CONSENSUS IN LANGUAGE DYNAMICS:
NAMING, CATEGORIZING AND BLENDING.

Vittorio Loreto[1,2] and Francesca Tria[2]
[1]Sapienza University of Rome and [2]ISI Foundation

ABSTRACT
Understanding the origins and evolution of language and meaning is currently one of the most promising areas of research in cognitive science. As challenging issue that touches on all aspects of cognition, it stimulates creative thinking and casts many fundamental issues in a new light. Recently, new theoretical and computational tools as well as synthetic modelling approaches have reached sufficient maturity to contribute significantly to the ongoing debate in cognitive science. In addition, unprecedented advances in information and communications technologies are enabling, for the first time, the possibility of precisely mapping the interactions, whether embodied and/or symbolic, of large numbers of actors, as well as the dynamics and transmission of information along social ties. The combination of these two elements is opening terrific new avenues for studying the emergence and evolution of languages, new communication and semiotic systems. As was the case with biology, new tools and methods can trigger a significant boost in the ongoing transition of linguistics into an experimental discipline, where multiple evolutionary paths, timescales and dependences on the initial conditions can be effectively controlled and modelled.

In this paper we shall review some of the progress made in the last few years and highlight potential future directions of research in this area. In particular, the emergence of a common lexicon, a shared set of linguistic categories and the emergence of duality of patterning will be discussed, as examples corresponding to the early stages of a language. We shall emphasize the cultural route to the emergence and evolution of language,

i.e., the idea that a community of language users can be seen as a complex dynamical system, which collectively solves the problem of developing a shared communication framework through the back-and-forth signalling between individuals. We shall illustrate a few examples where the predictions of synthetic modelling have been successfully compared with real data. Finally we shall discuss how new technologies and computational facilities are making available a huge amount of resources both as novel tools and data to be analyzed, allowing quantitative and large-scale analysis of the processes underlying the emergence of a collective information and language dynamics.

1. INTRODUCTION AND THEORETICAL BACKGROUND

Language dynamics is a rapidly growing field that focuses on all processes related to the emergence, evolution, change and extinction of languages. Recently, the study of self-organization and evolution of language and meaning has led to the idea that a community of language users can be seen as a complex dynamical system (Steels 2000), which collectively solves the problem of developing a shared communication framework through the back-and-forth signalling between individuals. From this perspective, language is thus seen as an evolving and self-organizing system, whose components are thus constantly being (re)shaped by language users in order to maximise communicative success and expressive power, while at the same time minimising effort. In this picture new words and grammatical constructions may be invented or acquired, new meanings may arise, the relation between language and meaning may shift (e.g., if a word adopts a new meaning), the relation between meanings and the world may shift (e.g. if new perceptually grounded categories are introduced). All these changes happen at the level of the individual as well as at the group level, the focus being on the interactions among the individuals, with both vertical (teacher-pupil) as well as horizontal (peer to peer) communications. Here communications acts are particular cases of language games, which, as already pointed out by Wittgenstein (Wittgenstein 1953), can be used to describe linguistic behaviour, even though they can include also non linguistic behaviour, such as pointing. Clark (Clark 1996) argues that language and communication are social activities - joint activities - that require people to coordinate with each other as they speak and listen. Language use is more than the sum of a speaker speaking and a listener listening. It is the joint action that emerges when speakers and listeners (Garrod & Pickering 2004), writers and readers perform their individual actions in coordination, as ensembles. Again language is not seen as an individual process, but rather as a social process where a continuous alignment of mental representations (Garrod & Pickering 2009) is taking place.

The landscape describing the large set of approaches to the study of language emergence and dynamics is extremely diversified, due to the flagrant complexity of a problem that can be addressed under many respects, with different methodologies, guided by often incompatible conceptual frameworks, and with different goals in mind. A useful way to gain insights into such a variegated world is therefore that of focusing on few dimensions that allow for a coarse categorization of the ongoing research (Haeger et al. 2009). In general it is possible to identify broad paradigms that frame the problem in a particular way, focusing on specific problems and addressing precise fundamental questions through concrete models and experiments (Nolfi & Mirolli 2009). Within each framework, then, the investigation can proceed through computational models, experiments with embodied agents, psychological experiments with human subjects and finally exploiting data made available by large information systems like the Web.

In this paper we shall mainly focus on the mathematical modelling of social phenomena. Statistical physics has proven to be a very fruitful framework to describe phenomena outside the realm of traditional physics (Loreto & Steels, 2007). The last years have witnessed the attempt by physicists to study collective phenomena emerging from the interactions of individuals as elementary units in social structures (Castellano, Fortunato & Loreto 2009). This is the paradigm of the complex systems: an assembly of many interacting (and simple) units whose collective (i.e., large scale) behaviour is not trivially deducible from the knowledge of the rules that govern their mutual interactions. This scenario is also true for problems related to the emergence of language.

From this new perspective, complex systems science turns out to be a natural ally in the quest for general mechanisms to understand the collective dynamics whereby conventions can spread in a population. In particular we shall focus on how conceptual and linguistic coherence may arise through self-organization or evolution, and how concept formation and expression may interact to co-ordinate semiotic systems of individuals. One of the key methodological aspects of the modelling activity in the domains of complex

systems is the tendency to seek simplified models to clearly pin down the assumptions and, in many cases, to make the models tractable from a mathematical point of view.

A crucial step in the modelling activity is represented by the comparison with empirical data. This comparison could help in checking whether the trends seen in real data are already compatible with plausible microscopic modelling of the individuals, or require additional ingredients. From this point of view the Web may be of great help, both as a platform to perform controlled online social experiments, and as a repository of empirical data on large-scale phenomena. In this way a virtuous cycle involving data collection, data analysis, modelling and predictions can be triggered, giving rise to an ever more rigorous and focused approach to language dynamics.

It is worth stressing how the contribution that physicists, mathematicians and computer scientists could give should not be considered as alternative to more traditional approaches. We rather think that it would be crucial to foster the interactions across the different disciplines cooperating with linguistics, by promoting scientific activities with concrete mutual exchanges among all the interested scientists. This would help both in identifying the problems and sharpening the focus, as well as in devising the most suitable theoretical concepts and tools to approach the research.

While the research field of semiotics may traditionally be considered a conceptual discipline, the cognitive turn has recently brought central semiotic questions and insights into the laboratories and a new discipline, dubbed experimental semiotics (Galantucci & Garrod 2010), is about to be born. A few important examples have already shown the viability of this approach: from coordination game with interconnected computers (Galantucci 2005, Selten & Warglien 2007) to experimental tests for Iterated Learning Models (Kirby, Cornish & Smith 2008).

But this is only the tip of a potentially huge iceberg. The new Information and

Communication Technologies (ICT) are opening terrific opportunities to monitor human actions both symbolic and embodied. This reverberates in the possibility to turn these new technologies in a living lab devoted to language dynamics and more generally to social sciences. In this framework the Web is playing a special role potentially very relevant for studies in language dynamics. Though only a few years old, the growth of the World Wide Web and its effect on the society have been astonishing, spreading from the research in high-energy physics into other scientific disciplines, academe in general, commerce, entertainment, politics and almost anywhere communication serves a purpose. Innovation has widened the possibilities for communication. Social media like blogs, wikis and social bookmarking tools allow the immediacy of conversation, with unprecedented levels of communication speed and community size. Millions of users now participate in managing their personal collection of online resources by enriching them with semantically meaningful information in the form of freely chosen tags and by coordinating the categories they imply. Wikipedia, Yahoo Answers and the ESP Game (Von Ahn & Dabbish 2004) are systems where users volunteer their human computation because they value helping others, participating in a community, or playing a game. These new types of communities are showing a very vital new form of semiotic dynamics. From a scientific point of view, these developments are very exciting because they can be tracked in real time and the tools of complex systems science and cognitive science can be used to study them. The Web is thus revealing yet another skin as a potential platform to run experiments where the cognitive and linguistic skills of humans have to be challenged and tested.

The outline of the paper is as follows. Sections 1 and 2 will be devoted to describe how a synthetic population of individuals can bootstrap increasing complex linguistics structures. We shall consider in particular how names and early syntactic structures (Sect. 1) and linguistic categories (Sect. 2) emerge. Section 3 will be devoted to challenging the theoretical predictions against real-world data. Section 4 will briefly describe the new opportunities offered by the new Information and Communication

Technologies to setup an experimental framework devoted to language dynamics. Finally the last section is devoted to drawing some conclusions.

1. NAMING

Let us start by considering the problem of Naming objects. This is probably the simplest, though non-trivial, problem when addressing language emergence. The problem we shall be considering can be posed in the following terms. How a population of individuals, without any pre-coordinated and shared linguistic structure, can bootstrap a shared set of names associated to a set of objects? Is this consensus always reached? Under which conditions? It is important to remark as these questions represent one specific instance of a more general question addressed when investigating social dynamics. How do the interactions between social agents create order out of an initial disordered situation? Order is a translation in the language of physics of what is denoted in social sciences as consensus, agreement, uniformity, while disorder stands for fragmentation or disagreement. It is reasonable to assume that without interactions, heterogeneity dominates: left alone, each agent would choose a personal response to a political question, a unique set of cultural features, his own special correspondence between objects and words. Still it is common experience that shared opinions, cultures, and languages do exist. The Naming Game was expressively conceived to explore the role of self-organization in the evolution of language (Steels 1995 & 1996) and it has acquired, since then, a paradigmatic role in the entire field of Semiotic Dynamics. The pioneering work (Steels 1995) mainly focused on the formation of vocabularies, i.e., a set of mappings between words and meanings (for instance physical objects). In this context, each agent develops its own vocabulary in a random and private fashion. Nevertheless, agents are forced to align their vocabularies, through successive conversation, in order to obtain the benefit of cooperating through communication. Thus, a globally shared vocabulary emerges, or should emerge, as a result of local adjustments of individual word-meaning associations. The communication evolves through successive conversations, i.e., events that involve a certain number of agents (two, in practical

implementations) and meanings. It is worth remarking that conversations are here particular cases of language games, which, as already pointed out by Wittgenstein (Wittgenstein 1953), are used to describe linguistic behaviour but, if needed, can also include non-linguistic behaviour, such as pointing.

This original seminal idea triggered a series of contributions along the same lines and many variants have been proposed along the years. It is worthwhile to mention here the work proposed in (Ke 2002), who focuses on an imitation model which simulates how a common vocabulary is formed by agents imitating each other either using a mere random strategy or a strategy in which imitation follows the majority (which implies non-local information for the agents). A further contribution of the mentioned paper is the introduction of an interaction model that uses a probabilistic representation of the vocabulary. The probabilistic scheme is formally similar to the framework of evolutionary game theory (Nowak et al. 1999, Nowak et al. 1999b), since a production matrix and a comprehension matrix is associated to each agent. Unlike the approach of Evolutionary Language Games, the matrices are here dynamically transformed according to the social learning process and the cultural transmission rule. A similar approach has been proposed in (Lenaerts 2005).

Here we discuss in details a minimal version of the Naming Game that results in a drastic simplification of the model definition, while keeping the same overall phenomenology. This version of the Naming Game is suitable for massive numerical simulations and analytical approaches. Moreover its extreme simplicity allows for a direct comparison with other models introduced in other frameworks of statistical physics as well as in other disciplines.

