SELF-ORGANIZATION AND EMERGENCE IN LANGUAGE
A CASE STUDY FOR COLOR

JOACHIM DE BEULE AND JORIS BLEYS

Artificial Intelligence Lab, Vrije Universiteit Brussel, Belgium
{*joachim,jorisb*}*@arti.vub.ac.be*

A computational language game model is presented that shows how a population of language users can evolve from a brightness-based to a brightness+hue-based color term system. The shift is triggered by a change in the communication challenges posed by the environment, comparable to what happened in English during the Middle English period in response to the rise of dyeing and textile manufacturing *c.* 1150–1500. In a previous model that is able to explain such a shift, these two color categorization strategies were explicitly represented. This is not needed in our model. Instead, whether a population evolves a brightness- or a hue-based system is an emergent phenomenon that depends only on environmental factors. In this way, the model provides an explanation of how such a shift may come about without introducing additional mechanisms that would require further explanation.

1. Introduction

There are thousands of languages in the world, differing not only in the words and grammatical elements of form they employ, but also in the conceptual distinctions they make explicit. Moreover, these languages are not static: in all of them a co-evolution of meaning and form continues to take place (see e.g. Fernandez-Ordonez (1999)). Still, there are (near-)universals of language and language evolution that require an explanation (Heine & Kuteva, 2002).

For many years, language universals were mainly ascribed to genetic predispositions (Chomsky, 1980; Pinker & Bloom, 1990; Hauser, Chomsky, & Fitch, 2002). Today, most are thought to be the result of environmental and cultural rather than genetic factors (Steels, 2000; Croft, 2007; De Beule, 2008). Nevertheless, there are still only very few demonstrations, in the form of a computational model for instance, that explain the appearance of complex regularities in the evolution of natural languages without ultimately relying on built in mechanisms. Iterated learning for example explains the emergence of compositionality by relying on a strongly developed capacity and 'desire' of language learners to generalize from examples, and on the assumption that language transmission is limited by a learning bottleneck (Kirby & Hurford, 2002; Smith, Kirby, & Brighton, 2003).

This paper presents a model that demonstrates the kind of conceptual refocusing that occurred in English between the Old English period *c.* 600–1150 and the

Middle English period *c.* 1150–1500, when the English color term system moved from being predominantly based on brightness to predominantly based on a combination of brightness and hue(Casson, 1997). Similar shifts also occured in other languages (MacLaury, 1992).

A previous model (Bleys & Steels, 2009) explained such shifts by introducing an additional level of competition and selection between two explicitly represented and predefined language strategies (Steels, 2009): one for a brightness- and one for a brightness+hue-based color term system. The current model shows that this extra level is not required. Rather, these language strategies emerge for free. The current model thus reproduces a non-trivial phenomenon without relying on additional mechanisms that in turn would require further explanation.

In the next section, the model is described in detail. It is essentially a prototype-based flexible extension to multiple dimensions of the model presented in (Puglisi, Baronchelli, & Loreto, 2008). It also bears similarities with the model described in (Wellens, Loetzsch, & Steels, 2008). The results section shows that the model can exhibit a shift from a brightness- to a hue-based color system when triggered by environmental and communicative challenges, similar to those experienced by speakers of Old English as a result of the development of dyeing and textile manufacturing. The discussion and conclusion section ends the paper.

2. The Model

The model is a language game model in which participants (or agents), each equipped with simple communication abilities, engage in repeated communicative interactions that eventually lead them to align their linguistic inventories. At each time step, two agents are randomly chosen from a population to play a game. They are presented with two distinguishable color samples, one of which is identified to the speaker agent as the topic of the game. The speaker agent then has one chance to describe the topic color, based on his current inventory of color terms and following a procedure described below; if the hearer agent correctly identifies the intended topic color, the game succeeds; otherwise the correct topic is revealed and the game fails.

Agents perceive the colors in the game as points in their private perceptual color space. The linguistic inventory of an agent simply consists of stored exemplar colors in its perceptual space, each linked to one or more color terms. All exemplars linked to the same term are said to form a *linguistic category* (Puglisi et al., 2008). If an exemplar is part of several linguistic categories (in other words, if it is linked to several color terms) then its main linguistic category is the one corresponding to the most recently added color term. To describe the topic color (i.e., find or create a color term to associate with the topic color), a speaker agent first determines which stored exemplar colors most closely match each of the new perceived colors. Standard Euclidian distance in color space is used for this. It then faces one of three conditions:

(a) If both perceived colors match the *same color exemplar* (and hence map to the same color term), the speaker agent stores the new color that is further away from the shared color exemplar as a new exemplar. This new exemplar is then associated with all the color terms of the original exemplar, as well as with a newly created term. Furthermore, the new term is also associated with all previously stored exemplar colors in the main linguistic category of the original exemplar that are closer to the new exemplar than to the other perceived color. As such a new linguistic category is formed for the new term.

