Degree of heterogeneity in the contexts of language users mediates the cognitive-communicative trade-off in semantic categorization

Vlad C. Nedelcu (v.c.nedelcu@sms.ed.ac.uk), Kenny Smith, Daniel Lassiter
School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, EH8 9AD

Abstract

It has been argued that patterns of cross-linguistic variation in the semantic categories labelled by individual words are a result of a trade-off between cognitive pressures (so as to be simple to learn and use) and communicative pressures (so as to be efficient in communication). However, the question of what exact mechanisms control this trade-off has been left largely unanswered. We argue that one factor could be the extent to which referential contexts at the level of local interactions are similar or different across users of a category system. To test this hypothesis we propose a hierarchical Bayesian model for communication in a multidimensional meaning space, in which agents actively consider spatial similarity relations during interaction. Our models predict that less variability in contexts across interactions induces categories with lower communicative cost, while more variable contexts across partners are more strongly associated with category systems with lower cognitive cost.

Introduction

Identifying the principles that govern how different lexical meanings can be grouped together into categories so as to be expressed by the same lexical item is a central issue for lexical typology (Koptjevskaja-Tamm et al., 2007). Despite different languages varying widely in their approach, crosslinguistic research has shown that this variation is far from arbitrary (Kemp et al., 2018). For example, in their study focusing on the categorization of cutting and breaking events, Majid et al. (2008) found that all the analyzed languages shared a small common set of dimensions that characterized the whole semantic space, with variation only in the exact number of categories and the placement of boundaries between them.

A naturally related problem has thus been to determine what causes this relatively constrained variation. It has been widely argued (see Kemp et al., 2018 for a review) that these restricted spaces are a result of meaning systems being universally constrained by two competing types of pressures: cognitive (i.e., to require minimal cognitive effort to be learned and used) and communicative (i.e., to be sufficiently informative for effective communication to be possible). There are now many domains where categorization has been demonstrated to reflect both types of factors, including color names (Zaslavsky et al., 2018), kinship terms (Kemp & Regier, 2012), and person systems (Zaslavsky et al., 2021).

However, the question of what exactly controls the trade-off between the cognitive and communicative factors so as to lead to the crosslinguistic variation that we find in attested category systems has been left largely unaddressed (Levinson, 2012). A more recent line of research (e.g., Zaslavsky et al., 2020) has proposed that differences in local communicative needs could have an important role in explaining how the two types of factors are weighted. Specifically, variation in categorization has been attributed to differences in concept-level need (i.e., how often different cultures need to communicate about some individual concepts within a domain), as well as domain-level need (i.e., how important a specific domain is relative to other domains for different cultures). But while conventions that are globally shared by a community, such as those of a category system, are expected to be shaped by the communicative needs of that community, these conventions have to first emerge over the shorter timescales of dyadic interaction. Indeed, communication partners are known to very rapidly form local conventions over interactions when it is communicatively relevant to do so (Clark & Wilkes-Gibbs, 1986). Crucially, because they emerge over shorter timescales, local conventions face unique contextual pressures, as dyadic interactions are always placed in a particular context, with that context supplying important information about the intended meaning of the speaker and about the individual needs of the listener. Thus, it may be valuable to expand any investigation into the role of communicative need in shaping semantic variation to cover not only the global but also the local level, where context plays a key role.

While context is often ignored in semantic categorization studies (Kemp et al., 2018), new evidence suggests that languages are shaped by their users' need to rapidly adapt to changing contexts (R. D. Hawkins et al., 2022). Recently, Nedelcu & Smith (2022) have shown through computational simulations that the languages evolving in communities where referential contexts constantly shift across partners will exhibit more compositionality than those of communities where contexts are stable across partners. In this paper, we use computational simulations to study whether this factor could mediate the trade-off between cognitive and communicative constraints in semantic categorization. We predict that less variability in context across interactions will induce categories that are more context-dependent and with lower communicative cost. Conversely, more variable contexts across partners should be associated with category systems that are shaped more strongly by the structure of the semantic space and the pressure of reducing cognitive cost.

