The Evolution of Lexical Usage Profiles in Social Networks

Gerhard Schaden
Université de Lille & CNRS UMR 8163 STL
gerhard.schaden@gmx.net

June 1, 2016

This paper investigates how network structure influences the outcomes of reinforcement learning in a series of multi-agent simulations. Its basic results are the following: i) contact between agents in networks creates similarity in the usage patterns of the signals these agents use; ii) in case of complete networks, the bigger the network, the smaller the lexical differentiation; and iii) in networks consisting of linked cliques, the distance between usage patterns reflects on average the structure of the network.

1 Introduction

1.1 Lexical Variation, Lexical Change

Lexical variation and change are pervasive phenomena in natural languages. Contrary to grammatical change — often assumed to be unidirectional, and predictable (see, e.g. Haspelmath, 1999) — lexical change is typically messy. It depends on specific historical contexts, new technological inventions, and probably, to a certain degree, also on random events.

This can be illustrated by one particular example of lexical shift and variation in French, namely *voiture*, which is today most commonly used to designate motorized vehicles for personal transportation. Notice that *voiture* is not a newly created lexical item; the word existed and referred in former times to (horse-drawn) carriages, a meaning that is still present in current French. After a period of variation at the beginning of the 20th century, when self-propelled vehicles were referred to as *voiture automobile*, *automobile* or *auto*, Metropolitan French settled for the older, and shorter, *voiture*.[1] German was confronted with the same problem and settled for *Auto*. The phenomenon was triggered by the introduction of a new technology, and is therefore an example of external change. The question arises as to whether this instance of lexical change should be considered a change in lexical *meaning*. After all, even though the prototypical use in the speech community is clearly different today (shifting from horse-propelled to self-propelled instances), it is not clear whether the meaning of the word as such has changed, since its former meaning remains available.

[1] See the *Trésor de la Langue française informatisé* (http://atilf.atilf.fr), entry VOITURE.

One may want to argue that initially, *voiture* had to extend its meaning, since it accommodated for a new referent. However, even this is far from clear: had something like *horse-drawn* been part of the original meaning, adding an adjective such as ''*automobile*'' should have led to a contradiction: no vehicle can be self-propelled and horse-drawn at the same time. Yet, be that as it may, it is clear enough that a semantic change may be under way, since horse-drawn carriages are extremely rare nowadays, and, should they disappear, the meaning of *voiture* would likely follow suit.

Summing up, the case of *voiture* illustrates a case that may very well not be a change in semantic meaning at all, triggered by a purely external cause, and possibly randomly. The question is whether we should bother to investigate such instances at all, given the importance of all sorts of language-external noise involved in it. Linguists have mainly focused on regular types of linguistic change (such as grammaticalization), where random processes can be by and large neglected.

Now, while it is true that individual cases of lexical *meaning* change often seem random, and may well turn out to be random, there are patterns to lexical change as well. For instance, change is not saltationist (a word meaning *horse* today is very unlikely to mean *pencil* tomorrow), and change generally occurs along patterns of polysemy. Other factors that have been identified are the imitation of prestigious individuals (see Chudek et al., 2012).

One of the main points that make such a pattern interesting and worthwhile to investigate is that, in order to ascertain what kinds of changes are *not* random, we need to understand the effects of randomness in language use and change, for instance, when, where and how it influences lexical change.[2] In genetics, the effect of *genetic drift* (random change, not affected by natural selection) is recognized to be important, and it would be similarly important to know the effects of random change on natural languages — a phenomenon that one may call *linguistic drift*. There have been efforts modeling phenomena of cultural evolution in terms of drift (see, e.g., Han and Bentley, 2003; Reali and Griffiths, 2009b; Griffiths and Reali, 2011; Bentley et al., 2011)[3], and the present paper aims to make a contribution towards furthering the comprehension of lexical drift.

Individually, instances of lexical meaning change may be anecdotic, but the phenomenon tells us something about the nature of language and how it is shaped: Language is a product of social consensus.

1.2 Learning as a Social Fact

Compared to our primate cousins, Humans are an intensely social and cooperative species. This has also consequences for language, and that on several levels: one basic fact on language is that it is acquired and used in social contexts (more precisely, in *social networks*).

The importance of social networks for natural language and its change has been recognized from early on. For instance, Hermann Paul wrote the following in his *Prinzipien der Sprachgeschichte*:

> Jede Veränderung des Sprachusus ist ein Produkt aus den spontanen Trieben der einzelnen Individuen einerseits und den [...] Verkehrsverhältnissen andererseits. [4]

[2] This was pointed out to me by Sylvain Billiard (p.c.).

[3] I would like to thank one of the anonymous reviewers for pointing this out.

[4] Paul (1995 : §25): *Any change in a language is a product of the spontaneous drives of individuals on the one hand, and the structure of lines of communication on the other.*

It is very tempting to interpret this quote in the light of recent theoretical developements, namely game theory (investigating the spontaneous drives of individuals) and network analysis (investigating the structure of lines of communication between individuals). In any case, linguists have started bringing these models together to study the interaction of language learning, change and diffusion (see Smith et al., 2003; Benz et al., 2006; Mühlenbernd and Franke, 2012). Fagyal et al. (2010) have shown that network structure is important for the fluctuations of norms within a language; whereas Pierrehumbert et al. (2014) provide a model showing that average-connected agents are the most likely source of successful innovation (as opposed to the highly connected members of an elite).

The basic aim of this paper is to investigate the implications of the fact that learning language necessarily takes place within a social setting. In order to do that, I will try to eliminate other potential influences — like individual preferences of agents, etc. — as much as possible.

One major question will be whether learning depends on the structure and size of the network the learners are engaged in. I will investigate what the impact of learning is in a network with respect to differenciation/homogeneity. The usual issue in research on learning in networks (e.g., DeGroot models, see Jackson, 2008 : 228ff.) is the emergence of consensus in an initially diverse population (this is also true to some degree for the question of the emergence of norms in Fagyal et al., 2010 and Pierrehumbert et al., 2014). The isssue I am interested in here is rather the emergence of (lexical) differenciation in an initially perfectly homogeneous population.

The study uses multi-agent simulations in circumstances where there is no *right* thing to learn, and thus, where there are no linguistic/social constraints on the use of variant forms that could alter the impact of pure randomness in a networked setting. All outcomes will be based on patterns of preferential use.

I will study not meaning change per se, but rather the changes in what I will call the *Lexical Usage Profiles* of networked agents. Assuming two words without any linguistic or social constraints that differentiate them in a given context, the Lexical Usage Profile is the probabilistic usage pattern associated by an Agent with these words (for instance, (s)he could use word1 with probability 0.4, and word2 with probability 0.6). This notion of Lexical Usage Profile will be made explicit below, in (2), p. 8. In a sense, this profile is about performance or pragmatics, and not about semantics or competence. However, it is easy to see that in a population where new agents enter the scene and old ones die, biased usage profiles might lead to the loss of one word, and therefore, in some configurations considered below, to meaning change.

An important background issue of this paper is the question of the importance of usage-based factors on language and language change (frequency, etc.). While I will not directly provide an answer to this issue, the results presented here suggest that they indeed matter.

The methodology at work in this paper uses multi-agent simulations. As pointed out by Gilbert (2008 : 3), ''*[i]n most of the social sciences, conducting experiments is impossible or undesirable*''. In order to see his point, consider the difficulty of isolating a relevant subsystem of persons. Even if this could be resolved, we would still need to set up a control group. Controling and isolating the relevant variables would be difficult at best, and a real-world experiment dealing with the linguistic evolution of a population in a given setting might last far too long to be sustainable.

