Title
The complexity of a language is shaped by the communicative needs of its users and by the hierarchical nature of their social inferences

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

Authors
Nedelcu, Vlad C
Smith, Kenny

Publication Date
2022

Peer reviewed

The complexity of a language is shaped by the communicative needs of its users and by the hierarchical nature of their social inferences

Vlad C. Nedelcu (v.c.nedelcu@sms.ed.ac.uk), Kenny Smith (kenny.smith@ed.ac.uk)
Centre for Language Evolution, University of Edinburgh

Abstract

Recent experimental and computational modelling work has found that languages are shaped by the referential context in which they operate. Wray & Grace (2007) argue that even compositionality, traditionally regarded as a universal and fundamental feature of human languages, may have only culturally evolved in response to changing social contexts. But how can the referential contexts of individual interactions come to shape the level of compositionality in the language of an entire community? To explore this question, we propose an iterated hierarchical Bayesian model that shows how partner-specific linguistic innovations can be generalized as community-wide features via a context-sensitive pathway. Our simulations show that the degree of compositionality that evolves in the language of a community depends on the communicative needs of its members, but also on the degree of user uncertainty over the nature of those needs, and on the level of heterogeneity in the community's needs.

Keywords: communication; context; compositionality; language; generalization; hierarchical modelling; pragmatics; cultural evolution; convention

Introduction

Human languages appear to strike a balance between communicative power, allowing us to convey the conceptual distinctions that matter to us, and relative ease of learning, allowing languages to be learned and transmitted across generations. The cultural evolutionary account (Christiansen & Chater, 2008) proposes that this balance is a result of linguistic transmission across multiple generations of users, which forces languages to adapt over time to the biases and constraints of language learners, and to the communicative needs of language users (Kirby et al., 2015). In this paper we introduce a computational model which allows us to explore how the same processes can lead to different linguistic design solutions in communities which differ in the communicative needs of their members or the heterogeneity of those needs. The model also offers an account of how innovations in language use determined by the referential context of local interactions can eventually become conventionalized into community-wide linguistic features.

Recent years have seen a surge in interest in studying how linguistic complexity could be shaped by the structure of the environment in which communication takes place, including socio-cultural factors such as community size (Nettle, 2012; Reali et al., 2018), geographic spread (Meir et al., 2012), degree and type of language contact (Lupyan & Dale, 2010), proportion of adult non-native speakers in a community (Lupyan & Dale, 2010; Bentz & Winter, 2014), and structure of the network of interactions (Trudgill et al., 2009). Wray & Grace (2007) posit that even some of the linguistic features that we regard as universal today, like compositionality (the fact that utterances are composed of parts, with the meaning of the whole being a function of the meaning of the parts and the way in which they are combined), may have only evolved in response to changing social structures. Specifically, they argue that the first language users lived in relatively small, geographically concentrated, and homogeneous communities, which allowed them to establish intimate connections and significant common ground between one another. As interactions would have taken place almost exclusively between members of the same community, languages would have been defined by opaque and irregular forms with high context-dependence. As such communities expanded and became more heterogeneous, communication with non-intimates would have become increasingly frequent, meaning that interlocutors could not consistently rely on common ground to make themselves understood, having to express their messages in a much more explicit and systematic manner, and to make as few assumptions as possible about their communicative partners. These conditions would have favored the evolution of transparent, regular, context-independent compositionality.

While some socio-cultural factors have received some attention in the modelling and experimental literature (as noted above), the process by which community-level differences in shared knowledge and reliance on context might lead to systematic differences in language complexity has received far less attention. Across two artificial language studies Winters et al. (2015, 2018) show that the amount of contextual information that a speaker can utilize to reduce referential uncertainty in a given situation influences the degree of compositionality in their language: less available context leads to more compositionality. However, these experiments only investigated dyads, where individuals interact repeatedly with a single partner, and are therefore unsuited to studying situations where speakers face continued uncertainty about their partners' communicative needs, e.g. in communities with heterogeneous communicative needs. Meanwhile, an account of how partner-specific knowledge can become conventionalized into population-level expectations through adaptation of the lexicon has been proposed by Hawkins et al. (2019) and