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

Additionally, R. X. Hawkins et al. (2018) point out that much existing research on the communicative-cognitive trade-off in semantic categorization consists of observations made from statistical analysis of linguistic data, or experiments involving descriptive or classification tasks. There is considerably less work on the mechanisms involved in the evolution of efficient category systems. Recent work using artificial language experiments and Bayesian modelling has investigated the role that cultural transmission (Carstensen et al., 2015), communication (Carr et al., 2017) and their interaction (Silvey et al., 2019) could play in this process. In contrast, the model that we propose in this paper shows how rapid partner adaptation of agents engaging in recursive social reasoning could result in similarly highly-structured categories.

Model

The basic problem addressed here is that of linguistic categorization: how do language users choose a label for a newly encountered meaning, and how does that label shift with changes in communicative need. Our model explores this problem in the context of an asymmetric reference game, where a speaker communicates with a number of listeners about meanings drawn from a shared domain. In this particular case, the domain consists of a structured meaning space with each meaning being composed of multiple continuous features. To succeed, agents need to coordinate on a categorization strategy for these meanings.

The reference game itself is preceded by a learning (i.e., pre-training) phase in which all agents observe the correct labels for a small subset of meanings so as to provide a shared but incomplete starting point that agents can leverage on during actual communication. The main, communicative component is split into a number of independent rounds in which the speaker and one of the listeners are presented with a context consisting of one target meaning and one or more distractor meanings, from which the target must be differentiated. The speaker knows the target and has to select a word for the listener to identify this target. At the end of each round the target is revealed to the listener and both agents are informed whether communication was successful. The speaker communicates with its partners in separate blocks, interacting with the first listener for a number of rounds, then moving on the next one, with this process repeating until the speaker has interacted with all listeners. Crucially, each listener is associated with a separate set of contexts from which one will be randomly sampled for each round. The extent to which these contexts are similar or different across listeners therefore influences whether the speaker faces an audience with homogeneous or heterogeneous communicative needs, in terms of the meanings that must be conveyed and which meanings must be differentiated linguistically. One of the key challenges that the speaker thus faces is to generalize its listener-specific categories across newly encountered listeners, so as to be flexible enough to adapt to the contexts of these subsequent listeners, while minimising disruption to established conventions.

RSA in a multidimensional meaning space

Our model is partly based on the Rational Speech Act (RSA) framework (Goodman & Stuhlmüller, 2013), which formalizes communication as recursive pragmatic reasoning: to express meaning m, a pragmatic speaker S_1 aims to produce an informative utterance u by reasoning about how a hypothetical literal listener L_0 would interpret u. A pragmatic listener L_1 similarly inverts its model of S_1 to infer meaning m.

$$S_1(u\,|\,m,c_k,l) \propto L_0(m\,|\,u,c_k,l)^{\alpha_S} \tag{1}$$

$$L_1(m\,|\,u,c_k,l) \propto S_1(u\,|\,m,c_k,l)^{\alpha_L}, \tag{2}$$

where c_k is the context of communication, l is the lexicon that an agent believes its partner is using, α_L and α_S control the optimality of the listener and speaker respectively.

In versions of RSA featuring structured meaning spaces (e.g. R. D. Hawkins et al. 2022), the relations between meanings are typically specified separately through the prior over lexicons. In our model, the structure of the meaning space is directly embedded into the inferences made by agents. Thus, L_0's choice whether to interpret word u as expressing meaning m depends not only on how well u describes m, but also on how well u describes meanings that are similar to m when compared to meanings that are less similar to m. The same logic also applies to S_0's choice whether to encode meaning m using word u.

$$L_0(m\,|\,u,c_k,l) \propto \begin{cases} \sum_{j=1}^{|M|} l_{uj}\, s_{mj}, & m \in c_k \setminus \{null\} \\ c_{null}, & m = null \\ 0, & m \notin c_k \end{cases} \tag{3}$$

$$S_0(u\,|\,m,l) \propto \sum_{j=1}^{|M|} l_{uj}\, s_{mj}, \tag{4}$$

where s_{mj} is the similarity between meanings m and j, l_{uj} is the entry in the lexicon that quantifies how well word u describes meaning j, c_{null} is the likelihood that word u describes the *null* meaning (see Representations subsection for more details on l_{uj} and c_{null}), $|M|$ and $|U|$ are the sizes of the meaning space and of the word space respectively.

