Colourful language and colour categories

Tony Belpaeme[1,2] Joris Bleys[1]

[1] Artificial Intelligence Lab
Vrije Universiteit Brussel
Pleinlaan 2, B-1050 Brussels, Belgium
{tony,jorisb}@arti.vub.ac.be

[2] University of Plymouth
School of Computing, Communication and Electronics
Portland Square, Plymouth, PL4 8AA, United Kingdom

Abstract

We investigate whether the universal character of colour categories can be explained as the result of a category acquisition process under influence of linguistic communication. A brief overview is presented of the different positions in explaining the mechanisms of colour category acquisition (or perceptual categories in general). We introduce a computational model to study the acquisition of colour categories with and without linguistic interactions. We present preliminary results, which are compared with recent results from the World Color Survey. We argue that combining biases from colour perception, perceptual categorisation and linguistic communication provides an alternative explanation for the nature of colour categories.

1 Introduction

For more than three centuries the precise nature of human colour categories has been one the most disputed topics among physicists, psychologists, cognitive scientists and anthropologists. Newton, already in the eighteenth century, wondering about the number of categories that could be discerned in the sunlight's spectrum, decided on the divine number of seven, thereby requiring a category called "indigo" that no-one had observed until then. Three centuries later much more precise data on colour categories is available and together with the data came a plethora of interpretations.

One of the most influential contributions is the monograph by Berlin and Kay (1969) in which they reported on the linguistic colour categories of 20 languages. Using naming experiments they elicited the colour categories of subjects and comparing the categories across different languages they noticed a remarkable cross-cultural correspondence. Until then the general consensus had been that colour categories were random for each culture, but Berlin and Kay's work rekindled the conviction that the universal character of colour categories could only be explained as being genetically determined.

In this paper we first summarise the results of the World Color Survey (WCS) (Kay et al., 1997, 2003) reported in (Kay and Regier, 2003). This work provides the strongest evidence yet of strong universal tendencies in colour naming in seperate languages. We give an overview of the different accounts which try to explain this universal character and then continue to present a computational model which tests whether linguistic relativism might be a viable candidate. We report several results from the simulation and compare these with the data from the WCS.

2 The World Color Survey

The WCS reports on colour naming experiments with speakers of 110 languages spoken in non-industrialised societies. The field data has been gathered in North and South America, Africa and South-East Asia.

In the study each subject is shown a series of 330 coloured chips drawn from the Munsell colour set[1]

[1]The Munsell Color Company (GretagMacbeth, New Windsor, NY) produces calibrated chips for art reproduction. The most saturated chips have been used by anthropologists to study colour categories since the 1950s.

(of which 320 chips show gradations of hues at different lightness, all at maximal saturation, and 10 chips show shades of grey, ranging from white to black) and asked to name each chip.

The analysis of the data proceeded as follows. For all subjects studied, the centroid was computed for every colour term they used. For this the Munsell colour values were converted to the CIE $L^*a^*b^*$ colour appearance model[2]. The term centroids were projected back onto the closest matching Munsell chip. For each language a chart can now be produced showing the average representation of all colour terms in that language.

To get a visual impression of the linguistic colour categories over all 110 languages, the centroids of all subjects of all languages can be combined into one single histogram (figure 1). The floor plane of the histogram corresponds to the ordered Munsell chart, with on one axis the hue value of the chip, ranging from red, over yellow, green, blue, to purple; and on the other axis the lightness of the chip (note that it does not display the counts for achromatic chips).

The histogram shows that the linguistic colour categories of different languages are not arbitrary; it clearly illustrates the universal character of colour categorisation. Peaks can be found at regions close to the English colour terms pink/red, brown, yellow, green, blue and purple.

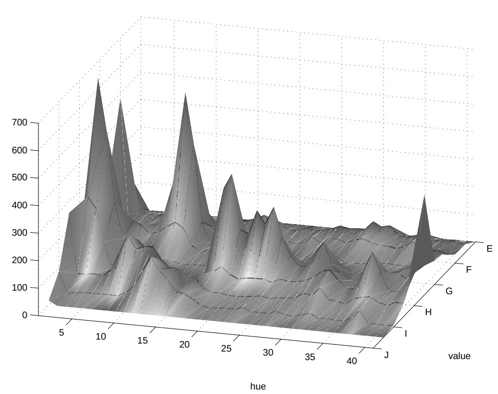

Figure 1: Histogram showing the linguistic colour categories for 110 languages spoken in non-industrialised societies (data from Kay and Regier, 2003).

