The Blending Game: emergence of duality of patterning in an interacting population.

Francesca Tria[1] & Vittorio Loreto[1,2]

May 22, 2012

[1]Institute for Scientific Interchange ISI
Via Alassio 11/c, 10126 Torino, Italy
[2]Sapienza University of Rome, Physics Dept.
Piazzale Aldo Moro 2, 00185, Roma, Italy

Abstract

Shared by all human languages and not found in any animal's communication systems, duality of patterning refers to the organization of the meaningful elements in a language (e.g., words) at two distinct levels. At a first combinatorial level, meaningless forms (referred to as phonemes) are combined into meaningful form. At a second compositional level, meaningful forms (referred to as morphemes) are composed into larger lexical units, the meaning of which is related to the meaning of the individual composing morphemes. The question remains wide open regarding how and why such a structure did emerge. Here we address this question in the framework of multi-agents modeling, where a population of individuals plays elementary language games aiming at a success in communication. We will focus on lexicon, though the compositional level involves syntax as well. We present in particular a Blending Game in which we let an initially blank slate population of individuals to bootstrap a lexicon using an open-ended set of forms in a large meaning space modeled as a graph. We do not make use of any notion of predefined linguistic categories, predefined meanings or relations between objects and forms, letting the graph to account for the semantic structure which organizes our conceptual understanding. The main result of this study is the identification of two main mechanisms leading to the emergence of duality of patterning. First, a minimal semantic mechanism for word coinage that exploits the structure of the world to compose individual forms into words. This mechanism feeds the combinatorial level of the lexicon and entails that part of the semantic structure percolates into the lexicon. Second, the existence of noise in the message transmission and the possibility for individuals of progressively increasing their level of comprehension each time they are exposed to each particular form. This mechanism accounts for the limited and low number of distinct forms present in the lexicon. These results demonstrate that the two sides of duality of patterning can emerge simultaneously as a consequence of a pure cultural dynamics in a simulated environment which contains meaningful relations, when a simple constraint on message transmission fidelity is also considered.

1 Introduction

A productive approach to the study of the emergence and evolution of language, that already led to a wealth of interesting results, is that of *language games* [1, 2]. The notion of language game dates back to Wittgenstein [1] and refers to communication acts and use of language of increasing complexity. In the language game approach one typically imagines a population of individuals who try to communicate about a topic through local, usually pairwise, interactions. As a consequence of these efforts, a communication system can be bootstrapped in a self-organized way, starting from a blank state condition for the individuals. The important point here is that no individuals can access the state of the population as a whole, neither can directly affect it, so that a shared agreement can emerge only trough repeated, local, interactions. From this perspective, a population of individuals can be viewed as a complex system and language as a complex emerging property. For a recent overview of the emerging field of Language Dynamics see [3].

The simplest, and thoroughly studied, example of language game is that of naming. The question in this case is: how a population of individuals can reach consensus on a common name for an object, without any central coordination? The *Naming Game* [4, 5, 6] dealt with this question, elucidating the conditions under which a common name is adopted by the whole population or, vice versa, a fragmentation appears, which are the time scales (as a function of the population size) over which agreement (or a fragmented stationary state) can be reached.

Language is of course much more complex than a simple process of naming. A second level of complexity is for instance tightly interlinked with the cognitive ability of categorization. A widely studied example in this direction is that of colors [7, 8, 9, 10, 11]. Colors can be adequately described in a three-dimensional and continuous space. For instance, the three coordinates can be chosen to be hue, saturation, and brightness. Despite the complexity of the definition of a color, a few basic color names, like red, green, blue, yellow, are typically sufficient to guarantee a good level of understanding. This means not only that a shared categorization is in place, but also that different individuals select very similar representatives for the same categories. In addition, color naming patterns exhibit structural regularities across cultures [12, 13]. The approach of language games allowed recently to elucidate many questions related to the emergence of agreement in naming colors [9, 11], the universality of the color categorization patterns across different cultures [14] as well as the origin of the hierarchy [12] of basic color names [15].

Further, languages exhibit complex syntactic and grammatical structures, and a major challenge of language dynamics is to account for their emergence [16]. Again, when syntax has to be considered, one cannot avoid to deal with 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 [17]. We make here a propaedeutic step towards the more complex goal of the study of the emergence of a complete syntax, considering the emergence of structures at the level of lexicon. In particular, we focus here on the so-called duality of patterning.

1.1 Duality of patterning

In a seminal paper, Charles Hockett [18] identified duality of patterning as one of the core design features of human language. A language exhibits duality of patterning when it is organized at two distinct levels [19, 20, 21]. At a first level, meaningless forms (typically referred to as phonemes) are combined into meaningful units (henceforth this property will be referred to as *combinatoriality*). For example, the English forms /k/, /a/, and /t/ are combined in different ways to obtain the three words /kat/, /akt/, and /tak/ (respectively written 'cat', 'act' and 'tack'). Because the individual forms in them are meaningless, these words have no relation in meaning in spite of being made of the same forms. This is a very important property, thanks to which all of the many words of the English lexicon can be obtained by relatively simple combinations of about forty phonemes. If phonemes had individual meaning, this degree of compactness would not be possible. At a second level, meaningful units (typically referred to as morphemes) are composed into larger units, the meaning of which is related to the individual meaning of the composing units (henceforth this property will be referred to as *compositionality*). For example, the meaning of the word 'boyfriend' is related to the meaning of the words 'boy' and 'friend' which composed it. The compositional level includes syntax as well. For example, the meaning of the sentence 'cats eat fishes' is related to the meaning of the words 'cats', 'eat', and 'fishes'. In this paper, for the sake of simplicity, we focus exclusively on the lexicon level. This has to be considered as a first step towards the comprehension of the emergence of complex structures in languages.

It has to be noted that duality of patterning is not a necessity, as lexicons could be organized exclusively at the combinatorial level. If this were the case, form similarities would very rarely correspond to similarities in meaning (and vice versa). However, lexicons of this kind do not exist and systematic relations between form and meaning are widespread. For example, the process through which the word 'bedroom' is created is fairly productive in English (e.g., 'bathroom', 'bedtime', etc.) as well as in many other languages. Indeed, even if languages vary considerably with respect to the extent to which they manifest composition at the lexical level, even languages in which composition is used the least - such as Classical Chinese [22] - have a great number of multimorphemic words and, in this sense, fully exhibit lexical duality of patterning. Hockett further pointed out that duality of patterning is not independent of productivity, listed as another core feature of human languages. Productivity is the ability of human beings of say things that are never been said and, equally importantly, to understand them. Productivity is in turn related to blending, that is the ability of "coining new utterances by putting together pieces familiar from old utterances, assembling them by patterns of arrangement also familiar in old utterances" [18]. Again, blending is present both at the lexicon and at the syntactic levels.

In the following, we will focus on the mechanisms that could lead to the establishment of duality of patterning in a lexicon. To be sure, there have been a number of previous works devoted to explain the emergence of combinatoriality and compositionality. A thorough review of the attempts presented in literature is far from the scope of the present paper, and to our knowledge no self-contained and critical collections of works in this area exist, though the community at large would greatly benefit of such a piece of information. Here we shall only focus on a few aspects which are relevant for our purposes.

It should be remarked that the two facets of duality of patterning have often been studied independently from each other. For instance, the origin of combinatoriality in speech sounds has been investigated in [23, 24], and in [25, 26] in an evolutionary perspective. Separate modeling efforts have been devoted to find evolutive advantages for the emergence of combinatoriality [27] and compositionality [28]. Emergence of compositionality at the sentence level have been studied for instance in [17], where it is shown how a structured language, as opposed to an holistic one, can emerge in a population of communicating individuals, when the scenes to be described feature themselves a structure. In [29] the co-evolution of compositionality and regularity is investigated through the simulation of a cultural process in which syntactic categories are gradually formed which mirror a set of predefined semantic categories.

One notable exception was presented in [30] where a recurrent neural network approach was used to model the emergence of syntax. In this case both combinatoriality and compositionality were addressed within the same framework. However, combinatoriality was not really an emergent property in Batali's work, the starting point being a set of four possible symbols to compose words: in this sense, combinatoriality was somehow an assumption of the model. Nevertheless, it was interesting the result that words shorten during the communication rounds thanks to an economy principle the individuals are supposed to follow. Batali also considered compositionality, showing that the emergent lexicon exhibits excess similarity between words corresponding to related meanings and something similar to a grammatical structure does emerge. In this case the emerging lexicon turned out to be always compositional, relying on a clear structure in the meanings space, that was divided into two categories. In particular Batali considers a predefined set of meaning categories, predicates and referents, each coded with specific entries of the meaning vector. This leads to a finite predefined set of meaning agents could play with.

It should also be remarked that often studies in this ares have been focused on evolutionary times scales (e.g., [31, 32, 33, 34, 35]), disregarding in this way the peer-to-peer negotiation taking place on cultural time-scales in large populations. In contrast, there is evidence suggesting that humans are capable of evolving languages with duality of patterning in the course of only one or two generations (consider for instance Nicaraguan Sign Language [36, 37] or the emergence of Pidgins and Creole languages [38]).