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

Hawkins et al. (2021). They formulate their theory in a hierarchical model, and also show (similarly to Winters et al., 2015) that conventions are sensitive to the context of communication. However, context-sensitivity is again only explored in dyads, as adaptation to a specific communicative partner, and the question of how these conventions would evolve when transmitted across generations is not addressed. Here we combine insights from these two approaches, manipulating reliability of context and speaker-specific adaptations in pair-based interaction while also modelling adaptation to higher-level, community-wide representations. This allows us to explore Wray & Grace's conjecture, and show how the complexity of linguistic systems responds to communicative need in homogeneous and heterogeneous communities. Our model predicts that the languages of communities with heterogeneous communicative needs evolve a high level of compositionality, and thus of complexity, even when the needs of all individual members of the community could be satisfied by a simpler language. Moreover, we find that speaker uncertainty over a partner's communicative context will gradually push the language towards more complexity than required.

Model

The agent-based model that we propose is hierarchical at its core, as it considers abstractions of socio-linguistic knowledge at two distinct levels: partner-specific and population-level (see Fig. 1 left). More specifically, an agent maintains a separate partner-specific representation of context for each communication partner, but a single overhypothesis about the distribution of types of contexts within the community, and a single community-wide distribution of possible languages. This setup allows an agent to dynamically adapt its language use to the communicative needs of individual partners, but also offers a mechanism for generalization of linguistic and contextual knowledge across different partners.

The components of our model reflect the mechanisms that act to shape these abstractions: language learning, language use, and language transmission (see Fig. 1 center). One iteration of our model is split into two main phases: learning and communication. A number of naive agents are initially trained on a set of meaning-utterance pairs, before being prompted to communicate about a series of referents. We implement learning as a process of Bayesian inference: agents observe linguistic data (i.e., how meanings are conveyed using utterances), then form a hypothesis about the language underlying those observations. This process of learning is influenced by a prior favouring simpler language types (see below). Communicative interactions between agents take place in the form of an asymmetric reference game, where each agent is assigned either the role of speaker or listener for the whole game. Each game is split into a number of independent rounds, in which the speaker is provided with a target meaning to convey, and the listener is confronted with a context consisting of that target plus a number of distractors, which put together form a communicative context. Language use is modelled as pragmatic reasoning following the Rational Speech Act framework (Goodman & Frank, 2016): speakers aim to produce utterances that are informative to listeners relative to their particular needs, and listeners are aware of this strategy when interpreting those utterances. At the end of each communication round, all agent are informed whether communication was successful, and update their expectations. Finally, language transmission is modelled following the iterated learning framework (Kirby, 2001): the agents in a generation observe data produced during the communication phase by speakers in the previous generation, while the data produced during their own interactions will be learned from by the following generation during the learning phase. This process will be repeated for a large number of iterations.

Representations

We represent languages in a very simple manner: a language is a set containing pairings of meanings and their associated signals. Meanings and signals are made up of smaller units: a meaning is a set of semantic features that characterize the concept being referred to, while a signal is an ordered sequence of characters. We chose minimal parameters for our experiments: meanings consist of two features each with two possible values (values of the first feature are drawn from $\{0, 1\}$, values of the second feature from $\{2, 3\}$, leading to 4 possible meanings) and signals are strings of length two composed from an alphabet of two possible characters, $\{a, b\}$, leading to 4 possible signals.

This simple representation leads to a set of 256 possible languages (mappings from all 4 meanings to a signal), which can be grouped into classes depending on their level of *simplicity* (following Smith et al., 2013). The highest on this scale are *degenerate* languages, which express all meanings using a single, completely ambiguous signal. These are followed by *one-feature* languages, which map all meanings sharing one of the two features to the same signal, and these in turn by *compositional* languages, which have consistent mappings for both features that make up the meaning. *Hybrid* languages, which have at least one ambiguous signal and mix the strategies used by the previous classes, sit on the scale just above *holistic* languages, which idiosyncratically map every meaning to a distinct signal, thus being the most complex. We assume that simpler language types have higher prior probability and are therefore more learnable. These language classes can also be assessed by their level of *expressivity* (i.e., their capacity to express meanings as unambiguously as possible): expressivity is approximately the inverse of simplicity, with the exceptions that hybrid languages are less expressive than compositional languages, and compositional and holistic languages are fully and equally expressive. Example languages are depicted in Fig. 2.

