a multi-agent model in which paradigms evolve via a PCFP mechanism.[1] An initial population of agents is created who all share the same

[1] Model code and a Wiki guide are available at bit.ly/ELXIV.

lexicon. In each iteration, a child population is created, all of whom begin life with an empty lexicon. During acquisition, children learn from adults, who utter wordforms including by using the PCFP (see below) to infer forms which the adults themselves have not heard. Finally, adults die, children become adults, and the next iteration begins with the creation of a new child population.

Simulations commence with an initial population of A adult agents, who share a randomised lexicon of L lexemes[2] $\{\text{LEX}_1, \text{LEX}_2, \ldots, \text{LEX}_L\}$. Each lexeme has a paradigm of C cells, whose exponents are represented as integers drawn from the range $\{1,2,\ldots V\}$. The initial lexicon contains I_{init} distinct inflection classes which may share some (but not all) exponents. The model runs for G iterations ('generations'). At each iteration, children during acquisition receive an input totaling W wordforms, transmitted in equal measure by Q randomly selected adult acquisition sources. If a child hears multiple wordforms for the same cell of a lexeme, it stores the first that it encounters. Adults sample wordforms to transmit as follows. A lexeme is sampled from among the lexemes known to the agent, according to a Zipfian distribution derived from the frequencies with which lexemes were heard by the agent as a child. A cell for that lexeme is chosen, again according to a Zipfian distribution. The exponent for the cell is retrieved from memory if the adult learned it as a child. Otherwise it is inferred by the PCFP process.

The PCFP The PCFP proceeds as follows. To infer the focal cell of the focal lexeme, (1) sample E *evidence lexemes*; (2) accord them a weight of $+1$ for each non-focal cell that matches the corresponding cell of the focal lexeme; (3) for each exponent value v in $\{1,2,\ldots,V\}$, assign it a weight equal to the summed weights of all evidence lexemes whose focal cell contains v. The selected exponent value is the one with the greatest weight; ties are broken randomly.

At step (1), the sampling of evidence lexemes is biased by lexemes' similarity in their extra-paradigmatic features. Moreover, the relative importance attached to each feature when calculating similarity is controlled by a feature *weight* which the agent *learns* at the conclusion of acquisition, as a function of the correlations it finds between a feature's values and lexemes' exponents.[3] As a language evolves, agents' feature weights—and the associated bias introduced into their PCFP solutions—will change over time, reflecting the shifting relationships between lexemes' unchanging extra-paradigmatic features and changing paradigms.

In this study, we investigate the effects of different kinds of correlations *in the initial population's lexicon* between extra-paradigmatic features and lexemes' ICs. Lexemes are given three features, allowing us to examine four conditions: *High*, in which one feature correlates exactly with ICs; *Medium*, in which one feature correlates perfectly with ICs for the first 50% of the lexicon; *Double medium*, which is identical to *Medium* but with the additional of a second feature that cor-

[2] In the remainder of this section, italicised single letters denote model parameters.

[3] Implemented using Random Forest Feature Importance. See model code for details.

relates perfectly with the ICs of the *second* 50% of the lexicon; and *Zero*, in which no feature correlates in these ways with initial ICs. Thus, in the *High, Medium* and *Double* conditions, whenever lexeme LEX_*i* is the subject of the PCFP, the set of evidence lexemes against which it is compared is more likely to be drawn from lexemes which shared LEX_*i*'s IC in the initial population's lexicon.

Table 2. Parameter definitions and values used. See main text for explanation.

Parameter		Value(s)	Parameter		Value(s)
PARADIGMS			TRANSMISSION		
Lexemes	L	200	Agents	A	4
Cells	C	5	Acquisition sources	Q	3
Possible cell values	V	5, 50	Generations	G	5,000
Inflection classes (initial)	I_{init}	5	Wordforms in acquisition	W	24,000
PCFP					
Evidence lexemes	E	10			
Extra-paradigmatic bias	B	High, Medium, Double, Zero			

4. Simulation Experiments and Results

Parameter settings investigated are shown in Table 2. Varied parameters were B (extra-paradigmatic bias) and V (range of available exponents), taking 4 and 2 values respectively for a total of 8 conditions. Each condition ran 30 times.

