Crucial factors in the origins of word-meaning

L. Steels (1,2), F. Kaplan (1), A. McIntyre (1) and J. Van Looveren (2)

(1) Sony Computer Science Laboratory - Paris
(2) VUB Artificial Intelligence Laboratory - Brussels
steels@arti.vub.ac.be, kaplan@csl.sony.fr, angus@csl.sony.fr, joris@arti.vub.ac.be

0.1. Introduction

We have been conducting large-scale public experiments with artificial robotic agents to explore what the necessary and sufficient prerequisites are for word-meaning pairs to evolve autonomously in a population of agents through a self-organized process. We focus not so much on the question of why language has evolved but rather on how. There are many good reasons to use language once it has come into existence; for example, for establishing and maintaining group coherence, for transmission of cultural knowledge like tool use, etc. But these reasons only explain why verbal behavior is reinforced, it does not explain how this verbal behavior might emerge or complexify. Our hypothesis is that when agents engage in particular interactive behaviors which in turn require specific cognitive structures, they automatically arrive at a language system. This interactive behavior should be a natural outgrowth of cooperative behavior (if this were not the case, the behavior would be unlikely to emerge), and the cognitive structures should be simple enough to have evolved for other purposes as well (pre-adaptation). Our main task is therefore to identify precisely what this behavior is and what cognitive structures are required for it.

Because we study this topic by performing experiments based on artificial systems, our work is comparable to other research attempting to understand the origins of semiotic systems through computer simulations (see (Hurford 1989) and (MacLennan 1991) for some of the earliest papers and a survey of other work in

(Steels 1997)). The research discussed in the present paper differs primarily from previous work in three respects: (1) The experiments are carried out by agents which have contact with the world through a sensori-motor apparatus. They are therefore grounded in reality. (2) The agents have no direct access to the meanings used by other agents; their only access is indirect, and comes in the form of feedback on interactions taking place in the environment. This is important, because simulations in which the agents are presented with both words and meanings during the lexicon-learning process do not allow certain observed phenomena of natural language such as polysemy or meaning evolution to arise. (3) We do not assume a prior repertoire of concepts given to the agents. Instead, agents must build up their conceptual repertoire in a co-evolutionary process simultaneous with the construction of their lexical system. One of the assumptions underlying the present work is linguistic relativism: the conceptual repertoire of the agents both influences and is influenced by the developing lexicon.

The experiment, known as the Talking Heads Experiment, employs a set of visually grounded autonomous robots into which agents can install themselves to play language games with each other. The language games are centred on scenes containing geometrical figures pasted on a white background (see Figure 1). The software system used has been implemented on top of a generic software infrastructure for exploring language games (McIntyre 1998). The language game played is called the guessing game. One agent plays the role of speaker and the other one then plays the role of hearer. Agents take turns playing games so all of them develop the capacity to be speaker or hearer. Agents start without any prior category set or lexicon, and learned knowledge always remains local to the agent. Agents are capable of segmenting the image perceived through the camera into objects and of collecting various sensory data about each object, such as the color (decomposed in the yellow-blue, red-green, brightness and saturation channels), average gray-scale, horizontal and vertical position, size, form, etc. The set of objects and their data constitute the context for a language game. The speaker chooses one

object from this context, hereafter described as the topic. The other objects form the background. The speaker then gives a linguistic hint to the hearer.

The linguistic hint is an utterance that identifies the topic with respect to the objects in the background. For example, if the context contains [1] a red square, [2] a blue triangle, and [3] a green circle, then the speaker may say something like "the red one" to communicate that [1] is the topic. If the context also contained a red triangle, he would have to be more precise and say something like "the red square". Of course, the Talking Heads do not say "the red square": they use their own language and concepts which are never going to be the same as those used in English. For example, they might say "malewina" to mean [UPPER EXTREME-LEFT LOW-REDNESS].[1]