In summary, we believe that all the works discussed above have shown the way forward to the understanding of the emergence of linguistic units composed by a blending of sub-units in human lexicons. Still some important further steps remained to be done and this is the aim of the work we present here. First of all we aim at explaining in an unitary framework the co-emergence of combinatoriality and compositionality. In addition, unlike previous approaches that looked for the emergence of meaning-symbols compositional mappings out of a small bounded set of predefined symbols available to the population, our approach adopts an open-ended set of forms and it does not rely on any predefined relations between objects/meanings and symbols. For instance we require combinatoriality to emerge out of a virtually infinite set of forms which are freely provided to a blank slate of individuals. Such set can only be limited by means of self-organization through repeated language games, the only purpose being that of communication. In addition, with our simple representation of the conceptual space, modeled as a graph, we do not hypothesize any predefined linguistic category or predefined meaning. This choice also allows to model the effect of differently shaped conceptual spaces and of conceptual spaces that may differ from individual to individual.

In order to address these issues we introduce a general modeling framework where the question of the emergence of lexicons featuring duality of patterning is addressed in a self-consistent way. We consider an initially *blank slate* population of individuals committed to bootstrapping a lexicon using an open-ended set of forms in a large conceptual space modeled as a graph. We show in particular that errors in communication as well as a blending repair strategy, sometime adopted by individuals when communication fails, can account for the emergence of compositional as well as combinatorial structures in the emerging lexicon, demonstrating in this way that duality of patterning can emerge via simple cultural processes. It is important to remark that, while duality of patterning is an emergent property of language at the population level, the framework we describe here reflects basic cognitive abilities of individuals and as such it is not necessarily able to bias the outcome of the evolution towards duality of patterning. Purely non-combinatorial or combinatorial but non-compositional lexicons can always emerge within our framework. Nonetheless, we show that, over a range of conditions, the pressure for communication at a population level leads to the emergence of compositional as well as combinatorial structures in the emerging lexicon.

The outline of the paper is as follows. In section 2 we introduce our modelling scheme, dubbed Blending Game. Section 3 reports the analysis of the properties of the emerging lexicon. In section 4 we discuss the role of different topologies of the conceptual space in order to test the robustness of our simulations with respect to the structure of the underlying network. In section 5 we consider the more general and perhaps more realistic case in which the conceptual space is not identical for each individual. We finally draw some conclusions in section 6.

2 The Blending Game

We here consider a population of N artificial agents committed to name M objects. We consider the objects to be named as nodes of a conceptual non-directional graph, where links represent conceptual relations between pairs of objects. We adopt the representation of the conceptual space as a graph [39, 40, 41, 42] as the simplest and most general way of introducing semantics, if we don't want to rely on predefined conceptual categorization. A precise and complete description of such a conceptual space being out of reach, we make very general hypotheses about the structure of the graph we use to model it, and we further check the robustness of our results with respect to different plausible choices.

For simplicity, we consider here a population of agents with an homogeneous structure, where each agent (individual) has the same probability of interacting with everyone else.

Starting from scratch and without any central coordination, the individuals perform pairwise language games aimed at naming the objects in their world. Each individual is characterized by its inventory or memory, i.e., a collection of lists of name-object associations that are empty at the beginning of the process and evolve dynamically as time progresses. As already introduced in language games [1] devoted to Naming [4, 5, 6] and Category formation [11, 14], at each time step two agents are randomly selected, one to play as Speaker (S), the other one as Hearer (H). S randomly chooses an object, named Topic (T), to discuss, his goal being that of uttering a *word* which enables H to correctly identify T.

In order to study the emergence of duality of patterning, we consider the possibility for any word to be composed either of a single form or of an ordered linear concatenation of forms[1]. The emergence of words composed of more than one form arises in a natural way through a mechanism of blending that acts as a repair strategy.

The main ingredients of the Blending Game are two. First, we introduce noise in comprehension and this applies to all games whenever a Hearer tries to understands a word uttered by a Speaker. This is an essential ingredient responsible for keeping the number of different forms shared by the whole population limited and low, without any a priori constraints on it. Second, along with the basic strategy of word creation, the game features a repair strategy that exploits the structure of the world to create new words. Sometimes the blending is independent of the meaning of these words, feeding the combinatorial level of the lexicon. On other occasions, blending involves words which are usually taken from the inventories of related objects (that is, objects connected by a link to the current topic), thus feeding the compositional level of the lexicon. Thanks to this compositional blending, part of the semantic structure which organizes the conceptual understanding of the world percolates into the lexicon.

2.1 Elementary interaction

S has to name the topic T.

(i) If S does not have already a word for it, she invents a word consisting of a single form which is novel for the entire population. We note that the constraint for the invented word to consist of a brand new form for the whole population corresponds to a virtually infinite, i.e., open-ended, repertoire of symbols. We made this (perhaps extreme) choice in order not to impose any a priori limit to the number of

[1] Here a form is intended as the meaningless unit of a signal.

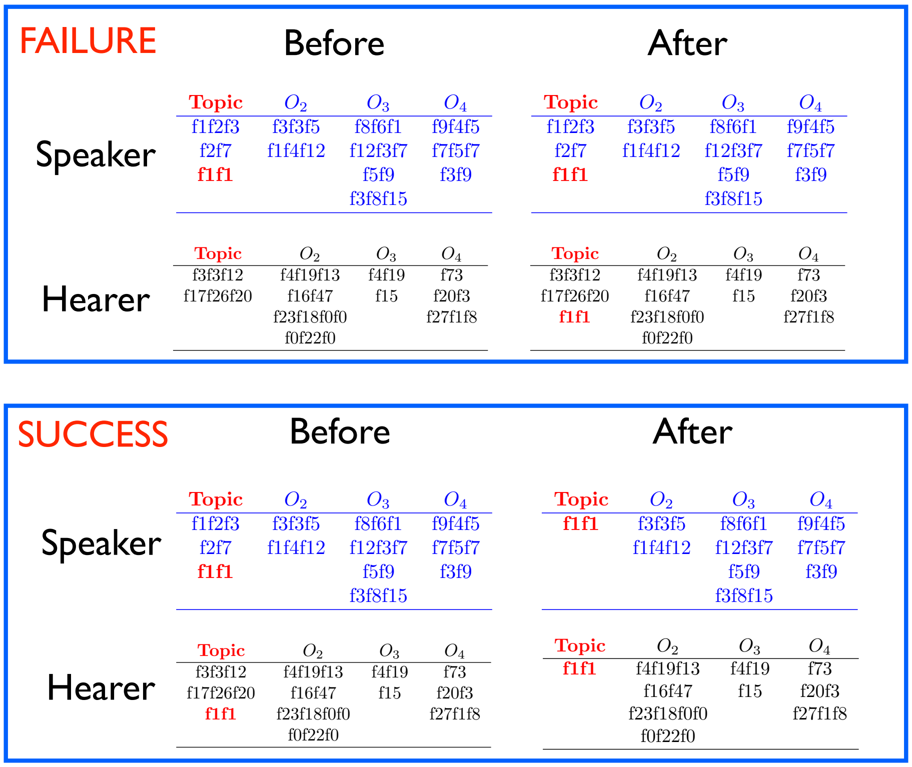

Figure 1: *Examples of Games. Top. Example of a failed game.* In this game the Speaker S selects the Topic and decides to utter the word *f1f1*. This word is unknown to the Hearer H since *f1f1* is not present in any of H's inventories. In this case the game is a failure and H adds the word *f1f1* to her inventory for the Topic. *Bottom. Example of a successful game.* In this game the Speaker S selects the Topic and decides to ls utter the word *f1f1*. This word is known to the Hearer H since *f1f1* is present in H's inventory for the Topic. In this case the game is a success and both S and H remove from the Topic's inventory all the competing words but *f1f1*.

possible symbols in the lexicon. We are interested in investigating under which conditions, even in this scenario, only a very limited number of different forms gets eventually fixed in the population's lexicon.

(ii) If S already possesses one or more words for T in its inventory, she chooses the last winning-word, if it exists (i.e., the word that allowed the last successful communication about T), or a random word otherwise.

H has to guess T.

As an essential ingredient of the model, we consider an initial imperfect understanding due to noise in communication. The guessing procedure consists thus of two parts: H has first to understand the word uttered by S, then she has to associate an object to that.

(i) H parses the heard word into its component forms, understanding independently each of them (for instance, if S utters the word $f_2 f_{16} f_1$, H tries to understand separately the forms f_2, f_{16} and f_1, recomposing again the word only at the end of the procedure). More precisely, H correctly understands each form with a time dependent probability:

$$P_t(f) = \left(1 - \exp\left(-\frac{n_t(f_i)}{\tau}\right)\right) \tag{1}$$

where $n_t(f_i)$ is the number of times H has heard the form f_i up to the current time t, and τ is a characteristic memory scale. If H does not understand a form, she replaces it with a random form from her inventory, i.e., a random form among all those composing any word she used to name any object. If her inventory is still empty, she replaces the form not understood with a brand new one, i.e., an invented form. In the end, the word understood by H is thus composed by the same number of forms (and in the same order) as in the uttered word, and it is equal to the uttered word except, possibly, for the misunderstood forms.

(ii) H checks whether she has already the word understood in her inventory, associated to any object. If so, H selects the corresponding object and if the word is associated with more than one object she chooses at random among them.