One possible solution to this problem is to experiment not on a group of people from the real world, but rather on a *model*. This has the advantages that the interaction of programmed agents does not pose ethical problems, and that properties of agents can be defined as needed. For our

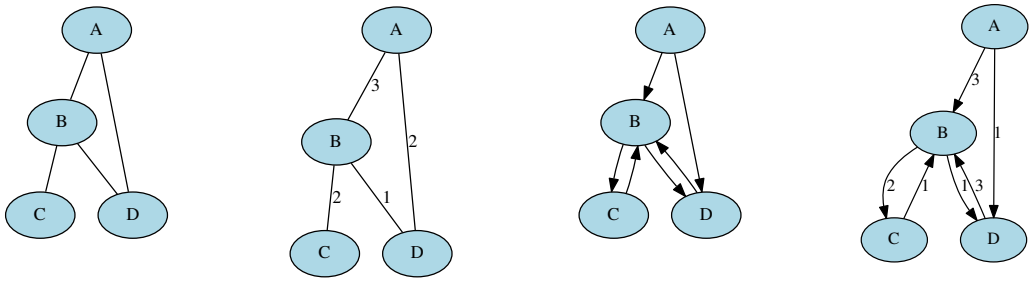

Figure 1: Different Kinds of Social Networks (from left to right): unweighted graph, weighted graph, unweighted digraph, weighted digraph

case, it means that we can create agents without any preferences, and which act truly at random.

However, multi-agent simulations are not without problems: firstly, we need to make sure that the model is close enough to the empirical phenomenon we are interested in, and secondly, that the model is simple enough to isolate causal factors. While ideally a model should conform to both requirements, these are also to some point inherently in conflict.

Let us now move to some required terminology with respect to social networks.

1.3 Social Networks

A social network is a structure representing agents and their interactions.[5]

Different kinds of networks model different types of interactions between agents. In some cases, we are only interested in the fact whether a link between two agents exists or not; in other cases, we want to compare the differing strengths of links between agents (unweighted vs. weighted networks). In some cases, a link is symmetrical (for instance, in friendships), in some cases, it is not (which opposes symmetrical *graphs* and directed *digraphs)*. The different combinations of these two properties are illustrated in figure 1.

Depending on the kind of interaction between agents one is interested in, different kinds of networks can be used. In our case, an act of communication between a speaker and a hearer is most easily modeled as a digraph, since it is an asymetrical kind of interaction, where participants have clearly defined, and non-identical, roles. In what follows, I will mostly use weighted digraphs, since adding weights provides an easy way of representing turns of speaking.

Apart from the kind of links between agents, the overall structure of a network will also turn out to be important. For our purposes here, we will examine more closely two types of structure, namely *complete networks* and *cliques*.

A complete network is a network where all possible edges are present. In other words, every node is connected to every other node. This is illustrated for a graph on the left diagram in figure 2.

[5] For a formal definition, see Jackson (2008 : 20ff.).

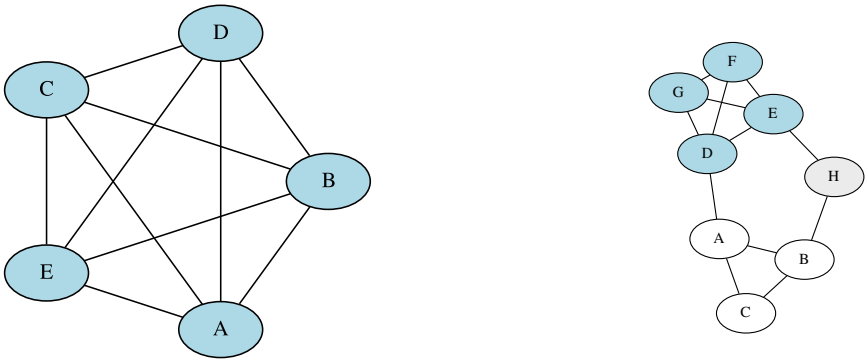

Figure 2: Structural Configurations of Networks: Complete Network and Cliques within a Network

The second type of configuration we will study are *cliques*, which are maximal completely connected subnetworks of a given network. In figure 2, on the right side, nodes ABC form one clique, and nodes DEFG form another. One can imagine a clique as being a complete network that is a subpart of another network.

In this paper, we will study reinforcement learning in complete networks (of varying sizes), and in cliques connected to other cliques by bridges, that is, links whose absence would lead the graph to decompose into several, unconnected components.

Having now introduced networks, the remainder of the paper is structured as follows. In section 2, we will introduce reinforcement learning and its effects on lexical use and meaning. In section 3, we will consider what might change if learning takes place in a complete network, and section 4 will investigate the outcome of reinforcement learning in the configuration of networks consisting of linked cliques.

2 Reinforcement Learning

Reinforcement Learning refers to a family of learning algorithms, and is one of the standard techniques in machine learning. It is modeled on behaviorist psychology, where learning is seen as the shift (typically: augmentation) of the probability of some behavior in a given agent.

2.1 Reinforcement Learning: Polya-Urns

Polya-urn processes provide a mathematical model of reinforcement learning.[6] It corresponds to a random-draw from an urn. Contrary to other types of urns, if the answer is correct, not only is the ball returned into the urn, but some quantity of the same balls will be added to the urn. Let us consider an example:

[6] This is not the only possible model of reinforcement learning, but it happens to be one of the simplest, and one that does not drive one alternative into virtual oblivion too fast.

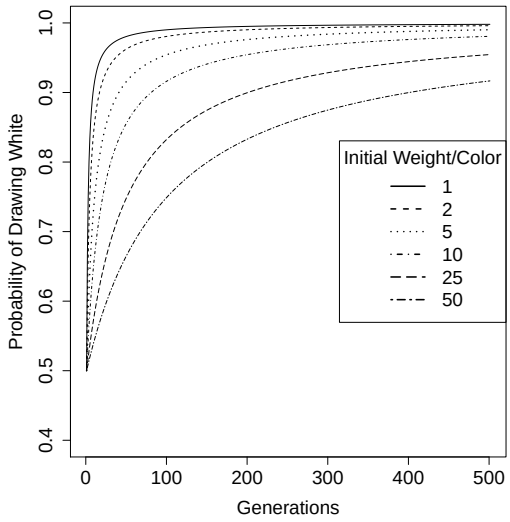

Figure 3: Speed of learning depends on Inclination Weights

URN_t
white:1
red:1

URN_{t+1}
white:2
red:1

At time t, corresponding to the beginning of the process, the urn contains one ball for each color. These are the *inclination weights*. Then, a ball is drawn at random (here, it happens to be the white ball). Assuming that white is the correct answer, white will be reinforced. This is done by adding an additional white ball — the *reward* — to the urn. The reward insures that the appropriate answers rises in probability. Whereas the initial probability of drawing white at t was only 0.5, this probability has risen to $0.\dot{6}$ at $t + 1$.

The system will converge in the limit to a point where the appropriate answer is the only one.[7]

As figure 3 shows, the speed of learning depends on the initial inclination weight (one may see higher inclination weights as higher confidence of the learner that both options are actually correct).

A general feature of reinforcement learning of this particular kind is that learning is initially fast, and then becomes gradually slower. That means that reinforcement learning obeys the *Law of Practice*, which is a general feature of human learning (see Skyrms, 2010 : 85). The particular version of Polya-urn-based reinforcement learning used here has the additional advantage that it is variation-friendly, that is, it will not drive into extinction one alternative too fast.[8]

Finally, reinforcement learning can provide a formalisation of *priming* and *entrenchment*.