[2]CIE $L^*a^*b^*$ is a perceptually uniform colour representation. It is a 3D colour space, in which the L^* dimension represents the lightness, and the a^* and b^* dimensions represent the chroma of the colour. A Euclidean distance function can be used to compute the perceptual distance between two CIE $L^*a^*b^*$ values.

3 Attempts at explaining universalism

The challenge now is accounting for colour naming universalism. The leading position has always been that colour categorisation results directly or indirectly from an innate endowment (Kay and McDaniel, 1978; Bornstein, 1985; Hardin, 1988; Shepard, 1992; Kaiser and Boynton, 1996). One hypothesis states that there exist basic colour categories that are explicitly related to the opponent colour processing in the human visual pathways. Psychological and neurophysiological data indeed points to an opponent character of human colour perception, with white contrasting with black, red with green and yellow with blue. All other basic categories —orange, brown, pink, purple and grey— can be deduced from these six primaries. Although this account has made it into textbooks (e.g. Crystal, 1997), some scholars still doubt that colour categories are unequivocally fixed by neural correlates (Saunders and van Brakel, 1997; Lucy, 1997; Jameson and D'Andrade, 1997) or that colour categories are universal at all (Roberson et al., 2000).

In the next section we will present a computational model to study if colour categories can be explained as a concept formation process which is under influence of language (or cultural exchange in general). It has been proposed by some that colour categories not only are associated with colour terms, but that colour terms also have an influence on the acquisition of colour categories (Gellatly, 1995; Davies and Corbett, 1997). This position has become known as the Sapir-Whorf hypothesis (Whorf, 1956).

4 The computational model

The computational model we use is based on a research methodology first proposed by Steels (1996a,b). Using this methodology Steels studied how meanings can be associated unambigiously with words. It was later extended for studying adaptive meanings and open lexica in (Steels, 1998; Belpaeme, 2001). The methodology relies on multi-agent simulations. Each agent is able to perceive, categorise its perceptions and lexicalise the resulting categories. We briefly present the internals of an agent:

Perception The perception of colours is modelled by relying on the properties of the CIE $L^*a^*b^*$ colour space (Fairchild, 1998). Agents are offered colour stimuli as RGB triplets, these are converted to CIE $L^*a^*b^*$ values. The conversion from RGB to CIE $L^*a^*b^*$ is given in the following equations. The conversion matrix is for PAL/SECAM viewing conditions, with $\gamma = 2.5$; the XYZ coordinates of the reference white are taken to be $[X_n Y_n X_n]^T = [0.950\ 1.000\ 1.089]^T$.

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.431 & 0.222 & 0.0202 \\ 0.342 & 0.707 & 0.130 \\ 0.178 & 0.0713 & 0.939 \end{pmatrix} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix}^{\gamma}$$

$$\begin{aligned}
L^* &= \begin{cases} 116\left(\frac{Y}{Y_n}\right) - 16 & \frac{Y}{Y_n} > \epsilon \\ 903.3\left(\frac{Y}{Y_n}\right) & \frac{Y}{Y_n} \leq \epsilon \end{cases} \\
a^* &= 500\left(\mathrm{f}\left(\frac{X}{X_n}\right) - \mathrm{f}\left(\frac{Y}{Y_n}\right)\right) \\
b^* &= 200\left(\mathrm{f}\left(\frac{X}{X_n}\right) - \mathrm{f}\left(\frac{Z}{Z_n}\right)\right) \\
\mathrm{f}(x) &= \begin{cases} x^{1/3} & x > \epsilon \\ 7.787x + 16/116 & x \leq \epsilon \end{cases} \\
\epsilon &= 0.008856
\end{aligned}$$

The CIE $L^*a^*b^*$ colour representation was designed to mimick human psychological colour experience, and therefore serves well as our colour perception model.

Categorisation To implement perceptual categorisation we resort to a point representation in the CIE $L^*a^*b^*$ space. Each colour category is a point in that space, and the membership function for a category is the Euclidean distance to that point.

Lexicalisation Colour categories can be associated with colour terms. The strength of the association is represented by a scalar value $s \in [0, 1]$. Colour categories can be associated with more than one word (thereby allowing synonymy) and words can be associated with more than one category (thereby allowing homonymy).