The sums in (3) and (4) quantify the evidence in favour of labeling meaning m with word w, calculated by summing the suitability of each meaning j to be expressed by w, weighted by how similar j is to m. Note than even though S_0 is not part of the RSA recursion (i.e., the starting point being L_0 as shown in (1)), we instead use S_0 for inference in the learning phase that precedes actual communication to reflect the fact that data used to "pre-train" agents in this phase has not been produced by a rational agent in a specific context (mimicking some experimental settings e.g., Winters et al. 2018)

The formulas for to the literal agents as well as our method for modelling similarity relations among meanings are based on the standard version of the generalized context model

(GCM) as described by Nosofsky (2011). In essence, the GCM is an exemplar-based model that classifies novel meanings based on their similarity to already seen exemplars, with categorization being seen as a process of mentally encoding meanings rather than naming them as in our case (see Malt et al. 1999 for a discussion of recognition vs. linguistic categories). As in the GCM, we represent individual meanings as points in a multidimensional conceptual space where each dimension corresponds to a different feature, with similarity between two meanings being calculated based on the distance between the two corresponding points in this space.

$$s_{ij} = e^{-a d_{ij}} \tag{5}$$

$$d_{ij} = \left(\sum_n w_n \left| x_{in} - x_{jn} \right|^r \right)^{\frac{1}{r}}, \tag{6}$$

where d_{ij} is the distance in conceptual space between meanings i and j, and a is the rate at which similarity between meanings declines with distance (we keep it set to 1 for the current paper); r is a parameter that controls the form of the distance metric, and w_n are importance weights for the dimensions of conceptual space (we set all $w_n = 1$).

Formula (6) uses the Minkowski distance metric, a generalization of the Euclidean and Manhattan distances. In psychological studies, the exact type of metric most closely resembling that of humans was found to depend on the nature of the stimuli (Shepard, 1987), with Euclidean distances performing better for unitary stimuli (e.g., colour) and Manhattan distances for complex stimuli (e.g., shape). Here, we consider only the latter type of stimuli and take $r = 1$.

Generalizing through hierarchical inference

With the agents' behaviour during communication set out, we next turn to the lexicon representations that agents rely on when encountering different partners, and how these are updated throughout the reference game. As noted, speakers must be able to balance the need for rapid generalization across partners with that for flexible adaption to a specific partner. Our solution is to adopt a hierarchical structure, with a speaker tracking a separate partner-specific lexicon $l^{(k)}$ for each listener k, and a single community-wide representation of the aspects of lexicon that are expected to be shared across all listeners λ. While listeners only need to track one representation l throughout the game (i.e., because they interact with the same speaker), we will consider for simplicity that they do the same kind of hierarchical inferences as speakers.

We use a continuous representation for our lexicons to capture the idea that the same word can express multiple meanings, and the same meaning can be expressed by multiple words, to different extents. Thus, λ and l are real-valued matrixes with $|U|$ rows and $|M|$ columns where each entry (i, j) quantifies how well word u_i applies to meaning m_j:

$$l = \begin{bmatrix} l_{11} & l_{12} & \dots & l_{1|M|} \\ l_{21} & l_{22} & \dots & l_{2|M|} \\ \vdots & \vdots & \ddots & \vdots \\ l_{|U|1} & l_{|U|2} & \dots & l_{|U||M|} \end{bmatrix}$$

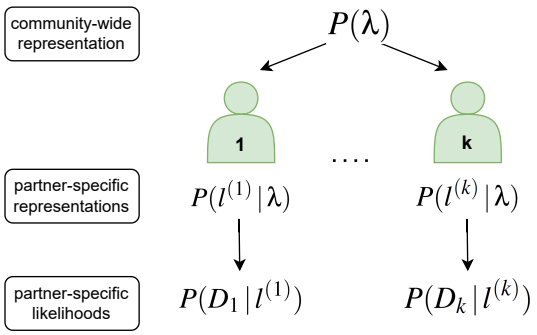

Figure 1: Schematic of hierarchical representations: each speaker tracks one partner-specific lexicon $l^{(k)}$ for each listener k, and one community-wide lexicon λ.