(b) If the perceived colors map to *different color exemplars but with identical main linguistic categories*, then a new term is associated with the exemplar closest to the topic color only if the agent's current self-estimated success rate is below 70%.[a] This ensures (relative) stability while retaining flexibility in case of a drop in success.

c If the two perceived colors map to *different color exemplars with different main linguistic categories*, then both colors may still be stored as new exemplars for their respective color terms if they are far away from other color exemplars (as determined by a constant parameter representing the smallest distinction in perceptual color space that agents are still able to perceive).

In each of these cases, the speaker then utters the color term now associated with the main linguistic category of the topic color.

The hearer agent's task is to choose the perceived color that is the best match for the term uttered by the speaker. It does so by first looking up which of its own color exemplars are closest to the two perceived colors and then checking whether either is associated with the uttered term. It then chooses the perceived color corresponding to whichever exemplar is linked to the speaker's term (and making a random choice if both or none of the exemplars are). The game succeeds if the hearer correctly identifies the topic color. The hearer agent may also add exemplars to its private perceptual color space using a procedure similar to that used by the speaker agent (steps (a)–(c) above), but taking into account that the topic color belongs to the linguistic category of the term used by the speaker).

Finally, if the game was a success, then the exemplar colors closest to the topic color in both agents are marked as belonging *only* to the linguistic category associated with the color term used in the game. This extra alignment mechanism helps to ensure that *shared* linguistic categories emerge as conventional collections of color exemplars all linked to the same color term in a meaningful and useful way. It extends the approach introduced in (Puglisi et al., 2008) to the case of multiple dimensions and applies it to regions in perceptual space determined by points rather than by explicit boundaries.

[a] Agents keep track of their personal success in games as a running average.

We conclude our description of the model with two remarks. First, note that although the model makes no explicit reference to the notion of *globally shared* linguistic categories, they still emerge as a result of agents' local attempts to improve their language game performance. This forms a first example of how the model explains a feature of language as an emergent phenomenon (Puglisi et al., 2008).

Secondly, the model also makes no reference to specific linguistic strategies. It works for any number of dimensions and has no prior preferences among them. Still, as will be shown in the next section, it is capable of supporting a shift from an initially stable brightness-based language system to a new, hue-based language system. Each of these systems is the result of the same "universal" strategy responding to changing conditions, where this universal strategy is simply that of self-organization.

3. Results

For convenience of presenting results, simulations were performed in two dimensions, called the *brightness* and *hue* dimensions, each ranging from 0 to 1. The minimal distance between colors in perceptual space still noticeable to agents was set to 0.05. We ran repeated simulations of the model consisting of two stages.

In Stage 1, the topic and other colors in each game were always set to be well apart in the brightness dimension (at least one-fifth of its range, or 0.20). No bias was employed in the hue dimension. Thus, this stage corresponds to Old English usage, in which people only communicated color distinctions when the brightness channel was salient. In the following section, it will be shown that this case predominantly leads to brightness-based color term systems.

In Stage 2, the bias is removed from the brightness dimension, and samples are distributed randomly across both brightness and hue dimensions. This corresponds to an increased need to communicate color distinctions even when differences in brightness are not salient. Such a change reflects the increased communication challenge that is hypothesized to have occurred due to the rise of dyeing and textile manufacturing. Consequently, a shift occurs from a now insufficient brightness-based system to a better adapted brightness+hue-based system.

3.1. *Stage 1*

Figure 1 on the left shows the the average success rate over time across all agents, along with the average number of color terms in the population. After first going through a phase of exploration in which new terms are proposed, the agents manage to reach a reasonable success rate ($\simeq 85\%$) with on average a rather small number of color terms ($\simeq 3.5$–recall that unused color terms may disappear during a successful game).

The left of Figure 2 shows the perceptual space of a typical agent after Stage 1, together with the exemplar colors amassed by the agent. Exemplar colors in the

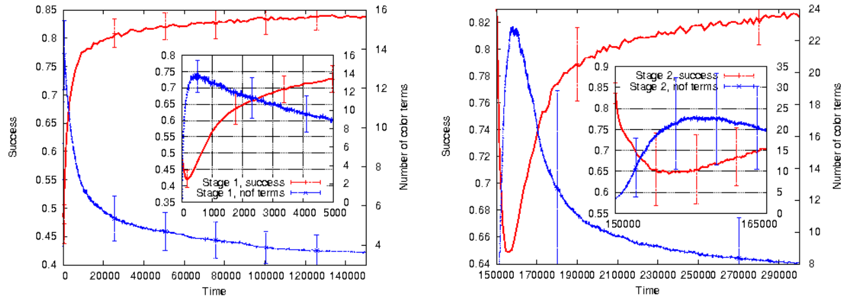

Figure 1. Average success rate and number of color terms in Stage 1 (left) and Stage 2 (right).