Additionally an interaction between two agents is defined, which serves to let the agents acquire a repertoire of colour categories and colour terms. The interaction implements horizontal transmission of lexical entries and categories. It consists of two components, a *discrimination game* and a *guessing game*, both described below.

4.1 The discrimination game

The discrimination game serves to build a repertoire of categories that allows an agent A to distinguish between stimuli. This pseudo code for the discrimination game is as follows.

Algorithm 1 Discrimination Game(A, $\mathcal{O}$)

1: Agent A chooses a topic o_t from the context $\mathcal{O} = \{o_1, \ldots, o_N\}$ containing N objects.
2: Agent A perceives each stimulus in the context by constructing an internal representation for it: $\{o_1, \ldots, o_{\mathcal{O}_N}\} \rightarrow \{r_1, \ldots, r_N\}$
3: For each internal representation r_i, the best matching category is found. This is the category which has the highest output for r_i of all the categories available in the category repertoire of the agent $A_{\mathcal{CR}}$ and which we will denote by c_i^{best}: $\{r_1, \ldots, r_N\} \rightarrow \{c_1^{best}, \ldots, c_N^{best}\}$
4: If the best matching category for the topic c_t^{best} is unique in $\{c_1^{best}, \ldots, c_{\mathcal{O}_N}^{best}\}$ the game succeeded, otherwise it has failed.

An agent is offered a number of objects, this is called the *context* $\mathcal{O}$. One of the object is the *topic* o_t, which the agent has to distinguish from the other objects in the context. For this, the agent first perceives all objects, which results in a number of *internal representation* r_i. Next, the internal representation are matched to categories. For example, if the agent has only one category, all representation of objects will be matched to that same category, making it impossible for the agent to distinguish betweem objects. However, as soon as the agent has more than one category, it can start distinguishing between objects. If the topic is matched with a category with which no other object matches, we say that the agent is able to "discriminate the topic from the context" and we call the discrimination game a success.

The discrimination game can fail in several ways: this is an opportunity to improve the agent's categorical repertoire. When the category repertoire $A_{\mathcal{CR}}$ is empty, a new category is created on the internal representation of the topic r_t. When no discriminating category could be found, there are two possible actions: (1) a new category is created on r_t or (2) the best matching category c_t^{best} is adapted to better represent the internal representation of the topic r_t, this is done by shifting c_t^{best} towards r_t. Option (1) is taken when the discriminative success of the agent is below a threshold $\theta_{adapt} = 0.95$, otherwise option (2) is taken.

4.2 The guessing game

The guessing game is played between two agents randomly chosen from the population: one acting as *speaker* (A_S) and the other as *hearer* (A_H). The pseudo code for the guessing game is as follows [3].

The speaker and hearer both observe the same context $\mathcal{O}$. The speaker knows what the topic o_t of the conversation is, and tries to linguistically communicate the topic to the hearer. For this the speaker first plays a discrimination game, if this succeeds the speaker looks up the word associated with the discriminating category. This word is then relayed to the hearer. The hearer looks up the category belonging to the word, and maps the category onto the objects in the context. It then points to the object which matches best with the category. Finally, the speaker reports if the hearer has pointed correctly to the topic. During the course of the guessing game, both agents adapt the strength s_{ij} between category c_i and word w_j according to the following equation (with $\delta = 0.1$).

$$\begin{cases} s_{ij} = \min\left(s_{ij} + \delta, 1\right) \\ s_{kl} = \max\left(s_{kl} - \delta, 0\right) \\ \quad \text{in row i and column j with } k \neq i, l \neq j \end{cases} \tag{1}$$

A categories is adapted by shifting the point representation of a category towards a representation r, as in eq. 2; α is a learning rate, set to 0.7.

$$c \leftarrow c + \alpha(r - c) \tag{2}$$

Of course, also the guessing game can fail at several ways. For each failure, an appropriate action is taken so that the agents will be more successful at communicating in future games.

- The speaker fails at the discrimination game: it adapts its categorical repertoire as described in 4.1.
- The speaker has no word associated with c_t^{best}: a new word is created and associated with an initial strength $s = 0.5$.
- The hearer does not know the word w: the speaker "points" at the topic and the hearer associates the word w with the category best matching the topic, with initial strength $s = 0.5$.
- The hearer fails to pick out the topic ($o_t \neq o_h$): the strength of the association between c_t^{best} and w is decreased by δ.