Even though an agent uses a separate l for each partner, due to our blocked structure of interactions, we only need to track two representations at any given point: that of the current partner and that of the community. Given a set of observed word-meaning pairs D_k from interacting with partner k, a joint inference is used for updating both representations:

$$P(\lambda, l^{(k)}|D_k) \propto P(D_k|l^{(k)}) P(l^{(k)}|\lambda) P(\lambda) \tag{7}$$

$$P(l^{(k)}|D_k) = \int_\lambda P(\lambda, l^{(k)}|D_k)\, d\lambda \tag{8}$$

$$P(\lambda|D) = \int_l P(\lambda, l|D)\, dl, \tag{9}$$

where $l = l^{(1)} \times l^{(2)} \times ... \times l^{(N)}$ represents the Cartesian product of the lexicons of all individual partners, and $D = \cup_k D_k$ represents all data observed so far across partners.

The prior term $P(l^{(k)}|\lambda) P(\lambda)$ establishes that interactions with novel partners will be first based on priorly formed beliefs about the aspects of lexicon that were commonly useful when interacting with previous partners. First, before any partner is encountered, entries in the community-level lexicon are sampled from independent Gaussians. The entries of the partner-level lexicons which are sampled each time a new partner is encountered will then be centered at the corresponding value of the community-level lexicon.

$$\lambda_{ij} \sim \mathcal{N}(0, 1) \tag{10}$$

$$l_{ij} \sim \mathcal{N}(\lambda_{ij}, 1) \tag{11}$$

The likelihood term $P(D_k|l^{(k)})$ enables an agent to gradually shift its lexicon towards one that is adapted specifically for its current partner. This likelihood is obtained by aggregating the likelihoods of individual word-meaning pairs as defined in formulas (1) and (3) for the listener and speaker respectively, with the speaker attempting to learn from the behaviour of the listener, and vice versa. A decay term is also added to allow agents to overcome early misunderstandings.

$$P(D_k \mid l^{(k)}) = \prod_{\tau=0}^{T-1} P(\{u,m,m',c_k\}_\tau \mid l^{(k)})^{\beta^{(T-\tau)}} \quad (12)$$

$$P_L(\{u,m,m',c_k\}_\tau \mid l^{(k)}) = S_1(u \mid m,c_k,l) \quad (13)$$

$$P_S(\{u,m,m',c_k\}_\tau \mid l^{(k)}) = L_0(m' \mid u,c_k,l), \quad (14)$$

where P_L, P_S represent the likelihood term for the listener and speaker respectively, $\{u,m,m',c_k\}_\tau$ is the τth data point from D_k ($|D_k| = T$), composed of the word used by the speaker u, the meaning chosen by the listener m (used to compute the speaker's posterior), the meaning intended by the speaker m' (used to compute the listener's posterior), and the context c_k. $\beta \in [0,1]$ controls the rate of decay, which we set to 0.9.

Simulations

For all our simulations, we instantiate a word space of size $|U| = 3$, and a meaning space consisting of two distinct features. We use four possible values for the first feature and three for the second feature, for a total of twelve distinct meanings (see Fig. 2). Each context has two of these meanings: a target and a distractor. However, we want to have a way to indicate that a word does not yet have any associated meaning, for which we add a separate *empty* meaning to our space (i.e., $|M| = 13$). We assign this meaning a small similarity to all other meanings in the space, so as to allow agents to eventually associate a meaning to a "meaningless" word. We opt for this solution rather than simply assigning meaningless words a flat distribution as it gives us better control over agents' preference of expressing a novel meaning with an existing word or with a new one. Here, by setting a small similarity to other meanings, we consider a preference for using existing words, with novel ones being adopted by agents only if existing ones express highly dissimilar meanings. Separate from the meaning space, we also add a *null* meaning to all contexts, which can be interpreted as a reference failure on the part of the listener. This effectively offers listeners the option of not choosing any of the meanings in a context, should all of them be unlikely to be described by the speaker's chosen word. Since words have a continuous representation, this null meaning also prevents the listener from distinguishing between two meanings using noisy rather than meaningful differences (e.g., in a context with meanings m_1 and m_2, where word u_1 describes m_1 with probability $l_{11} = 0.02$ and m_2 with probability $l_{12} = 0.002$).