The Minimal Naming Game

The simplest version of the Naming Game (Baronchelli 2006) is played by a population of N agents trying to bootstrap a common vocabulary for a certain number M of objects present in their environment. The objects can be people, physical objects, relations, web sites, pictures, music files, or any other kind of entity for which a population aims at reaching a consensus as far as their

naming is concerned. Each player is characterized by an inventory of word-object associations he/she knows. All the inventories are initially empty (t = 0). At each time step (t = 1, 2, ..) two players are randomly picked and one of them plays as speaker and the other as hearer.
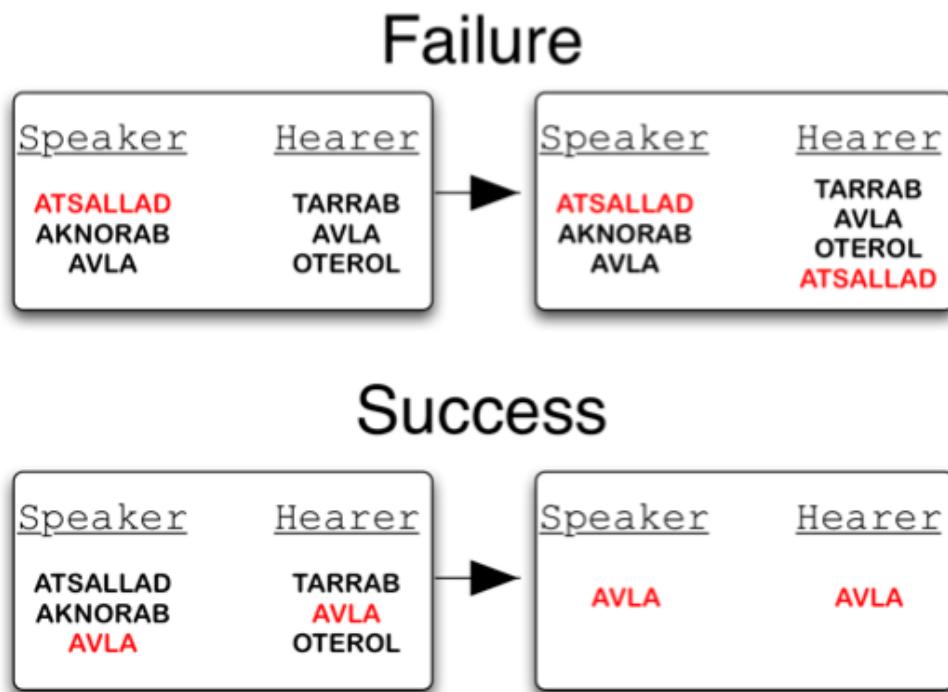

Figure 1. Naming Game. Examples of the dynamics of the inventories in a failed (top) and successful (bottom) game. The speaker selects the word highlighted. If the hearer does not possess that word he includes it in his inventory (top). Otherwise both agents erase their inventories only keeping the winning word (bottom).

Their interaction obeys the following rules (see Figure 1):

- The speaker selects an object from the current context;
- The speaker retrieves a word from its inventory associated with the chosen object, or, if its inventory is empty, invents a new word;
- The speaker transmits the selected word to the hearer;
- If the hearer has the word named by the speaker in its inventory and that word is associated to the object chosen by the speaker, the interaction is a success and both players maintain in their inventories only the winning word, deleting all the others;
- If the hearer does not have the word named by the speaker in its inventory, or the word is associated to a different object, the interaction is a failure and the hearer updates its inventory by adding an association between the new word and the object.

The game is played on a fully connected network, i.e., each player can, in principle, play with all the other players, and makes two basic assumptions. One assumes that the number of possible words is so huge that the probability of a word to be re-invented is practically negligible (this means that homonymy is not taken into account here, though the extension is trivially possible). As a consequence, one can reduce, without loss of generality, the environment as consisting of only one single object ($M = 1$).

A third assumption of the Naming Game consists in assuming that the speaker and the hearer are able to establish whether a game was successful by subsequent actions performed in a common environment. For example, the speaker may refer to an object in the environment he wants to obtain and the hearer then hands the right object. If the game is a failure, the speaker may point (non-verbal communication) or get the object himself so that it is clear to the hearer which object was intended.

Macroscopic analysis of the Minimal Naming Game reveals that consensus is

always reached going through three different phases. Very early, pairs of agents play almost uncorrelated (and unsuccessful) games and both total number of words in all the inventories and the number of distinct words increase linearly over time. In the second phase the success probability is still very small and agents' inventories start correlating, the total number of words presenting a well identified peak. The process evolves with an abrupt increase in the number of successes and a further reduction in the numbers of both total and different words. Finally, the dynamics ends when all agents have the same unique word and the system is in the attractive convergence state. It is worth noting that the developed communication system is not only effective (each agent understands all the others), but also efficient (no memory is wasted in the final state). The system undergoes spontaneously a disorder/order transition to an asymptotic state where global coherence emerges, i.e., every agent has the same word for the same object. It is remarkable that this happens starting from completely empty inventories for each agent. The asymptotic state is one where a word invented during the time evolution took over with respect to the other competing words and imposed itself as the leading word. In this sense the system spontaneously selects one of the many possible coherent asymptotic states and the transition can thus be seen as a symmetry breaking transition.

It is important to focus on the scaling behaviour of the relevant quantities, i.e., how they depend on the population size. Let us consider two relevant quantities: the convergence time and the maximal size of all the inventories. The first quantity tells us how long is the convergence process as measured in terms of games per player. It turns out that in a fully connected graph the convergence time scales as:

$$t_{conv} \propto N^{3/2}.$$

The maximal size of all the inventories is a very important quantity to evaluate the typical size of the inventories. In its turn the typical size of an inventory is reflecting the size of the memory required to each individual, i.e., an estimate

of his/her cognitive requirements. In a fully connected graph it turns out that the maximal size of all the inventories scales as:

$$N_{word}^{max} \propto N^{3/2},$$

which implies that, at the maximum, the average size of each individual inventory scales as the square root of the population size. This is not a realistic outcome since this implies that the memory size should diverge with the population size. Let us now consider what happens when one considers different interaction topologies with respect to the fully connected graph.

Role of the interaction topology

Social networks play an important role in determining the dynamics and outcome of language change. The first investigation of the role of topology was proposed in 2004 in a seminal paper presented at the 5th Conference on Language evolution, Leipzig (Ke et al. 2008). Since then, many approaches focused on adapting known models on topologies of increasing complexity: regular lattices, random graphs, scale-free graphs, etc.

The Naming Game, as described above, is not well-defined on general networks. When the degree distribution is heterogeneous, it does matter if the first randomly chosen agent is selected as a speaker and one of its the neighbour as the hearer or viceversa: high-degree nodes are in fact more easily chosen as neighbours than low-degree vertices. Several variants of the Naming Game on generic networks can be defined. In the direct Naming Game (reverse Naming Game) a randomly chosen speaker (hearer) selects (again randomly) a hearer (speaker) among its neighbours. In a neutral strategy one selects an edge and assigns the role of speaker and hearer with equal probability to one of the two nodes (Dall'Asta et al. 2006).

On low-dimensional lattices each agent can rapidly interact two or more times with its neighbours, favouring the establishment of a local consensus with a high success rate, i.e. of small sets of neighbouring agents sharing a common

unique word. Later on these "clusters" of neighbouring agents with a common unique word undergo a coarsening phenomenon (Baronchelli et al. 2006) with a competition among them driven by the fluctuations of the interfaces (Bray 1994). The coarsening picture can be extended to higher dimensions and the scaling of the convergence time has been conjectured as being $t_{conv} \propto N^{1+\frac{1}{d}}$, where d ≤ 4 is the dimensionality of the space. On the other hand, the maximum total number of words in the system (maximal memory capacity) scales linearly with the system size, i.e., each agent uses only a finite capacity. In summary, low-dimensional lattice systems require more time to reach the consensus compared to mean-field, but a lower use of memory.

Let us now considers what happens on Small-World topologies (Watt & Strogatz 1998). The effect of a small-world topology has been investigated in (Dall'Asta et al. 2006b) in the framework of the Naming Game (Baronchelli et al. 2006) and in (Castelló et al. 2006) for the AB-model (see section below). Two different regimes are observed. For times shorter than a cross-over time, $t_{cross} = O(N / p^2)$, one observes the usual coarsening phenomena as long as the clusters are typically one- dimensional, i.e., as long as the typical cluster size is smaller than $1/p$. For times much larger than t_{cross}, the dynamics is dominated by the existence of short-cuts and enters a mean-field like behaviour. The convergence time is thus expected to scale as $N^{3/2}$ and not as N^3 (as in d = 1). Small-world topology allows thus to combine advantages from both finite-dimensional lattices and mean-field networks: on the one hand, only a finite memory per node is needed, in opposition to the $O(\sqrt{N})$ in mean-field; on the other hand the convergence time is expected to be much shorter than in finite dimensions. In (Castelló et al. 2006) it has been studied the dynamics of the AB-model on a two-dimensional small world network. Also in this case a dynamical stage of coarsening is observed followed by a

fast decay to the A or B absorbing states caused by finite size fluctuations.

We finally conclude this panoramic about different topologies briefly mentioning about the behaviour of the Naming Game on complex networks, and referring to (Dall'Asta et al. 2006) for an extensive discussion. It turns out that the convergence time scales with an exponent compatible with 3/2 both Erdös-Renyi (ER) (Erdös & Renyi 1959, 1960) and Bárabasi-Albert (BA) (Bárabasi & Albert 1999) networks. The scaling laws observed for the convergence time are a general robust feature and they are not affected by further topological details, such as the average degree, the clustering or the particular form of the degree distribution.

Is consensus always reached?

A variant of the Naming Game has been introduced with the aim of mimicking the mechanisms leading to opinion and convention formation in a population of individuals (Baronchelli et al. 2007). In particular a new parameter, β ($\beta = 1$ corresponding to the Naming Game), has been added mimicking an irresolute attitude of the agents in making decisions. β is simply the probability that in a successful interaction both the speaker and the hearer update their memories erasing all opinions except the one involved in the interaction (see Figure 1). This negotiation process, as opposed to herding-like or bounded confidence driven processes, displays a non-equilibrium phase transition from an absorbing state in which all agents reach a consensus to an active (not-frozen as in the Axelrod model (Axelrod 1997)) stationary state characterized either by polarization or fragmentation in clusters of agents with different opinions. Figure 2 moreover shows that the transition at β_c is only the first of a series of transitions: when decreasing $\beta < \beta_c$, a system starting from empty initial conditions self-organizes into a fragmented state with an increasing number of opinions. At least two different universality classes exist, one for the case with two possible opinions and one for the case with an unlimited number of opinions. Very interestingly, the model displays the non-equilibrium phase transition also on heterogeneous networks, in contrast with other opinion-

dynamics models, like for instance the Axelrod model (Klemm et al. 2003), for which the transition disappears for heterogeneous networks in the thermodynamic limit.
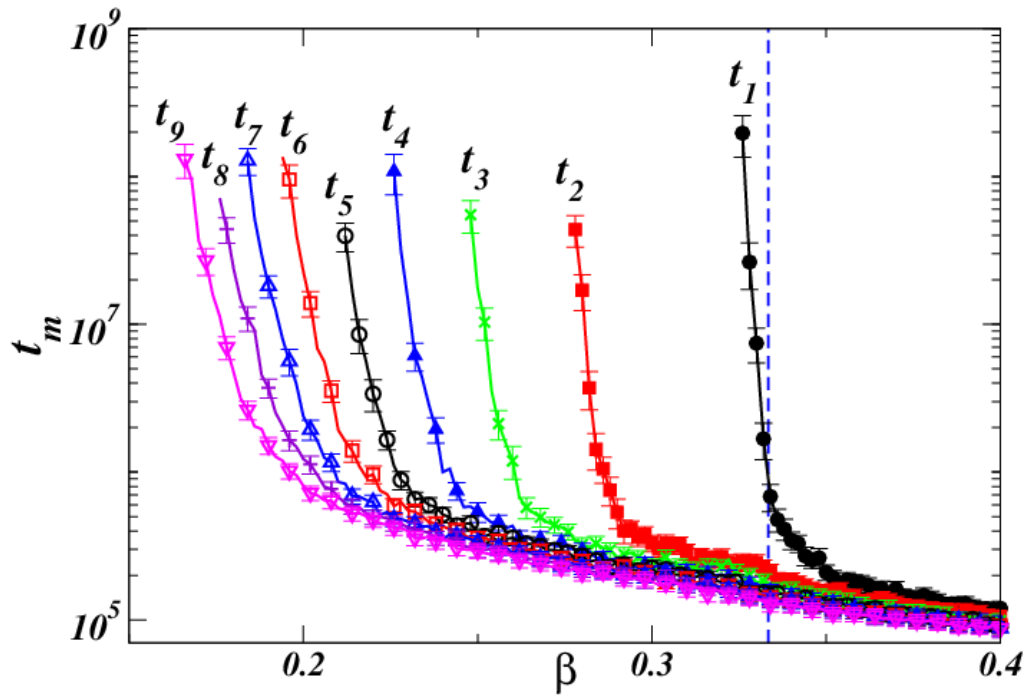

Figure 2. Phase transitions in the Naming Game. Time t_m required to a population on a fully-connected graph to reach a (fragmented) active stationary state with m different opinions. For every m > 2, the time t_m diverges at some critical value $\beta_c(m) < \beta_c$.