[3] DG stands for discrimination game.

When the guessing game is successful the speaker and hearer both increase the strength of the association between the categories used and the communicated word[4].

5 Experimental results

As input to the agents we use two different sets of colour data. One set, called the *random* set, contains random colours generated by drawing colours from the RGB colour solid and then converting them to CIE $L^*a^*b^*$. The other set, called the *nature* set, draws colours from digital photographs of natural scenes. The difference between both is that the *random* set contains a uniform distribution of colours, while the *nature* set contains a skewed distribution with an abundance of low-saturated colours and few high-saturated colours. The purpose of having two data sets is to study the effect of the environment on the acquisition of colour categories.

For reference the results from the WCS (Kay and Regier, 2003) are repeated in figure 2 now a contour plot of figure 1. The locations of English colour terms are added for reference.

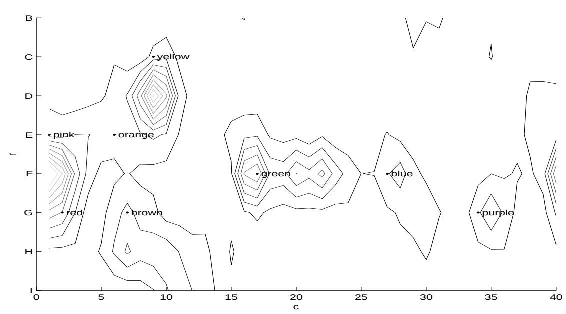

Figure 2: Contour plot of the WCS data.

Four types of simulations have been run. DGRAN: discrimination game where agents are fed *random* data. DGNAT: discrimination game where agents are fed *nature* data. GGRAN: guessing game where agents are fed *random* data. And GGNAT: guessing game where agents are fed *nature* data.

Each type of simulation has a population of 10 agents and has been run 105 times[5]. The results presented for each type of simulation are the sum of these 105 runs.

[4] More details, specifically on the implementation of the update rules, can be found in (Bleys, 2004; Steels and Belpaeme, 2005)

[5] One could think of these 105 runs as hundred different artificial societies.

Algorithm 2 Guessing Game(A_S, A_H, O)

speaker A_S		hearer A_H
chooses topic o_t		
plays DG for o_t		
DG succeeds and returns c^S		
finds term w for c^S		
utters w	$\rightarrow w \rightarrow$	hears w
		finds category c^H for w
		finds o_h closest to c^H
sees o_h	$\leftarrow o_h \leftarrow$	points to o_h
if hearer guessed right, then $o_t = o_h$		
update s^S_{cw} using eq. 1		
points to o_t	$\rightarrow o_t \rightarrow$	sees o_t
		updates s^H_{cw} using eq. 1
		adapts category c^H to r_t using eq. 2

Figures 3, 4, 5 and 6 show contour plots of histograms collecting the colour categories of 10×105 agents. A first observation is that each type of simulation cuts up the colour continuum in a number of peaks: colour categories are not randomly constructed (if they would be, the histogram should not have any peaks).

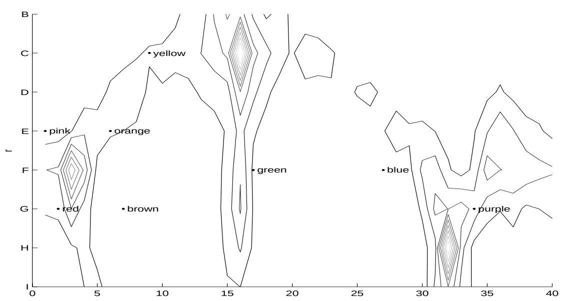

Figure 3: Contour plot of DGRAN results.

Two biases, present in all four simulations, are quite influential. On the one hand, the psychological colour space —modelled by the CIE $L^*a^*b^*$ colour space— puts constraints in the location of the categories (the colour space is shaped like two bumpy cones connected to each other at their base). The second bias is formed by the property of categories to be maximally distinctive. Both biases act together so that colour categories are in a way "pushed" towards locations where they are maximally distinctive and where they form a stable configuration. Colour categories are stable when they are located in places where shifting the colour category would result in a lower discriminative or communicative success.