All these properties are desirable and can explain why reinforcement learning is popular, and why we want to adopt it in our simulations.

[7] Although, in this setting, the initial weight of 1 is never removed, as time progresses, and if reinforcement is consequent, the probability of drawing white will approach 1, since $\lim_{n\to\infty} \frac{n}{n+1} = 1$.

[8] In other words, it has a bias against regularization.

2.2 Reinforcement Learning When There is Nothing To Learn

The setup I will consider here is a configuration where there is nothing to learn, that is, where both answers (or, more generally, all answers) are appropriate, and will be reinforced. This is something Skyrms (2010) has explored in some detail.[9]

From a linguistic point of view, this might model a situation where two words are absolute synonyms. However, such a strong constraint is not necessary; it suffices that the context does not support a difference between words (see Schaden, 2014). In such a case, the frequencies may shift around wildly, and chance (together with the impact of reinforcement learning) is the only influence on the system.

As was shown in Schaden (2014), the outcome of the process is dependent on the ratio between inclination weights and the reward. The higher the inclination weights with respect to the reward, the more the probabilities of drawing one word or the other cluster around 0.5 — which means that both words are used with roughly equal frequency. Yet, the smaller the inclination weights with respect to the reward, the more probability mass will cluster towards 0 and 1. This means that one or the other of the two words will have become dominant.

This pattern is interesting because the farther we move away from the area of the central 0.5 frequency, the more a word is fragile with respect to environmental changes or pragmatic reinterpretations.

Up to now, we have considered a word to be a monolithic entity, with a clear and unified sense. However, this may not be very realistic an assumption if we take into account polysemy, and it is not convenient for our simulations.

2.3 Simulating Change of Internally Differentiated Concepts

In what follows, I will assume internally differentiated concepts, and more precisely, something like Pustejovsky's *qualia-structure*. Pustejovsky's basic assumption is that the (linguistic) meanings of words are far more complicated than what is often assumed in (formal) semantics, and that many elements of what is often thought of as *encyclopedic knowledge* are actually part of the linguistic meaning of a word. For instance, Pustejovsky assumes that words denoting objects have four *qualia*, or submeanings[10], which specify the denoted objects shape, material, function, etc. While one does not need to accept this specific framework for dealing with polysemy, the basic ontological commitment my simulation requires is the existence of independently ponderable submeanings.[11]

This move has several reasons: firstly, it considerably facilitates dealing with shifts of prototypical meaning. And secondly, from the empirical point of view, meaning shifts generally follow patterns of polysemy, and thus, some theory of polysemy (like the generative lexicon) is required anyway.

I will take a very simplified approach to polysemy. In the simulations, I will merely assume that every word consists of four[12] independently ponderable lexical meanings, as illustrated in (1):

[9] As pointed out by one of the reviewers, such models are also related to the literature on regularization or absence of regularization of linguistic forms (see, e.g., Reali and Griffiths, 2009a; Pijpops et al., 2015).

[10] I will henceforth use the word ''*quale*'' as a synonym for *submeaning*.

[11] As far as I see, this commitment is not tied to a particular framework in the generative or functional tradition.

[12] The precise number of submeanings has no real importance, and is rather a choice of convenience. I have adopted four submeanings because these are the number of qualias in Pustejovsky (1995).

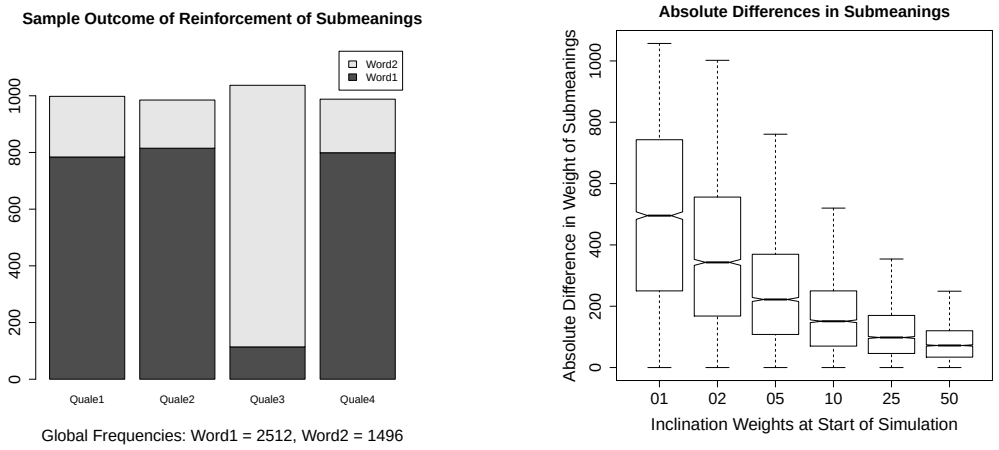

Figure 4: Outcomes of Reinforcement Process

(1)

	Word1	Word2
submeaning1	type-1: n_1	type-1: n_5
submeaning2	type-1: n_2	type-1: n_6
submeaning3	type-1: n_3	type-1: n_7
submeaning4	type-1: n_4	type-1: n_8

Each submeaning forms an independent Polya-Urn. Nature chooses some submeaning at random, and a submeaning has to be matched by either Word1 or Word2. The chosen word will then be reinforced at that particular submeaning/quale.

The outcome of such processes is shown in figure 4, where I have plotted on the left side a sample outcome with an initial weight of 1. As one can see here, the speaker will use Word1 with a frequency of around 0.75 for submeanings 1,2 and 4, whereas with respect to submeaning 3, Word2 is dominant in about the same proportion. On the right side, I plotted a more global view in the 1000 simulation trials of what happens with different kinds of inclination weights with respect to the absolute differences in weights in submeanings.

I should stress once again that what has changed throughout these trials is not the lexical meaning in itself, but rather the frequency of usage associated with each word at a given submeaning. The outcome of reinforcement learning is an array of numbers, representing the (probabilistic) *Lexical Usage Profile* of an agent. This is illustrated in (2).[13]

(2)

	W1Q1	W1Q2	W1Q3	W1Q4	W2Q1	W2Q2	W2Q3	W2Q4
Ag 1	1000	1000	1000	1000	1000	1000	1000	1000
Ag 2	2000	2000	2000	1	1	1	1	2000
Ag 3	1800	200	1800	200	1800	200	1800	200

[13] The numbers in (2) have been chosen for illustration purposes, and are *extremely* unlikely to show up in an actual simulation.

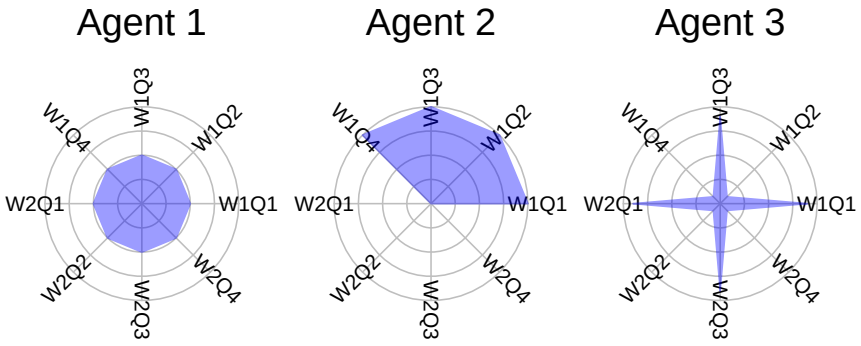

Figure 5: Spiderplots of the Lexical Usage Profiles of the 3 Agents in (2)

In order to get the probability of choosing a word for a specifing submeaning (say Quale1), we need to compare the weights of that quale in Word1, with the weight in Word2.[14]. Agent1 illustrates a case where all submeanings of both words have been reinforced exactly to the same degree. Therefore, at the next turn, Agent1 will use any word for any submeaning with a probability of 0.5. Agent2 illustrates the case where for submeanings 1–3, Word1 has always been reinforced; whereas for submeaning4, Word2 has been exclusively reinforced. Therefore, Agent2 will choose near deterministically (with a probability of $\frac{2000}{2001} \approx 0.9995$) Word1 for submeanings1–3, and with the same probability Word2 for submeaning4.