Speakers also maintain partner-specific representations of the type of context t_i for each communicative partner i that they encounter. In the setup of our communication game, an agent's type of context determines the semantic features in which the referents in any given context can differ for that

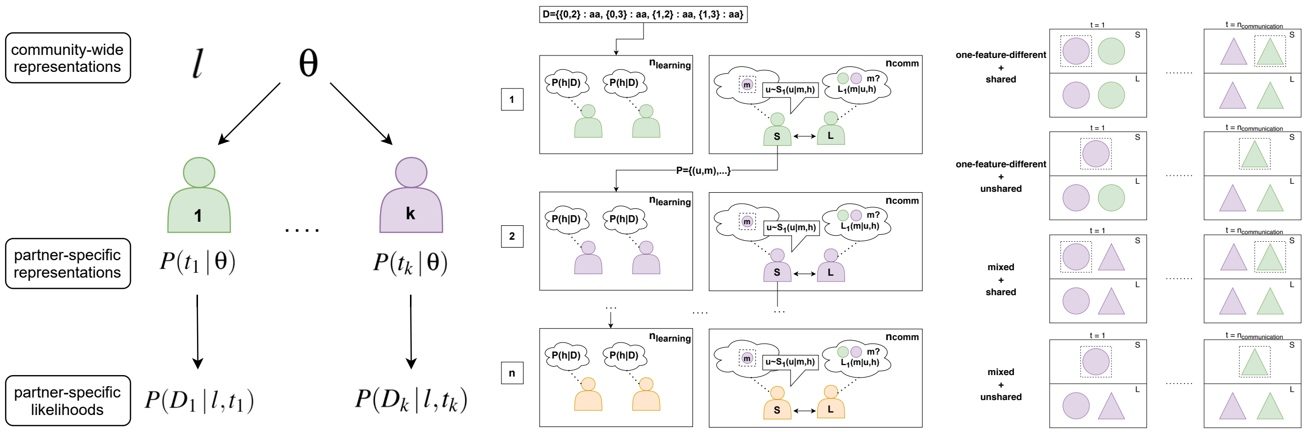

Figure 1: (left) Schematic of hierarchical representations: each agent maintains one partner-specific representation of context for each partner (t_k), one overhypothesis about the distribution of contexts across the community (θ), and one community-wide distribution of languages (l). (center) Overview of a simulation: n generations of agents are arranged in a chain; in the learning phase (left side), agents learn from a set of meaning-signal pairs D for $n_{learning}$ rounds; in the communication phase (right side) agents play the reference game for $n_{communication}$ rounds, and their utterances will be learned from by the next generation. (right) Examples of communication phase setups for our four settings: a phase consists of multiple rounds of communication, and each square represents the available contextual information from the perspective of the speaker (upper box) and listener (lower box), for one round; we represent referents as geometrical shapes, with shape and colour as the two features.

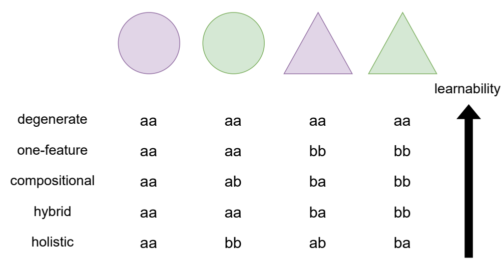

Figure 2: Example languages for all our five classes, showing how meanings (depicted as coloured shapes with two shapes and two possible colours) map to signals.

agent, with this aspect being fixed for the whole duration of a game. In a *one-feature-different* context-type, a single feature is sufficient for discriminating among all the referents in a context (e.g. potential referents in a context might differ in their first feature), and that feature is furthermore consistent across all rounds. Following Winters et al. (2018), a *mixed* context-type is one where the meanings in any given context also differ in one feature, but that feature is not fixed across the interaction and differs on a round-by-round basis (e.g. the relevant feature might be feature one in the first round, and feature two in the second round). We represent contexts as sets of meanings, and define a context-type as the set of all contexts that it includes (see Fig. 1 right). Uncertainty about a partner-specific context-type t_i is parameterized by θ, the overhypothesis about the general distribution of possible types of contexts across the community: $P(t_i|\theta)$.