Given the high number of total parameters and our use of continuous representations, exact inference is not tractable, so we resort to variational inference techniques to derive predictions. We use a mean field approximation as implemented in PyMC, which assumes that the random variables of the posterior distribution can be split into independent partitions. We obtain the parameters of each approximated posterior by maximizing the evidence lower bound (ELBO) objective over 10,000 iterations, and we use 10,000 samples to calculate the agents' marginal predictions.

Setup and evaluation methods

We ran simulations where we set up one speaker in a sequence of reference games with four listeners, each of these four having their own set of four contexts. Before the reference game starts, all agents play 30 learning rounds where they observe each of the following word-meaning associations 10 times: $(u_1,m_1),(u_2,m_9),(u_3,m_{empty})$ and update their hypothesis as in (12), but with $P(\{u,m,m'\}_\tau \mid l^{(k)})$ set to $S_0(u \mid m,l)$, as data used for "pre-training" is not generated by a rational agent and is not placed in any context. Throughout the game, the speaker interacts with each listener for 40 rounds, each of the listener's contexts being sampled 10 times. We generate these context sets using the following procedure: first, we fix the contexts that make up the set of the first partner as shown in Fig. 2, then a separate set is created for each of the other partners by mutating the original one. We define two types of mutation steps: an addition step, which adds a new random context to the current set, and a removal step, which removes a randomly selected context. As we want to generate mutations that have the same size as the original set, we alternate between a removal step and an addition step. The number of steps to obtain a mutated set depends on the type of audience that the speaker is facing, with fewer steps associated with audiences that have more homogeneous communicative needs, and more steps associated with audiences that have more heterogeneous communicative needs. We show results for a predominantly homogeneous audience, where 2 mutation steps are done to obtain each set of contexts, and a predominantly heterogeneous audience, where 8 steps are done instead.

We are interested in how category systems emerging from our model will vary in two key variables: efficiency in transmitting information (as a measure of communicative function) and convexity (as a measure of cognitive simplicity).

We measure the communicative efficiency of a system by assessing how well aligned it is with a theoretical optimally efficient system – one that assigns meanings into categories solely based on the distinctions that need to be made for communicative success, using a minimal number of categories and without regard to the similarity space. Finding the optimal system for a set of contexts is equivalent to solving a vertex coloring problem using a minimal number of colors, where each meaning is represented by a vertex in the graph and an edge is added between each of the meanings that need to be distinguished. To calculate the degree of alignment between an observed system and the optimal system, we use the adjusted Rand index for measuring similarity between two systems with hard categories (i.e., with clearly defined boundaries) and we also use our own variation of the index, which is adapted specifically for soft categories. Essentially, the original Rand index (Rand, 1971) calculates the proportion of agreements over the total pairs by the two systems. The adjusted version (Hubert & Arabie, 1985) corrects for the chance grouping of elements which depends on the number and size of categories. To adapt this measure for soft categories, instead of making categorical judgements about