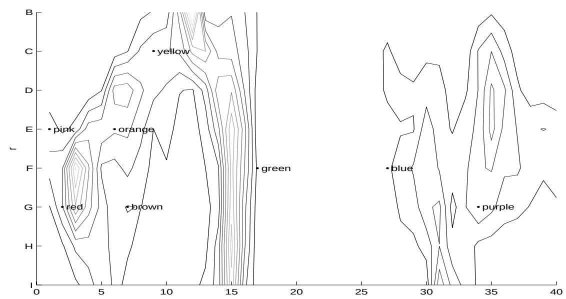

Figure 4: Contour plot of DGNAT results.

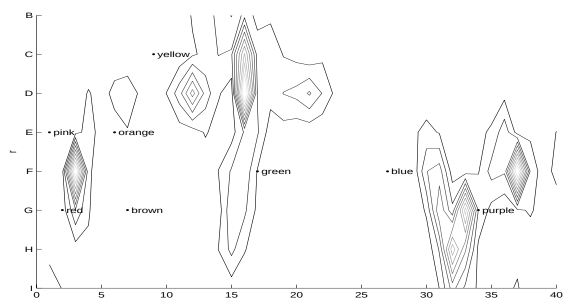

Figure 5: Contour plot of GGRAN results.

In this sense, all four simulations return colour categories that retain all properties of human perceptual categories. However, the purpose of our study is to see whether acquiring colour categories with an additional bias formed by linguistic communication would result in categories that are more human-like. Figures 5 and 6 when compared to figure 2 give a

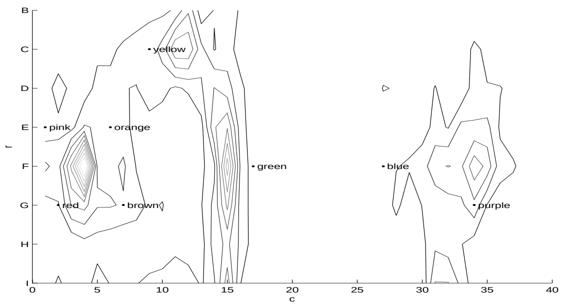

Figure 6: Contour plot of GGNAT results.

qualitative impression, but a measure is needed compare the histograms quantitavily. Eq. 3 computes the sum of squared pair-wise differences between two histograms h and h' (with h_i being the bin at index i in histogram h).

$$d\left(h, h'\right) = \sum_i \left(h_i - h'_i\right)^2 \tag{3}$$

Table 1 shows the comparison between the histograms obtained from the simulations and the WCS data. According to the measure we use, the DGNAT simulation resembles human colour categories most. However, also the DGRAN and GGNAT data have a similar distance to the human data. Only the GGRAN data seems to be off, why remains eludes us at the moment.

$d(h, WCS)$	
DGRAN	0.00973
DGNAT	0.00842
GGRAN	0.0145
GGNAT	0.00994
WCS	0

Table 1: Sum of squared differences between simulation histograms and WCS data (lower values correspond to a higher similarity).

6 Discussion

The computational models that are presented here implement a view on colour categorisation which contrasts with the innatist viewpoint on colour categories. We have shown how agents can acquire a set of categories that is sufficient to discriminate colours, and in the case of the guessing game simulations, the agent acquire colour categories that not only discriminate well, but also communicate well.

The categories resulting from the simulations are qualitatively similar to human colour categories: they take up regions in the colour space that correspond well to the WCS data. We have not been able to show that the influence of communication on category formation results in radically different categories. This might however be due to the limitations of our analysis. The sum of squared distances measure might not be suited to compare two-dimensional histograms. For example, if two identical histograms are compared, but one is shifted relative to the other, the sum of squared distances measure will return a low value; this is not desired.

Future analysis will point out if there exist measures which might give a better impression of the similarity of histograms. One alternatively could be to extract the peaks of the histograms and compare the using a certain distance measure[6].

Acknowledgements

The authors wish to thank Terry Regier and Paul Kay for the helpful discussion and for making the World Color Survey data available. Tony Belpaeme is a postdoctoral fellow of the Flemish Fund for Scientific Research (FWO Vlaanderen). The research reported was in part sponsored by the FWO Verbal Communication on Humanoid Robots project.