Rather than detail the hierarchical aspects from the start, we will first describe a non-hierarchical, single-partner variant, which we use to evaluate our model's predictions against the experimental results of Winters et al. (2015), and to ground our main model. We will then introduce the extensions that make up the hierarchical, multiple-partner variant.

Learning phase

At the start of a generation, each agent in the population updates its posterior distribution over the hypotheses $H = (l, t)$ by learning from a set of pairings of signals and their associated meanings D. The source of the data set is either the previous generation's produced output during the communication phase or, for the first generation, meaning-utterance pairs drawn from a fully degenerate language.

$$P(l, t \mid D) \propto \prod_{(m,u) \in D} S_0(u \mid m, l) P(l) P(t) \tag{1}$$

$$S_0(u \mid m, l) = \begin{cases} 1 - \varepsilon_{err}, & (m, u) \in l; \\ \frac{\varepsilon_{err}}{|sgn| - 1}, & (m, u) \notin l \end{cases} \tag{2}$$

where S_0 gives the probability that a literal speaker using language l will produce utterance u to convey meaning m; $P(l)$ and $P(t)$ are the priors over languages and context-types respectively; ε_{err} is the probability that the literal speaker makes a mistake and chooses a signal which is not associated with the target referent in their language (which we set to 0.06), $|sgn|$ is the number of signals in a language (in our case 4).

Note that in this phase agents do not distinguish the hypotheses in terms of context-type, as learning happens in a context-free setting, and similarly we assume that learners interpret utterances in a literal rather than pragmatic way (following e.g. Kirby et al., 2015).

Communication phase

In an individual communication round, the pragmatic speaker samples a hypothesis from its distribution, which contains the language that the agent will use for communication, l, and the context-type that it assumes the listener faces, t. Next, the speaker chooses an utterance u to express the given meaning m, by sampling from a distribution S_1: the speaker reasons about how a literal listener would interpret each possible utterance in one of its possible contexts, then selects utterances proportionally to the probability that they will be correctly interpreted by the listener. As the speaker cannot know the specific context of the listener, it will have to be informative on average with respect to the set of contexts $c_i \in C(t)$ where $C(t)$ is the set of all contexts that are part of context-type t.

$$\begin{aligned} S_1(u \mid m,t,l) &\propto L_0(m \mid u,t,l) \\ &\propto \sum_{c_i \in C(t)} \frac{1}{|C(t)|} L_0(m \mid u,c_i,l) \end{aligned} \tag{3}$$

$$L_0(m \mid u,c_i,l) = \begin{cases} \frac{1-\varepsilon_{err}}{|p_u|}, & (m,u) \in l \quad and \quad m \in c_i; \\ \frac{\varepsilon_{err}}{|mng|-|p_u|}, & (m,u) \notin l \quad and \quad m \in c_i; \\ 0, & m \notin c_i \end{cases} \tag{4}$$

where $|p_u|$ is the number of meanings that map to utterance u in language l and are part of context c_i, ε_{err} is the probability that the literal listeners makes a mistake (set to 0.06); $|mng|$ is the number of meanings that are part of context c_i.

After the listener observes the speaker's utterance u, it samples its own language l from its posterior distribution, then applies pragmatic inference to guess the intended meaning m given its context c, by sampling from distribution L_1.

$$L_1(m \mid u,c,l) \propto S_1(u \mid m,c,l)\, P(m) \tag{5}$$

where $P(m)$ is a flat prior over the meaning space.