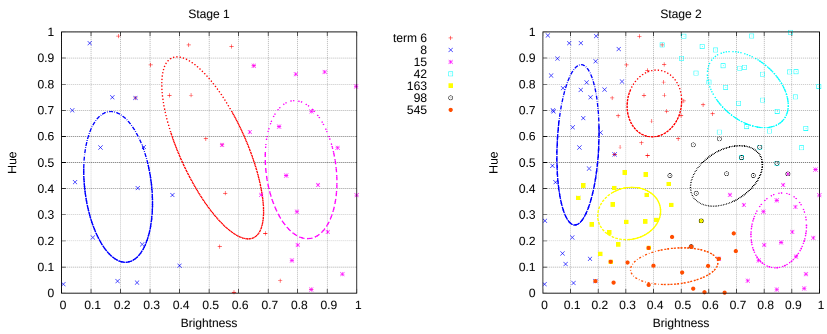

Figure 2. Example of color exemplars and (approximated) linguistic categories in the perceptual color space of an agent after Stage 1 (left) and Stage 2 (right), showing a shift from a brightness-based to a brightness+hue-based color system.

same linguistic category are shown using the same color. This agent's system can be considered brightness-based: its color categories are insensitive to hue (i.e., they all span the entire range of hue values) while they effectively partition the brightness dimension.

To provide a more precise quantitative basis for this measure, we performed a principal component analysis of the exemplar colors in each category. A linear approximation of the category was then obtained as an ellipse with major and minor axes along the principle components and with major and minor radii proportional to the standard deviations of the exemplars along them. These ellipses are also shown in Figure 2. A color category can now be defined to be brightness-based if its approximating ellipse is oriented along the brightness dimension and has a large eccentricity (in other words, if its shape is stretched out along the brightness dimension).

To obtain an "average category" that can be used to compare these values across the different stages (i.e., for different kinds of environmental bias arising from the distribution of color samples), we can now also average over all color categories from different agents and different runs, or rather over their elliptic directions and eccentricities. Figure 3 shows three average categories each based on 100 simulations involving five agents. In each of the three runs, a different environmental bias was employed. The large ellipse corresponds to the end of Stage 1

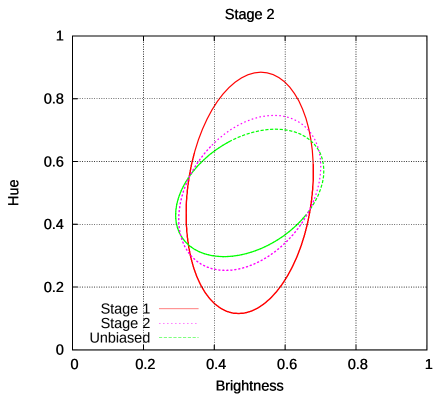

Figure 3. Average (linear approximations of) linguistic categories resulting from different environmental biases. The large red ellipse corresponds to a brightness-based system induced when the presented pairs of colors are required to differ by at least 0.2 in brightness (Stage 1). The middle ellipse shows how the average category shifts toward being based on both brightness and hue in response to a removal of the bias (Stage 2.) Finally, the inner ellipse corresponds to skipping Stage 1 (i.e., to having no environmental bias from the beginning) and is included for reference.

and is clearly brightness-based: it extends over the entire hue dimension, and thus can distinguish only between color percepts based on brightness. Its major axis lies along the direction $(\text{Brightness} = 0.96, \text{Hue} = 0.10)$ and its eccentricity is 0.89. It can be compared with the smallest ellipse, which represents the average category arising in a completely unbiased environment, effectively skipping Stage 1. Its major direction is $(0.49, 0.45)$, meaning that it does not distinguish color percepts based specifically on either brightness or hue.

From these results, we can conclude that after Stage 1, agents have evolved an efficient color term system, on average containing around three to four mainly brightness-based color terms.

3.2. *Stage 2*

Figure 1 on the right shows how the success rate and the number of color terms change after the switch to an unbiased environment. As new challenges are encountered, the success rate starts to fall, and new terms are tried out. Then alignment dynamics take over until a successful color system is established again, but with on average around twice as many terms ($\simeq 8$).