If H correctly guesses T, the game is a *success* and both agents delete all but the winning word from their inventories for T. If this is not the case, the game is a *failure* and H updates her inventory for T by adding the word understood (see figure 1).

At odds with the standard Naming Game, in case of failure, here S has a second (and last for the current interaction) chance to communicate about T, adopting a *blending repair strategy.*

S names the topic T by adopting the blending repair strategy.

S chooses two objects in the world. If T has at least two neighbors with non-empty inventories for S, she chooses among them, otherwise she considers two random objects. Having chosen the two objects, S chooses randomly a word from the inventory of each of them (for instance $f_{12}f_{18}f_2$ from the first object and $f_{13}f_1f_{22}f_3$ from the second). She then composes the new word for T by taking the initial part (of random arbitrary length) from one (e.g., $f_{12}f_{18}$) and the final part (again of random arbitrary length) from the other (e.g., f_3), keeping the order of the parent words in the composed word (in our example she would utter $f_{12}f_{18}f_3$).

When the blending repair strategy is adopted, the same communication and understanding procedure as before is considered and the game is declared a *success* or a *failure* with the same criteria as in the first communication attempt. Thus, in case of success both agents delete all their words from their inventories relative to T, but the winning one. In case of failure, H inserts also this second word, as she understands it, in its inventory for the topic T.

Figure 2 synthetically summarizes the structure of the game.

3 Properties of the emerged lexicon

All the results we present below are obtained with a conceptual space modelled as a Erdős -Rényi (ER) random graph [43, 44]. We will later see that the main results are not affected, both from a

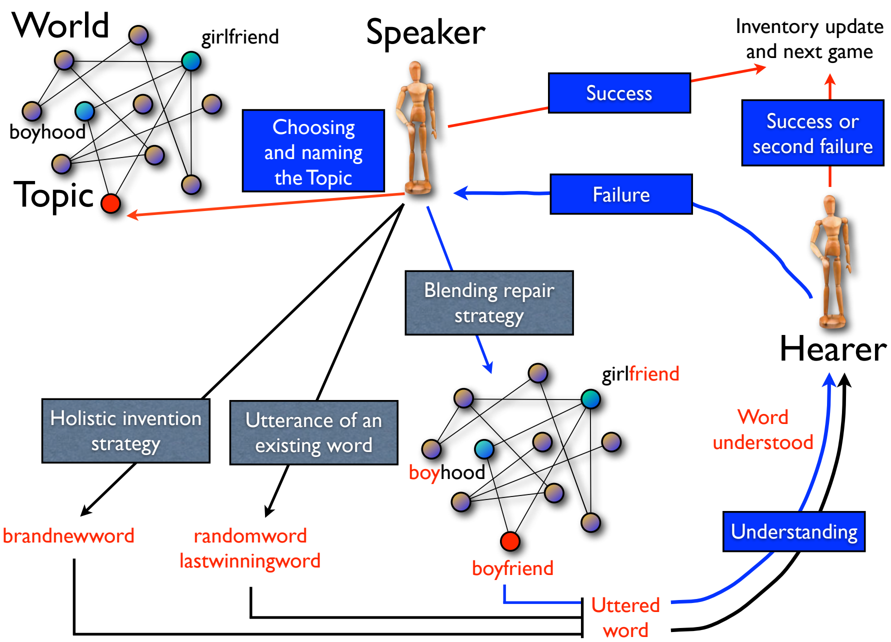

Figure 2: *Diagram of the Blending Game.* In each game a Speaker and a Hearer are chosen. The Speaker chooses a Topic and utters a name for it (i.e., it produces one or more meaningless forms, consisting of unique identifiers, e.g., f_0, f_1, etc.). If her inventory for the Topic is empty she invents a brand new word, i.e., a word never appeared in the whole population for all the possible objects. Alternatively, she chooses one of the existing word (the last-winning one). The hearer has to understand the uttered word and to relate it to an object on the basis of the content of her previously recorded information (her repertoire). The understanding phase is ruled by the parameter τ, a characteristic memory scale introduced through eq. 1. If the communication fails the Speaker adopts a blending repair strategy which consists in inventing a new word composed by reusing parts of already proposed words, preferably associated with object nearest neighbors of the chosen topic. In the example depicted, *boyfriend* is composed out of *boyhood* and *girlfriend*. If also after the adoption of the repair strategy the communication fails, the Hearer updates her inventory with the word just heard. If the game is a success, the inventories of both the Speaker and the Hearer for the Topic are cleared except for the successful word.

qualitative and quantitative point of view, by the particular structure of the conceptual graph. We will report examples by using: (i) uncorrelated random scale-free (UCM) graphs [45]; (ii) an experimentally derived word-association graph, namely the South Florida Free Association Norms [46]. We will give below details on both the structures.

All the results that follow are averaged over different (100) realizations of the process on the same conceptual graph. We checked that a further average on different graphs with the same statistical properties (e.g., ER random graphs with the same link probability p_{link}) produces fluctuations that remain within the reported error bars.

3.1 Success rate, synonymy and homonymy

Let us considering the dynamics of the Blending Game by looking at the time evolution of the success rate and of the homonymy. The communicative success starts from zero and progressively increases, leading eventually to a fully successful shared communication system with an "S"-shaped time behaviour. At the same time, both homonymy is defeated after a transient phase in which it spreads widely.

It is interesting to note the role of the blending mechanism in defeating homonymy. Let us consider the standard Naming Game with more than one object, in which invention always happen with a brand new word, that is a word never used before by any agent to name any object. In this case homonymy cannot be created and the inventories corresponding to different objects are totally uncorrelated. Let us now introduce in this dynamics noise in communication. We denote this model Naming Game with noise in communication. When the Hearer does not correctly understand an heard form, she can borrow a form from any of her inventories. In this way the inventories associated to different objects starts to correlate. In this case, homonymy can arise, and if it gets fixed in the whole population, i.e., all the agents have the same unique name for more than one object, it will last forever. At the same time, success cannot be rigorously one, since a finite probability of misunderstanding exists, due to this ambiguity generated by the homonymy. On the contrary, the blending mechanism provides a

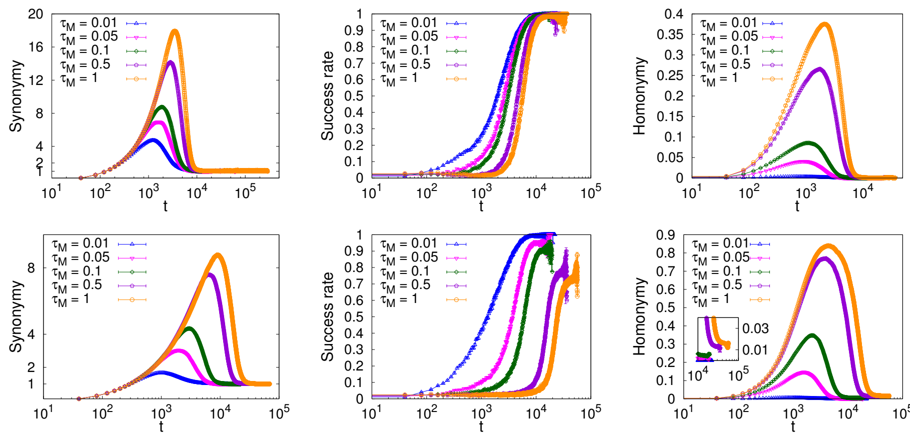

Figure 3: From left to right: *Homonymy, success rate and synonymy as a function of time* (number of games), averaged over sliding time windows, for the Blending Game (top) and the Naming Game with noise in communication (bottom). Homonymy is here defined as the number of pairs of objects that have at least an associated word in common in an individual inventory, divided by the number of pairs $M(M-1)/2$ and averaged over all the agents. Synonymy is defined as the average number of words associated to each object and averaged over all the agents. We define here a normalized link probability for the ER random graph as $p_M = \frac{p_{link}}{p^*}$, where $p^* = \frac{\log M}{M}$ is the threshold above which the whole graph in connected with probability one in the infinite size limit ($M \to +\infty$). Similarly, a normalized time scale parameter is defined as $\tau_M = \frac{\tau}{M}$. Results are reported for $p_M = 0.5$ and for different values of the learning parameter τ_M. The number of agents and the number of objects in the environment are fixed respectively to $N = 10$ and $M = 40$ and results are averaged over 100 realizations of the dynamics on the same objects graph (see main text for further details).