Based on the linguistic hint, the hearer tries to guess what topic the speaker has chosen, and he communicates his choice to the speaker by pointing to the object. A robot points by transmitting the direction in which it is looking. The game succeeds if the topic guessed by the hearer is equal to the topic chosen by the speaker. The game fails if the guess was wrong or if a failure occurred earlier in the game (for instance, because the speaker was unable to categorize or describe the topic). If the speaker has no adequate category to discriminate the topic from the other objects, a new category is generated by a learning process. If the speaker has no word to express a desired category, it constructs a random string and associates that in its lexicon with this category. If communication fails, the speaker gives an extra-linguistic hint by pointing to the topic it had in mind, and the hearer tries to guess what the possible meaning of the speaker's word might have been in the

[1] Meanings can be composed of several "atomic" meanings when there are no atomic meanings that have enough discriminative power to identify the topic on their own. This means that the topic has to possess all of the properties listed in the meaning, while none of the objects in the background may have all of properties. There is no relation between parts of the meaning and parts of the word; words are never decomposed in this system.

present context. This meaning is then stored in the hearer's lexicon. At no point is any global knowledge stored anywhere.

The robots are located in different places in the world (Paris, Brussels, Tokyo, Antwerpen, Lausanne, Amsterdam etc.) and are connected through the Internet. Agents travel from one body to another through an `agent teleportation' infrastructure, although they can only interact when being at the same physical site. The teleportation infrastructure uses the Internet as a way to transfer the software states of the agents from one location to another. Some of the installation sites have been at public places: art galleries, museums, and conferences. We estimate that close to 300,000 people have seen the experiment live, most of them at the Paris science museum, le *Palais de la découverte*. Agents are initially launched by human users. Through a web-page (<http://talking-heads.csl.sony.fr/>), anyone can follow the experiment and interact with the agents. Through the web interface, human users may teach words to their agents and thus influence the evolving lexicon. In this way we have also been able to explore human influence on the emerging artificial language. We estimate that between 10,000 and 15,000 people have also visited the experiment's web site and close to 6,000 agents have been launched by sometimes very active human users.

Figure 1. Typical example of a Talking Heads setup. Two steerable cameras are connected to computer equipment and oriented towards a white board on which geometric figures are pasted.

The first `Talking Heads' experiment ran for 4 months during the summer of 1999 and showed the validity of the mechanisms that were used for the agent architecture and of the interaction patterns and group dynamics of the agents. A shared lexicon and its underlying conceptual repertoire emerged after a few days, enabling successful communication by the agents about the scenes before them. In total, 400,000 grounded games were played. The population of agents rose to just under 2000, increasing steadily over the period of the experiment. Despite the many perturbations due to grounding, intermittent

technical failures, a continuous influx of new agents entering the population, and unpredictable human interaction, the lexicon was maintained throughout the period. A total of 8000 words and 500 concepts were created, with a core vocabulary consisting of 100 basic words expressing concepts like up, down, left, right, green, red, large, small, etc. A second experiment cycle was begun at the end of January 2000 and continued until August 2000. After a difficult initial period, in which an overly high agent influx prevented a shared language from establishing itself, a successful language nevertheless emerged (Kaplan 2001).

The success of the experiment comes from a specific self-organizational dynamic which assumes a positive feedback loop between use and success. Agents keep in their memories scored associations between words and meanings. The score reflects the expectation that the word has a given meaning (one can also say that it reflects the probability that the word will be used with this particular meaning in the group). When a game succeeds, i.e. when the hearer can correctly identify the topic selected by the speaker through the utterance, speaker and hearer each increase the score of the used association and decrease the score of competing associations. When a game fails, speaker and hearer decrease the score of the association that was used. This dynamics, which are similar to those adopted in other simulation work (e.g. (Oliphant, 1996)), have a winner-take-all structure (Figure 2), in the sense that synonymy (one meaning/many words) is damped.

Figure 2. Evolution of word-meaning relations for the same meaning over a period of 90,000 games (as taken from the experiment). The graph shows the average use of all the words per 10 games. One word (*wogglesplat*) comes out as the winner.