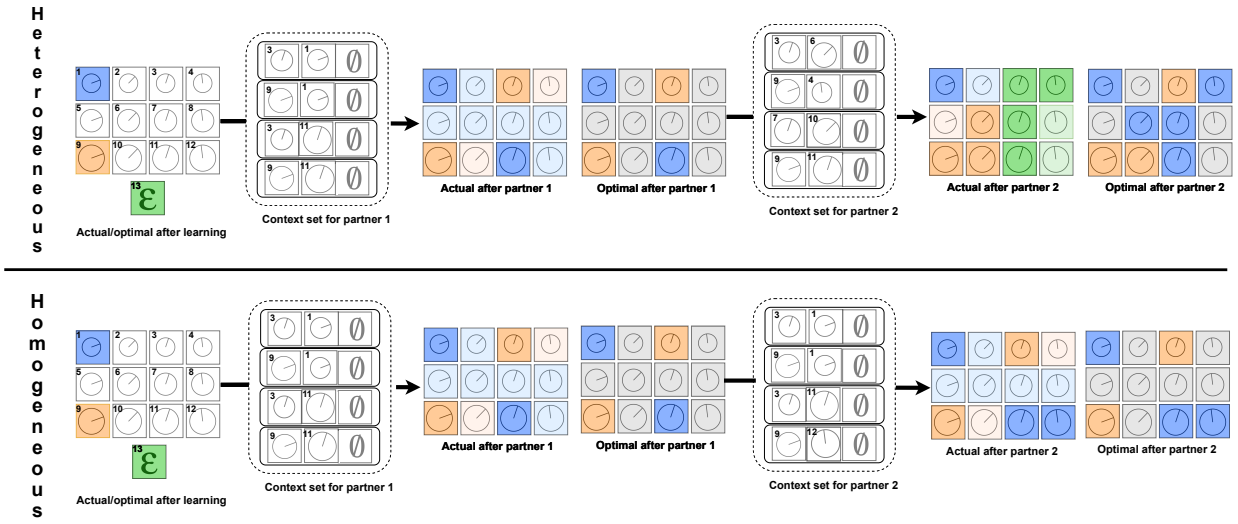

Figure 2: Sample simulation showing how the optimal and actual lexicons of the speaker change after communication with each of the first two partners in the heterogeneous and homogeneous conditions. For each round, one of the four contexts in the set associated with a specific partner is randomly sampled. A lexicon is represented here as a large rectangle with its twelve meanings being arranged in accordance with the structure of the semantic space. We represent meanings using the so-called "Shepard circles" (Shepard, 1964) where the two continuous features are the radius of the circle (X axis) and the angle at which the line is oriented (Y axis). The three labels are represented with blue, orange, and green, and the faded colours in the actual lexicons indicate the labels with the highest probability for currently unseen meanings.

whether a pair of meanings (a,b) is or is not in the same category in system σ, we calculate a score for each option:

$$In^{\sigma}_{ab} = \sum_{i=1}^{|U|} l_{ia}\, l_{ib} \tag{15}$$

$$Out^{\sigma}_{ab} = \sum_{i=1}^{|U|} \sum_{j=1}^{|U|} l_{ia}\, l_{jb} - In^{\sigma}_{ab}, \tag{16}$$

Essentially, In^{σ}_{ab} is calculated by multiplying the suitability of the same word i to describe each of the two meanings, then summing across all words in the lexicon. Out^{σ}_{ab} is calculated similarly, but instead considers all combinations of distinct words for describing the two meanings.

Our convexity measure is based on the intuition that the categories of more convex systems should form tightly clustered regions in the meaning space. Thus, a category u is more convex the more likely it is to contain similar meanings:

$$Convexity^{\sigma}_{u} = \frac{\sum_{i=1}^{|M|} \sum_{j=1}^{|M|} s_{ij}\, l_{ui}\, l_{uj}}{\sum_{i=1}^{|M|} \sum_{j=1}^{|M|} l_{ui}\, l_{uj}} \tag{17}$$

The denominator in formula (17) is used for normalization, as some categories will contain more meanings than others. Finally, we calculate and average this score over all categories $u \in U$ to obtain the convexity of system σ.

Results: communication accuracy

We found highly similar patterns when comparing how communicative success develops in the homogeneous and heterogeneous conditions (Fig. 3A). When the speaker starts interacting with the first listener, the community-wide and partner-specific lexicon expectations will be uninformative, so the only meanings that the speaker will label with confidence will be those observed during the learning phase. As the two agents start observing their partner's behaviour, they will gradually form conventions for referring to the meanings that require distinguishing, with communicative accuracy increasing as a result. When the speaker switches partners, its beliefs about the community-wide lexicon will be updated to incorporate information extracted from the previous interaction, so the speaker's lexicon for the second partner will be weakly biased towards that of the first. The speaker will thus be able to express meanings with more confidence and consistency, allowing it to more easily form conventions with its new partner. Thus, the peak success rate obtained with the previous partner will be more easily reached and eventually outperformed. This trend continues over future interactions, as the speaker gets increasingly better at communicating the types of distinctions that were most useful across partners.