Figure 2 shows how the color system from Stage 1 (left) has changed in Stage 2 (right). The intermediate ellipse in Figure 3 shows the average linguistic category after Stage 2. Clearly, a shift has occurred from a brightness-based color term system to a more neutral system based on the full color space: the linguistic categories from Stage 1 have become sensitive also to hue (i.e., they no longer span the entire range of hue values) and new, additional categories have emerged to fill the resulting gaps.

4. Discussion and conclusion

We have presented a computational language game model that is capable of showing effects similar to those observed in natural languages.

Many of the assumptions made in the model are vast oversimplifications of reality, and in particular may not properly model the details of perceptual color space.[b] Despite these simplifications and the relatively small number of assumptions it makes, it can still explain how color systems based on brightness or hue can emerge in a biased environment, without any built-in biases toward either kind of system or strategy. Instead, environmental changes in the kinds of color distinctions necessary to succeed at the language game are sufficient to push the model toward more useful strategies. These results are qualitatively the same as what happened in English during the Old English to Middle English color system shift.

The model does assume (models of) language users that are willing to play a game, which captures aspects of daily life communications between human interlocutors. It therefore takes seriously the fact that language serves a function and is best studied within the context of its use, in particular as it is subjected to the alignment dynamics induced by local, communicative usage interactions. Such interactions may be sufficient to account for the emergence of useful language strategies (such as the brightness- and hue-based color systems demonstrated here) as well as other universals of language, even in the absence of strong genetic biases.

Acknowledgments

The authors thank Nancy Chang for her incredibly detailed and invaluable comments that made it possible to considerably improve the paper. Joachim De Beule is funded by the ComplexDis European sixth framework project (FP6-2005-NEST-PATH). Joris Bleys is funded through a fellowship of the Institute of

[b] For example, D'Andrade and Egan (1974) show that brightness is cognitively more salient than hue, a result that could interfere with results presented in this paper. Note, however that our model could easily be adapted to investigate the impact of such differences in salience.

the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen).

References

Bleys, J., & Steels, L. (2009). Linguistic selection of language strategies, a case study for color. In *Proceedings of the 10th european conference on artificial life (lncs).* Springer-Verlag.

Casson, R. (1997). Color shift: evolution of english color terms from brightness to hue. In C. Hardin & L. Maffi (Eds.), *Color categories in thought and language* (p. 224239). Cambridge: University Press, Cambridge.

Chomsky, N. (1980). *Rules and representations.* Oxford, UK: Blackwell.

Croft, W. (2007). Language structure in its human context: new directions for the language sciences in the twenty-first century. In P. Hogan (Ed.), *Cambridge encyclopedia of the language sciences.* Cambridge: Cambridge University Press.

D'Andrade, R., & Egan, M. (1974). The color of emotion. *American Ethologist*, *1*, 49-63.

De Beule, J. (2008). The emergence of compositionality, hierarchy and recursion in peer-to-peer interactions. In A. D. M. Smith, K. Smith, & R. F. i Cancho (Eds.), *Proceedings of the 7th international conference on the evolution of language* (p. 75-82). World Scientific.

Fernandez-Ordonez, I. (1999). Le'ısmo, laísmo, loísmo: estado de la cuestion. In I. Bosque & V. Demonte (Eds.), *Gramatica descriptiva de la lengua espanola* (Vol. 1, p. 13191390). Espasa Calpe, Madrid: Real Academia Éspanola.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*, 1569-1579.

Heine, B., & Kuteva, T. (2002). *World lexicon of grammaticalization.* Cambridge University Press, Cambridge, MA.

Kirby, S., & Hurford, J. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (p. 121-148). London: Springer Verlag.

MacLaury, R. E. (1992). From brightness to hue: An explanatory model of color-category evolution. *Current Anthropology*, *33*(2), 137-186.

Pinker, S., & Bloom, P. (1990). Natural languages and natural selection. *Behavioral and Brain Sciences*, *13*, 707-784.

Puglisi, A., Baronchelli, A., & Loreto, V. (2008). Cultural route to the emergence of linguistic categories. *PNAS*, *105*(23), 7936-7940.

Smith, K., Kirby, S., & Brighton, H. (2003). Iterated learning: a framework for the emergence of language. *Artificial Life*, *9*(4), 3710386.

Steels, L. (2000). Language as a complex adaptive system. In M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. J. Merelo, & H.-P. Schwefel (Eds.), *Proceedings of parallel problem solving from nature vi.* Berlin, Germany: Springer.

Steels, L. (2009). Can evolutionary linguistics become a science? *Journal of Evolutionary Linguistics*, *1*(1).

Wellens, P., Loetzsch, M., & Steels, L. (2008). Flexible word meaning in embodied agents. *Connection Science*, *20*(2), 173-19.