Referential feedback

Finally, after each interaction the speaker and listener receive feedback on the success of that interaction (i.e. did the listener correctly identify the target meaning?) and update their posterior by learning from their partner's behaviour. For the listener, the data used for the update depends on communication success: if the listener identified the correct meaning, it learns from the pair (u,m), otherwise it learns from all the other possible pairings of u and a referent from $c \setminus m$.

Thus, after each interaction with its partner k, a listener will update its distribution by considering how likely each of the meaning-utterance pairs in the set of observations from all rounds of interaction up to that point D_k would have been to be expressed by a pragmatic speaker that uses language l in context-type t_k. The inference will also incorporate the posterior update from the learning phase P_{learn}.

$$P(l,t_k \mid D_k) \propto P_{learn}(l,t_k) \prod_{(m',u) \in D_k} S_1(u \mid m',c,l) \tag{6}$$

Similarly, the speaker will update its distribution over languages after observing the listener's behaviour. However, the feedback given to the speaker has one more crucial function: it is used in inferring the context-type of the current listener, as the speaker reasons about how likely the pair would have been to be correctly interpreted by the literal listener for each of the contexts that form a particular context-type.

$$P(l,t_k \mid D_k) \propto P_{learn}(l,t_k) \prod_{(m',u) \in D_k} L_0(m' \mid u,t_k,l) \tag{7}$$

The speaker will use this information in the next round, when it will again have to sample a context-type.

Simulation 1: interaction with a single partner

We present results for the non-hierarchical model where the speaker at each generation interacts with a single listener for 90 rounds, and 60 randomly-selected meaning-utterance pairs are transmitted to the next generation. To explore the effects of speaker uncertainty about listener context (following e.g. Winters et al., 2015) we present results for two variants of the model. In the *shared context* version, we assume that the speaker has direct access to the listener's context: since the speaker knows the context in which their utterance will be interpreted, it is trivial to identify which distinction is sufficient for successfully conveying the intended meaning on a round-by-round basis. We compare results to the *unshared context* model, as described above, where the speaker has uncertainty both about the context-type of the listener and which specific context the listener faces on any given round.

If access to the context is shared and referents can be differentiated using the same feature across all rounds (Fig. 3A), conveying only that feature will be enough to guarantee successful communication. Consequently, we see that the initial degenerate language is rapidly abandoned and one-feature languages, which are simple yet communicatively functional in one-feature contexts, becoming dominant. In the mixed context-type condition (Fig. 3B), using the previous strategy no longer guarantees success, which eventually leads to the emergence of compositional and hybrid languages. We note that the initial emergence of one-feature languages can be attributed to them having a significantly higher prior, while also guaranteeing success in communication for over half of the possible contexts. Another interesting observation is that the partially-compositional hybrid languages consistently outperform compositional languages throughout our simulations, despite having a worse prior and not being fully expressive. This can be explained by their much higher number and, consequently, their much larger total prior probability, coupled with the relative simplicity of the task: agents coordinate on a hybrid language early on and the benefits of full compositionality are too small to cause a subsequent re-coordination.

When context is not shared, speakers must infer the context-type of the listener over the course of communication. Critically, this introduces an additional level of uncertainty as to what context-type the speaker should design their

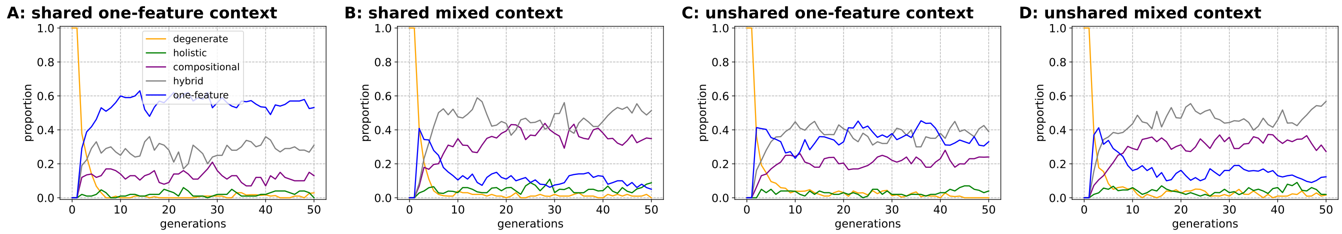

Figure 3: Simulation results for the single-partner variant showing how the posterior distribution of language classes across a population evolves over the first 50 generations of a 200 generation simulation, averaged over 100 separate runs.