A situation in which one word form may have multiple meanings (polysemy) occurs naturally in a semiotic system as soon as hearers have to guess the meaning of unknown words. A word can usually have more than one meaning in a given situation and therefore there is no guarantee that the hearer infers the same meaning

as the one intended by the speaker. But here too a damping effect occurs. Meanings that are compatible with the same situations will remain entangled until clear situations arise where they are different. This was for example the case for the word "*bozopite*" (Figure 3). There are two competing meanings: large area (large) and large width (wide). These meanings co-occur often because objects that are large in area typically also have a large width. However when there are enough situations where the two are incompatible (for example because the object is very tall but not very wide), disambiguation starts and is enforced by the positive feedback loop. Figure 3 also illustrates that lexicons are never completely stable. Evolution continues to take place mostly under the influence of various sources of randomness that are unavoidable in real world communication (Steels and Kaplan 1998a), (Steels and Kaplan 1998b).

Figure 3. Evolution of word-meaning relations for the same word "bozopite" over a period of 250,000 games (as taken from the actual experiment). The graph shows the average use of all the meanings per 10 games. There is a struggle between two meanings – large area and large width – until one emerges as the winner.

Apart from the damping of synonymy and polysemy, we also see a natural selection towards those concepts and words that are most stable, in the sense that they work perceptually across different environments and are more reliable for picking out the topic from the other objects in the context. Thus very fine-grained distinctions based on subtle light variations would arise in the system but would have a hard time to propagate to the rest of the population because light conditions vary, even when the same set up is viewed from different angles. The net effect of these tendencies was that the total set of words became progressively restricted to a core of hundred words, with a key vocabulary of 8 words referring to colors (red, green, blue and bright) and positions (left, right, up, down). Figure 4 shows the number of words in frequent use as the experiment progresses. Steels and Kaplan

(Steels and Kaplan 1999) discusses in more detail the semiotic dynamics that we have seen emerging.

Figure 4. Evolution of word diversity. After an initial period in which many words were used, the system stabilizes around a kernel of about a hundred words which are sufficient to deal with the situations encountered.

The goal of the remainder of this paper is to identify the factors that we found to be crucial for the success of these experiments. These can be grouped into two subsets: internal factors relating to the individual architecture of the agents and external factors relating to the group dynamics and the environments encountered. We report not only those factors that we explicitly incorporated in the experiment but also the ones that we expressly omitted in order to prove that they are not needed.

We believe that the same factors must have been in place to allow the origins of human language. Indeed, the role of simulations and experiments is to show the strength and limitations of certain theoretical models (Kaplan 1998). The fact that human language first arose in the very distant past means that direct observation of the origins of human language is impossible; approaches based on simulation thus offer perhaps the best option for a rigorous evaluation of theoretical models.

The simulation described here deals with formation of a lexicon. Human languages are characterized by having complex grammatical conventions. Although we have been working on the problem of how grammar may arise in a setup similar to that of the Talking Heads experiment (Steels 1998), this issue is not discussed in the present paper. In any case, a firm vocabulary must be in place before more complex grammatical constructions can be envisioned.

0.2. Internal Factors

0.2.1 Agents must be able to engage in coordinated interactions.

This means that they must be able to have shared goals and a willingness to cooperate. To enable coordinated interaction, each agent must be able to follow a script of actions in agreement with a shared protocol, and have a way to see whether the goal of the interaction has been satisfied. In our experiment, we simply assumed these obvious requirements and explicitly programmed into each agent the scripts achieving the desired cooperative interaction. Emergence of cooperation is not addressed in our research, but it is addressed in other work (see e.g. (Lindgren and Nordal, 1991)). The emergence of communication as part of cooperation has also been addressed by a number of researchers (see e.g. (Noble and Cliff 1996)). On the other hand, the emergence of shared interaction scripts is – as far as we know – an open problem, not only for human communication but for other forms of social interaction – such as physical cooperation in a shared task – as well..

0.2.2. Agents must have parallel non-verbal ways to achieve the goals of verbal interactions.

The communicative goal of the agents in our experiment is to draw attention through verbal means to an object in a visually-perceived reality. There are of course many other things humans do with language but this is surely an important one and a prerequisite for more sophisticated verbal exchanges, such as requests for action that require reliable identification of the objects involved in the action. We have found that it is crucial that the agents have a non-verbal way to achieve this goal: by pointing, gaze following, grasping, etc. This alternative way must be sufficiently reliable, at least initially when the system is bootstrapping from scratch. Once the language system is in place however, external behavioral feedback is less crucial or may be absent altogether. A non-verbal means of communication is a necessity if the hearer has no 'telepathic capacity' to know what meanings are intended by the

speaker, no prior innate categories (as in many other experiments in language evolution), or any other way to guess what might be the meaning independently from language.