Symmetry breaking: a controlled case

We concentrate now on a simpler case in which there are only two words at the beginning of the process, say A and B. This is a very interesting case since it highlights the close connection with other modelling schemes introduced for opinion dynamics and language competition. In this case the population can be divided into three classes: the fraction of agents with only the word A, n_A, the fraction of those with only the word B, n_B, and finally the fraction of agents with both words, n_{AB}. The description of the time evolution of the three species is straightforward:

$$\begin{cases} \dfrac{dn_A}{dt} = -n_A n_B + n_{AB}^2 + n_A n_{AB} \\[2mm] \dfrac{dn_B}{dt} = -n_A n_B + n_{AB}^2 + n_B n_{AB} \\[2mm] \dfrac{dn_{AB}}{dt} = 2n_A n_B - 2n_{AB}^2 - (n_A + n_B)n_{AB} \end{cases}$$

This system of differential equations is deterministic. It presents three fixed points in which the system can collapse depending on initial conditions. If at time zero $n_A(t=0) > n_B(t=0)$ $[n_B(t=0) > n_A(t=0)]$ then at the end of the evolution we will have the stable fixed point $n_A = 1[n_B = 1]$and, obviously, $n_B = n_{AB} = 0$ $[n_A = n_{AB} = 0]$. If, on the other hand, we start from a perfectly symmetrical situation $n_A(t=0) = n_B(t=0)$, then the equations lead to $n_A = n_B = 2n_{AB} = 0.4$. The latter situation is clearly unstable, since any external perturbation would make the system fall in one of the two stable fixed points. Indeed, it is never observed in simulations, due to stochastic fluctuations that in all cases determine a symmetry breaking forcing a single word to prevail.

The system of equations above however, are not only a useful example to clarify the nature of the symmetry breaking process. They also describe the interaction among two different populations that converged separately on two distinct conventions. In this perspective, they predict that the population whose size is larger will impose its conventions. In the absence of fluctuations, this is

true even if the difference is very small: B will dominate if $n_B(t=0) = 0.5 + \varepsilon$ and $n_A(t=0) = 0.5 - \varepsilon$, for any $0 < \varepsilon \leq 0.5$ and $n_{AB}(t=0) = 0$. Data from simulations shows that the probability of success of the convention of the minority group n_A, decreases as the system size increases, going to zero in the thermodynamic limit (N → ∞). A similar approach has been proposed to model the competition between two languages in the seminal paper (Abrams & Strogatz 2003). It is worth remarking the formal similarities between modelling the competition between synonyms in a Naming Game framework and the competition between languages: in both cases a synonym or a language are represented by a single feature, e.g., the characters A or B, for instance, in equations above. The similarity has been made more evident by the subsequent variants of the model introduced in (Abrams & Strogatz 2003) to include explicitly the possibility of bilingual individuals. In particular in (Wang & Minett 2005 – Minett & Wang 2008) deterministic models for the competition of two languages have been proposed which include bilingual individuals. In (Castelló et al. 2006) a modified version of the Voter model including bilinguals individuals has been proposed, the so-called AB-model. In a fully connected network and in the limit of infinite population size, the AB-model can be described by coupled differential equations for the fractions of individuals speaking language A, B or AB that are, up to a constant normalization factor in the time-scale, identical to corresponding equations for the Naming Game. In (Castelló et al. 2009) it has been shown that the Naming Game and the AB-model are equivalent in the mean field approximation, though the differences at the microscopic level have non-trivial consequences. In particular, the consensus-polarization phase transition taking place in the Naming Game (see section above) is not observed in the AB-model. As for the interface motion in regular lattices, qualitatively, both models show the same behaviour: a diffusive interface motion in a one-dimensional lattice, and a curvature driven dynamics with diffusing stripe-like metastable states in a two-dimensional one. However, in comparison to the Naming Game, the AB-model dynamics is shown to slow down the diffusion of such configurations.

Towards syntax: the case of duality of patterning

Before closing this section we only briefly mention a promising research direction inspired to the Naming Game and devoted to understanding the origin of progressively more complex syntactic structures. The first step has concerned the origin of duality of patterning at the lexicon level. The lexicons of human languages organize their units at two distinct levels (Hockett 1960 & 1960b). At a first combinatorial level, meaningless forms (typically referred to as phonemes) are combined into meaningful units (typically referred to as morphemes). Thanks to this, many morphemes can be obtained by relatively simple combinations of a small number of phonemes. At a second compositional level of the lexicon, morphemes are composed into larger lexical units, the meaning of which is related to the individual meanings of the composing morphemes. This duality of patterning is not a necessity for lexicons and the question remains wide open regarding how a population of individuals is able to bootstrap such a structure and the evolutionary advantages of its emergence. In (Tria et al. 2012) this question is addressed in the framework of a multi-agents model, where a population of individuals plays simple naming games in a conceptual environment modeled as a graph. Through an extensive set of simulations we demonstrated the existence of two sufficient conditions for the emergence of duality of patterning in a pure cultural way. The first condition is represented by a *noisy communication*, i.e., a constraint on the fidelity of message transmission. No predefined relations between objects/meanings and forms are hypothesized and we adopted a virtually infinite, i.e., open-ended, repertoire of forms. Despite this freedom, the number of different forms that get eventually fixed in the population's lexicon is kept limited by the constraint on transmission fidelity. The second sufficient condition is what we dubbed a *blending repair strategy* that allows to overcome errors in communication by allowing the creation of new words, crucially exploiting a shared conceptual representation of the environment. New words in the lexicon can be created in two ways. They can be holistically introduced as brand new forms or constructed through a blending strategy that combines and re-uses forms taken from other object's names. At the individual level, the mechanism of blending is thus introduced not as a necessity but as a

possibility to exploit when the first communication attempt resulted in a failure. We note that a blending strategy exists in natural languages (e.g., SMOG = SMoke + fOG). The blending strategy we refer to here, however has to be thought as a general mechanism by which different bits of words are put together through blends, compounds or other morphological structures. Interestingly, endowing individuals with a blending ability is necessary but not sufficient in order to observe a lexicon featuring duality of patterning. For instance combinatorial abilities are observed also in nonhuman primates (e.g. monkeys and great apes) though they still appear not having triggered the emergence of duality of patterning (Ottenheimer 2009). Compositional lexicons turn out to be faster to lead to successful communication than purely combinatorial lexicons, suggesting that meaning played a crucial role in the evolution of language.

Two crucial manipulations in the game were (i) the degree of transmission fidelity and (ii) the density of the network representing semantic relations among the objects. Combinatoriality, meant as both forms reusing and economy, is only found when the transmission fidelity is sufficiently low. With a high degree of understanding, the number of distinct forms composing the emerged lexicon turns out to be high with respect to the number of objects to be named (in particular, higher that the number of objects), and the resulting lexicon features an extremely low level of combinatoriality. Conversely, an high degree of noise leads to an high level of compactness and combinatoriality. These results suggest that combinatoriality enhances message transmission in noisy environments (Nowak et al. 1999) and emerges as a result of the need of communicative success. In contrast, the level of compositionality is not strongly affected by the level of noise, but strongly depends on how much the conceptual space is structured. In particular, the lexicons developed by the agents exhibited clear signs of compositionality when the networks representing semantic relations among the objects were neither too sparse nor too dense. This can be understood as follows: compositionality does emerge if we are able on the one hand to find common features in different objects, on the other hand to make distinctions so that not

all the objects are equally related to each other. Thus, compositionality emerges as a consequence of the organization of our conceptual space (Gärdenfors 2004, Collins & Loftus 1975).

It is important to stress how this work only considers compositionality in lexicon. In order to properly include syntax, one needs to endow the conceptual space with a more sophisticated structure, where nodes could represent nouns/concepts and links could encode semantic relations or specific actions and roles. From this perspective, the work presented in (Tria et al. 2012) can be thought as a first step of investigation on the emergence of duality of patterning through a pure cultural way. A next step could focus on investigating how a population of individuals can bootstrap a language where categorization and syntax both emerge as a pure outcome of communication efforts. When syntax has to be considered, one cannot forget to link this problem to the emergence of linguistic categories. One has, for instance, to distinguish between nouns and predicates, and if a noun is subject or complement, and more generally one has to be able to express actions and relations. We think there are at least two ways to face the problem in the framework of language games. The simpler way is to consider predefined linguistic categories and look at how a population of individuals is able to bootstrap higher order linguistic structures given for granted the underlying conceptual categorization. Personally, we are however more keen to follow a different route, where categories themselves do emerge as a result of repeated communication acts. In order to do so a richer conceptual space than the one considered in (Tria et al. 2012) has to be taken into account. For instance, one could consider a more complex conceptual space where nodes represent again nouns/concepts and links encode the semantic relations or specific actions agent/patient. In this framework one could investigate which kind of linguistic structures would emerge and whether holistic or compositional strategies will be eventually successful. A very interesting question would be for instance whether specific structures will emerge to name nodes and links. This would correspond to a first step towards the emergence of linguistic categories, i.e., the onset of syntactic structures.