A convenient way of plotting such Lexical Usage Profiles (which I will use below for comparison) are spiderplots, as illustrated in figure 5 for the three agents in (2).

3 Learning in Networks

So far, we have considered learning in isolation of a social setting. This might represent a single agent remaining all alone. Alternatively, it could also represent an agent in a network where all signaling is public, that is, where all communication is targeted to all other members in a network, and where the speaker reinforces himself just like all the other hearers do. These are, however, very unrealistic assumptions, since not all communication is directed to the general public. Therefore, the question arises whether the fact of agents being members of a network will change the dynamics and outcome of the learning — just as the inclination weight of the agents did before.

I will show that this is precisely the case.

3.1 Learning in Networks: General Algorithm

The algorithm used for the simulations in this paper is round-based. It can be illustrated as follows.

1. Take some weighted digraph.
2. Choose an edge uniformly at random.

[14] More precisely, in order to know what the probability is to draw W1 for submeaning Q1, we need to calculate $\frac{W_1Q_1}{W_1Q_1+W_2Q_1}$, which, for Agent1 in (2), is: $\frac{1000}{1000+1000} = \frac{1}{2}$.

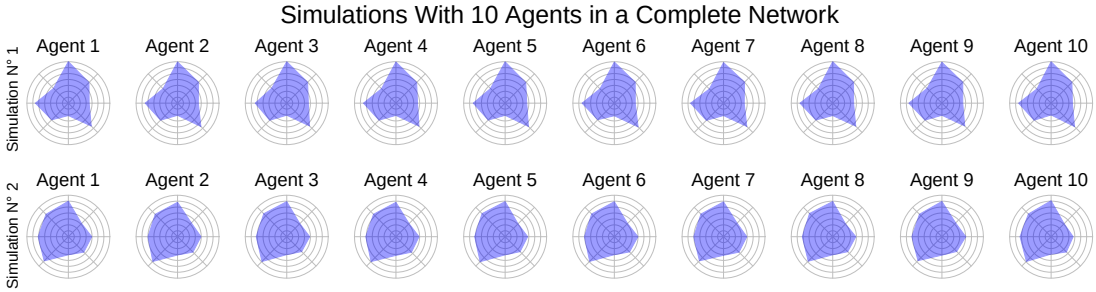

Figure 6: Contact Creates Similarity

3. The agent located at the tail node of the edge signals to the agent located at the head node of the edge, according to the signal chosen uniformly at random by Nature, and according to the tail node agent's weights for the chosen submeaning.
   a) update the lexical representation of the head node agent.
   b) update the lexical representation of the tail node agent (if self-reinforcement is in place).
4. Decrease the weight of the edge by 1[15]
   a) if the weight of the edge=0 after decrementing, remove it;
   b) otherwise, repeat the procedure from step 2 with the resulting network.
5. Repeat until no edge remains; end of turn.
6. Perform 1–5 k times on the weighted digraph

Since this algorithm depends strongly on random elements (order of the edges, choice of the submeanings, probabilistic choice of the word by signaler), every simulation has been repeated 1000 times, in order to obtain a representative sample of possible outcomes.

3.2 Learning in Complete Networks

The first setup we will consider is signaling in complete networks, where all edge-weights are the same. Therefore, every agent will signal and receive signals exactly the same number of times.

One of the clearest, and maybe, at first sight, most surprising consequences of agents learning in a complete network is that it creates uniformity among the agents' lexical usage profiles. This is illustrated in figure 6 for a complete network with 10 agents. What the final profiles look like cannot be predicted, since the process is guided by chance; however, the outcome that the members of a complete network all look alike is a constant pattern, regardless of the number of agents in the network.

Figure 6 presents this in a rather intuitive way; we will look below for a way of better measuring similarity/dissimilarity across agents in a network. Interestingly, it turns out that there is an interaction between the size of the network and the degree of differentiation that is created.

To start with, let us define the differentiation of two words with respect to a given submeaning, or a ponderable lexical component (henceforth abbreviated as *lc* and subscripted with an index):

[15] Decreasing the edge-weight by one (as well as removing 0-weighted edges) makes sure that an edge-weight of n corresponds to n interactions between the agents.

(3) Lexical differentiation between Word1 and Word2 with respect to some ponderable submeaning $lc_i =_{\text{def}} |(\text{lc}_i(\text{Word1}) - \text{lc}_i(\text{Word2}))|$
the absolute difference between the ponderable lexical components of a word

Let me repeat why lexical differentiation is interesting for us in the context of lexical meaning change over time. While in our setup no (sub)meaning will ever be completely forgotten (because technically, no ball will ever be removed from the Polya-urn), the same is not true in real life. Clearly, a word that is hardly ever used for some meaning is not very likely to make it into the next generation's lexicon. Let us now consider two extreme cases in relation with lexical differentiation: first, where lexical differentiation is zero, and another one where it approaches infinity. A lexical differentiation of zero means that both words have the same weight for the given submeaning, which means that they are equally likely to be chosen in order to describe the state of affairs corresponding to this submeaning. This corresponds to a situation where both words are used with the same frequency. Therefore, transmission to the next generation for both words with respect to the submeaning can be taken for granted.

Let us consider the case where lexical differentiation is very large. Under such circumstances, one word is chosen nearly all the time, and the other is hardly ever (or never) recruited to express that particular submeaning. This means that there is little chance for a new member of the language community to be ever exposed to that word for that meaning, and, therefore, to learn that it can be used to express this meaning. Thus, while lexical differentiation does not directly correspond to meaning change, strong lexical differentiation can be seen as a prerequisite or as a fore-runner for meaning change (that is, change not only affecting performance-related aspects of language use, but competence).

We had seen above that — in a non-networked setting — the differentiation between the two words depended on the inclination weights. It turns out that there is an impact of network size as well. If we keep stable the number of reinforcements per agent,[16] the bigger the network, the less differentiated the submeanings of Word1 and Word2.

Furthermore, there is also a clearly discernable influence of self-reinforcement: at constant network size, if the speaker reinforces his own lexical representation after speaking, lexical differentiation is bigger than without.

Why should this be so? Remember from what we have seen in section 2, that in cases of reinforcement learning, the initial tendency is very important, and in the beginning, the system quickly shifts away from the fifty-fifty initial conditions. Consider for instance a three-agent complete network, where speakers reinforce themselves. After the first signaling turn, two out of the three agents will have one word reinforced at the same submeaning. So, the next time that submeaning is drawn, there is a $\frac{2}{3}$ chance that the speaker has already been reinforced for one of the two words, and if one out of these two agents is drawn, the probability that the same word will be reinforced will be $\frac{2}{3}$. On the other hand, there is a $\frac{1}{3}$ chance that the not-reinforced agent will be drawn, and then, the chance that the other word will be reinforced is $\frac{1}{2}$.