utterance for. In a one-feature-different context-type condition, speakers are thus encouraged to produce less context-dependent utterances, resulting in a higher proportion of more expressive systems than in the case of a shared context (Fig. 3C). Similarly, in the mixed context-type condition (Fig. 3D), the added level of uncertainty causes a slightly higher proportion of less expressive languages to evolve alongside the still dominant compositional and hybrid languages. These results generally match the experimental findings of Winters et al. (2015).

For statistical analysis, we coded each generation of each run as either featuring a one-feature language type or a compositional/hybrid one, depending on which had the higher posterior probability. We ran a mixed effects logistic regression, discarding the first 50 generations of each run for burn-in. We included a by-run random intercept, and used context-type (one-feature-different/mixed) and sharedness of context (unshared/shared) as fixed effects with interaction. We found a significant effect for context-type (b=1.08, SE=0.036, p<.001), as well as for the interaction between the two factors (b=-0.39, SE=0.036, p<.001), the latter showing that in an unshared setting, one-feature contexts evolve significantly less one-feature languages than would be expected given the independent contribution of context-type. As expected, the effect of sharedness alone was found to not be significant.

Multi-partner hierarchical extensions

In the multi-partner setting, each generation contains a single speaker who is paired with a fixed number of distinct listeners (i.e., here we show results for 3 listeners) and is prompted to interact with each of them in blocks: after interacting with a listener for 30 rounds, the speaker updates its overhypothesis over the space of context-type representations, then moves on the next listener. This process repeats until the speaker has interacted with all other partners. At the end of the communication game, two thirds of the data pairs produced during all interactions are randomly selected and transmitted to the next generation (i.e., 60 pairs). The blocked approach allows us to only track one partner-specific context-type at any point, that of the current partner, as the hyper-parameter θ will encapsulate all the relevant information of previous blocks[1].

Given an agent's observations D_k from interacting with partner k, updates of both its partner-specific beliefs, as well as its community-wide beliefs, are done using a single joint hierarchical inference:

$$P(l, t_k, \theta \mid D_k) \propto P(D_k \mid l, t_k)\, P(l)\, P(t_k \mid \theta)\, P(\theta) \tag{8}$$

In the above formulation, the prior term $P(l)\,P(t_k \mid \theta)\,P(\theta)$ captures the idea that interactions with a novel partner will be initially guided by a priori beliefs of the context-types found across the community and of the language used within the community. Conversely, the likelihood term $P(D_k \mid l, t_k)$ allows an agent to adapt to a partner's specific needs, once enough evidence of what those needs are has been gathered. To compute the likelihood, the speaker will reason about how likely a literal listener is to interpret utterance u as conveying meaning m under different (l, t_k), for each pair (m, u) in D_k: $P(D_k \mid l, t_k) \propto \prod_{(m,u)\in D_k} L_0(m \mid u, t_k, l)$. Likewise, the listener will consider the probability that a pragmatic speaker produces utterance u to express meaning m within its context c: $P(D_k \mid l, t_k) \propto \prod_{(m,u)\in D_k} S_1(u \mid m, c, l)$.