2. BOOTSTRAPPING LINGUISTIC CATEGORIES

In this section we shall focus on a more complex consensus problem concerning how a population of individuals, without any pre-coordinated and shared linguistic structure, can bootstrap a shared set of linguistic categories. Categories are fundamental to recognize, differentiate and understand the environment. From Aristotle onwards, the issue of categorization has been subject to strong controversy in which purely cultural negotization mechanisms (Wittgenstein 1953 – Whorf 1956) competed with physiological and cognitive features of the categorizing subjects (Rosch 1973). A recent wave in cognitive science has induced a shift in viewpoint from the object of categorization to the categorizing subjects: categories are culture-dependent conventions shared by a given group. From this perspective, a crucial question is how they come to be accepted at a global level without any central coordination.
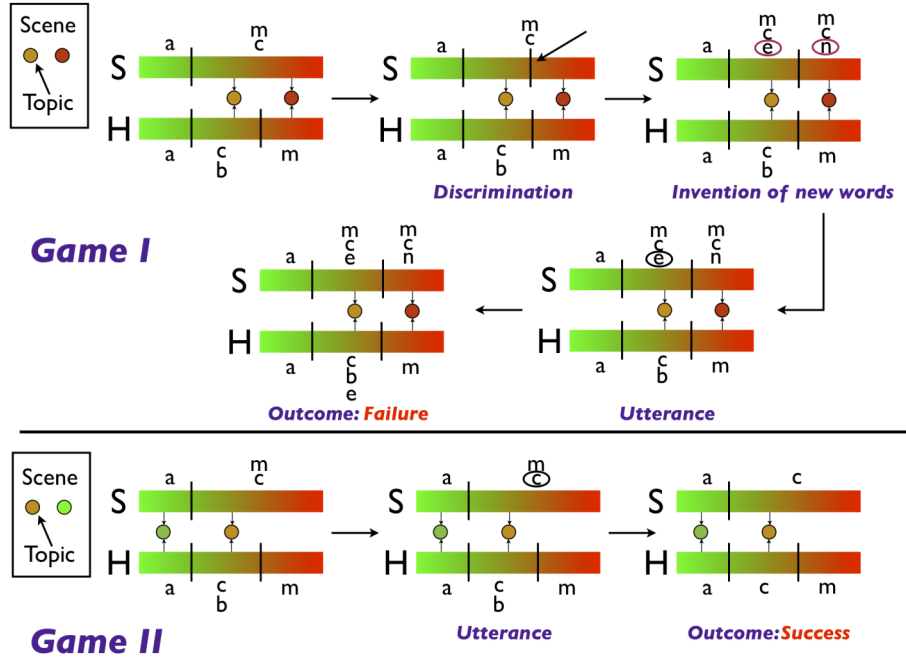
Figure 3. Basic rules of the Category Game. A pair of examples representing a failure (Game I) and a success (Game II), respectively. In a game, two players (S denoting the speaker and H denoting the hearer) are randomly selected from the population. Both the players are presented with a scene with two objects and the speaker selects the topic for the subsequent communication. In Game I, since the two objects belong to the same perceptual category of the speaker, the speaker has to discriminate her perceptual space by creating a boundary at the middle of the segment containing the two objects (marked by the bold black arrow). The two new categories formed after discrimination inherit the words-inventory of the parent perceptual category (here the words "m" and "c"); in addition, a different brand new word is invented for each of the two categories (words "e" and "n" marked by colored circles). Subsequently, the speaker browses the list of words associated to the perceptual category

containing the topic (i.e., "m", "c" and "e" here). At this point, there can be two possibilities: if a previous successful communication has occurred with this category, the last winning word is chosen; alternatively, the last word invented is selected. For the current example, the speaker chooses the word "e" (marked by the black circle here), and transmits it to the hearer. The outcome of the game is a failure since the hearer does not have the word "e" in her inventory associated with the topic. Finally, the speaker unveils the topic, in a non-linguistic way (e.g., by pointing at it), and the hearer adds the new word to the word inventory of the category corre- sponding to the topic. In Game II, the topic that the speaker chooses is already discriminated. Therefore, the speaker verbalizes it using the word "c" (which, for example, is possibly the winning word in the last successful communication concerning that category). The hearer knows this word and can therefore point to the topic correctly, thereby leading to a successful game. Both the players dispose all competing words for the perceptual category corresponding to the topic except "c". In general, if there are ambiguities (e.g., the hearer finds the word uttered to be linked to multiple categories containing an object), they are resolved by making an unbiased random choice of one of the categories.

In this paper we shall focus on the so-called Category Game (Puglisi et al. 2008), a scheme where an assembly of individuals with basic communication rules and without any external supervision may evolve an initially empty set of categories, achieving a non-trivial communication system. Its basic purpose is to examine how a population of interacting individuals can develop, through a series of language games, a shared form-meaning repertoire from scratch and without any pre-existing categorization. The model involves a set of N artificial agents committed to the task of categorizng a single analogical perceptual channel (e.g., the hue dimension of the color spectrum), each stimulus being represented as a real-valued number ranging in the interval [0, 1). We identify categorization as a partition of the [0, 1) interval (representing the perceptual channel of the agents) into discrete sub-intervals which are denoted as perceptual categories. Each individual has a dynamical inventory of

form-meaning associations linking perceptual categories (meanings) to words (forms), denoting their linguistic counterpart. The perceptual categories as well as the words associated to them co-evolve dynamically through a sequence of elementary communication interactions, usually referred to as games. In a sense the Category Game concerns the Naming of an unknown and evolving number of objects represented by the different partitions of the perceptual space.

All the players are initialized with only the trivial [0, 1) perceptual category that has no name associated to it. In each step, a pair of individuals (one playing as speaker and the other as hearer) is randomly selected from the population and presented with a new "scene", i.e., a set of M ≥ 2 objects (stimuli) where each object is a real number in the [0, 1) interval. The speaker discriminates the scene and names one object (i.e., the topic) and the hearer tries to guess the topic from the name. A correct guess results in a successful communication. Based on the outcomes of the game, the two individuals update their category boundaries and the inventory of the associated words. A detailed description of the game is provided in Fig. 3. The perceptive resolution power of the individuals limits their ability to distinguish between the objects in the scene that are too close to each other in the perceptual space. In order to take this factor into account, no two stimuli appearing in the same scene can be at a distance closer than a given value, denoted as $d_{min}(x)$ where x can be either of the two. This function, usually termed as the Just Noticeable Difference (JND), encodes the finite resolution power of human vision by virtue of which the artificial agents are not required to distinguish between stimuli that a human eye cannot differentiate.

Dynamical properties of the Category Game

In the Category Game dynamics it is possible to distinguish two different phases. In the first regime, the number of perceptual categories increases (see dashed lines in Fig. 4c) due to the pressure of discrimination, and at the same time many different words are used by different agents for naming similar

perceptual categories. This kind of synonymy reaches a peak and then dries out (as displayed in Fig. 4a), in a similar way as in the Naming Game described above. This kind of synonymy reaches a peak and then drops in a fashion similar to the well-known Naming Game. A second phase starts when most of the perceptual categories are associated with only one word. During this phase, words are found to expand their dominion across adjacent perceptual categories (solid lines in Fig. 4c). In this way, sets of contiguous perceptual categories sharing the same words are formed, giving raise to what we define as "linguistic categories "(see Fig. 5). The coarsening of these categories becomes slower and slower, with a dynamical arrest analogous to the physical process in which supercooled liquids approach the glass transition (Mézard et al. 1987). In this long-lived almost stable phase, (Mukherjee et al. 2011) usually after 10^4 games per player, the linguistic categorization pattern has a degree of sharing between 90% and 100%; success is measured by counting in a small time window the rate of successful games (Fig. 4b), while the degree of sharing of categories is measured by an overlap function, which measure the alignment of category boundaries (both for perceptual or linguistic ones), displayed in Fig. 4d: for a mathematical definition of this function see (Puglisi et al. 2008). The success rate and the overlap both remain stable for $10^5 - 10^6$ games per player: we consider this pattern as the "final categorization pattern" generated by the model, which is most relevant for comparison with human color categories (see below).

We can thus identify two main outcomes of the Category Game. On the one hand the emergence of a hierarchical category structure made of two distinct levels: a basic layer, responsible for fine discrimination of the environment, and a shared linguistic layer that groups together perceptions to guarantee communicative success. Remarkably, the emergent number of linguistic categories in this phase turns out to be finite and small, as observed in natural languages, even in the limit of an infinitesimally small length scale dmin, as opposed to the number of the underlying perceptual categories which is of
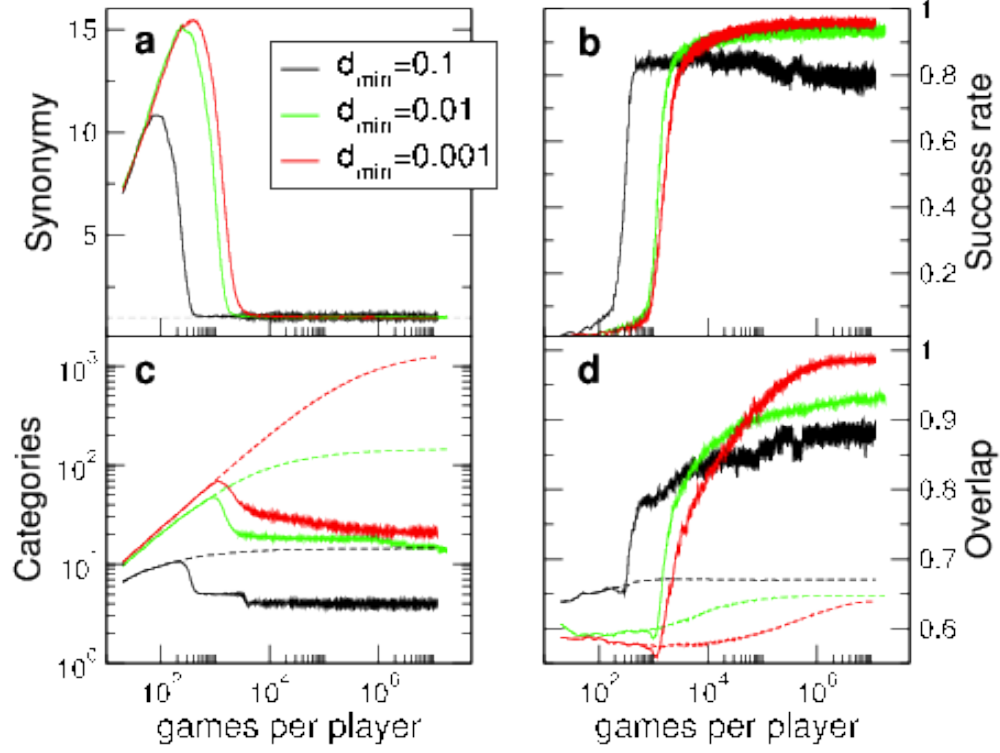
order $1/d_{min}(x)$.

Figure 4. Time evolution of the Category Game. We considered here a population of N = 100 individuals a and a flat (constant) JND function $d_{min}(x) = d_{min}$ with different values of d_{min}: a) Synonymy, i.e., average number of words per category; b) Success rate measured as the fraction of successful games in a sliding time windows games long; c) Average number of perceptual (dashed lines) and linguistic (solid lines) categories per individual; d) Averaged overlap, i.e., alignment among players, for perceptual (dashed curves) and linguistic (solid curves) categories.