Let us now consider the same process in a network with 10 agents, and without self-reinforcement. After one submeaning has been chosen for one word, there is only $\frac{1}{10}$ chance that the next time the submeaning is drawn, the speaker will already have been reinforced at this place. So, there are

[16] That means, if we consider cases where an agent invariably receives in one simulation run n reinforcements, whether he is in a network with 2 other or 9 other agents.

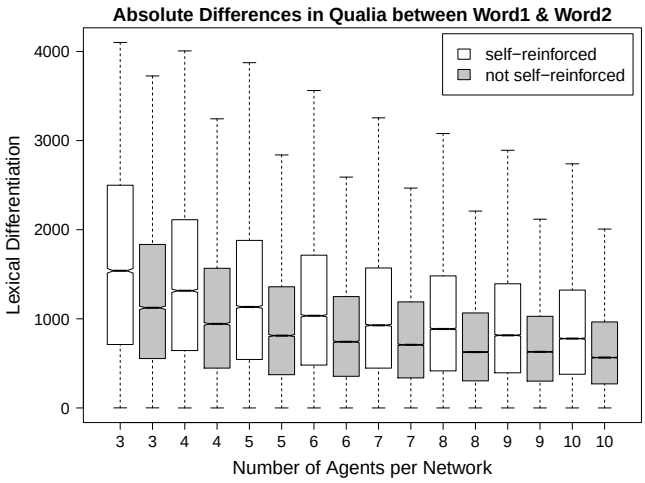

Figure 7: Increasing Network Size and Decreasing Lexical Differentiation

good chances (namely $\frac{9}{10} \times \frac{1}{2}$) that reinforcement ends up targetting the other word, which will lead to a smaller overall lexical differentiation.

The effect of network size (and self-reinforcement) is thus the following: the bigger the network, and without self-reinforcement, the smaller the probability of choosing in all agents consistently the same word for a given submeaning. Thus, network size (in the case of complete networks) does influence the kind of lexical differentiation one can expect.

This also means that smaller-sized (complete) networks have a tendency to evolve faster, and to show bigger degrees of lexical differentiation, while bigger networks show a tendency to maintain the initial status quo, and have less tendency to show extreme degrees of lexical differentiation.

Can we extrapolate this result directly to human networks, and actual languages spoken in real-life communities? Although there are claims about (what one can interprete as) the relation between network structure and language structure (see, e.g., Kusters, 2003, 2008), the present result does not carry over to existing human communities that easily, because these tend to have a different network structure.

According to Sutcliffe et al. (2012), a typical human being is engaged in the following circles of network-relations (with their approximate basic sizes): (i) the active network of individuals that we know as persons and with whom we have reciprocated, personalized relationships that have a history (of about 150 individuals; known in the ethnographic litterature as *clan*) (ii) the affinity group (or *band*, of around 50 individuals) (iii) the sympathy group, with whom we have routine ties (numbering 12-15 individuals) (iv) the support clique, that is, people with which we have intimate ties (4-5 individuals). Thus, the overall network in which a human being evolves is not a complete network, which simply means that not all people that an individual has ties with are acquainted with each other.

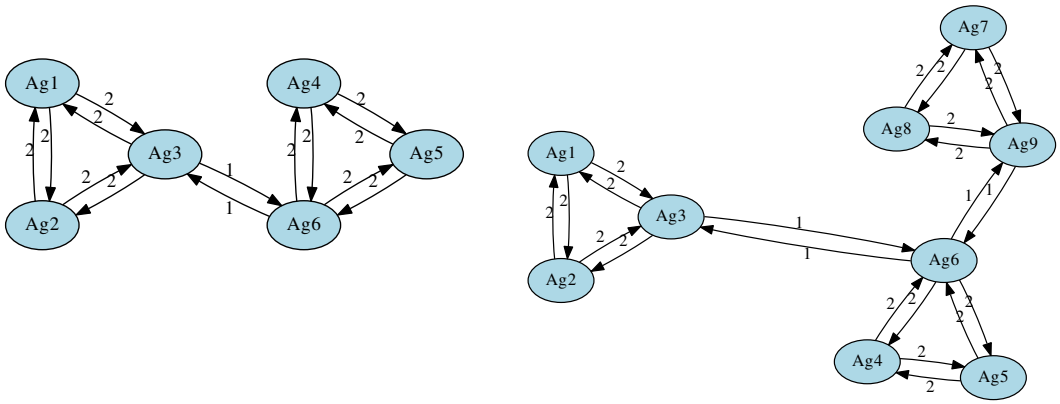

Figure 8: Cliques of Equal Size, Linked by Bridges

4 Learning in Linked Cliques

While I will not investigate naturalistic networks, we can draw the conclusion that ''normal'' language communities are not built up as complete networks (at least for not acutely endangered languages). However, cliques are important subnetworks, and Mühlenbernd and Franke (2012) have shown that in language acquisition, they have an important role to play.

I have investigated a very simple setup, where two or three same-sized cliques are linked by bridges (see figure 8). The edges within the clique are stronger than the edges between cliques. This can be seen as the effect of more costly interaction between cliques, than within cliques; one simple way of imagining this scenario are islands, where contact across islands is more limited than contact within islands.

In our setup, there is always only one agent that is in contact with the other clique(s). Now, the question is: does lexical differentiation mirror the network structure?

We will first look at cases with two linked cliques, and then, at three linked cliques.

4.1 Cases with Two Linked Cliques

Let us start by taking a very impressionistic first look at the lexical usage profiles graphed as spiderplots, as in figure 9. Each line depicts a single simulation run, where the first three agents belong to one clique, and the other three to another. It is quite obvious that within a clique, the agents always have very similar lexical usage profiles, whereas, if we compare two agents coming from different cliques, there is not necessarily any similarity. Notice however, that there is no intrinsic tendency to establish differentiation across cliques; the profiles in simulation run n° 4 are extremely similar in both cliques.

Yet, if we want to see the bigger picture, we need to have a better means of establishing a lexical proximity metric. We will take a very simple measure for the lexical distance between two words: we will simply add up the lexical differentiation (i.e., the absolute difference) for each pondered lexical component (or *lc*):