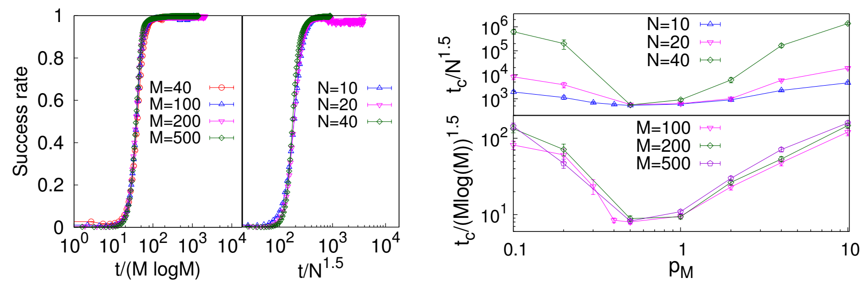

Figure 4: *Success rate and convergence time for different values of the conceptual space and population sizes. Left*: The success rate is reported as a function of time (number of games) for different values of M (left) and N (right), with $p_M = 1$ and $\tau_M = 1$. The curves for different values of M collapse if time is rescaled with a factor $M \log M$, while the curves for different values of N collapse if time is rescaled with a factor $N^{1.5}$, as in the original Naming Game [5]. *Right*: Time needed for the system to reach convergence for different values of M (bottom) and for different values of N (top), as a function of p_M and with $\tau_M = 1$. We rescaled the curves in order to superimpose them on the values of p_M where the convergence time is minimal.

way to repair homonymy. In case of failure in understanding, in fact, the Speaker can invent a new word through the blending procedure. In this way homonymy can be defeated and agents can achieve a perfect success in communication. In figure 3 (left) we report the time evolution of homonymy both for the Blending Game (top) and for the Naming Game with noise in communication (bottom). As for the communicative success, it starts from zero and progressively increases, leading eventually to a fully successful shared communication system with an "S"-shaped time evolution (figure 3, center). Synonymy (figure 3, right) follows a time behaviour very similar to that of homonymy, with the difference that it is defeated both in the Blending Game and in the Naming Game with noise in communication. In both cases the population reaches consensus on naming the different objects.

In figure 4 we show how the time evolution of the success rate scales with the number of objects M and with the population size N. We look both to the scaling of the onset of success in communication (the vertical part of the "S"-shaped curve), corresponding to the peak in the homonymy and synonymy curves, and to the scaling of the convergence time t_c, i.e., the time needed by the population

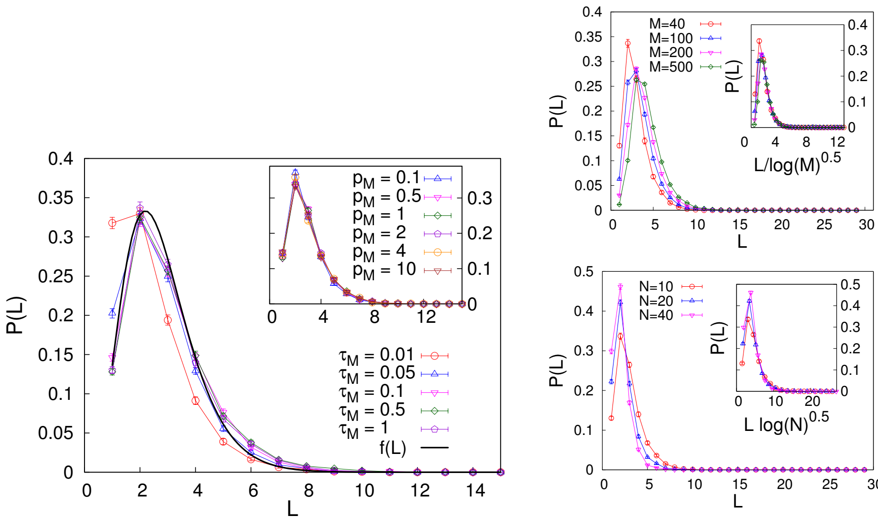

Figure 5: *Left*: in the main figure the distribution of word length L for different τ values is reported, fixing $p_M = 0.5$. $f(L) = aL^b c^L$ is the observed empirical function [47] (solid black line, with fitting parameters $a = 0.74$, $b = 3.7$, $c = 0.18$). In the top inset we show the same distribution fixing $\tau_M = 1$ for different values of p_M. Note that here the curves overlap, indicating that the word length distribution does not depend on the objects graph connectivity. The number of agents and the number of objects in the environment are fixed respectively to $N = 10$ and $M = 40$. *Right*: the same distribution for different values of M (*top*) and for different values of N (*bottom*), with $p_M = 1$ and $\tau_M = 1$. In the inset of the two figures we report the collapse of the distributions by rescaling the words length L respectively by $\log(M)^{0.5}$ and by $\log(N)^{0.5}$. This indicates a very weak dependence of the average word-length both on the size M of the conceptual space and on the population size N.

to reach consensus on naming all the M objects. Let us call t_o the time at which the transition from a success rate close to zero to a success rate close to one occurs. We find $t_o(M) \sim M \log(M)$, while we recover the same behaviour as in the Naming Game for the scaling with the population size: $t_o(N) \sim N^{1.5}$. The convergence time t_c features a stronger dependence both on M ($t_c \sim (M \log(M))^{1.5}$) and on N. Interestingly, in the latter case, the same dependence as in the Naming Game ($t_c(N) \sim N^{1.5}$) is recovered only for intermediate values of the link probability p_{link}. We will see that these values of p_{link} correspond to an intermediate level of structure in the conceptual space, leading to the emergence of compositionality in the emerging lexicon.

3.2 Words length distribution

Due to the blending procedure, the lexicon shared by the population contains words that are composed by several forms. A typical word in the lexicon is for instance $f_{23}f_{18}f_0f_0$, composed by the elementary forms f_{23}, f_{18} and f_0. We here consider the length of a word as the number of forms composing it. In figure 5, the words length distribution in the emerging lexicon is reported. We observe that the observed limitation on the word length is an outcome of the negotiation dynamics, emerging without any explicit constraints on it. The distribution features a typical shape that has been observed in human languages [47], well fitted by the function $f(x) = ax^bc^x$, which corresponds to the Gamma distribution when the parameters are suitably renamed [47]. Figure 5 (left) shows the word length distribution for different values of the rescaled memory parameter $\tau_M = \tau/M$ and of the rescaled graph connectivity p_M (see the caption of figure 5 for the definition). While the distribution is not affected by the graph connectivity, a very light dependence on τ_M is observed. Further, the peak of the histogram moves very slowly when changing the number M of objects to be named (figure 5, top right) or the number N of agents in the population (figure 5, bottom right). The average word length in the emerging lexicon is thus very stable when changing the parameters of the model, remaining finite and small, and comparable with the length of words in human languages.

3.3 Frequency-rank distribution of elementary forms

As a deeper investigation of the properties of the emerged lexicon, we consider the frequency-rank distribution of the different forms composing the words (figure 6). As in the case of the word length distribution, the frequency-rank distribution for forms does not depend on p_{link} (see figure 6 top right). However, we note in this case a clear dependence on the memory parameter τ_M. In particular, the higher τ (for M fixed), i.e., the lower the fidelity in communication, the smaller the number of distinct forms on which the agents eventually find agreement. Since the invention rate of new forms does not depend on τ, the effect of a high τ is that of strengthening the selection process, reducing in this way the number of forms that get fixed in the population. The dependence of the frequency-rank

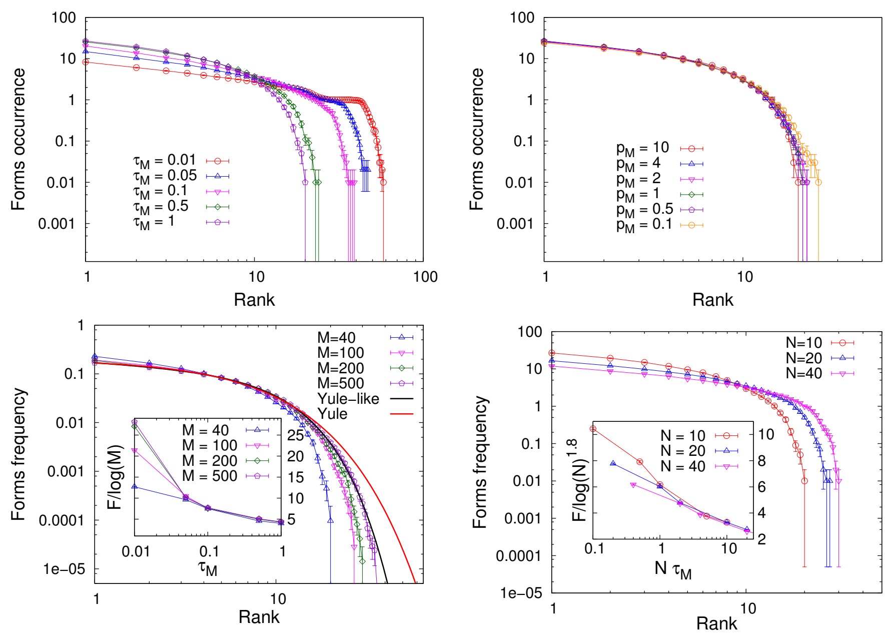

Figure 6: *Top left*: frequency-rank distribution for elementary forms is shown for different values of the parameter τ_M keeping fixed $p_M = 0.5$, $N = 10$ and $M = 40$. *Top right*: the same distribution fixing $\tau_M = 1$ and for different p_M, showing that the distribution of elementary forms does not depend on the objects graph connectivity. *Bottom left*: frequency-rank distribution for elementary forms for different values of M, with $p_M = 1$ and $\tau_M = 1$. In the inset we show the number of distinct forms F in the emerged lexicons as a function of τ_M and for $p_M = 1$, rescaling the curves in order to let them overlap for high value of τ_M. When τ_M is not high enough, the noise ceases to be relevant and both the absolute number of distinct forms increases and the dependence on M becomes stronger. We also report two fits obtained with a Yule distribution (red solid line) $f(R) = aR^b c^R$, which has been hypothesized to reproduce the actual distribution in human languages [48], and a Yule-like (black solid line) distribution, $f(R) = \alpha \exp -\beta R^\gamma R^\delta$ [49]. The two functions coincide for $\gamma = 1$. The Yule function seems to fit better and better our distributions as M increases, while a finite-size distribution is better fitted by a Yule-like distribution. *Bottom right*: frequency-rank distributions for elementary forms for different values of N, again with $p_M = 1$ and $\tau_M = 1$. In the inset again the number of distinct forms F in the emerged lexicons as a function of τ_M and for $p_M = 1$, again rescaled in order to overlap for high values of τ_M.

distribution, as well as of the number of distinct forms (F) in the emerged lexicon, on the number of objects M and on the population size N is very weak (figure 6, bottom), again, as in the case of the words length distribution, pointing to the possibility of comparing the obtained results with human languages properties. Indeed, it is worth noticing that for large values of τ, the frequency-rank distribution we observe is remarkably similar to the corresponding distribution observed in human languages [48] for which a Yule-like distribution [49] has been hypothesized (see figure 6, bottom left, and the figure caption).