Another important feature of the Category Game concerns its long-time behaviour. Human languages evolve continuously, and a puzzling problem is how to reconcile the apparent robustness of most of the deep linguistic structures we use with the evidence that they undergo possibly slow, yet ceaseless, changes. Is the state in which we observe languages today closer to what would be a dynamical attractor with statistically stationary properties or rather closer to a non-steady state slowly evolving in time? The Category Game allows to address this question in the framework of the emergence of shared linguistic categories in a population of individuals interacting through language games. The observed emerging asymptotic categorization, which we shall see can be successfully tested against experimental data from human languages, corresponds to a metastable state where global shifts are always possible but progressively more unlikely and the response properties depend on the age of the system. This aging mechanism exhibits striking quantitative analogies to what is observed in the statistical mechanics of glassy systems. We argue that this can be a general scenario in language dynamics where shared linguistic conventions would not emerge as attractors, but rather as metastable states.
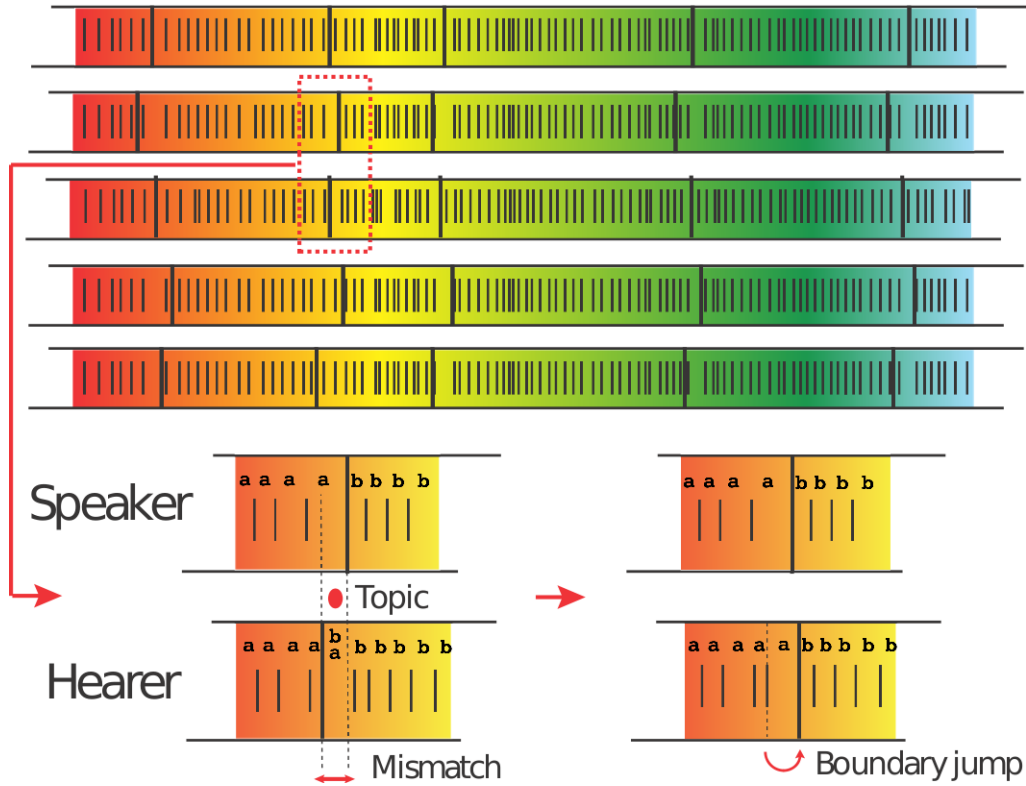
Figure 5. Typical long-time configuration of five representative agents in the population. For each agent perceptual and linguistic categories (separated by short and long bars, respectively) are shown. The highlighted portion of two agents illustrates an instance of a successful game in a so-called mismatch region between the linguistic categories of the two agents associated with the words "a" and "b". The hearer - in a previous game - learned the word "a" as a synonym for the perceptual category at the leftmost boundary of the linguistic category "b". During the game the speaker utters "a" for the topic; as a result the hearer deletes "b" from her inventory, keeping "a" as the name for that perceptual category, moving de facto the linguistic boundary.

3. COMPARISON WITH REAL-WORLD DATA

An important counterpart of the modelling activity should be the comparison with empirical or experimental data. When theoretical modelling is coupled with a serious data analysis activity devoted to the discovery of emergent features, it can result in a virtuous loop, where measures inspire modelling schemes, model analysis suggests new measures and observations, which in turn allow the evaluation and refinement of models. Traditionally language dynamics suffered for the lack of extensive datasets mainly due to the difficulty of monitoring in a systematic and reproducible way the emergent steps of new linguistic features or new languages altogether. The situation is being radically changing in the last few years, Information and Communication Technologies playing a major role in this revolution. In this section we shall describe a few examples where the predictions of the modelling schemes described above have been tested against experimental data from human languages. Next section will be devoted to illustrate the new perspective to conceive and run synthetic experiments aimed at investigating specific aspects of the emergence and evolution of language.

Naming syntactic structures in Twitter

Despite a fast growing literature about the emergence of social conventions in online social networks, not many results are available whenever one is interested in the evolution of linguistic conventions. Without the aim of being exhaustive, here we only mention a recent contribution that attracted our attention given the similarity of the dynamics described to that of the Minimal Naming Game described above. In a recent paper (Kooti et al. 2012) investigated the emergence of social conventions in Twitter. Twitter (twitter.com) is an online social networking and microblogging service that enables its users to send and read text-based posts of up to 140 characters,

known as "tweets". A specific action on Twitter is the so-called retweet, i.e., the action by which a user makes a post simply re-posting content that has been posted by another user. A retweet is signaled by a specific symbol and what Kooti and collaborators have investigated is the history of the different conventions users adopted for a retweet along the whole Twitter history. The fairly complete dump of the Twitter history allowed the researchers to monitor who first invented each given symbol for a retweet and when each symbol was adopted by which user. It is thus possible to trace the time evolution of each competing convention. They monitored in particular the following conventions (in order of introduction): *via*, *HT*, *Retweet*, *Retweeting*, *RT*, *R/T* or the recycle icon. Despite being invented at different times and having different adoption rates, only two variations came to be widely adopted: *RT* and *via*. It is important to notice that the most popular convention does not coincide with the first invented. In addition an actual competition of different conventions took place since it is not rare the situation in which users adopted, in their Twitter activity, several conventions at different times and sometimes at the same time. These last two elements make this dynamics quite close to the dynamics of the Naming Game. Also in the Naming Game the winning convention is not always the first invented. In addition each player is adopting, in the Naming Game, several conventions at the same time or in its own lifetime, before converging on a final choice. Finally also the Naming Game predicts the possibility of the coexistence of several conventions in a quasi-stationary state. It would be quite interesting to extend this research in two directions. First, monitoring the time evolution of different conventions and syntactic structures in Twitter in order to test the robustness of the results. Second, perform a thorough quantitative comparison of the Naming Game predictions with the data presented in (Kooti et al. 2012).

Universality in colour naming

A large amount of data on colour categorization was gathered in the World Colour Survey (Berlin & Kay 1969, Kay & Regier 2003), in which individuals belonging to different cultures had to name a set of colours. The results of the

analysis of the categorization patterns obtained in this way have had a huge impact not only on such areas as Cognitive Science and Linguistics, but also Psychology, Philosophy and Anthropology (see for example, Lakoff 1987, Gardner 1985, Deacon 1998). The main finding is that colour systems across language are not random, but rather exhibit certain statistical regularities, thus implying that the classical theory of categorization, dating back to the work of Aristotle and claiming the arbitrariness of categorization, had to be reconsidered (Gardner 1985). In this section, we describe how the Category Game model described above can be used to run a *Numerical* World Colour Survey (NWCS) and point out that, remarkably, the synthetic results obtained in this way agree quantitatively with the experimental ones (Baronchelli et al. 2010, Loreto et al. 2012).

The World Color Survey

P. Kay and B. Berlin (Berlin & Kay 1969) ran a first survey on 20 languages in 1969. From 1976 to 1980, the enlarged World Colour Survey was conducted by the same researchers along with W. Merrifield and the data are public since 2003 on the website http://www.icsi.berkeley.edu/wcs. These data concern the basic colour categories in 110 languages without written forms and spoken in small-scale, non-industrialized societies. On average, 24 native speakers of each language were interviewed. Each informant had to name each of 330 colour chips produced by the Munsell Colour Company that represent 40 gradations of hue and maximal saturation, plus 10 neutral colour chips (black-gray-white) at 10 levels of value. The chips were presented in a predefined, fixed random order, to the informant who had to tag each of them with a ``basic colour term'' is her language. Berlin and Kay's established the universal presence of a special subset of color names which they called the "basic colour names". These are the most salient and frequently used colour words across the majority of the world's languages. They represent the following eleven English color names: black, white, red, green, yellow, blue, brown, orange, purple, pink and gray. Berlin and Kay found that these names have prototype properties which means that there is usually one name that best

represents a color while other colors that are progressively more dissimilar with this color become less good examples for the name. They also found that the number of basic color names range from 2 to 11 across the world's languages, of course with exceptions like Russian and Hungarian that have 12 basic names.

A very important result, always due to Berlin and Kay (Kay & Regier 2003), emerged from a quantitative statistical analysis of the World Colour Survey, proving that the colour naming systems obtained in different cultures and language are in fact not random. Through a suitable transformation Berlin and Kay identified the most representative chip for each colour name in each language and projected it into a suitable metric colour space (namely, the CIEL*a*b colour space). To investigate whether these points are more clustered across languages than would be expected by chance, they defined a dispersion measure on this set of languages L_0 as:

$$D(L_0) = \sum_{l,l^* \in L_0} \sum_{c \in l} min_{c^* \in l^*} \, dist(c,c^*)$$

where l and l* are two different languages, c and c* are two basic colour terms respectively from these two languages, and $dist(c,c^*)$ is the distance between the points in colour space in which the colours are represented. To give a meaning to the measured dispersion $D(L_0)$, Kay and Regier created ``new'' datasets $L_i$ ($i = 1, 2, \ldots, 1000$) by random rotation of the original set $L_0$, and measured the dispersion of each new set $D(L_i)$. The human dispersion appears to be distinct from the histogram of the ``random'' dispersions with a probability larger than 99.9\%. As shown in Figure 3a of (Kay & Regier 2003), the average dispersion of the random datasets, $D_{neutral}$, is 1.14 times larger than the dispersion of human languages. Thus, human languages are more clustered, i.e., less dispersed, than their random counterparts and universality does exist.

A third and a totally unexpected finding of Berlin and Kay concerns the existence of a hierarchy of colour names. They observed that if a language encodes fewer than eleven names, then there are strict limitations on which names it may encode. The typological regularities observed by them can be summarized by the following implicational hierarchy:

$$\begin{bmatrix} white \\ black \end{bmatrix} < \begin{bmatrix} red \end{bmatrix} < \begin{bmatrix} green \\ yellow \end{bmatrix} < \begin{bmatrix} blue \end{bmatrix} < \begin{bmatrix} brown \end{bmatrix} < \begin{bmatrix} purple \\ pink \\ orange \\ gray \end{bmatrix}$$

where for distinct color names a and b, the expression a < b signifies that a is present in every language where b is present but not vice versa. Based on the above observation, the authors further theorize that, as languages evolve, they acquire the new basic color names in a fixed chronological sequence of the form

Stage I: dark-cool and light-warm
Stage II: red (including all shades of violet)
Stage III: either green or yellow
Stage IV: both green and yellow
Stage V: blue
Stage VI: brown
Stage VII: purple, pink, orange, or gray

It is worth noticing that Stage I is not referring to the emergence of the two achromatic colors "black" and "white", rather it refers to a division of the perceptual space that has nothing to do with the chromatic properties of light, being based exclusively on the light intensity. Ratliff writes (Ratliff 1976) that the well-known studies of Dani color terms by Eleanor Heider-Rosch and Donald Olivier (Heider-Rosch & Olivier 1972) "put the question of

psychophysiological bases of the two color terms of Stage I into better perspective. These terms appear to be panchromatic, more or less equivalent to the general panchromatic English terms dark and light or dull and brilliant rather than equivalent to the specific achromatic terms black and white. Although the Dani color terms do include chromatic colors, and do have attributes of coolness and warmth, the division between them appears to be based mainly on brightness."

The Numerical World Colour Survey

The key aspect of the statistical analysis described above is the comparison of the clustering properties of a set of true human languages against the ones exhibited by a certain number of randomized sets. In replicating the experiment it is therefore necessary to obtain two sets of synthetic data, one of which must have some human ingredient in its generation. The idea put forth in (Baronchelli et al. 2010) is to act on the d_{min} parameter of the Category Game, describing, as discussed in the previous section, the discrimination power of the individuals to stimuli of a given wavelength. In fact, it turns out that human beings are endowed with a d_{min}, the ``Just Noticeable difference" or JND, that is not constant, but rather is a function of the frequency of the incident light (see Fig. 6)[1]. Technically, psychophysiologists define the JND as a function of wavelength to describe the minimum distance at which two stimuli from the same scene can be discriminated (Bedford & Wyszecki 1958, Long et al. 2006). The equivalence with the d_{min} parameter is therefore clear and different artificial sets can be created. On the one hand *human* categorization patterns are obtained from populations whose individuals are endowed with the rescaled human JND (i.e., $d_{min}(x)$). On the other hand *neutral* categorization patterns are obtained from populations in which the individuals have constant JND $d_{min} = 0.0143$, which is the average value of the human JND (as it is projected on the [0,1) interval, Fig. 6).