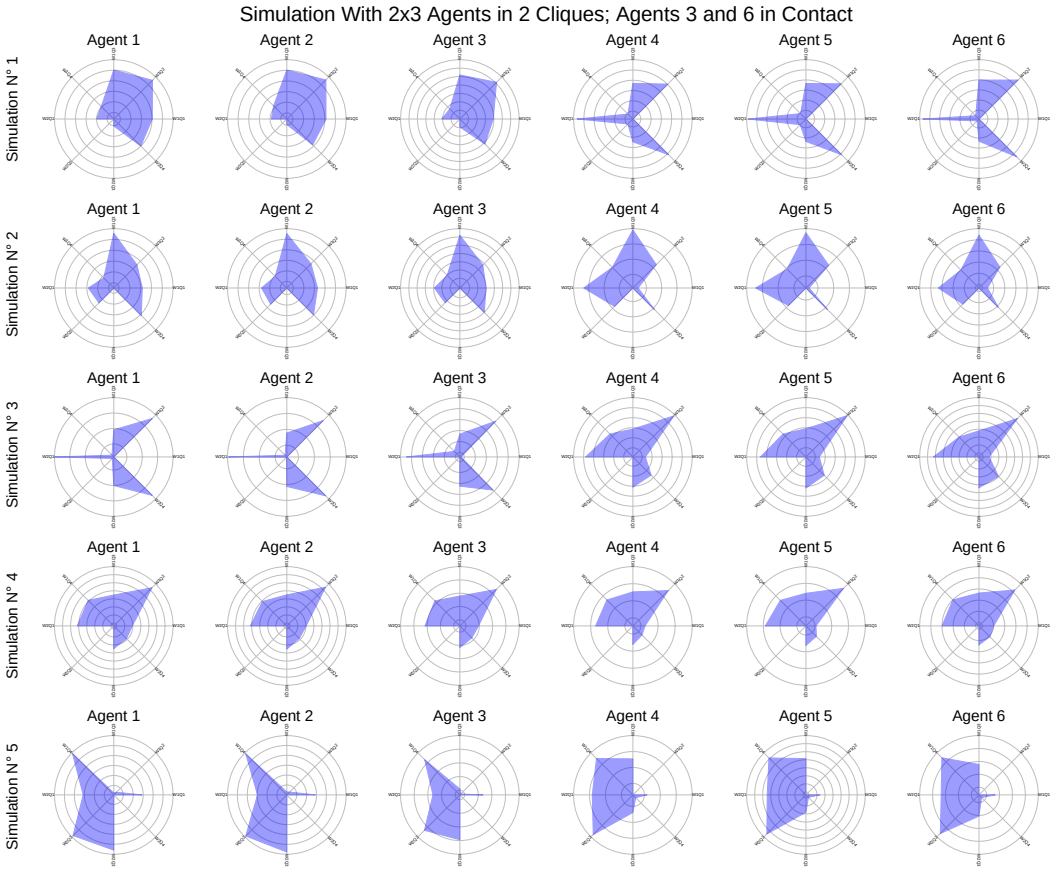

Figure 9: Agents Within Cliques Look Alike; Agents in Different Cliques Mostly Differ

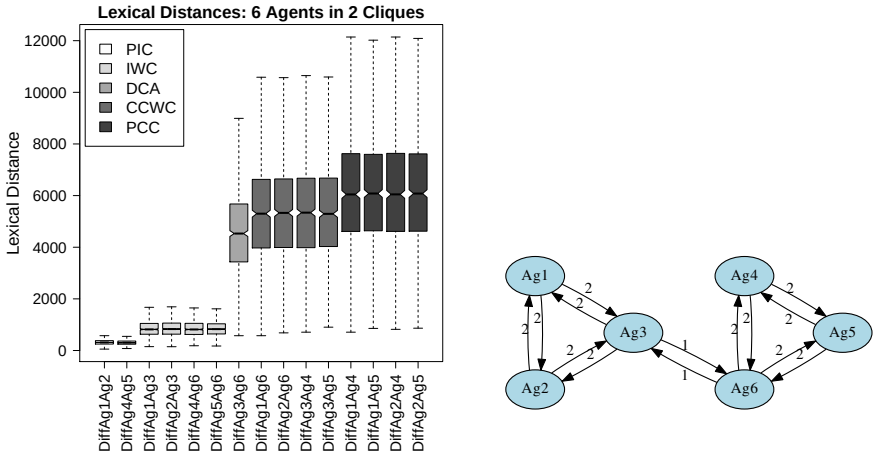

Figure 10: Lexical Distance Mirrors Network Structure

$$\mathrm{LD}(\mathrm{Ag}_1, \mathrm{Ag}_2) = \sum_{i=1}^{k} |lc_i(Ag_1) - lc_i(Ag_2)|$$

Let us reexamine our first two agents from (2), on p. 8.

(4)

	W1Q1	W1Q2	W1Q3	W1Q4	W2Q1	W2Q2	W2Q3	W2Q4
Ag 1	1000	1000	1000	1000	1000	1000	1000	1000
Ag 2	2000	2000	2000	2000	1	1	1	1
$\lvert lc_i(Ag_1) - lc_i(Ag_2)\rvert$	1000	1000	1000	1000	999	999	999	999

Summing the numbers from the last line gives us $\mathrm{LD}(\mathrm{Ag}_1, \mathrm{Ag}_2) = 7996$. Notice that this way of calculating distance is sensitive to the number of reinforcements an agent has received. For instance, if an agent A was identical to agent B, but if all its ponderations where ten times higher, there would be a great lexical distance, even though their observable behavior would be identical (at least, if ponderations are sufficiently big not to be influenced too much by the adding of 1). Therefore, the ponderations need to be scaled in case of unequal number of reinforcements per agent.[17]

In figure 10, we have represented the outcomes of 1000 simulation trials with two 3-agent cliques,

[17] The actual ponderation I used was to multiply each term of the formula by the minimum amount of global reinforcements an agent of the network had received, divided by the global amount of reinforcements that particular agent had received, or:

$$\mathrm{LD}(\mathrm{Ag}_o, \mathrm{Ag}_p) = \sum_{i=1}^{k} \left| lc_i(Ag_o) \times \frac{\mathrm{mgr}}{\sum_{i=1}^{m} lc_i(\mathrm{Ag}_o)} - lc_i(Ag_p) \times \frac{\mathrm{mgr}}{\sum_{i=1}^{m} lc_i(\mathrm{Ag}_p)} \right|, \text{ where}$$

$$\mathrm{mgr} = \min\left(\sum_{i=1}^{m} lc_i(\mathrm{Ag}_1), \ldots, \sum_{i=1}^{m} lc_i(\mathrm{Ag}_j)\right)$$

This way, if the agent has received only the minimum amount of reinforcements, the number of reinforcements at his/her lexical components will not be altered; if (s)he has been reinforced more often, that number will be scaled down.

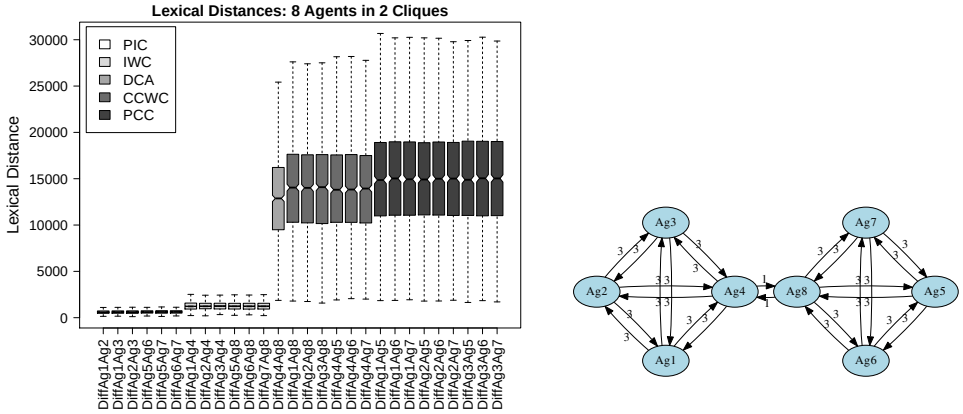

Figure 11: Lexical Distance Mirrors Network Structure II

with an in-clique edge-weight of 2 (the underlying network is depicted on the right). Different colors represent different types of distances between agents. The first one, labeled PIC for Purely In-Clique, corresponds to the distances between agents in the same clique, and where no agent is in direct contact with a member of the other clique. The next category is labeled IWC — In-clique With Contact — and concerns agents in the same clique, but where one agent is in direct contact with the other clique. On average, the lexical distance is larger than in the PIC-group, but the distance is small, with little spread, compared to the remaining categories. The next category is DCA (or Direct Contact Agents, here Agent3 and Agent6), who belong to different cliques but who are in direct contact. Their lexical distance is much bigger, with much greater spread, but it is on average less than the distance between other cross-clique relationships. CCWC (Cross-clique With Contact) combines cases where agents belong to different cliques, do not have direct contact with one another, but where one of the agents is in direct contact with the other clique. On average, their lexical distance is greater than with direct contact agents, but it is less than if none of the two agents had any contact whatsoever with the other group (which is the category PCC, or Purely Cross-Clique).