The outcomes of primary interest are the change in number of ICs, the degree to which different agents conform in the ICs in their grammars, and the inter-predictability between cells in agents' grammar. To examine these we measure five population-wide statistics every 50 generations: (1) *IC Diversity*, the total number of distinct ICs in the grammars of the population of agents (lexemes with missing wordforms are ignored); (2) *IC Disparity*, defined as $1 - A/D$ where A is the average number of distinct ICs for individual agents and D is IC Diversity; (3) *Conditional entropy*, the mean conditional entropy between pairs of cells in an agent's paradigm system, averaged across all agents; (4) *Mutual information*, the mean mutual information between pairs of cells in an agent's paradigm system, averaged across all agents; (5) *Maximum prediction weight*, the weight of the feature with respect to which the agent's selection of evidence lexemes is most strongly biased. Figure 1 visualises the four measures across 5,000 generations.

The results show initial, transient spikes in IC Diversity, IC Disparity and conditional entropy as the system departs from its initialisation state (first-generation agents share the same lexicon). Then, under all conditions, we find broad falls in IC Diversity and IC Disparity, falls in conditional entropy, and falls in mutual information as the newly-formed ICs begin to collapse together. The analogising effect of the PCFP is bringing different lexemes into line with one another. In almost all cases, each statistic takes a larger value in the 50-exponent case (blue lines). More exponents entail greater entropy and information because there

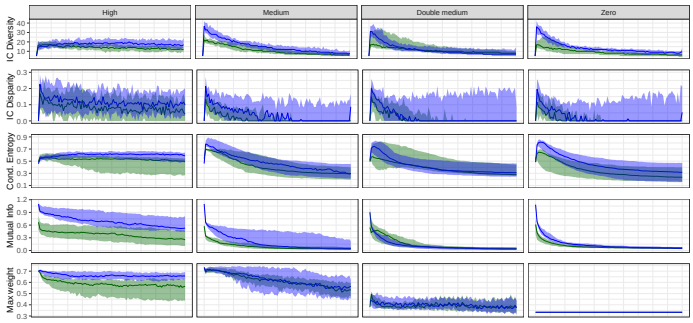

Figure 1. Five measures of ICs over 5,000 generations, observed at intervals of 50 generations. Conditions with 50 available exponents are in blue, with 5 in green. Each condition was run 30 times. Lines show medians, ribbons show 80% of variation. Each column of panels shows one bias level.

is a smaller chance of the same exponents being used across different ICs. And more exponents entail greater IC Diversity and Disparity because there is a greater source of variation with which to recombine existing ICs into new ones.

Differences in different bias conditions (i.e. across different columns in figure 1) can be understood as follows. First note all measures' uniqueness in the *High* condition. In *High*, evidence lexemes are very likely to be drawn from the same initial IC as the focal lexeme. Consequently, initial groups of lexemes will stick together through generations: they constitute reliable clusters exerting an internal gravitational pull, preventing individual lexemes from escaping to form new classes. Mutual information remains high because the initial ICs are mostly conserved. The other conditions show accelerated declines in conditional entropy and mutual information. Inflection classes are collapsing together across the lexicon; because the feature correlations are not strong enough to generate insular clusters, lexemes from different initial classes might be used to inform the PCFP for any given lexeme. Second, note the difference between the *Medium* and *Double* conditions. On the face of it, these patterns are paradoxical: the *Double* condition contains greater correlation across the lexicon, so one would expect its statistics to resemble those of the *High* condition. But its IC Disparity and mutual information are more similar to the *Zero* condition. How can more correlation across multiple features accelerate collapse, when maximum correlation in a single feature retards collapse? The answer is revealed by the 'Max weight' statistic, measuring how strongly an agent favours one set of extra-paradigmatic features over the others. In the *High* condition, agents favour the strongly-correlated feature. In the *Medium* condition, there is no strongly correlated column, but agents favour the medium-correlation feature. This leads to classes beginning to 'cluster' according to those features (even though they do not perfectly correspond to initial ICs), thus the collapse (decrease in IC Disparity and mutual information) is somewhat slowed.

By contrast, in the *Double* condition, **agents cannot strongly favour any single feature**; their assignment of lexemes to groups is pulled in two different directions indicated by the two distinct medium-correlation features. Two medium-strength correlations that pull in different directions cancel each other out; as a result, the statistics in this condition more resemble those of the *Zero* condition, where there is no information by which to favour certain lexeme clusters at all. These results indicate that IC stability may be affected in strikingly different ways by correlations which agents detect between inflection classes and extra-paradigmatic features, such as lexical semantics or phonological stem shape.