Results: transmission optimality and convexity

For both conditions, we show alignment scores between the theoretical optimally efficient community-wide lexicon and

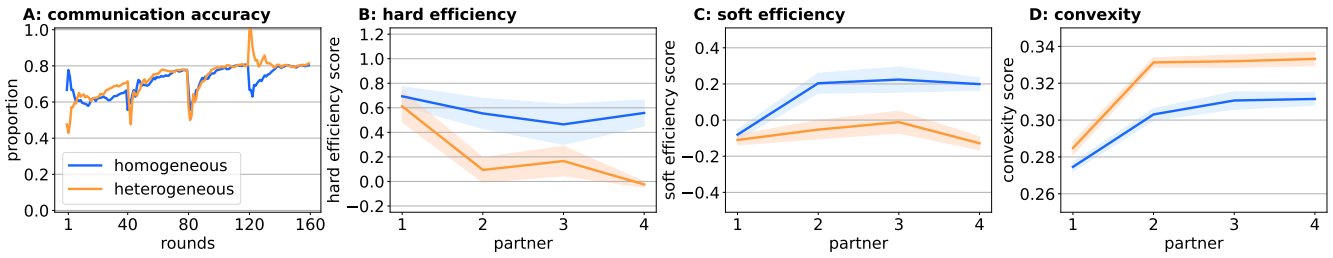

Figure 3: Simulation results showing how the rate of communicative success (A), degree of hard efficiency (B), degree of soft efficiency (C), and level of convexity (D) develop over the course of a speaker's interaction with four different partners with either homogeneous (blue line) or heterogeneous (orange line) communicative needs, averaged over 15 separate runs. The shaded regions show the standard errors for each of the two conditions.

actual lexicon (λ) of the speaker after each partner, using the standard adjusted Rand index (Fig. 3B, henceforth we refer to it as *hard efficiency*), and our version adapted for soft categories (Fig. 3C; henceforth we refer to it as *soft efficiency*). To obtain a hard categorization from our model, we simply assign each meaning to the label with the highest probability.

Since the context set that we assign for the first listener requires a relatively straight-forward solution, we expect category systems resulting from this interaction in both conditions to generally group meanings in an optimal manner. We observe in Fig. 3B that our expectations are indeed correct. However, this symmetricity disappears in the systems emerging after subsequent partners: with a homogeneous audience hard efficiency stays fairly constant across partners, while with a heterogeneous one it drastically falls during interaction with the second partner and remains significantly lower for the rest of the simulation. Fig. 3C tells a similar story from a different perspective. While initial lexicons are likely to group meanings in a communicatively optimal way, confidence in this groupings in low in the absence of evidence from multiple partners, so soft efficiency scores start at a low value. We see scores significantly increase over a homogeneous audience, but stay relatively invariant over a heterogeneous one. Note that an optimal system as measured by the soft efficiency score would assign the entire probability mass of its labels to the meanings that need to be communicated, so systems are expected to score much lower on this measure.

To understand what is causing this pattern of results, recall that after interacting with just one partner, meanings in the speaker's lexicon are unstable and prone to relabelling. This is especially true for meanings without neighbours in the semantic space that share the same label, as a meaning is more likely to be expressed by a label that also expresses similar meanings (see formula (3)). But if the same distinctions need to be made with multiple partners, confidence in the previously established label-meaning associations is repeatedly reinforced. Thus, with a homogeneous audience, the speaker's actual lexicon gradually becomes more aligned with the optimal one, as the optimal lexicon only slightly changes across partners (for an example see Fig.2, homogeneous).