---

[1] The attention is here on the human Just Noticeable Difference for the hue, see (Baronchelli et al. 2010).
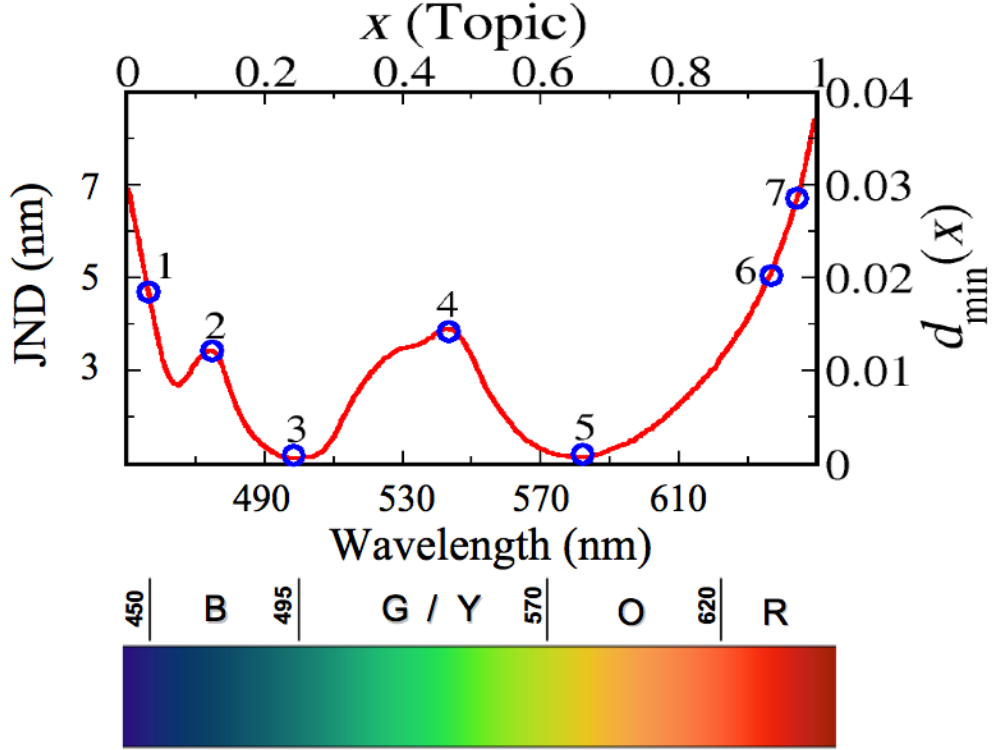
Figure 6. The Just Noticeable Difference (JND) function. The wavelength change in a monochromatic stimulus needed to elicit a particular JND in the hue space. For the purpose of the Category Game, $d_{min}(x)$ and topic respectively refers to the JND and the monochromatic stimulus rescaled within the $[0 : 1]$ interval. The blue circles represent the centers of seven regions (to be used later in the article) that can be together expressed as a vector $\vec{c}$ with entries $(c_1, c_2, \cdots, c_7)$. The specific values for these entries are $\vec{c} = (0.0301, 0.125, 0.250, 0.465, 0.66015, 0.925, 0.970)$. Each entry, in turn, respectively

corresponds to a wavelength (nm) that can be written (approx.) as (445, 475, 500, 545, 585, 635, 645).

In analogy to the WCS experiment, the randomness hypothesis in the NWCS for the neutral test-cases is supported by symmetry arguments: in neutral simulations there is no breakdown of translational symmetry, which is the main bias in the *human* simulations.
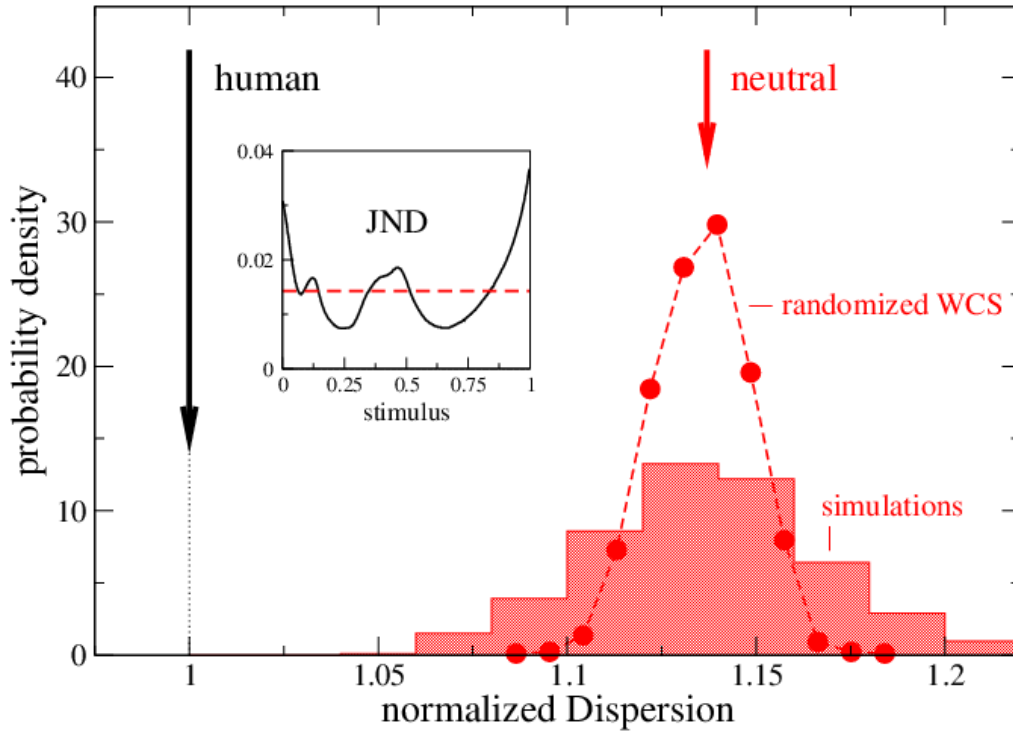

Figure 7. Neutral worlds, $D_{neutral}$, (histogram) are significantly more dispersed than human worlds, D_{human}, (black arrow), as also observed in the WCS data (the filled circles extracted from (Kay & Regier 2003) and the black arrow). The abscissa is rescaled so that the human D (WCS) and the average ``human worlds'' D both equal 1. The histogram has been generated from 1500 neutral

worlds, each made of 50 populations of 50 individuals, and M=2 objects per scene. Categorization patterns have been considered after the population had evolved for a time of 10^6 games per agents. The inset figure is the human JND function (adapted from (Long et al.2006)). On the vertical axis: the probability density $\rho(x_i)$ equals the percentage $f(x_i)$ of the observed measure in a given range $\left[x_i - \Delta/2, x_i + \Delta/2\right]$ centring around x_i divided by the width of the bin Δ, i.e., $\rho(x_i) = f(x_i)/\Delta$. This procedure allows for a comparison between the histogram coming from the NWCS (Baronchelli et al. 2010) and that obtained in the study on the WCS (Kay & Regier 2003), where the bins have a different width.

Thus, the difference between human and neutral data originates from the perceptive architecture of the individuals of the corresponding populations. A collection of human individuals form a human population, and will produce a corresponding human categorization pattern. In a hierarchical fashion, finally, a collection of populations is called a *world*, which in (Baronchelli et al. 2010) is formed either by all human or by all non-human populations. To each world it corresponds a value of the dispersion D defined above, measuring the amount of dispersion of the languages (or categorization patterns) belonging to it. In the actual WCS there is of course only one human World (i.e., the collection of 110 experimental languages), while in (Baronchelli et al. 2010) several worlds have been generated to gather statistics both for the human and non-human cases.

The main results of the NWCS are presented in Figure 7. Since the dispersion D depends on the number of languages, the number of colours, and the space units used, every measure of D in the NWCS is normalized by the average value obtained in the human simulations, and every measure of D from the WCS experiment is divided by the value obtained in the original (non-randomized) WCS analysis (as in (Kay & Regier 2003)). Thus, both the average of the *human worlds* and the value based on the WCS data are

represented by 1 in Figure 7. In the same plot, the probability density of observing a value of D in the *neutral world* simulations is also shown by the red histogram bars. Finally, the Figure contains also the data reported in the histogram of the randomized datasets in Figure 3a of (Kay & Regier 2003), whose abscissa is normalized by the value of the non-randomized dataset and frequencies are rescaled by the width of the bins.

Figure 7 illustrates the main results. The Category Game Model informed with the human $d_{min}(x)$ (JND) curve produces a class of worlds that has a dispersion lower than and well distinct from that of the class of worlds endowed with a non-human, uniform $d_{min}(x)$. Strikingly, moreover, the ratio observed in the NWCS between the average dispersion of the ``neutral worlds'' and the average dispersion of the human worlds is $D_{neutral} / D_{human} \cong 1.14$, very similar to the one observed between the randomized datasets and the original experimental dataset in the WCS.

These findings are important for a series of reasons. First of all, it is the first case in which the outcome of a numerical experiment in this field is comparable at any level with true experimental data. Second, as discussed above, the results of the NWCS are not only in qualitative, but also in quantitative agreement with the results of the WCS. Third, the very design of the model suggests a possible mechanisms lying at the roots of the observed universality. Human beings share certain perceptual bias that, even though are not strong enough to deterministically influence the outcome of a categorization, are on the other hand capable of influencing category patterns in a way that becomes evident only through a statistical analysis performed over a large number of languages. This explanation for the observed universality had already been put forth based on theoretical analysis (see, for instance (Deacon 1998, Christiansen & Chater 2008), but the NWCS represents the first numerical evidence supporting it.

The hierarchy of colour names

We mentioned above the finding about the existence of an implication hierarchy of color names. In the framework of the Category Game it is has been possible to provide a possible explanation of the origin of such a hierarchy. In particular it has been observed (Loreto et al. 2012) that a clear hierarchy for colour names is found to naturally emerge, in the framework of the Category Game, through purely cultural negotiations among a population of co-evolving agents, each endowed with the human Just Noticeable Difference (JND) function (Bedford & Wyszecki 1958, Long et al. 2006). In particular, a hierarchy emerges that ranks different colour names with respect to the time needed for them to get fixed in a population.