The same picture obtains if we look at the outcome of 1000 simulation trials with two 4-agent cliques, with an in-clique edge-weight of 3, as in figure 11. We have the same pattern of small lexical distance within a clique, and greater distance across-cliques, with graduations given the amount of contact individual agents had.

4.2 Cases with Three Linked Cliques

In the cases with two linked cliques, the main difference was between intra-clique and inter-clique relations. Does this carry over to cases with three linked cliques? Figure 12 shows that the cross-clique category no longer behaves in a homogeneous way, and that there is another factor that has to be taken into account, namely distance between the cliques.

The key observation is the following: the clique formed by Agents 1–3 is in direct contact with the clique formed by Agents 4–6, but they are in touch with the clique of agents 7–9 only through Agent6. Once we order by clique distance and by the type of contact, things fall into place, and we

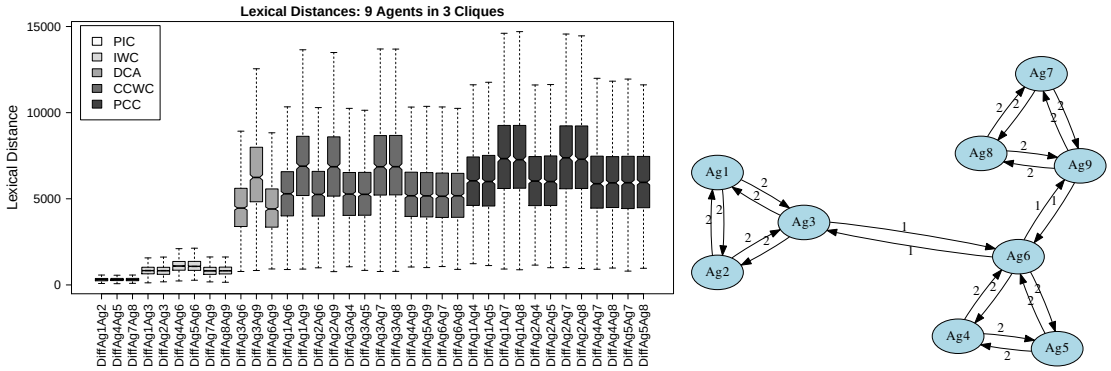

Figure 12: 3 Linked Cliques — In-Clique vs. Cross-Clique

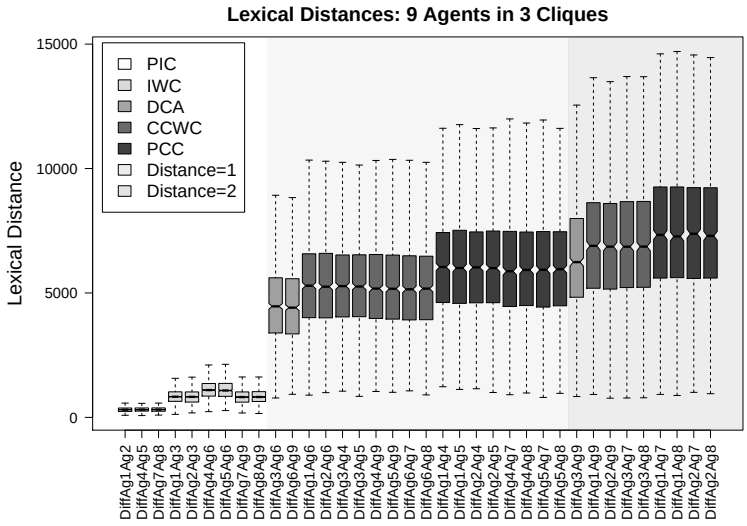

Figure 13: 3 Linked Cliques – Lexical Distance Mirrors Network Distance

get once again the familiar, ascending pattern: in-clique distance is lowest; agents that are in direct contact, but belong to adjacent cliques have relatively lower lexical distance than those that have no contact at all.

While the picture seems rather clear in cliques with few agents, the bigger the number of agents per clique, the less clear the difference in distance between different cliques that are or are not directly adjacent. This is illustrated for a network consisting of 3 cliques with 6 agents per clique in figure 14.[18]

[18] The effect would be more pronounced with even more agents per clique. However, there are limits to reasonable visualisation. Since we are doing pair-wise comparison (excluding comparisons of agents with themselves, which would trivially yield a distance of 0), for a network of n agents, we have $\frac{n(n-1)}{2}$ comparisons. This increase is exponential, and would mean that for a network of 30 agents (10 agents per clique in 3 cliques), we would have $\frac{30*29}{2} = 435$ pair-wise comparisons to plot.

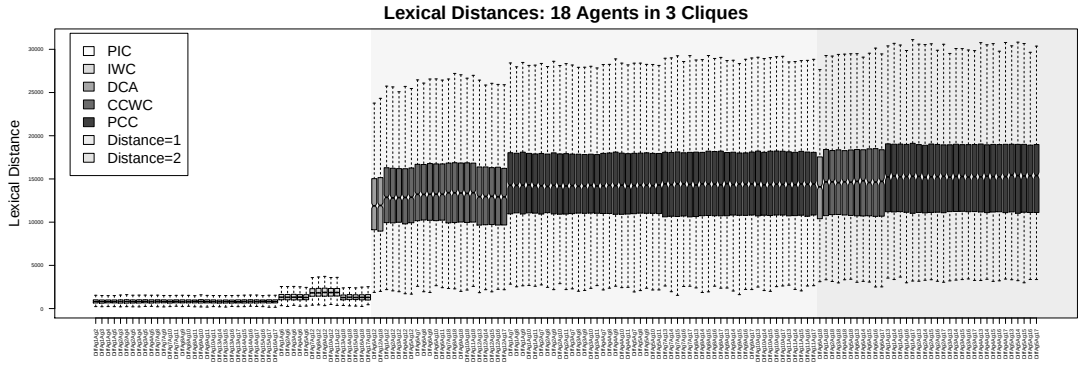

Figure 14: 3 Linked Cliques – Lexical Distance and Network Distance

What could be the reasons for this outcome? My guess would be as follows.[19] The more agents per clique, the smaller the impact of the cross-clique contacts. Therefore, the impact of the contact with another clique simply gets swamped by the mass of reinforcements that happen within the clique. It is probably this (limited, but continuous) cross-clique contact that reduces the lexical distance between adjacent cliques. If agents live forever, and clique size becomes very big, I would therefore expect adjacency between cliques to get less and less important.

The pattern obtained in this section for smaller clique-sizes seems to be able to account for phenomena like the development of dialect continua, where linguistic differences are small in adjacent populations, but get more and more important as populations become separated. But can we extrapolate this result to human societies? Arguably, clique size is not unlimited in human agents (see Sutcliffe et al., 2012), and we are clearly not immortal. Therefore, the basic results obtained in the simulations should carry over to real populations. However, I should iterate once again the central finding when studying complete networks: while contact creates similarity, absence of contact does not create per se dissimilarity; it only makes it improbable that, in the absence of selectional processes, two populations develop in the same direction. Therefore, even if the simulations predict that actual populations that are connected by bridges should be divergent at opposite ends of a network most of the time, this is not a necessary outcome.