3.4 Combinatoriality

We now introduce a measure of combinatoriality to quantify the property of a communication system to combine and re-use a small number of elementary forms to produce a large number of words. Following the idea in [50], we introduce a real-valued quantity ranging in the interval $[0:1]$ that quantifies how frequently forms recur in the emerged lexicon, according to the following formula:

$$C = \frac{\sum_i (m(f_i) - 1)}{(M-1)F}\,, \tag{2}$$

where the sum runs over all the F distinct forms present in the emerged lexicon and $m(f_i)$ is the number of distinct objects whose name includes the form f_i. The term $m(f_i) - 1$ takes into account only the forms that are used to name at least two objects, i.e., only the forms that are actually re-used. M is the number of objects to be named. The results for the combinatoriality are reported in figure 7 (left) as a function of τ_M and for different values of p_M. Again, a negligible dependence on p_{link} is found, while, as in the case of the frequency-rank distribution, a clear dependence on τ_M is found, the maximal combinatoriality occurring for high values of τ_M. This can be understood if one thinks that for a perfect level of understanding there is no selective pressure acting on the different forms and many distinct forms are eventually fixed in the lexicon with a small re-use rate, i.e., little combinatoriality. In a sense, the limit of small τ_M is the *holistic* limit of our model, i.e., the limit in which forms stand for the meaning as a whole and have no meaningful subparts. In our case the word holistic refers both to

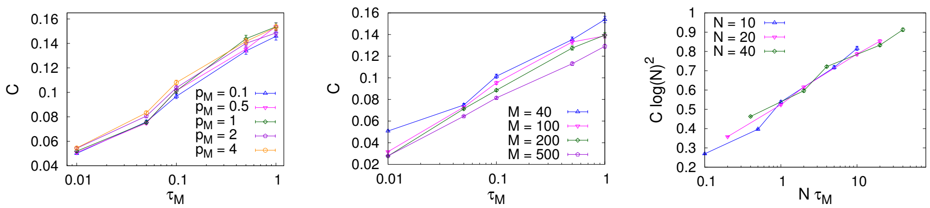

Figure 7: *Left*: combinatoriality C (see the text for definition) for different values of p_M as a function of τ_M. The number of agents and the number of objects in the environment are fixed respectively to $N = 10$ and $M = 40$. *Center*: combinatoriality C for different values of M as a function of τ_M and for $p_M = 0.5$. Although the curves are very noisy, the value of the combinatoriality is very weakly dependent on M. *Right*: combinatoriality C for different values of N as a function of τ_M, rescaled with the number N of agents, and for $p_M = 0.5$.

the case where every word of the lexicon is composed by a single form or cases where several forms are needed for a word but none of them is re-used in the whole lexicon. Summarizing, when the effort for understanding new forms is sufficiently high, one finds, at the lexical level, features similar to the ones observed in real languages, such as the word length distribution and the number and the frequency of use of the different forms. In this perspective, combinatoriality emerges as a workaround to overcome the problem of noisy communication. In figure 7 (center and right) we also show the dependence of the combinatoriality on the number of objects M and on the population size N, again highlighting an extremely weak dependence.

3.5 Compositionality

Let us now turn to the compositional aspects of the lexicon, the aim here being that of establishing whether, in the emerged lexicon, words for semantically related concept are expressed through morphologically similar words.

Here we measure the semantic similarity of two objects in terms of their distance on the graph describing the conceptual environment. In addition, we need to define a measure of morphological similarity between words. To this end we introduce a Master-Mind-like (MM) measure. Given two words w_1 and w_2, each composed of a certain number of forms, the Master-Mind-like (MM) measure of form similarity is defined as follows: after the two words have been aligned, either making the left-end or the right-end coincide, we sum 1 whenever the two words share the same form in the same position and 0.5 for the same form in different positions. The MM measure will be the maximum between the left-end and the right-end alignments. The MM measure conveys the idea that meaningful forms are often included in words as a suffix or a prefix, and in general in a well defined position. However, the results obtained turn out to be stable against different measures of words similarity (refer to [51]).

As a measure of compositionality, we measure the *excess similarity* of words used to name related objects (at low distance in the graph) when compared to the similarity of randomly chosen words. In order to do that, we consider the average difference between the similarity between each pair of words as a function of the distance of the corresponding objects in the conceptual graph, and the same value computed in the random case, obtained by reshuffling the associations between words and objects. In figure 8 (left), we report the excess similarity for a fixed value of τ_M and several values of p_M as a function of the topological distance d on the conceptual graph. The inset reports the same measure for a fixed value of p_M and different values of τ_M. Compositionality is evident in the figure: the more closely related the words, the higher the excess similarity. At odds with the above studied properties of the lexicon, the excess similarity only weakly depends on τ_M, while strongly depends on p_M. This indicates that a percolation of the organization of the world into the lexicon is possible when the world has a non-trivial semantic structure, i.e., in our case when p_{link} is different from zero and from one. In the former case no relation between objects exists, while in the latter case all the objects are equally related (all are at distance one in the graph). Diluted graphs are more prone to induce compositionality. As for combinatoriality, the dependence of compositionality on the size M of the conceptual space and on the number N of agents in the population is extremely weak (figure 8 right, top and bottom).

3.6 Summary of the results

We analyzed the main properties of the emerged lexicon as functions of the two parameters of the model, the graph connectivity

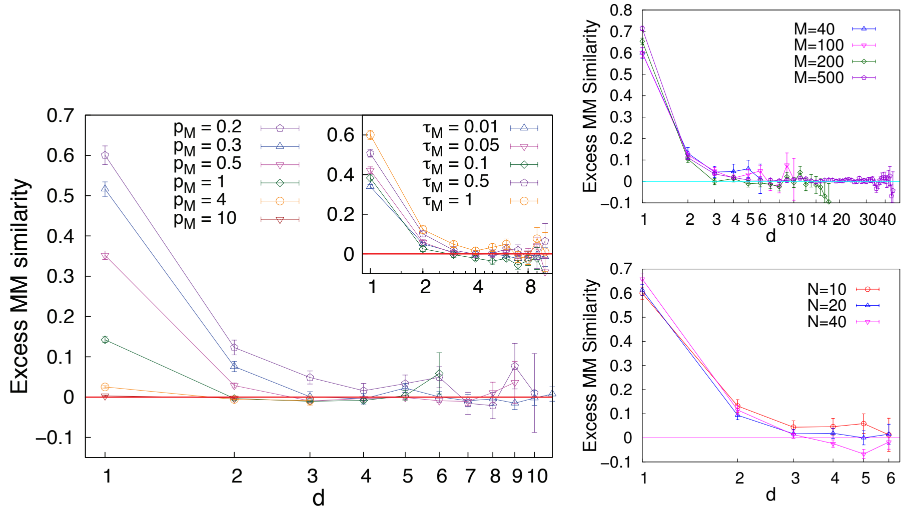

Figure 8: *Excess Master-Mind-like similarity of words as a function of the distance d of the corresponding objects on the graph.* The difference between the actual similarity and its random value computed on a reshuffled communication system is shown (see text for details). A decrease in the excess similarity as a function of the topological distance d is the signature of the emergence of compositionality; in particular, compositionality implies higher similarity among words which are closer in the semantic space. The topological distance on the object graph is our proxy for the semantic relatedness. The MM similarity (see text) is adopted here to compute similarity between words. *Left*: results are reported for $N = 10$ and $M = 100$. Results are shown for different values of the objects graph connectivity p_M, keeping fixed $\tau_M = 1$ (main figure) and for different values of τ_M keeping fixed $p_M = 0.2$ (inset). *Right*: results for different values of M (top) and for different values of N (bottom), for $\tau_M = 1$ and $p_M = 0.2$. Again, the results depend very weakly both on M and on N, the excess similarity becoming slightly more pronounced as M and N increase.

p_{link} and the memory scale τ, as well as a function of the number of objects M on the conceptual space and the number of agents N in the population. We found that properties of the lexicon related to the combinatoriality, namely the words length distribution, the frequency of use of the different forms and a measure for the combinatoriality itself, reflect both qualitatively and quantitatively the corresponding properties as measured in human languages, provided that the memory parameter τ is sufficiently high, that is that a sufficiently high effort is required in order to understand and learn brand new forms. Conversely, the compositional properties of the lexicon are related to the parameter p_{link}, that is a measure of the level of structure of the conceptual graph. For intermediate and low values of p_{link}, semantic relations between objects are more differentiated with respect to the situation of a more dense graph, in which every object is related to anyone else, and compositionality is enhanced. In summary, while the graph connectivity strongly affects the compositionality of the lexicon, noise in communication strongly affects the combinatoriality of the lexicon.