Let us first focus on the frequency of access to higher levels of linguistic categorization as a function of the local value of the JND (see (Loreto et al. 2012) for further details). To this end, we computed the extent of the emergent agreement (i.e., match) at different regions of the perceptual space. The notion of *match* is as follows. A match region $match(i, j)$ for a pair of agents i and j is the sum of the lengths of all the regions in their perceptual space where both of them have the same most relevant name. The most relevant name is either the one used in a previous successful communication or the newly invented name in case the category has just been created due to a discrimination event . Note that this is a quantitative measure of the amount of agreement between the agent pair. The match of the whole population is simply:

$$\frac{2\sum_{i=1}^{N}\sum_{j=i+1}^{N} match(i, j)}{N(N-1)}$$

In Figure 6, the blue circles indicate the centers c_i of seven such regions (i.e., the points of inflection in the JND function) that we choose to calculate the so-called "regional" agreement. We define a region by the length spanning the interval $\left[c_i - d_{min}(c_i), c_i + d_{min}(c_i)\right]$, where $d_{min}(c_i)$ is the y-value corresponding to

the x-value c_i (see Fig. 6). In Fig. 8(a) and (b), we respectively show, for N = 500 and 700, the regional agreement for these seven regions. The plots clearly signal that consensus emerges first in regions corresponding to high values of d_{min} (e.g., region 6 and 7) while it occurs later in regions corresponding to very low d_{min} (e.g., region 3 and 5). Most strikingly, if the regions are arranged according to the time (i.e., t/N) to reach a desired level of consensus (say a match value of 0.1) then they get organized into a hierarchy (Fig. 8(c) and (d)) with [red, (magenta)-red], [violet], [green/yellow], [blue], [orange] and [cyan] (or [cyan] and [orange] as is usually observed for secondary basic color names) appearing in this order. This result is strikingly similar to that reported in (Berlin & Kay 1969). Further, the data points for the fixation times are observed to obey a simple functional form, $Ae^{-\alpha t}$, where A and α are non-zero positive constants (gray lines in Fig. 8(c) and (d)). This amounts to say that the fixation time for specific primary colors at the population level diverges log- arithmically with the resolution power $1/d_{min}$. Though this specific prediction cannot be checked with the currently available data, it is reminiscent of the logarithm law which is typically associated to human perception. Error bars in Fig. 8(c) and (d), representing the intrinsic variability of fixation times in different simulations, are important to explain the slight fluctuations in the color name hierarchy as observed in the World Colour Survey across different cultures.
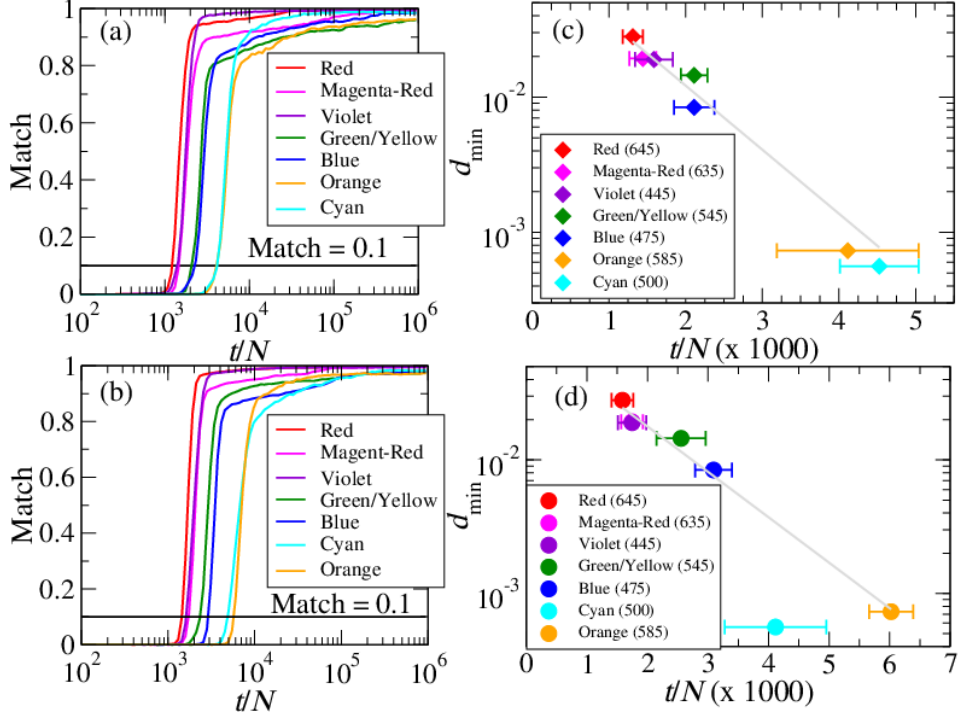
Figure 8 Agreement emergence. Emergence of the agreement in the population. Match for (a) N = 500 and (b) N = 700 in the seven regions marked in Fig. 6. For better visualization, each curve is plotted in a color that best represents the corresponding region in the hue space (see Fig. 6). The time (i.e., t/N) for (c) N = 500 and (d) N = 700 to reach a desired consensus (match = 0.1) versus the value of d_{min} corresponding to the seven regions. The results present an average over 60 simulation runs. In both the plots the approximate wavelength (nm) associated with each colored data point is mentioned within the parenthesis. Error bars are drawn according to the variance of the distribution of consensus times in the different simulations. The gray lines in both the plots represent a fit of the respective data with an exponential function of the form $Ae^{-\alpha t}$ (see text for more details).

It is important to observe how the similarity of the ranking of fixation times obtained in the framework of the Category Game with that observed in the framework of the World Colour Survey is not the outcome of a pure coincidence. It turns out that only a right choice of JND function, coupled with the language game dynamics, can reproduce the colour hierarchy observed across human languages. The Supporting Information of (Loreto et al. 2012) reports the outcomes of two additional experiments performed by substituting the human JND with a flat and an inverse JND. In none of these two cases the hierarchy obtained from the World Colour Survey could be reproduced.

In summary these results show that a simple negotiation dynamics, driven by a weak non-language specific bias, namely the frequency dependent resolution power of the human eye, is sufficient to guarantee the emergence of the hierarchy of color names getting so arranged by the times needed for their fixation in a population. The observed hierarchy features an excellent quantitative agreement with the empirical observations, confirming that the theoretical modeling in this area has now attained the required maturity to make significant contributions to the ongoing debates in cognitive science. Our approach suggests a possible route to the emergence of hierarchical colour categories: the color spectrum clearly exists at a physical level of wave-lengths, humans tend to react most saliently to certain parts of this spectrum often selecting exemplars for them, and finally comes the process of linguistic colour naming, which adheres to universal patterns resulting in a neat hierarchy of the form obtained here. These intuitions are of course not a novelty (see for instance (Evans & Levinson 2009)); however, we provided a theoretical framework where the origin of the colour hierarchy, as well as its quantitative structure, could be explained and reproduced through a purely cultural route driven, on its turn, by a non-language-specific property of human beings.

It should be remarked that, despite the striking universal character of the colour hierarchy, fluctuations exist across different languages as for the precise order in which colour names got fixed in each language. In the framework of

the Category Game this phenomenon is naturally explained as a consequence of the unavoidable stochasticity of the underlying cultural negotiation dynamics (Puglisi et al. 2008). The error bars in the fixation time of each specific colour term in Fig. 8 specifically support this picture.

4. AN EXPERIMENTAL FRAMEWORK

We already mentioned how the Web is acquiring the status of a platform for social computing, able to coordinate and exploit the cognitive abilities of the users for a given task, and it is likely that the new social platforms appearing on the Web will rapidly become a very interesting laboratory for social sciences in general (Lazer et al. 2009), and for studies on language emergence and evolution in particular. These recent advances are enabling for the first time the possibility of collecting the interactions of large numbers of people at the same time and observing their behaviour in a reproducible way. In particular, the dynamics and transmission of information along social ties can nowadays be the object of a quantitative investigation towards a comprehension of the processes underlying the emergence of a collective information and language dynamics.

A very original example is represented by Amazon's Mechanical Turk (MT) (https://www.mturk.com/mturk/welcome), a crowdsourcing web service that coordinates the supply and the demand of tasks that require human intelligence to complete. It is an online labour market in which users perform tasks, also known as Human Intelligence Tasks, proposed by "employers" and are paid for this. Salaries range from cents for very simple tasks to a dollar or more for more complex ones. Examples of tasks range from categorization of images, the transcription of audio recordings to test websites or games. MT is perhaps one of the clearest examples of the so called crowdsourcing and thousands of projects, each fragmented into small units of Work, are performed every day by thousands of users. MT has opened the door for exploration of processes that outsource computation to humans. These human computation processes

hold tremendous potential to solve a variety of problems in novel and interesting ways. Thanks to the possibility of recruiting thousands of subjects in a short time, MT represents a potentially revolutionary source for conducting experiments in social science (Chilton et al. 2009, Paolacci et al. 2010). It could become a tool for rapid development of pilot studies for the experimental application of new ideas. As a starting point for this new idea of experiments, the blog http://experimentalturk.wordpress.com/ already presents a review of the results of a series of classic game theoretical experiments carried out on MT (Suri & Watts 2011).

Despite its versatility (Chilton et al. 2009) MT has not been conceived as a platform for experiments. This is the reason why it is important to develop a versatile platform to implement social games. Here the word game is intended as an interaction protocol among a few players implementing a specific task and it is used as a synonym of experiment. The development of such web games has to take into account the following points: (i) the running applications must be modular, so that they can interact with different services and interfaces and can be interchangeable; they must be event-driven in order to ease the real time interactions between users and have to possibly interact with social networks and cloud services through their own APIs; (ii) the transactions between synchronous (i.e., real time) and asynchronous mode should be the most transparent as possible; (iii) the cross-platform web-based graphical interface, either ajax, flash or java, must be differently designed according to the client platform (e.g. desktops, smart-phones, tablets, etc.); (iv) the hosting infrastructures have to be care- fully designed to manage an expected heavy load and to process and store the relative amount of data. The advantage of this kind of experiments is that every useful piece of information and detail of the evolution will be fully available and leveraged for benchmarking as well as for the modelling activity. Moreover the effects of social interactions can be observed with a larger statistical basis and in a more controlled environment. A first prototype of such a platform is already available and it has been dubbed Experimental Tribe (www.xtribe.eu) (ET) (Cicali et al. 2011). ET is intended as a general-purpose platform that allows

the realization of a very large set of possible games. It has a modular structure through which most of the complexity of running an experiment is hidden in a complex Main Server and the experimentalist is left with the only duty of devising the experiment as well as a suitable interface for it. In this way most of the coding difficulties related to the realization of a dynamic web applications are already taken care by the ET Server and the realization of an experiment should be as easy as constructing a webpage with one the main utilities for it (e.g., googlesite).

5. SHORT SUMMARY

Before concluding we wish to thank Dr. Peng Gang and Dr. Shi Feng for the kind invitation to contribute to this festschrift devoted to William (Bill) Wang. It is with great joy that we wrote this contribution since Bill has been pioneering the field of language dynamics and most of what we did in the last few years probably wouldn't be there without the initial seed ideas Bill first planted. As a pioneer in evolutionary studies of language Bill always put a special emphasis on a multi-disciplinary perspective with a particular sensibility for pushing forward synthetic modelling approaches as well as for drawing parallels with biological principles. On top of this Bill has always been a friend and a passionate and generous scholar, always ready to guide us with his kind, wise and deep way to transfer his knowledge.

In this paper we made an attempt to summarize the efforts we made in the last few years to elucidate the processes underlying the emergence of shared linguistic structures in a population of individuals. We presented in particular models of increasing complexity devoted to reproduce in a synthetic way the processes leading to naming objects, bootstrapping a system of linguistic categories and to the emergence of simple syntactic structures such as duality of patterning in lexicons. Along with the theoretical models we presented the results of the comparison of the theoretical predictions with available empirical data. This helps in grounding the hypotheses made against well known facts as well as in triggering a virtuous loop in which new theoretical predictions foster

the gathering of new datasets which in turn stimulates further thinking about the underlying mechanisms and so on. We highlighted the need for the field of language dynamics to complete its transformation in an experimental discipline by paying an increasing attention to the comparison with real-world data as well as to the protocols to conceive and run suitable experiments. To this end the Web, and more generally the new Information and Communication Technologies (ICT), could be the new frontier to run linguistics motivated experiments, by leveraging on the possibility to coordinate and exploit the cognitive abilities of the users for a given task. It is very likely that the new social platforms appearing on the Web, could rapidly become a very interesting laboratory for social sciences in general and for studies on language emergence and evolution in particular.