To conclude this section, let us take stock of what this paper has shown, and to what degree this is relevant for the sociology and long-term consequences of agents learning a language in a social network. The experimental set-up is highly idealized, and has considered agents without desires, aims, or preferences, whose lexical choices are exclusively determined by the result of reinforcement learning, and are random. The experimental conditions also investigated only highly simplified types of network configurations, which do not apply to humans living ‘‘in the

[19] Ideally, if we could study the limit-behavior of this system mathematically, this issue would be simply a matter of calculation. However, as it stands, the simulation has properties which make it difficult to deal with mathematically, the most severe being that it is not a Markov-chain process. This means that the future state of the system cannot be wholly determined from the current state alone, and without knowing how the present state has been attained. As pointed out by Sylvain Billiard (p.c.), there is a way of transforming the experimental setup into a Markov-chain process, and preliminary results indicate that at least for complete networks, the outcomes are the same. However, it is not clear as of yet whether this identity of outcomes carries over to arbitrary networks.

wild''. Yet, these idealizations provided us with a way of examining the impact of network structure, which was the only element that varied through the different experimental conditions.

Thus, this paper has shown that the following patterns obtain in relation with linked cliques: i) within-clique differences are smaller if none of the agents involved has direct contact with members of other cliques; ii) within-clique differences with a contact agent are smaller if the contact agent is in contact with only one member of another clique; iii) cross-clique differences are smallest when both agents are in direct contact with one another; and iv) cross-clique differences increase with the distance between the cliques the agents belong to.

5 Conclusions and Perspectives

In this paper, I have studied the importance of network structure in acquisition and use of lexical items. We have seen that the size of complete networks has an impact on the outcome of learning, and more precisely, on lexical differentiation. We have also seen that network structure and distance have a direct repercussion on the similarity profiles of the agents in the network (dialect continua, etc.).

As such, this paper has tried to provide a contribution to the phenomenon of lexical drift, that is, lexical change that is not caused by any type of functional or adaptive pressure, as driven by the structure of the social network the learners are engaged in. Other kinds of influences are certainly at work as well, as for instance production and learning biases. Therefore, the results obtained here can provide us with a benchmark against which we can evaluate the claims that some other influence had an impact on certain types of historical developments, and which are not reducible to mere randomness.

An important limitation of the present paper concerns the types of network investigated, which were of a highly idealized sort. In future work, it will have to be checked whether and to what degree these results carry over to networks satisfying the characteristics of standard real-life Human communication networks (as investigated by Sutcliffe et al., 2012).

Acknowledgements

I had the opportunity to present preliminary versions of this paper at the *Séminaire des linguistes* of the UMR 8163 STL, and at the Yearly Meeting of the Belgian Linguists' Association, and I would like to thank the organizers and the audience for their feedback and criticism. I would also like to thank the two anonymous reviewers, as well as Sylvain Billiard and Cédric Patin for their comments on previous versions of this paper. Finally, Katia Paykin helped me with my English. None of them should be assumed to agree with anything I wrote in this paper; all errors and omissions are mine alone.

The simulations have been performed with **sbcl** Common Lisp, using the **graph**-library by Eric Schulte (`https://github.com/eschulte/graph`). Networks have been drawn with **graphviz** (Gansner and North, 2000). Data analysis has been performed with **GNU R**.

References

Bentley, R. A., P. Ormerod, and S. Shennan (2011). Population-level neutral model already explains linguistic patterns. *Proceedings of the Royal Society B: Biological Sciences 278*(1713), 1770–1772.

Benz, A., G. Jäger, and R. van Rooij (Eds.) (2006). *Game Theory and Pragmatics*. New York: Palgrave MacMillan.

Chudek, M., S. Heller, S. Birch, and J. Henrich (2012). Prestige-biased cultural learning: Bystander's differential attention to potential models influences children's learning. *Evolution and Human Behavior 33*(1), 46–56.

Fagyal, Z., S. Swarup, A. M. Escobar, L. Gasser, and K. Lakkaraju (2010). Centers, peripheries, and popularity: The emergence of norms in simulated networks of linguistic influence. In *University of Pennsylvania Working Papers in Linguistics*, Volume 15, Chapter 10, pp. 81–90. Philadelphia.

Gansner, E. R. and S. C. North (2000). An open graph visualization system and its applications to software engineering. *Software — Practice and Experience 30*(11), 1203–1233.

Gilbert, N. (2008). *Agent-Based Models*. Thousand Oaks, CA: SAGE Publications.

Griffiths, T. L. and F. Reali (2011). Modelling minds as well as populations. *Proceedings of the Royal Society of London B: Biological Sciences 278*(1713), 1773–1776.

Han, M. W. and R. A. Bentley (2003). Drift as a mechanism for cultural change: An example from baby names. *Proceedings of the Royal Society B: Biological Sciences 270*(Suppl. 1), 120–123.

Haspelmath, M. (1999). Why is grammaticalization irreversible? *Linguistics 37*(6), 1043–1068.

Jackson, M. O. (2008). *Social and Economic Networks*. Princeton: Princeton University Press.

Kusters, W. (2003). *Linguistic Complexity. The Influence of Social Change on Verbal Inflection*. Ph. D. thesis, Universiteit Leiden, Leiden.

Kusters, W. (2008). Prehistoric and posthistoric language in oblivion. In R. Eckardt, G. Jäger, and T. Veenstra (Eds.), *Variation, Selection, Development. Probing the Evolutionary Model of Language Change*, Trends in Linguistics, pp. 199–218. Berlin: Mouton de Gruyter.

Mühlenbernd, R. and M. Franke (2012). Signaling conventions: Who learns what where and when in a social network. In T. C. Scott-Philips, M. Tamariz, E. A. Cartmill, and J. R. Hurford (Eds.), *The Evolution of Language: Proceedings of EvoLang9*, pp. 242–249. Singapore: World Scientific.

Paul, H. (1995). *Prinzipien der Sprachgeschichte* (9 ed.). Tübingen: Niemeyer.

Pierrehumbert, J. B., F. Stonedahl, and R. Daland (2014). A model of grassroots changes in linguistic systems. *CoRR abs/1408.1985*.

Pijpops, D., K. Beuls, and F. V. de Velde (2015). The rise of the verbal weak inflection in germanic. an agent-based model. *Computational Linguistics in the Netherlands Journal 5*, 81–102.

Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge: MIT Press.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reali, F. and T. L. Griffiths (2009a). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition 111*, 317–328.

Reali, F. and T. L. Griffiths (2009b). Words as alleles: connecting language evolution with Bayesian learners to models of genetic drift. *Proceedings of the Royal Society of London B: Biological Sciences 277*(1680), 429–436.

Schaden, G. (2014). Markedness, frequency and lexical change in unstable environments. In J. Degen, M. Franke, and N. Goodman (Eds.), *Proceedings of the Formal & Experimental Pragmatics Workshop*, Tübingen, pp. 43–50. ESSLLI.

Skyrms, B. (2010). *Signals. Evolution, Learning, & Information*. Oxford: Oxford University Press.

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

Sutcliffe, A., R. I. M. Dunbar, J. Binder, and H. Arrow (2012). Relationships and the social brain: Integrating psychological and evolutionary perspectives. *British Journal of Psychology 103*(2), 149–168. contains estimates of different types and sizes of networks humans are engaged in.