5. Conclusions and Future Directions

In this paper we presented a model of evolution in systems of inflectional classes, which are known to act as complex cultural attractors in languages around the world. Framing the study in terms of Cultural Attractor Theory, we modeled the reconstruction of ICs in terms of the paradigm cell filling problem (PCFP), which has attracted considerable recent attention in linguistics. We ran simulations chosen to highlight a key potential parameter of variation in the cognitive, inferential processes that underlie the PCFP and which we therefore expect to play a role in shaping the cultural attractor landscapes which ICs traverse. We find that multiple sources of analogical inference can prevent distinct categorisation, because those sources can conflict with each other. Therefore, explaining how agents categorise by analogy is not just a case of quantifying correlational relationships, but understanding the different sources of those correlations and their interactions.

Future elaborations of our approach include adding the effects of noise and channel biases during transmission, and transforming interaction into a **communication game** involving costs for communicative failure. More complex storage of heard exponents could allow agents to track the grammatical **variation** they encounter. Further possibilities are implemented in our model, but have not been explored in this study. Our model design enables agents to **converse as adults**, leading to paradigm gaps being filled via horizontal inheritance among a peer generation. Recent evidence (e.g. Raviv, Meyer, & Lev-Ari, 2020) underlines the importance of **population demographics**, e.g. population size and network structure (which our model implements), and overlapping, non-synchronous generations. Additional parameter values of the morphological system to be examined include **larger paradigms** with more lexemes, more cells, and multiple exponents per cell. Finally, it will be important to examine exogenous impacts on paradigm systems such as sound change and borrowing of lexicon from other languages; our model implementation anticipates the addition of these extensions.

Acknowledgements

The authors wish to thank three anonymous reviewers for helpful feedback. This work was supported by British Academy Global Professorship GP300169 to ER

and British Academy Newton International Fellowship NIF23\100218 to SB.

References

Ackerman, F., Blevins, J. P., & Malouf, R. (2009). Parts and wholes: Patterns of relatedness in complex morphological systems and why they matter. In J. P. Blevins & J. Blevins (Eds.), *Analogy in grammar: Form and acquisition* (pp. 54–82). Oxford: Oxford University Press.

Ackerman, F., & Malouf, R. (2013). Morphological Organization: The Low Conditional Entropy Conjecture. *Language*, *89*(3), 429–464.

Ackerman, F., & Malouf, R. (2015). The No Blur Principle Effects as an Emergent Property of Language Systems. In *Proceedings of the 41st Annual Meeting of the Berkeley Linguistics Society* (pp. 1–14). Berkeley: BLS.

Albright, A., & Hayes, B. (2002). Modeling English past tense intuitions with minimal generalization. In *Proceedings of the ACL-02 workshop on morphological and phonological learning* (pp. 58–69). Philadelphia: Association for Computational Linguistics.

Blevins, J. P., Milin, P., & Ramscar, M. (2017). The Zipfian Paradigm Cell Filling Problem. In F. Kiefer, J. Blevins, & H. Bartos (Eds.), *Perspectives on Morphological Organization* (pp. 139–158). Leiden: Brill.

Carstairs-McCarthy, A. (1987). *Allomorphy in Inflexion.* London: Croom Helm.

Claidière, N., & Sperber, D. (2007). The role of attraction in cultural evolution. *Journal of Cognition and Culture*, *7*(1-2), 89–111.

Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive science*, *31*(3), 441–480.

Hayes, B., Zuraw, K., Siptár, P., & Londe, Z. (2009). Natural and Unnatural Constraints in Hungarian Vowel Harmony. *Language*, *85*(4), 822–863.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.

Maiden, M. (2018). *The Romance Verb: Morphomic Structure and Diachrony.* Oxford: Oxford University Press.

Maiden, M. (2020). The verbs 'rain' and 'snow' in Gallo-Romance, and other morphological mismatches in diachrony. In *Variation and Change in Gallo-Romance Grammar.* Oxford: Oxford University Press.

Raviv, L., Meyer, A., & Lev-Ari, S. (2020). The role of social network structure in the emergence of linguistic structure. *Cognitive Science*, *44*(8), e12876.

Round, E. R., Beniamine, S., & Esher, L. (2021). The role of attraction–repulsion dynamics in simulating the emergence of inflectional class systems. In *International Symposium of Morphology 2021.* (http://nabil.hathout.free.fr/ISMo2021)