However, a speaker encountering partners with highly varying contexts will instead have to continuously adapt its lexicon to support communication of an increasing number of distinctions. This will make it harder for the speaker to maintain categories with idiosyncratic structure, such as those where neighbouring meanings fall into distinct categories, because on the one hand, meaning-label association are not being reinforced, while on the other hand each meaning is being continuously 'pulled' by neighbors from other directions towards their label. To solve this tension, the structured meaning space will favor the formation of tight clusters (see Fig.2, heterogeneous). These clusters will gradually grow in size and act as strong attractors for meanings with weak label associations, eventually causing the lexicon to stabilize. This effect can be also seen in Fig. 3D, which compares convexity rates across partners between the lexicons that emerge in the two conditions: we see that in the heterogeneous condition the systems that eventually become stable have a significantly higher convexity than in the homogeneous condition.

Conclusion

What exactly controls the trade-off between communicative and cognitive pressures on the semantic categories labelled by single words? We suggest that one factor might be the extent to which communicative contexts at the level of local interactions are similar or different across users of a category system. We propose a hierarchical model for communication in a multidimensional meaning space to test this hypothesis. Our model predicts that if a speaker faces a more homogeneous audience with many meaning distinctions that are common across partners, those distinctions will be conventionalized in the lexicon and reinforced over time. Thus, the emerging lexicon will primarily reflect the communicative needs of its users. Conversely, with a more heterogeneous audience where communicatively relevant meaning distinctions are rapidly changing, our model predicts that labels will evolve to form tight clusters in the similarity space. Such a lexicon is shaped to a greater extent by the pressure to reduce cognitive cost, as it better reflects the structure of the semantic space rather than the needs of individual users.

Acknowledgements

This research received funding from a University of Edinburgh College of Arts, Humanities and Social Sciences Research Award (held by V. Nedelcu).

References

Carr, J. W., Smith, K., Cornish, H., & Kirby, S. (2017). The cultural evolution of structured languages in an open-ended, continuous world. *Cognitive science*, *41*(4), 892–923.

Carstensen, A., Xu, J., Smith, C., & Regier, T. (2015). Language evolution in the lab tends toward informative communication. In *Cogsci.*

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39.

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in cognitive science*, *5*(1), 173–184.

Hawkins, R. D., Franke, M., Frank, M. C., Goldberg, A. E., Smith, K., Griffiths, T. L., & Goodman, N. D. (2022). From partners to populations: A hierarchical bayesian account of coordination and convention. *Psychological Review*.

Hawkins, R. X., Franke, M., Smith, K., & Goodman, N. D. (2018). Emerging abstractions: Lexical conventions are shaped by communicative context. In *Cogsci.*

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of classification*, *2*, 193–218.

Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, *336*(6084), 1049–1054.

Kemp, C., Xu, Y., & Regier, T. (2018). Semantic typology and efficient communication. *Annual Review of Linguistics*, *4*, 109–128.

Koptjevskaja-Tamm, M., Vanhove, M., & Koch, P. (2007). Typological approaches to lexical semantics.

Levinson, S. C. (2012). Kinship and human thought. *science*, *336*(6084), 988–989.

Majid, A., Boster, J. S., & Bowerman, M. (2008). The cross-linguistic categorization of everyday events: A study of cutting and breaking. *Cognition*, *109*(2), 235–250.

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language*, *40*(2), 230–262.

Nedelcu, V. C., & Smith, K. (2022). The complexity of a language is shaped by the communicative needs of its users and by the hierarchical nature of their social inferences. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 44).

Nosofsky, R. M. (2011). The generalized context model: An exemplar model of classification. *Formal approaches in categorization*, 18–39.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, *66*(336), 846–850.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of mathematical psychology*, *1*(1), 54–87.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323.

Silvey, C., Kirby, S., & Smith, K. (2019). Communication increases category structure and alignment only when combined with cultural transmission. *Journal of Memory and Language*, *109*, 104051.

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

Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, *115*(31), 7937–7942.

Zaslavsky, N., Kemp, C., Tishby, N., & Regier, T. (2020). Communicative need in colour naming. *Cognitive neuropsychology*, *37*(5-6), 312–324.

Zaslavsky, N., Maldonado, M., & Culbertson, J. (2021). Let's talk (efficiently) about us: Person systems achieve near-optimal compression. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 43).