Using the joint formulation for the posterior update comes with two important benefits. First, at any point throughout the reference game, it allows the agent to sample a hypotheses which is specifically adapted to the needs of a given partner, by marginalizing over the community-wide representation.

$$P(l, t_k \mid D_k) = \sum_{\theta} P(l, t_k, \theta \mid D_k) \tag{9}$$

Second, its hierarchical nature offers a simple pathway of transferring previous experience to an interaction with a novel partner, by marginalizing over the rest of the parameters. This is essential for updating the two community-wide representations (l and θ) at the end of a communication block.

$$P(l \mid D) = \sum_{t} P(l, t \mid \cup_k D_k) \tag{10}$$

$$P(\theta \mid D) = \sum_{l,t} P(l, t, \theta \mid \cup_k D_k) \tag{11}$$

$$t = t_1 \times t_2 \times \ldots \times t_N \tag{12}$$

[1]We actually restrict the range of possible distributions for θ to a discrete set of only 10 possible distributions. This was done to ensure that exact inference for obtaining all the posteriors in our model remained possible. However, an alternative Dirichlet-Multinomial model that approximates some of the posteriors using Markov chain Monte Carlo is currently being explored.

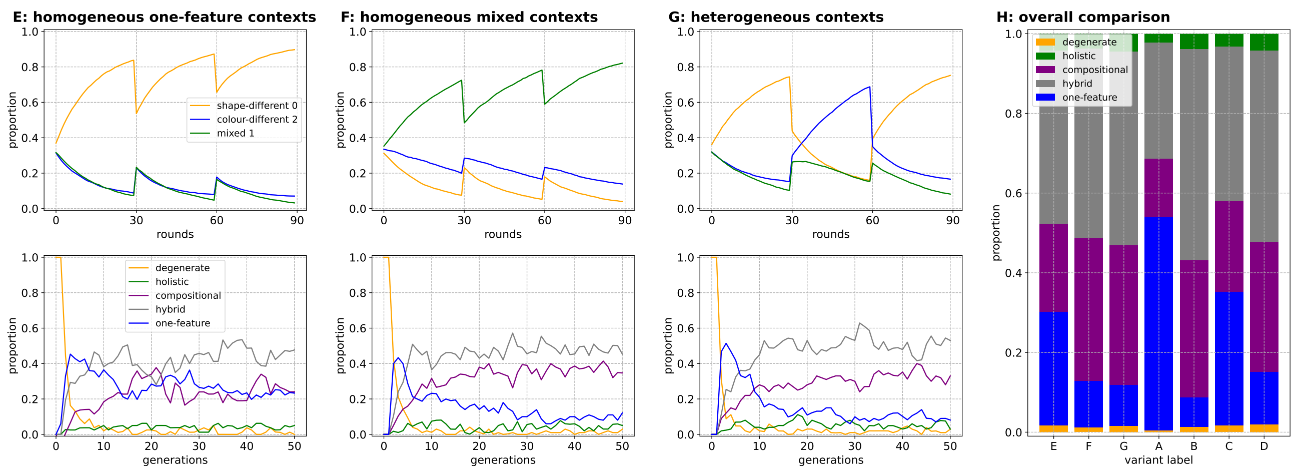

Figure 4: (E, F, and G) Simulation results for the multi-partner model: top row shows how uncertainty over context-types develops over the course of interactions with the three partners; bottom row shows how the posterior distribution of language classes across a population evolves over a chain of 50 generations, averaged over 100 separate runs. (H) Comparison between the posterior distribution of language classes across generations 50 to 200 of all described conditions, averaged over 100 runs.

Simulation 2: interaction with multiple partners

We report results for three conditions, all involving unshared contexts: two homogeneous conditions where all partners a speaker encounters have the same context-type, and need the same feature(s) to be encoded linguistically for disambiguation (e.g., only shape, only colour, or both), and a heterogeneous condition testing the interesting case in which all partners have one-feature contexts, but differ in whether they need the first or second feature encoded for disambiguation.

We first look at how speaker uncertainty over context-types develops over the course of an interaction with multiple partners. When the speaker starts interacting with its first partner, both the community-wide hyperprior and the partner-specific prior will be uninformative, so the speaker's utterances will initially be designed for a context-type chosen at chance. As the speaker starts observing its partner's behaviour, it will rapidly start inferring the correct context-type for that interaction. Crucially, the speaker's beliefs about the community-wide distribution of context-types will also be updated to incorporate any information extracted from the listener's behaviour, and this overhypothesis will provide a weak bias towards the previous partner's context-type when interacting with the second partner. If this partner meets the speaker's previously established beliefs (Fig. 4E, 4F), the bias towards a skewed distribution of context-types will become stronger. However, in case of a partner with different communicative needs than its predecessor (Fig. 4G), the bias will shift towards a non-homogeneous community-level distribution.

