CONVERGENT CULTURAL EVOLUTION OF CONTINUERS (*mmhm*)

MARK DINGEMANSE[*1,2], ANDREAS LIESENFELD[1], and MARIEKE WOENSDREGT[2]

[*]Corresponding Author: mark.dingemanse@ru.nl
[1]Center for Language Studies, Radboud University
[2]Max Planck Institute for Psycholinguistics

Continuers —words like *mm, mmhm, uhum* and the like— are among the most frequent types of responses in conversation. They play a key role in joint action coordination by showing positive evidence of understanding and scaffolding narrative delivery. Here we investigate the hypothesis that their functional importance along with their conversational ecology places selective pressures on their form and may lead to cross-linguistic similarities through convergent cultural evolution. We compare continuer tokens in linguistically diverse conversational corpora and find languages make available highly similar forms. We then approach the causal mechanism of convergent cultural evolution using exemplar modelling, simulating the process by which a combination of effort minimization and functional specialization may push continuers to a particular region of phonological possibility space. By combining comparative linguistics and computational modelling we shed new light on the question of how language structure is shaped by and for social interaction.

1. Introduction

Social interaction is characterised by people exchanging short bursts of articulatory activity organised in turns (Sacks, Schegloff, & Jefferson, 1974). Evidence that linguistic resources are organized in response to this fact is the existence of a class of responsive items variously called response tokens, backchannels, or continuers (Yngve, 1970; Jefferson, 1985). CONTINUERS are items that occur at the boundaries of turn-constructional units and that "demonstrate both that one unit has been received and that another is now awaited" (Goodwin, 1986). Common examples of continuers so defined are forms like *mm, uh huh, yeah* in English (Indo-European) or *umm, mm, ing* in Anal Naga (Tibeto-Burman) (Fig. 1).

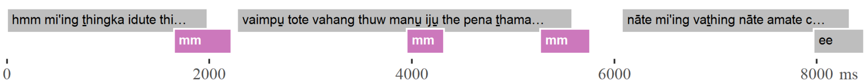

Figure 1.: Conversation in Anal Naga (Tibeto-Burman, India), showing how continuers like *mm* support another participant's production of multiple longer turns and often occur in streaks of 3+ items, a fact we use to identify candidate tokens in a language-agnostic way (data from Ozerov, 2018).

Continuers have been studied in conversation analysis and psycholinguistics (Yngve, 1970; Jefferson, 1985; Ward, 2006; Bavelas, Coates, & Johnson, 2000), but work on their cultural evolution is scarce. Yet if interaction is the primary ecology of language in use (Schegloff, 2006) and if utterances are the main target of selection in cultural evolution (Croft, 2000), then continuers, as stand-alone utterances directly exposed to the exigencies of conversation, provide a crucial locus for studying how language is shaped by interaction. We propose that key properties of items like this can be explained by thinking of them as a distinctive stratum of vocabulary (Ward, 2006) adapted to the ecology of turn-by-turn interaction.

Continuers are highly frequent words whose conversational ecology is characterized by frequent overlap and a strong skew in the amount of talk contributed by each participant (Fig. 1; cf. Rühlemann, 2018). This places special requirements on their shape (Gardner, 1997). Optimal continuers are (i) easy to plan and produce, (ii) unobtrusive, and (iii) sufficiently distinct from regular words to be seen as ceding the conversational floor. While the first criterion may be ascribed to frequency-driven effort minimization, the other two are likely to push continuers into a particular part of the possibility space. In cultural evolutionary terms, these are selective pressures that over diachronic time are likely to result in convergent solutions across unrelated languages (Dingemanse, Torreira, & Enfield, 2013).

Here we test this proposal by combining two methods. Cross-linguistic comparison provides us with primary data on the interactional ecology of continuers across languages, and can show whether indeed continuers, more than some other words, display tell-tale signs of being adapted to this ecology. Computational modelling provides us with ways to formalize and test the causal account, enabling us to see whether the proposed convergence of form can arise from cultural evolution given selective pressures of effort minimization and conversational ecology. While we focus on continuers, the results have wider relevance for language evolution in at least two ways. They point to interactional infrastructure as a key causal locus affecting the cultural evolution of linguistic items (Enfield, 2014); and they put the spotlight on metacommunication, one of the major advances in the evolution of communication (Bateson, 1972).