Evidence for Claim 2.2. comes from two sources. (1) In the experiment, the hearer physically points to the object that the speaker indicated through verbal means and the speaker physically points to the object when the speaker made the wrong guess. Pointing is done by moving the camera in the direction of the object and zooming in on the object. We have found that major problems occurred during certain phases of the experiment in which the calibration of the pointing behavior was off due to physical movement of the robot beyond our control. In these cases the feedback introduced random errors and destroyed the communication system in place. (2) We also carried out some simulation experiments to explicitly test the importance of non-verbal interaction (see Figure 5) (Steels and Kaplan 1998a). In these experiments the accuracy of non-verbal feedback is a parameter that can be varied. We see that in Phase 1, no language formed due to a very high randomness in non-verbal interaction. In Phase 2, we decreased this randomness and a shared lexicon formed. In Phase 3, we increased the randomness of non-verbal feedback and observed that communicative success (and lexical stability) remained at very high level. In other words, the conventions, once established, are strongly enough ingrained in the population to overcome pointing randomness. To determine the referent of an utterance, the hearer uses both knowledge of the language and knowledge about the situation. Rather than simply choosing the word with the highest score, the best interpretation in the current situation will be chosen.

Figure 5. Simulation experiments with 20 agents and 10 meanings for 20,000 games. The x-axis plots consecutive games and the y-axis average success per 10 games. We see three phases depending on a randomness factor ET.

0.2.3 Agents must have ways to conceptualize reality and to acquire these conceptualizations, constrained by the semantic concepts expressed in the emerging lexicon and the types of situations they encounter.

Obviously, conceptualization precedes verbalization. A word like "left" does not refer to a specific object, but assumes that objects are conceptualized based on their horizontal position and that the object is assumed to be in one subregion along this dimension. So words express categories as opposed to names of specific situations. We believe that it is unrealistic to assume that the repertoire of concepts is fixed and given in advance (as is often done in simulations) because agents will always be confronted with new situations and new tasks. So there must be a concept acquisition process coupled to the language process. Which concept acquisition process is used is not critical. It can be based on neural network techniques like perceptron-style feedforward networks, Kohonen networks (Bishop 1995), or symbolic machine learning techniques (Mitchell 1997). We have used a specific method based on the learning of decision trees, more specifically binary discrimination trees (Figure 6) in a selectionist fashion (Steels 1997). Although there are many possible mechanisms for concept acquisition, we found some important constraints for this process:

The concept formation processes of the agents must be based on similar sensory channels and result in similar although not necessarily equal conceptual repertoires.

We have incorporated this constraint at present by giving each agent the same low level sensory apparatus (vision) and by assuming the same sort of binary discrimination trees for the agent's conceptual repertoires. There are other researchers who have used significantly different types of agents (Yanco and Stein 1993) so that this constraint appears to be less strict, however the vocabularies and concept repertoires of the agents have been too small in these experiments to pose a

serious search problem when agents have to guess the meaning of unknown words. Conceptualization schemes based on randomly-structured discrimination trees, prototypes or neural networks are also adequate for finding a distinctive conceptualization but we found that they result in larger differences between the repertoires of the agents, making it more difficult to achieve lexical coherence in the population. Coherence may still be achieved, as shown in the experiments of Vogt (Vogt 1999) which use a prototype-based categorization, but such systems tend to be less effective than the ones used in this experiment, even when the concept repertoire is small enough to render the search problem negligible.

Figure 6. Example of a part of the conceptual repertoires of a single agent. Each tree divides a sensory channel into subregions, associating a concept with each region.

The conceptualization for a particular situation must be constrained to be similar so that the agents have a reasonable chance at guessing the conceptualization that a speaker may have used.