4 Role of the topology of the conceptual space

We here introduce networks with different topologies in order to test the robustness of our simulations with respect to the structure of the underlying network.

4.1 Uncorrelated random scale-free networks

Each node i of a network is first characterized by its degree k_i (number of links) and a first characterization of the network properties is obtained by the statistical distributions of the nodes' degree, $P(k)$. In order to quantify the topological correlations in a network, two main quantities are usually measured. The clustering coefficient c_i of a node i measures the local cohesiveness around this node [52]. It is defined as the ratio of the number of links between the k_i neighbors of i and the maximum number of such links, $k_i(k_i-1)/2$. The clustering spectrum measures the average clustering coefficient of nodes of degree k, according to

$$C(k) = \frac{1}{N_k} \sum_i \delta_{k,k_i} c_i \, . \tag{3}$$

Moreover, correlations between the degrees of neighboring nodes are conveniently measured by the average nearest neighbors degree of a vertex i, $k_{nn,i} = \frac{1}{k_i} \sum_{j \in \mathcal{V}(i)} k_j$, where $\mathcal{V}(i)$ is the set of neighbors of i, and the average degree of the nearest neighbors, $k_{nn}(k)$, for vertices of degree k [53]

$$k_{nn}(k) = \frac{1}{N_k} \sum_i \delta_{k,k_i} k_{nn,i}. \tag{4}$$

In the absence of correlations between degrees of neighboring vertices, $k_{nn}(k)$ is a constant. An increasing behavior of $k_{nn}(k)$ corresponds to the fact that vertices with high degree have a larger probability of being connected with large degree vertices (assortative mixing). On the contrary, a decreasing behavior of $k_{nn}(k)$ defines a disassortative mixing, in the sense that high degree vertices have a majority of neighbors with low degree, while the opposite holds for low degree vertices [54].

The results presented in the previous sections were obtained by considering the homogeneous Erdős -Rényi random graph [43, 44], in which nodes are linked with a uniform probability p_{link}. In this case, the graph features a small diameter and a small clustering coefficient, and the degree distribution is homogeneous and binomial. The specific properties of the graph depend on p_{link}. In particular if M is the number of nodes, for $p_{link} > log(M)/M$ the graph will almost surely be connected.

We consider now a random scale-free network obtained from the uncorrelated configuration model (UCM) [45]. UCM graphs present a broad degree distribution $P(k) \sim k^{-\gamma}$ and are constructed in such a way to avoid two- and three-vertex correlations, as measured by the average degree of the nearest neighbors $k_{nn}(k)$ and the clustering coefficient of the vertices of degree k, respectively. The average degree distribution is finite for the values of the exponent $\gamma > 2$, and the second moment of the distribution in finite for $\gamma > 3$. We consider here two values for the degree distribution exponent, one below and one above the latter threshold: $\gamma = 2.5$ and $\gamma = 3.5$. We find that all the considered observables do not depend on the value of γ and, moreover, show the same qualitative (and in most cases quantitative) behaviour observed for an homogeneous Erdős -Rényi random graph with the same number M of nodes (see Figure 9).

4.2 South Florida Free Association Norms

In this section we present the results obtained when considering a conceptual graph taken from a real-world experiment, namely the South Florida Free Associations Norms [46]. South Florida is the outcome of a great effort started back in 1973 and lasted almost thirty years. The persons taking part to the experiment were presented with input words ('cues') and had to give another word as answer, following a free association. The issued word is called 'target'. The database consists of roughly 5000 words and 700000 associations. Each dataset yields a graph whose nodes are the words and whose edges correspond to the associations between words made by the players/subjects. Each edge is moreover weighted by the number of times that the corresponding association has been made. The network formed by these associations, the Word Association Graph (WAG), is directed (one target word being an answer to a cue, the links are obviously directed from cue to target) [55]. In this experiment we consider the South Florida Word Association Graph as a proxy of the conceptual graph. In doing this we disregard the directed character of the original graph as well as its weights. Figure 10 reports the results for the words length distribution, the frequency-rank distribution and the Excess Master-Mind similarity. The results are perfectly online with those obtained with random graphs and UCM graphs, pointing to the robustness of our results with respect to the choice of the structure of the conceptual space.

5 Differences in the conceptual space of different individuals

We here consider the more general and perhaps more realistic case in which the conceptual space is different for each individual. We still consider static conceptual spaces, referring to a future work for the possibility of co-evolving conceptual spaces and language. However, differences in the conceptual spaces of different individuals are taken into account, modeled as differences in the connection between the

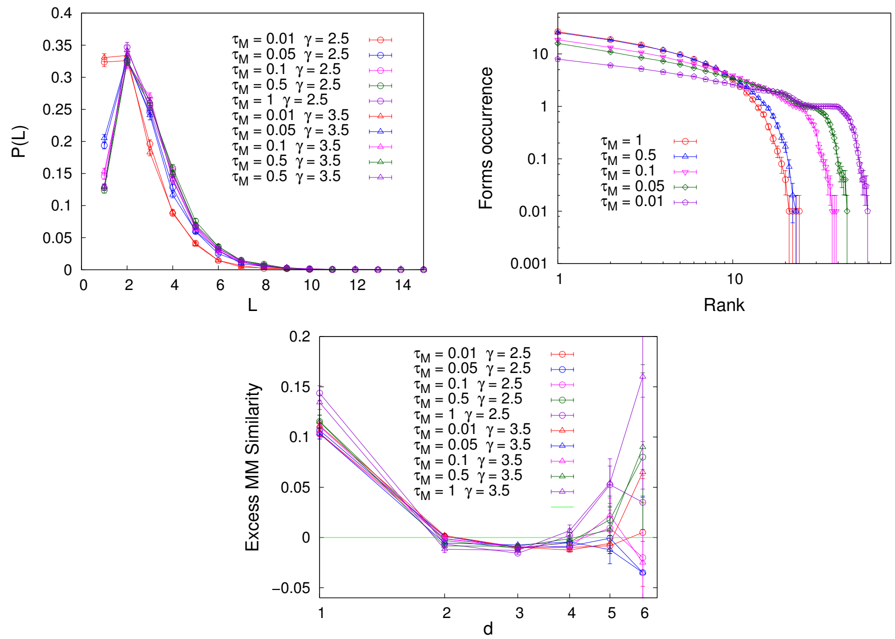

Figure 9: *Dependence of the graph structure* In this figure we report the results obtained by considering a different structure of the conceptual space. We considered in particular random scale-free networks obtained from the uncorrelated configuration model (UCM) [45], characterized by a degree distribution $P(k) \sim k^{-\gamma}$. In our case we used $\gamma = 2.5$ and $\gamma = 3.5$. *Top Left. Word length distribution.* The distributions of words length for different τ and for the two values of the degree distribution exponent γ are reported. As observed in the text, the word length distribution (and thus the average word length) does not depend on γ and is perfectly comparable to that obtained when considering Erdős-Rényi random graphs. *Top Right. Frequency-rank distribution for elementary forms.* The frequency-rank distribution for elementary forms is shown again for different values of the parameter τ and $\gamma = 2.5$. Again the results are perfectly comparable to those obtained when considering Erdős-Rényi random graph. *Bottom. Excess similarity of words as a function of the distance of the corresponding objects on the graph.* The excess MM similarity (see text) for different values of τ and for the two values of γ. Provided a non trivial structure of the world is preserved, the results do not depend on the actual value of the γ exponent. We here considered a graph with $M = 100$ nodes in order to have a greater variability. Note that the increase in excess similarity at large distances is an artifact, as the large error bars indicate, of the small number of objects at those distances in the graph. All the above results are averaged over 100 realizations of the process on the same graph with population size $N = 10$ and number of objects $M = 40$.