2. Methods

Observational. We curate transcribed corpora of natural conversation in 67 spoken languages of 28 phyla around the world (Fig. 2). Crucially, we do not search for continuers by form, but define their conversational ecology as a sequential pattern, making it possible to identify candidate tokens in a language-agnostic, non-circular way. We look for streaks of non-unique conversational turns that occur in frequent alternation with unique turns by other participants. Using a minimum streak length of 3, we identify the top 5 candidate continuer tokens per language. We also collect two contrasting sets of linguistic items for comparison. First a set we call *discontinuers*, defined as the top 5 highest frequency stand-alone turn for-

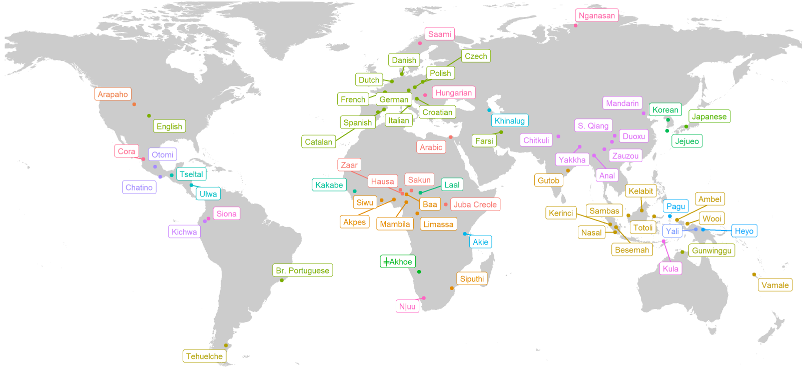

Figure 2.: The sample incorporates conversational corpora of 67 spoken languages from 28 distinct phyla, with 32 languages from 12 phyla studied in more detail.

mats that do not occur in streaks like continuers do. The interactional work done by these items is almost the converse of continuers: they do not cede the floor but take it as they redirect, resume, or repair a conversation. This means they are not subject to the same selective pressures. Second a set of *top tokens*, defined as the top 5 highest frequency tokenised words in the language (this provides a baseline comparison for frequency-driven reduction effects). Since audio processing is ongoing, most observations reported here are based on written annotations.

Computational. To formalise our causal account of how continuers are shaped by social interaction, we use exemplar models. Such models allow us to simulate biases affecting the shape of culturally evolving items over repeated production and perception cycles (Winter & Wedel, 2016), and have been used to study how language adapts to interaction (Roberts & Levinson, 2017). We start by reproducing a model from Wedel (2012) that aims to capture the cultural evolution of words in an abstract phonological space under various pressures. To simulate the case of continuers, we adjust the model in three ways. First, we add an overarching category distinction between a set of words that together form the 'regular vocabulary', and a set (here consisting of a single word) that constitutes the 'continuer' category. This allows us to put continuers under different pressures from regular vocabulary. Second, we increase the pressure for minimising production effort for continuers relative to regular words. Third, we remove the pressure for reuse of phonetic features for continuers. For regular words, this pressure pushes the system to make use of a limited set of phonetic features, leading to a restricted phoneme inventory, rather than a proliferation of somewhat distinct sounds. For continuers, we remove this pressure in order to allow a distinct stratum to emerge. Data and code for observational and computational methods are available through OSF: https://osf.io/v6kfn.

3. Results

3.1. *Observational*

Our language-agnostic sequential search method allows us to identify candidate continuers in all 67 languages in the sample. Figure 3 shows candidate continuers in their natural conversational ecology in 10 unrelated languages. Already these examples suggest the close intertwining of conversational habitat and continuer form. In line with their function of displaying recipiency in minimally obtrusive ways, continuers are shorter than the turns they are interspersed with, often occur in full or partial overlap, and appear to be phonologically quite minimal.