Even if there is a more or less shared repertoire, there will still be many possible ways to conceptualize reality. For example, in a particular scene containing a red triangle to the left and a blue square to the right, three distinctions – red versus blue, triangle versus square and left versus right – are all adequate. We found that if the search space for possible meanings in a given situation is too large, the agents do not manage to reach a highly coherent lexicon.

In the Talking Heads experiment, we have reduced the search space in two ways. First, by using saliency: sensory differences that stand out more will be preferred for conceptualizing the scene, thus reducing the search space for the meaning of unknown words. Thus if there are three objects all with strongly different colors in the scene, two on the left and one on the right, then color will be preferred to position because it is more distinctive. The second constraint comes from taking the lexicon into account for conceptualization. When there are two

concepts which are equally salient, but one has a stronger lexicalisation (i.e. a word with a higher score) than the other, then the first one is chosen. The latter leads to a steadily increasing coherence in the ontological repertoires of the agents, and thus shows how linguistic relativism is possible.

0.2.4 Agents must have ways to recognize word forms and reproduce them.

This is quite obvious, because otherwise words would be confused all the time. In the Talking Heads experiment we have simply given the agents the capability to recognize and reproduce each others' word forms perfectly. We have also done some simulation experiments (Steels and Kaplan 1998a) to test the validity of Claim 2.4. by introducing an error rate on the transmission of signals (see Figure 7). When this error rate is too high the communication system does not get off the ground (Phase 1). When the error rate is lowered, so that word forms are better recognized, we see a communication system forming (Phase 2). When the error rate is increased again, we see that communicative success decreases (Phase 3) but still stays very high. We see again that once a lexical system is in place it can overcome the randomness inherent in verbal communication.

Figure 7. The x-axis plots for the number of games and the y-axis the communicative success. A parameter has been introduced to vary the accuracy with which signals are recognized and reproduced by agents.

0.2.5 Agents must have the ability to discover and use the strongest associations (between words and meanings) in the group.

The associative memory of an agent must be two-way (from words to meanings and meanings to words), must handle multiple competing associations (one word-many meanings, one meaning-many words), and must keep track of a

score that represents how well the association has been doing based on their own past experience. When a decision must be made (which word to use, which meaning to prefer), there is an internal competition between different associations in which the one with the highest score wins. There are still many possible ways to achieve each of these behaviors. For example, the score updating can be based on use and success, or on a simple score that goes up and down with every usage, or on more sophisticated mechanisms.

Simulation experiments have been conducted to test each of these claims and explore different variations (Kaplan 1999). A very important outcome of these experiments is that statistical learning is not sufficient to bootstrap a coherent mapping between words and meanings. Such a mapping needs to be efficient for communicating, which means that if, for instance, a meaning is associated with several words, all these words need to be decoded in the same meaning. Agents need to test whether the associations they use actually lead to success in communication. If agents are only using statistical learning they have no reason to agree on an efficient mapping. Their mapping need not even be internally coherent.

Our experiments also show that the convergence towards a shared mapping between words and meanings is a lot easier when the number of words available is greater than the number of meanings. This assumption is realistic in the case of the emergence of human language, because languages typically allow for an open-ended set of potential words within a given phonological framework, but was not made in some other work (c.f. (Oliphant 1996) in the sense that it assumed a finite set of words fixed a priori.

Assuming that humans possess a two-way associative memory of this kind makes sense in a wider evolutionary context. Such a memory would be useful for many other tasks as well, such as associating physical locations with sources of food. It is therefore reasonable to assume that cognitive structures of this type are likely to evolve, and that they need not do so solely and specifically for language purposes. This justifies the preadaptation hypothesis.