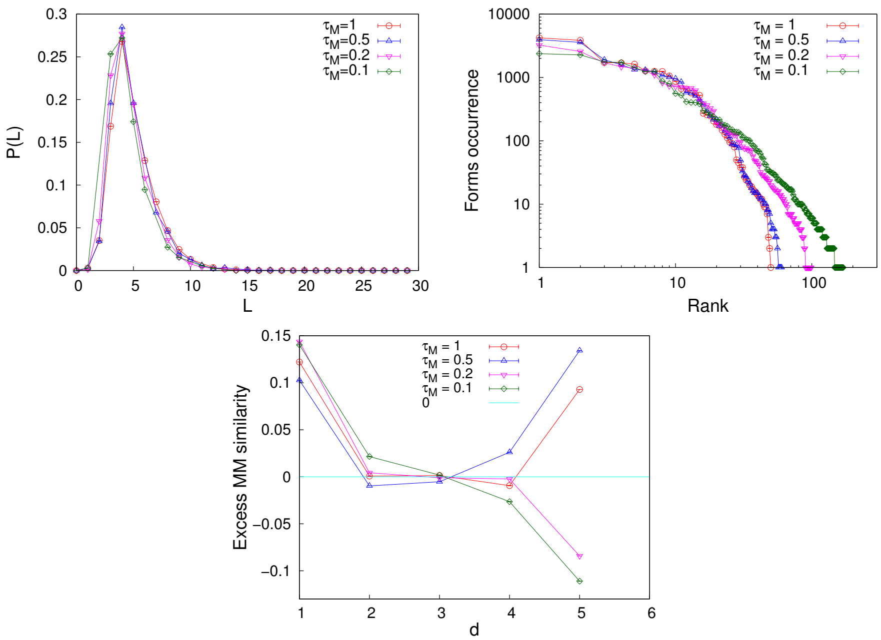

Figure 10: *Results for the South Florida Word Association Graph* In this figure we report the results obtained by considering the South Florida Word Association Graph as a proxy for the conceptual space. *Top Left. Words length distribution.* The distributions of words length for different τ are reported. The words length distribution (and thus the average word length) is perfectly comparable to that obtained when considering Erdős-Rényi random graphs. *Top Right. Frequency-rank distribution for elementary forms.* The frequency-rank distribution for elementary forms is shown again for different values of the parameter τ. Again the results are perfectly comparable to those obtained when considering Erdős-Rényi random graph. *Bottom. Excess similarity of words as a function of the distance of the corresponding objects on the graph.* The excess MM similarity (see text for details) for different values of τ. Note that the increase in excess similarity at large distances is an artifact of the small number of objects at those distances in the graph. All the above results are obtained with a population size $N = 10$ and number of objects $M = 40$.

objects to be named. More in particular, we start from an Erdős -Rényi random graph with a given link probability p_{link}, that act as a template. For each individual, we then reshuffle each link of the template graph with a probability p_{err} in the following way: we disconnect two random objects that were connected and we connect two random objects that were disconnected. In this way, the new graphs share the same statistical properties of original one. The difference between the conceptual spaces of different individuals is thus modulated by the value of p_{err}. In figure 11 we show results for the compositionality in the emerged lexicon for different values of p_{err}. We observe that provided the different graphs are not completely uncorrelated, a certain degree of compositionality still emerges. As expected, compositionality is a decreasing function of p_{err}.

6 Discussion and conclusion

In this paper we focused on the origin of duality of patterning at the lexicon level. We addressed this question 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 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 condition* is 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 as an option to be exploited when the first communication attempt resulted in a failure.

The blending strategy we refer to here must be thought as a gen-

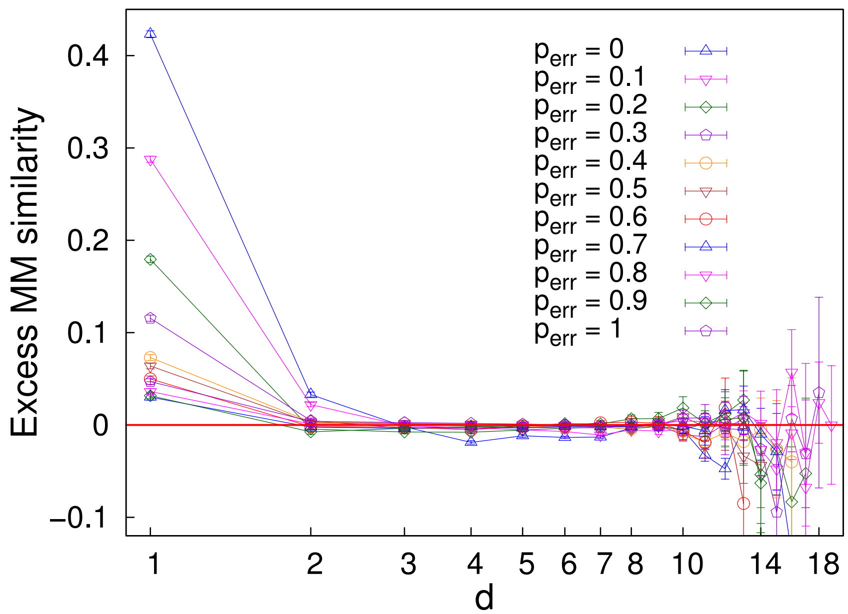

Figure 11: *Excess Master-Mind-like similarity of words as a function of the average distance d of the corresponding objects on the different graphs, for different values of p_{err}.* A decrease in the excess similarity as a function of the topological distance d is the signature of the emergence of compositionality. Here d indicates the distance between two objects in a graph averaged over all the conceptual graphs of the different individuals in the population. Results are averages over 100 realization of the process and are shown for the parameters values: $N = 10$, $M = 100$, $p_M = 0.5$ and $\tau_M = 1$. We note that compositionality is a decreasing function of the graphs dissimilarity parametrized by p_{err}. We also note a saturation for $p_{err} = 1$ on a value slightly higher than zero. This is a finite-size effect, indicating that a relation between language and meaning is still present when looking at the conceptual graph of some individuals. We would recover compositionality strictly zero when $p_{err} = 1$ in the infinite population ($N \to +\infty$) limit.

eral 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 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 [56]. In our modelling scheme duality of patterning emerges only when the two conditions mentioned above are paralleled by specific requirements for the structure of the conceptual space.

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.

Let us first consider combinatoriality. Combinatoriality, meant both as forms reuse and economy, does not always emerge in our modeling framework. A significant level of noise is crucial for it to emerge. When the level of understanding is almost perfect (extremely low noise), the number of distinct forms composing the emerged lexicon turns out to be very high with respect to the number of objects to be named (in particular, higher than the number of objects), the lexicon featuring a very low degree of combinatoriality (the fact that the combinatoriality is not strictly zero depends on the specific definition of combinatoriality we introduced). Conversely, a high level of combinatoriality as well as a low number of distinct forms in the emerging lexicon occur when the communication is noisy. Moreover, it has to be noted that we always start from a truly potentially infinite set of forms, without imposing any constraints on the words length. Both the limited number of forms we find in the emerging lexicon and the limited length of words are outcomes only of the communication process. It is important to emphasize that all these statements are not only qualitative. Rather, they are suitably quantified in terms of the parameter τ. These results suggest that combinatoriality enhances message transmission in noisy environments [27] and emerges as a result of the need of communicative success.

Let us now consider compositionality. Again a compositional lexicon does not always emerge and, as already noted, the blending repair strategy is not enough for compositionality to emerge. 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 the agents 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 [40, 57].

In summary, the ensemble of our results points to the following scenario. Combinatoriality seems to reflect the effort of communicating in a noisy environment, while compositionality seem to reflect the organization of our conceptual understanding. As further analysis, we manipulated the type of semantic network as well as the number of objects and agents in the simulations, showing the robustness of the modelling approach we are proposing with respect to parameter manipulations.

These results are important because they demonstrate for the first time that the two sides of duality of patterning can emerge simultaneously as a consequence of a purely cultural dynamics in a simulated environment which contains meaningful relations. In addition, the relevance of the interplay between the structure of the conceptual space and simple communication strategies of humans has been highlighted. Additionally, the study provided a number of measures which capture basic linguistic properties of the emerged languages. In other words, the cultural modeling of language emergence has begun to produce predictions which are amenable to testing with realistic samples of natural languages [58, 59].

Before concluding let us now briefly discuss some generalizations of the actual model that will be addressed in future works. First of all, it is interesting to consider the co-evolution of the conceptual spaces of each individual and language. The crucial question is: how language itself shapes (and is shaped by) our conceptual understanding of the world? We saw in the present model that language is shaped by the way we conceptualize the world, the next step will be investigating to which extent the conceptual understanding of different individuals are affected by repeated language games. Second, we here considered the blending strategy as an additional opportunity for the speaker to achieve success in communication. However, the ability of blending plays also a crucial role in understanding: humans are able not only to speak about concepts they never expressed, but also to understand a concept they never heard before. An interesting generalization of the present model would be that of investigating the role played by this generalization ability in achieving success in communication also when a low level of noise is always present.

Further, the model should be generalized in order to include syntax. In particular, the crucial step would be investigating how a population of individuals can bootstrap a language where categorization and syntax both emerge as a pure results of communication efforts.

A very last observation concerns the similarity between semantically related words. Semantically related words are not necessarily related in forms, and only in special cases, even if numerous, they are. Nevertheless, an excess similarity between words used to name related meanings do exists in natural languages and is statistically relevant. A very interesting direction in order to compare models outcomes with properties of natural languages is, for instance, that of looking at word-associations experiments. In this experiments human beings are asked to write a word in response to another word given as input. This simple task allows for the construction of the so-called Word Association Graphs and many examples already exist [55]. A Word Association Graph represents a very good proxy for the structure of a semantic network and an excellent test-bed for grounding theoretical predictions.

7 Acknowledgements

It is pleasure to warmly thank Bruno Galantucci with whom part of this work has been carried out.

References

[1] L. Wittgenstein, Philosophical Investigations. (Translated by Anscombe, G.E.M.), Basil Blackwell, Oxford, UK, 1953.