Figure 3.: Candidate continuers in 10 unrelated languages (A) shown in their natural ecology (B, annotations as in the original data), with spectrograms and pitch traces of representative tokens made using the Parselmouth interface to Praat (Jadoul et al., 2018; Boersma & Weenink, 2013) (C).

Out of the full sample there are 32 languages (of 12 phyla) with large enough corpora to yield sufficient examples of all three groups of items to be compared (we find a total of 118 continuers, 104 discontinuers, and 160 top tokens, which works out to per-language averages of 4, 3, and 5 respectively). The observations below relate to this subset of data. Because the work is ongoing and we aim for more direct comparisons of acoustic and phonetic features, we do not run inferential statistics on the preliminary observations presented here.

Top tokens are an order of magnitude more frequent than either continuers or discontinuers. They are short (2.28 characters) and relatively phonemically simple (1.8 distinct characters per token). This provides us with baseline expectations

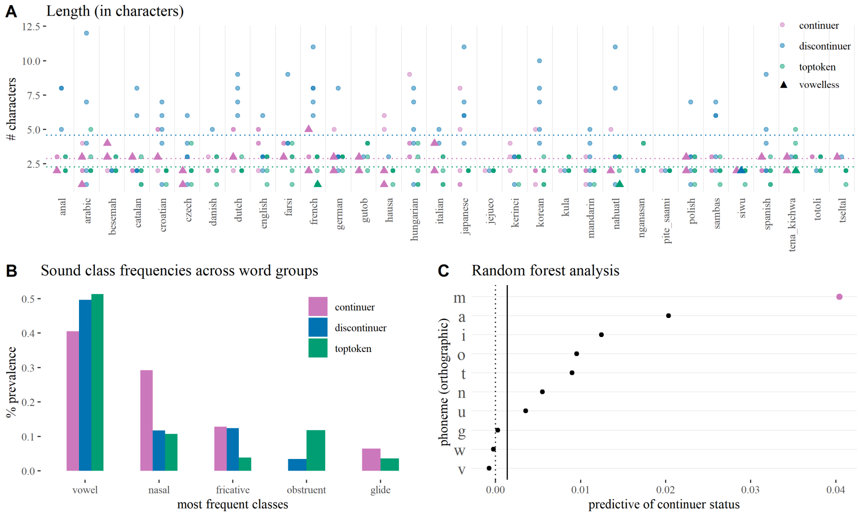

Figure 4.: **A**. Length of tokens for continuers, discontinuers and top tokens. Dotted lines mark group averages; items marked △ are vowelless. **B**. Frequencies of major sound classes across types. Vowel nuclei occur across types, but continuers stand out for their preference for nasals. **C**. Random forest analysis for 118 continuer forms in 32 spoken languages showing the top 10 most predictive phonemes (out of 29 attested). Vertical black line: absolute value of the least predictive phoneme.

for the other groups. Discontinuers are twice as long (4.1 characters) and more phonemically diverse (3.0 distinct characters per token), in line with expectations if length is affected by frequency and phonology is arbitrary. Continuers appear to be phonemically shorter than discontinuers yet longer than top tokens (2.8 characters), and as phonemically simple as top tokens (1.8 distinct characters per token).

Continuers appear to stand out in several ways. On average, one continuer token per language (32/118) is transcribed as a vowelless token like *hm, mhm, mm* (Fig. 3 and Fig. 4A), against only 1/104 discontinuers and only 3/160 top tokens. Further, 47% of continuer tokens (56/118) feature a nasal, against 32% of discontinuers (33/104)) and 24% of top tokens (39/160). Both trends are seen in a tally of phoneme classes across word types (Fig. 4B). A random forest analysis of all continuer forms together shows that 'm' is highly predictive of continuer status across languages (Fig. 4C). Further, 25% of continuer tokens (29/118) feature full or partial reduplication as in *mhm, hm hm, uhum, un un* (see also Fig. 3A), against only 10% of discontinuers (11/106 tokens) and 1% of top tokens (1/160).