In the setups where the speaker's partners are homogeneous in facing one-feature or mixed contexts (Fig. 4E, 4F), the hierarchical model shows a similar pattern of results to the single-partner simulations: we see the emergence of languages that best compromise between ease of learning, adaptation to the homogeneous context-type, and robustness to uncertain inferences. However, for multiple partners with one-feature contexts, more compositionality emerges than in the case of a single partner, as the successive swapping of partners causes more uncertainty over context-types, thus encouraging agents to use more context-independent utterances (Fig. 4H, compare C and E). If the speaker interacts with partners with heterogeneous context-types, we see additional effects of this heterogeneity (Fig. 4G): compositional and hybrid languages become dominant, as speakers must compromise on using a language that encodes both features, thus maximizing the communicative success across partners.

We ran a similar statistical analysis as in the single-partner case, but additionally used Helmert coding for the condition factor, as it has 3 levels. The distribution of languages in the homogeneous mixed and heterogeneous conditions were not significantly different (b=0.03, SE=0.063, p=.674), while the homogeneous one-feature condition differed significantly from the other two conditions (b=0.30, SE=0.035, $p<.001$).

Conclusion

What mechanisms allow the contexts of individual interactions to shape the level of complexity in the language of an entire community? Under our account, this is a result of the hierarchicality of social inferences: higher-level, community-wide linguistic features emerge as generalizations of lower-level, partner-specific strategies. Using our model, we show that the level of compositionality that evolves in the language of a community depends on the communicative needs of its members, but also on the degree of speaker uncertainty over the nature of those needs. We also test the theory proposed by Wray & Grace for the emergence of compositionality, and find that compositionality can emerge in communities where simpler languages would satisfy the individual needs of its users, if the community's needs as a whole are heterogeneous.

Acknowledgements

This research received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (Grant 681942, held by K. Smith), and from a University of Edinburgh College of Arts, Humanities and Social Sciences Research Award (held by V. Nedelcu).

References

Bentz, C., & Winter, B. (2014). Languages with more second language learners tend to lose nominal case. In *Quantifying language dynamics* (pp. 96–124). Brill.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and brain sciences*, *31*(5), 489–509.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in cognitive sciences*, *20*(11), 818–829.

Hawkins, R. D., Franke, M., Frank, M. C., Goldberg, A. E., Smith, K., Griffiths, T. L., & Goodman, N. D. (2021). From partners to populations: A hierarchical bayesian account of coordination and convention. *arXiv preprint arXiv:2104.05857*.

Hawkins, R. D., Goodman, N. D., & Goldstone, R. L. (2019). The emergence of social norms and conventions. *Trends in cognitive sciences*, *23*(2), 158–169.

Kirby, S. (2001). Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, *5*(2), 102–110.

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

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PloS one*, *5*(1).

Meir, I., Israel, A., Sandler, W., Padden, C. A., & Aronoff, M. (2012). The influence of community on language structure: evidence from two young sign languages. *Linguistic Variation*, *12*(2), 247–291.

Nettle, D. (2012). Social scale and structural complexity in human languages. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1597), 1829–1836.

Reali, F., Chater, N., & Christiansen, M. H. (2018). Simpler grammar, larger vocabulary: How population size affects language. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1871), 20172586.

Smith, K., Tamariz, M., & Kirby, S. (2013). Linguistic structure is an evolutionary trade-off between simplicity and expressivity. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 35).

Trudgill, P., et al. (2009). Sociolinguistic typology and complexification.

Winters, J., Kirby, S., & Smith, K. (2015). Languages adapt to their contextual niche. *Language and Cognition*, *7*(3), 415–449.

Winters, J., Kirby, S., & Smith, K. (2018). Contextual predictability shapes signal autonomy. *Cognition*, *176*, 15–30.

Wray, A., & Grace, G. W. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, *117*(3), 543–578.