ACKNOWLEDGEMENTS

REFERENCES

ABRAMS Daniel M. and STROGATZ Steven H. 2003. Modelling the dynamics of language death. *Nature* 424: 900.

AXELROD Robert 1997. The Dissemination of Culture: A Model with Local Convergence and Global Polarization. *J. Conflict Resolut*. 41(2): 203-226.

BÁRABASI Albert-László & ALBERT Reka 1999. Emergence of scaling in

random networks. *Science* 286: 509.

BARONCHELLI Andrea & FELICI Maddalena & LORETO Vittorio & CAGLIOTI Emanuele & STEELS Luc 2006. Sharp transition towards shared vocabularies in multi-agent systems. *J. Stat. Mech.*, P06014.

BARONCHELLI Andrea & DALL'ASTA Luca & BARRAT Alain & LOREYO Vittorio 2006. Topology Induced Coarsening in Language Games. *Phys. Rev. E* 73: 015102.

BARONCHELLI Andrea & DALL'ASTA Luca & BARRAT Alain & LORETO Vittorio 2007. Non-equilibrium phase transition in negotiation dynamics. *Phys. Rev. E* 76: 051102.

BARONCHELLI Andrea & GONG Tao & PUGLISI Andrea & LORETO Vittorio 2010. Modelling the emergence of universality in colour naming patterns. *Proc. Natl. Acad. Sci. USA (PNAS)* 107: 2403-2407.

BEDFORD RE & WYSZECKI GW 1958. Wavelength discrimination for point sources. *JOSA* 48(2): 129-130.

BERLIN Brent & KAY Paul 1969. Basic Color Terms. Berkeley: University of California Press.

BRAY Alan 1994. Theory of phase-ordering kinetics. *Adv. Phys.* 43(3): 357-459.

CASTELLANO Claudio & FORTUNATO Santo & LORETO Vittorio 2009. Statistical physics of social dynamics. *Reviews of Modern Physics* 81: 591–646.

CASTELLÓ Xavier & EGUÍLUZ Víctor M. & SAN MIGUEL Maxi 2006.

Ordering dynamics with two non-excluding options: bilingualism in language competition. *New J. Phys.* 8(12): 308.

CASTELLÓ Xavier & BARONCHELLI Andrea & LORETO Vittorio 2009. Consensus and ordering in language dynamics. *The European Physical Journal B - Condensed Matter and Complex Systems* 71(4): 557-564.

CHILTON Lydia B. & SIMS Clayton T. & GODMAN Max & LITTLE Greg & MILLER Robert C. 2009. Seaweed: a web application for designing economic games. In Proceedings of the ACM SIGKDD Workshop on Human Com- putation, HCOMP '09, pages 34–35, New York, NY, USA.

CHRISTIANSEN Morten H. & Chater Nick 2008. Language as shaped by the brain. *Behavioral and Brain Sciences* 31(5): 489-509.

CICALI Claudio & GRAVINO Pietro & LORETO Vittorio & PAOLACCI Gabriele & SERVEDIO Vito D.P. & TRIA Francesca & WARGLIEN Massimo 2011. Experimental tribe: a general platform for web-gaming and social computation, Proceedings of the NIPS Workshop on Computational Social Science and the Wisdom of Crowds (http://people.cs.umass.edu/~wallach/workshops/nips2011css/).

CLARK, Herbert H., 1996. Using Language, Cambridge University Press, UK.

COLLINS Allan M. & LOFTUS Elizabeth F. 1975. A spreading-activation theory of semantic processing. *Psychological Review* 82(6): 407–428.

DALL'ASTA Luca & BARONCHELLI Andrea & BARRAT Alain & LORETO Vittorio 2006. Nonequilibrium dynamics of language games on complex networks. *Phys. Rev. E* 74(3): 036105.

DALL'ASTA Luca & BARONCHELLI Andrea & BARRAT Alain &

LORETO Vittorio 2006b. Agreement dynamics on small-world networks. *Europhys. Lett.* 73(6): 969-975.

DEACON Terry W. 1998. The symbolic species: The co-evolution of language and the brain. Norton & Company, New York.

ERDÖS Paul & RÉNYI Alfred 1959. On Random Graphs I. *Publ. Math. Debrecen* 6: 290.

ERDÖS Paul & RÉNYI Alfred 1960. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* 7: 17.

EVANS Nicholas, & LEVINSON Stephen C. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32:429–448.

HAEGER et al. 2009. What can mathematical, computational and robotic models tell us about the origins of syntax?, book chapter in "Biological Foundations and Origin of Syntax", ed. D. Bickerton and E. Szathmáry. Strüngmann Forum Reports, vol. 3. Cambridge, MA: MIT Press.

GALANTUCCI Bruno & GARROD Simon 2010. Experimental Semiotics: A new approach for studying the emergence and the evolution of human communication. Special Issue of *Interaction Studies* 11(1).

GALANTUCCI Bruno 2005. An Experimental Study of the Emergence of Human Communication Systems. *Cognitive Science* 29(5): 737–767.

GÄRDENFORS Peter 2004. Conceptual spaces as a framework for knowledge representation. *Mind and Matter* 2: 9–27.

GARDNER H. 1985. The Mind's New Science: A History of the Cognitive Revolution. Basic Books, New York.

GARROD, Simon & Pickering Martin J., 2004. Why is conversation so easy? *Trends in Cognitive Sciences* 8(1): 8 –11.

_____. 2009. Joint Action, Interactive Alignment, and Dialog. *Topics in Cognitive Science* 1(2): 292–304.

HEIDER-ROSCH E. & OLIVIER D.C. 1972. The structure of the color space in naming and memory for two languages. *Cognitive Psychology* 3:337-354.

HOCKETT Charles F. 1960. Logical considerations in the study of animal communication, *Americal Institute of Biological Sciences*. 392-430.

HOCKETT Charles F. 1960b. The origin of speech. *Scientific American* 203: 88-96.

KAY Paul & REGIER Terry 2003. Resolving the question of color naming universals. *Proc. Natl. Acad. Sci. USA (PNAS)* 100(15): 9085-9089.

KE Jinyun, MINETT James, AU Ching-Pong & WANG William S-Y 2002. Self-organization and selection in the emergence of vocabulary. *Complexity* 7(3): 41–54.

KE Jinyun & GONG Tao & WANG William S-Y 2008. Language change and social networks. *Comm. Comput. Phys*. 3(4): 935—949. Originally presented at the 5th Conference on Language evolution, Leipzig, Germany, March 2004.

KIRBY Simon & CORNISH Hanna H. & SMITH Kenny 2008. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proc. Natl. Acad. Sci. USA (PNAS)* 105(31): 10681–10686.

KLEMM Konstantin & EGUÍLUZ Víctor M. & TORAL Raúl & SAN
MIGUEL Maxi 2003. Nonequilibrium transitions in complex networks:
A model of social interaction. *Phys. Rev. E* 67(2): 026120.

KOOTI Farshad & CHA Meeyoung & GUMMADI Khrishna P. & MASON
Winter 2012. The Emergence of Conventions in Online Social Networks,
*Proceedings of the 6th International AAAI Conference on Weblogs and
Social Media (ICWSM)*, Dublin, Ireland.

LAKOFF George 1987. Women, fire, and dangerous things: What categories
reveal about the mind. University of Chicago Press.

LAZER David et al. 2009. Computational Social Science. *Science* 323: 721–
723.

LONG F. & YANG Z. & PURVES D. 2006. Spectral statistics in natural
scenes predict hue, saturation, and brightness. *Proc. Natl. Acad. Sci. USA
(PNAS)* 103(15): 6013-6018.

LORETO Vittorio & STEELS Luc 2007. Social dynamics: the emergence of
language. *Nature Physics* 3:758–760.

LORETO Vittorio & MUKHERJEE Animesh & TRIA Francesca 2012. On
the origin of the hierarchy of color names. *Proc. Natl. Acad. Sci. USA
(PNAS)* xxx:xxxx (published ahead of print April 16).

LENAERTS Tom & JANSEN Bart & TULYS Karl & DE VYLDER Bart
2005. The evolutionary language game: An orthogonal approach. *J.
Theor. Bio.* 235(4): 566-582.

MÉZARD Marc & PARISI Giorgio & VIRASORO Miguel Angel 1987. Spin
glass theory and beyond. World Scientific lecture notes in physics.

World Scientific New York

MINETT James W. & WANG William S.Y. 2008. Modeling endangered languages: the effects of bilingualism and social structure. *Lingua* 118(1). 19-45.

MUKHERJEE Animesh, TRIA Francesca, BARONCHELLI Andrea and PUGLISI Andrea 2011. Aging in language dynamics. *PloS ONE* 6(2): 316677.

NOLFI Stefano & MIROLLI Marco Eds. 2009. Evolution of Communication and Language in Embodied Agents, Springer.

NOWAK Martin & KRAKAUER David 1999. The evolution of language. *Proc. Natl. Acad. Sci. USA (PNAS)* 96: 8028-8033.

NOWAK Martin & PLOTKIN Joshua B. & KRAKAUER David 1999b. The evolutionary language game. *J. Theor. Bio.* 200(2): 147-162.

NOWAK Martin & KRAKAUER David & DRESS Andreas 1999. An error limit for the evolution of language. *Proceedings of The Royal Society of London Series B, Biological Sciences* 266: 2131-2136.

PAOLACCI Gabriele & CHANDLER Jesse & PANAGIOTIS G. Ipeirotis 2010. Running experiments on amazon mechanical turk. *Judgment and Decision Making*, 5:411–419.

PUGLISI Andrea & BARONCHELLI A. & LORETO V. 2008. Cultural route to the emergence of linguistic categories. *Proc. Natl. Acad. Sci. USA (PNAS)* 105: 7936.

RATLIFF F 1976. On the psychophysiological bases of universal color terms.

Proceedings of the American Philosophical Society 120:311–330.

ROSCH Eleanor 1973. Natural categories. *Cognitive Psychology* 4:328-350.

SELTEN Reinhard & WARGLIEN Massimo 2007. The emergence of simple languages in an experimental coordination game. *Proc. Natl. Acad. Sci. USA (PNAS)* 104(18): 7361–7366.

STEELS Luc 1995. A self-organizing spatial vocabulary. *Artif. Life* 3(2): 319-332.

STEELS Luc 1996. Self-Organizing Vocabularies. Proceeding of the Fifth International Workshop on the Synthesis and simulation of Living Systems, C. Langton and T. Shimohara Eds., The MIT Press, Pages 179-184.

STEELS, Luc, 2000. Language as a Complex Adaptive System, in: M. Schoenauer (ed.): Proceedings of PPSN VI (Lecture Notes in Computer Science), Berlin (Germany).

SURI Siddharth & WATTS Duncan J. 2011. Cooperation and Contagion in Networked Public Goods Experiments. *PLoS ONE* 6(3): e16836.

TRIA Francesca & GALANTUCCI Bruno & LORETO Vittorio 2012. Naming a structured world: a cultural route to duality of patterning. *PLoS ONE* xx(x):xxx-xxx.

VON AHN Luis & DABBISH Laura 2004. Labeling images with a computer game. CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems, Vienna, Austria, pages 319–326.

WANG William S-Y. & MINETT James W. 2005. The invasion of language: emergence, change and death. *Trends Ecol. Evol* 20(5): 263-269.

WATTS Duncan J. & STROGATZ Steven H. 1998. Collective dynamics of 'small-world' networks. *Nature* 393: 440-442.

WHORF Benjamin Lee 1956. Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf. MIT Press.

WITTGENSTEIN, Ludwig, 1953. Philosophical Investigations. (Translated by Anscombe, G.E.M.) Basil Blackwell, Oxford, UK.