While our main focus here has been on the 32 languages for which a comparison of the three sets is possible, the prevalence of nasals, vowelless tokens and reduplication in continuers is also seen in the remaining 35 languages, and also if we only look at 1 language per phylum (as reported in the online materials).

3.2. *Computational*

Figure 5 shows that whereas regular words (blue, yellow, green and orange) end up in arbitrarily different positions in the possibility space across different runs of the simulation, continuers (purple triangles) consistently end up in the same region: the centre of the graph which is the region of least effort. We also see a mild effect of this on regular words, which across independent simulation runs tend to avoid the very centre of the space as a result of the anti-ambiguity bias.

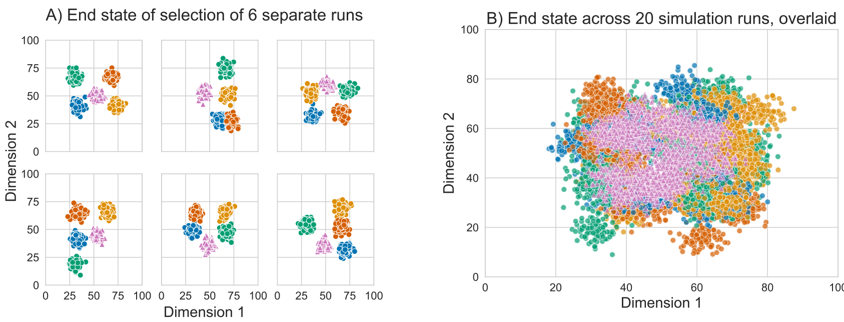

Figure 5.: Plots of where in a phonetic possibility space different words end up after 10,000 rounds of interaction, across 20 independent simulation runs (each cloud of 100 exemplar dots/triangles represents a single word at round 10,000 of a single simulation run). Blue, yellow, green and orange are regular words; purple is the continuer word. On each independent simulation run, all words are initialised at randomly selected positions in the space. **A** shows a selection of 6 separate simulation runs chosen for illustrative purposes (showing how regular words end up in different positions); **B** shows the end-state of all 20 simulation runs overlaid. Parameter settings: (i) minimal effort bias 3 times as strong for continuer word (G=1250) than for regular vocabulary words (G=5000), and (ii) the bias for reuse of features (i.e. segment-similarity bias) is not applied to the continuer category.

The simulation results show that while most words in most languages can differ arbitrarily, a distinct and cross-linguistically similar stratum of vocabulary can emerge under the cumulative effect of relatively small biases over many communicative events. They also show that commonality does not exclude diversity: while continuers are probabilistically pulled towards a similar part of the space, individual languages can still be organized in language-specific ways.

4. Conclusions

We have presented a first investigation into the cultural evolution of continuers. Cross-linguistic evidence suggests that spoken languages make available highly similar forms for this function. While frequency-driven reduction may partly account for the minimal shape of continuers, this is not the full story: more frequent words are still shorter on average, yet they are more phonemically diverse and less skewed in terms of word forms. Horizontal and vertical diffusion may be responsible for some similarities, but are unlikely to explain the prevalence of similar forms in over 20 independent language families. The crosslinguistic evidence is in line with the proposal that conversational infrastructure can drive convergent cultural evolution. Computational modelling supports this account by showing how a combination of effort minimization and a relaxation of the pressure for the reuse of phonetic features can push continuer-like words to the same low-effort yet distinctive part of the possibility space across languages.

The work reported here is ongoing and we are aware of a number of limitations. While written annotations are already telling, they reduce fluid signals to discrete categories, so the next step is to derive more fine-grained measures of similarity and dissimilarity from richer phonemic representations and audio signals (Fig. 3C). Further, continuers are often realized multimodally with blinks and nods; however, the subset of conversational corpora for which video is available is still limited. Finally, we have focused on spoken language, not because sign languages do not have continuers (Mesch, 2016), but because the few sign language corpora available do not systematically annotate items like this.