[2] L. Steels, Modeling the cultural evolution of language, Physics of Life Reviews 8 (4) (2011) 339–356.

[3] V. Loreto, A. Baronchelli, F. e. Tria, Language Dynamics, Vol. 15, Advances in Complex Systems, 2012.

[4] L. Steels, A self-organizing spatial vocabulary., Artificial Life 2 (3) (1995) 319–332.

[5] A. Baronchelli, M. Felici, E. Caglioti, V. Loreto, L. Steels, Sharp transition towards shared vocabularies in multi-agent systems, Journal of Statistical Mechanics P06014.

[6] A. Baronchelli, V. Loreto, L. Steels, In-depth analysis of the Naming Game dynamics: the homogeneous mixing case, Int. J. Mod. Phys. C 19 (5) (2008) 785–812.

[7] G. Lakoff, Women, fire, and dangerous things: What categories reveal about the mind, University of Chicago Press, Chicago, 1987.

[8] J. Taylor, J. Taylor, Linguistic categorization, Oxford University Press New York, 2003.

[9] L. Steels, T. Belpaeme, Coordinating perceptually grounded categories through language: A case study for colour, Behav. Brain Sci. 28 (04) (2005) 469–489.

[10] N. Komarova, K. Jameson, L. Narens, Evolutionary models of color categorization based on discrimination, J. Math. Psych. 51 (6) (2007) 359–382.

[11] A. Puglisi, A. Baronchelli, V. Loreto, Cultural route to the emergence of linguistic categories, Proc. Natl. Acad. Sci. USA 105 (23) (2008) 7936.

[12] B. Berlin, P. Kay, Basic Color Terms, Berkeley: University of California Press, 1969.

[13] F. Ratliff, On the psychophysiological bases of universal color terms, Proceedings of the American Philosophical Society 120 (1976) 311–330.

[14] A. Baronchelli, T. Gong, A. Puglisi, V. Loreto, Modeling the emergence of universality in color naming patterns, Proceedings of the National Academy of Sciences 107 (6) (2010) 2403–2407.

[15] V. Loreto, A. Mukherjee, F. Tria, On the origin of the hierarchy of color names, Proc. Natl. Acad. Sci. USA 109 (2012) 6819–6824.

[16] H. Jaeger, A. Baronchelli, E. Briscoe, C. M. H., G. T., Jaeger, G. S. Kirby, N. Komarova, P. J. Richerson, L. Steels, J. Triesch, What can mathematical, computational and robotic models tell us about the origins of syntax?, in: B. D., E. Szathamary (Eds.), Biological Foundations and Origin of Syntax, Strungmann Forum Reports, vol. 3. Cambridge, MA: MIT Press, 2009.

[17] J. D. Beule, B. K. Bergen, On the emergence of compositionality, in: Proceedings of the 6th International Conference on the Evolution of Language, 2006, pp. 35–42.

[18] C. F. Hockett, The origin of speech, Scientific American 203 (1960) 88–96.

[19] C. F. Hockett, Logical considerations in the study of animal communication, Animal sounds and communication (1960) 392430.

[20] A. Martinet, Elements of General Linguistics (Translated by E. Palmer, 1964), University of Chicago Press, 1960.

[21] S. Altmann, C. Hockett, T. Sebeok, A note on design features, Indiana University Press, 1968, pp. 61–72.

[22] E. G. Pulleyblank, Outline of Classical Chinese Grammar, Vancouver: UBC Press, 1995.

[23] P.-Y. Oudeyer, The self-organization of speech sounds, Journal of Theoretical Biology 233 (3) (2005) 435–449.

[24] P.-Y. Oudeyer, The self-organization of combinatoriality and phonotactics in vocalization systems., Connect. Sci. 17 (3-4) (2005) 325–341.

[25] W. Zuidema, B. de Boer, The evolution of combinatorial phonology, Journal of Phonetics 37 (2) (2009) 125–144.

[26] B. De Boer, W. Zuidema, Multi-agent simulations of the evolution of combinatorial phonology, Adaptive Behavior - Animals, Animats, Software Agents, Robots, Adaptive Systems 18 (2010) 141–154.

[27] M. A. Nowak, D. Krakauer, A. Dress, An error limit for the evolution of language, Proceedings of The Royal Society of London. Series B, Biological Sciences 266 (1433) (1999) 2131–2136.

[28] J. P. M.A. Nowak, V. Jansen, The evolution of syntactic communication, Nature 404 (2000) 495–498.

[29] T. Gong, J. W. Minett, W. S.-Y. Wang, Computational simulation on the co-evolution of compositionality and regularity, in: Proceedings of the 6th International Conference on the Evolution of Language, 2006, pp. 99–106.

[30] J. Batali, Computational simulations of the emergence of grammar, Approaches to the Evolution of Language Social and Cognitive Bases 405 (1998) 405–426.

[31] H. Brighton, Compositional syntax from cultural transmission, Artificial Life 8 (1) (2002) 25–54.

[32] K. Smith, S. Kirby, H. Brighton, Iterated learning: A framework for the emergence of language, Artificial Life 9 (4) (2003) 371–386.

[33] P. Vogt, The emergence of compositional structures in perceptually grounded language games, Artificial Intelligence 167 (1-2) (2005) 206–242.

[34] P. Vogt, Overextensions and the emergence of compositionality, in: Proceedings of the 6th International Conference on the Evolution of Language, 2006, pp. 364–371.

[35] S. Kirby, H. Cornish, K. Smith, Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language, Proc. Natl. Acad. Sci. USA 105 (31) (2008) 10681–10686.

[36] A. Senghas, M. Coppola, Children creating language: how nicaraguan sign language acquired a spatial grammar, Psychological Science 12 (2001) 323–328.

[37] A. Senghas, S. Kita, A. Ozyurek, Children creating core properties of language: Evidence from an emerging sign language in nicaragua, Science 305 (5691) (2004) 1779–1782.

[38] S. Mufwene, The ecology of language evolution, Cambridge Univ Pr, 2001.

[39] J. F. Sowa, Conceptual structures: information processing in mind and machine, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1984.

[40] P. Gärdenfors, Conceptual spaces as a framework for knowledge representation, Mind and Matter 2 (2004) 9–27.

[41] R. V. Solé, B. Corominas, S. Valverde, L. Steels, Language Networks: their structure, function and evolution, Trends in Cognitive Sciences.

[42] P. Wellens, M. Loetzsch, L. Steels, Flexible word meaning in embodied agents, Connection science 20 (2-3) (2008) 1–18.

[43] P. Erdös, A. Rényi, On random graphs i, Publ. Math. Debrecen 6 (1959) 290.

[44] P. Erdös, A. Rényi, On the evolution of random graphs, Publ. Math. Inst. Hung. Acad. Sci. 7 (1960) 17.

[45] M. Catanzaro, M. Boguna, R. Pastor-Satorras, Generation of uncorrelated random scale-free networks, Physical Review E (Statistical, Nonlinear, and Soft Matter Physics) 71 (2) (2005) 027103.

[46] D. Nelson, C. McEvoy, T. Schreiber, The university of south florida free association, rhyme, and word fragment norms, Behavior Research Methods 36 (2004) 402–407, 10.3758/BF03195588.
URL `http://dx.doi.org/10.3758/BF03195588`

[47] B. Sigurd, M. Eeg-Olofsson, J. Van Weijer, Word length, sentence length and frequency - zipf revisited, Studia Linguistica 58 (1) (2004) 37–52.

[48] Y. Tambovtsev, C. Martindale, Phoneme frequencies follow a yule distribution, SKASE Journal of Theoretical Linguistics 4 (2) (2007) 1–11.

[49] G. U. Yule, The statistical study of literary vocabulary, Cambridge University Press, 1944.

[50] B. Galantucci, C. Kroos, T. Rhodes, The effects of rapidity of fading on communication systems, Interaction Studies 11 (1) (2010) 100–111.

[51] F. Tria, B. Galantucci, V. Loreto, Naming a structured world: a cultural route to duality of patterning, PLoS ONE (2012) xxx–xxx.

[52] D. J. Watts, S. H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (6684) (1998) 440–442. doi:10.1038/30918.

[53] R. Pastor-Satorras, A. Vázquez, A. Vespignani, Dynamical and correlation properties of the Internet, Physical Review Letters 87 (25) (2001) 258701.

[54] M. E. J. Newman, Assortative Mixing in Networks, Physical Review Letters 89 (20) (2002) 208701+.

[55] P. Gravino, V. Servedio, A. Barrat, V. Loreto, Complex structures and semantics in free word association, Advances in Complex Systems.

[56] H. Ottenheimer, The Anthropology of Language: An Introduction to Linguistic Anthropology, 2nd ed., Wadsworth, Cengage Learning, 2009.

[57] C. AM, L. EF, A spreading-activation theory of semantic processing, Psychological Review 82 (6) (1975) 407–428.

[58] S. L. M. Dunn, S.J. Greenhill, R. Gray, Evolved structure of language shows lineage-specific trends in word-order universals, Nature 473 (2011) 79–82.

[59] G. Lupyan, R. Dale, Language structure is partly determined by social structure, PLoS ONE 5 (1) (2010) e8559.