References

Tony Belpaeme. Simulating the formation of color categories. In Bernhard Nebel, editor, *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'01)*, pages 393–398, Seattle, WA, 2001. Morgan Kaufmann, San Francisco, CA.

Tony Belpaeme and Joris Bleys. Explaining universal colour categories through a constrained acquisition process. *Adaptive Behavior*, 2005. Submitted.

Brent Berlin and Paul Kay. *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley, CA, 1969.

Joris Bleys. The cultural propagation of colour categories: Insights from computational modelling. Master's thesis, Artificial Intelligence Lab, Vrije Universiteit Brussel, 2004.

[6]This is further explored in (Belpaeme and Bleys, 2005)

Marc H. Bornstein. On the development of color naming in young children. *Brain and Language*, 26:72–93, 1985.

David Crystal. *The Cambridge Encyclopedia of Language*. Cambridge University Press, Cambridge, MA, 2nd edition, 1997.

Ian R.L. Davies and G.G. Corbett. A cross-cultural study of colour grouping: Evidence for weak linguistic relativity. *British Journal of Psychology*, 88:493–517, 1997.

Mark D. Fairchild. *Color Appearance Models*. Addison-Wesley, Reading, MA, 1998.

Angus Gellatly. Colourful Whorfian ideas: Linguistic and cultural influences on the perception and cognition of colour, and on the investigation of them. *Mind and Language*, 10(3):199–225, 1995.

Clyde L. Hardin. *Color for Philosophers: Unweaving the Rainbow*. Hacket, Indianapolis, 1988.

Kimberley Jameson and Roy D'Andrade. It's not really red, green, yellow, blue: an inquiry into perceptual color space. In C.L. Hardin and L. Maffi, editors, *Color categories in thought and language*, pages 295–319. Cambridge University Press, Cambridge, 1997.

Peter M. Kaiser and Robert M. Boynton. *Human Color Vision*. Optical Society of America, Washington DC, 1996.

Paul Kay, Brent Berlin, Luisa Maffi, and William Merrifield. Color naming across languages. In C.L. Hardin and L. Maffi, editors, *Color Categories in Thought and Language*. Cambridge University Press, Cambridge, 1997.

Paul Kay, Brent Berlin, Luisa Maffi, and William R. Merrifield. *The World Color Survey*. Center for the Study of Language and Information, Stanford, 2003. ISBN 1575864150.

Paul Kay and Chad McDaniel. The linguistic significance of the meanings of basic color terms. *Language*, 54(3):610–646, 1978.

Paul Kay and Terry Regier. Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences*, 100(15):9085–9089, 2003.

John A. Lucy. The linguistics of "color". In Clyde L. Hardin and Luisa Maffi, editors, *Color Categories in Thought and Language*, pages 320–346. Cambridge University Press, Cambridge, 1997.

Debi Roberson, Ian Davies, and Jules Davidoff. Color categories are not universal: replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, 129(3):369–398, 2000.

Barbara A.C. Saunders and Jaap van Brakel. Are there nontrivial constraints on colour categorization? *Behavioral and Brain Sciences*, 20(2):167–228, 1997.

Roger N. Shepard. The perceptual organization of colors: An adaptation to regularities of the terrestrial world? In J. Barkow, L. Cosmides, and J. Tooby, editors, *Adapted Mind*, pages 495–532. Oxford University Press, Oxford, 1992.

Luc Steels. Perceptually grounded meaning creation. In M. Tokoro, editor, *Proceedings of the International Conference on Multiagent Systems (ICMAS-96)*, pages 338–344, Menlo Park, CA, 1996a. AAAI Press.

Luc Steels. Self-organizing vocabularies. In C. Langton and T. Shimohara, editors, *Proceedings of the Conference on Artificial Life V (Alife V) (Nara, Japan)*, Cambridge, MA, 1996b. The MIT Press.

Luc Steels. The origins of ontologies and communication conventions in multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 1:169–194, 1998.

Luc Steels and Tony Belpaeme. Coordinating perceptually grounded categories through language. A case study for colour. *Behavioral and Brain Sciences*, 2005. Accepted as target article. Available from http://arti.vub.ac.be/˜tony.

Benjamin Lee Whorf. *Language, Thought and Reality: selected writings of Benjamin Lee Whorf*. The MIT Press, Cambridge, MA, 1956. Edited by Carrol, J.B.