We have combined linguistic evidence with computational modelling to formalise and test a proposed process of conversational ecology driving convergent cultural evolution. The observational evidence suggests that continuers are pushed to a particular part of the possibility space of linguistic forms in similar ways across unrelated languages. The computational modelling shows that selective pressures enacted over cumulative cultural evolution can produce such a pattern. While modelling does not deliver definitive answers about the history of natural languages, it does shed light on the probability of explanations, and has the added virtue of requiring clarity in formulating assumptions and predictions. The findings have implications for models of cultural evolution and for our understanding of the fundamentals of human interactional infrastructure.

Acknowledgements

We thank Andy Wedel for providing insight into his original models and Jacqueline van Arkel for help in implementing the computational model. Funding comes from Dutch Research Council (NWO) Vidi grant 016.vidi.185.205 to MD.

References

Bateson, G. (1972). *Steps to an ecology of mind; collected essays in anthropology, psychiatry, evolution, and epistemology.* University of Chicago Press.

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology*, *79*(6), 941–952.

Boersma, P., & Weenink, D. (2013). *Praat: Doing phonetics by computer.*

Croft, W. (2000). *Explaining Language Change: An Evolutionary Approach.* Harlow: Pearson Education Limited.

Dingemanse, M., Torreira, F., & Enfield, N. J. (2013). Is "Huh?" a universal word? Conversational infrastructure and the convergent evolution of linguistic items. *PLOS ONE*, *8*(11), e78273.

Enfield, N. J. (2014). *Natural Causes of Language: Frames, biases and cultural transmission.* Berlin: Language Science Press.

Gardner, R. (1997). The Conversation Object Mm: A Weak and Variable Acknowledging Token. *Research on Language & Social Interaction*, *30*(2), 131–156.

Goodwin, C. (1986). Audience diversity, participation and interpretation. *Text*, *6*(3), 283–316.

Jadoul, Y., Thompson, B., & De Boer, B. (2018). Introducing parselmouth: A python interface to praat. *Journal of Phonetics*, *71*, 1–15.

Jefferson, G. (1985). Notes on a systematic Deployment of the Acknowledgement tokens 'Yeah' and 'Mmhm'. *Papers in Linguistics*, *17*(2), 197–216.

Mesch, J. (2016). Manual backchannel responses in signers' conversations in Swedish Sign Language. *Language & Communication*, *50*, 22–41.

Ozerov, P. (2018). *A community-driven documentation of natural discourse in Anal, an endangered Tibeto-Burman language.* (Tech. Rep.). Endangered Languages Archive.

Roberts, S. G., & Levinson, S. C. (2017). Conversation, cognition and cultural evolution: A model of the cultural evolution of word order through pressures imposed from turn taking in conversation. *Interaction Studies*, *18*(3), 404–431.

Rühlemann, C. (2018). TCU-initial backchannel overlap in storytelling. *Narrative Inquiry*, *28*(2), 257–279.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, *50*(4), 696–735.

Schegloff, E. A. (2006). Interaction: The Infrastructure for Social Institutions, the Natural Ecological Niche for Language, and the Arena in which Culture is Enacted. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition, and human interaction* (pp. 70–96). Oxford: Berg.

Ward, N. (2006). Non-lexical conversational sounds in American English. *Pragmatics & Cognition*, *14*, 129–182.

Wedel, A. (2012). Lexical contrast maintenance and the organization of sublexical contrast systems. *Language and Cognition*, *4*(4), 319–355.

Winter, B., & Wedel, A. (2016). The Co-evolution of Speech and the Lexicon: The Interaction of Functional Pressures, Redundancy, and Category Variation. *Topics in Cognitive Science*, *8*, 503–513.

Yngve, V. (1970). On getting a word in edgewise. In *Papers from the Sixth Regional Meeting, Chicago Linguistic Society* (pp. 567–578).