It is perhaps important also to point which factors we did NOT incorporate in the experiment because we feel that they are superfluous:

a. Theory of mind. There is a widespread belief that verbal communication requires a strong theory of mind of the other agents before verbal interactions are possible. In our experiment, this is not the case, although for more sophisticated language games (such as for reference to abstract entities or belief-states) it is obviously required. To support the emergence of a language directly grounded in visual and motor interaction, it is sufficient that agents follow specific protocols of interaction. They do not need to know why these protocols are successful. (Just as a child does not need explicit knowledge of theories of physics to throw a ball but just has to acquire the appropriate behaviors compatible with these laws.)

b. Prior concepts. Another widely-made assumption is that concepts (particularly the perceptually grounded concepts that are the focus of our experiment) need to be shared prior to and independent of language. For some cognitive researchers this implies that they are innate (Fodor 1998). For others, it suggests that they are acquired through a universal inductive mechanism that yields the same concepts for all agents (Harnad 1990). We do not assume a prior set of categories in our experiments and in fact believe this to be impossible given the adaptive nature of verbal communication. Instead we have set up a strong interaction between language acquisition and concept formation: The repertoire of categories develops in a selectionist fashion under pressure from the language and concepts which have no success in verbal interaction are not encouraged.

c. Telepathy. We have not assumed that agents have a way of knowing what meaning the speaker transmits independently of language. Although non-verbal communication, similarity of sensors, shared history of past experiences, saliency, etc. help to restrict the set of possible meanings, the hearer can only guess what the speaker meant. Neither have we assumed that agents have exactly the same perception. Usually raw perception and consequently-derived sensory features are

different. Equal perception is of course an unrealistic assumption for embodied agents because each agent sees the scene from a different point of view.

0.3. External factors

0.3.1. There must be sufficient group stability to enable a sufficient set of encounters between agents.

We have found that if there is a too rapid in- and outflux of agents, a lexicon will collapse because there is not enough time for new members to acquire the conventions (so they build their own) and older members leave too quickly so that there is no memory in the population of the existing conventions. The exact critical levels of the fluxes depend on the size of the population and on the complexity of the environment. Evidence for Claim 3.1. is borne out by simulation experiments, as shown in Figure 8, concerned with varying the increase in population. In Phase 1 no new agents enter or leave and complete communicative success and lexical coherence is reached. In Phase 2, there is a replacement of 1 agent every 100 games. This lowers success and coherence but the population can cope. In Phase 3, the replacement rate is 1 agent every 10 games. The lexicon collapses. The cumulative change measure increases each time a meaning is coded by a new word in the global lexicon. When no new words are invented for existing meanings the cumulative change is zero. Figure 8 shows that the lexicon is transmitted without alteration in the first two phases but changes rapidly in the last phase.

Figure 8. The graph shows the communicative success of the emergent lexicon and the lexicon change over time in a population of 20 agents over 15,000 games and with different population renewal rates.

0.3.2. Initial group size should not be too large, so that there are enough encounters between the same individuals.

Obviously when the group size is large, it is going to be more difficult to establish lexical coherence. This is partly because there is a much greater chance that new words will be created by individuals or groups of individuals before the mass effect of self-organization can have an impact, and partly because a minimum number of interactions is required: due to the limited number of sites available, the more agents there are the fewer games they have in the same time period. Once a lexicon is in place however, there can be an almost unbounded increase in the population. In the second run of the Talking Heads experiment, a very large population was created, causing significant difficulties for language formation. In the first experiment the population grew more gradually so that after a week a solid lexicon had already emerged and remained in place for the remaining months of the experiment, despite an agent population that ultimately contained some thousands of agents.

Each agent site typically had a population that focused on that site (as directed by their users). Because these agents had more interaction with each other than with agents at other sites sub-lexicons formed that were unique to the site. The geographical separation was never total because other agents (again as directed by their users) traveled a lot in between different sites. In another series of experiments with spatially distributed language games, phenomena familiar in studies of language contact started to appear when the importance of this geographical separation was reduced; under such circumstances, the language from the largest population tends to dominate (Steels and McIntyre 1998).

0.3.3 There must be sufficient environmental stability and different degrees of complexity.

The environments encountered by the agents and perceivable by the agents through their sensory apparatus must have certain invariant structural properties so that concepts can form and word-meaning pairs can settle. This does not mean that the environment needs to be closed (indeed it should not be if we want to be realistic), nor even that the sensory space should be closed (new sensory routines surely develop in children even after they have acquired their first words).

We found that if the agents encounter only complex scenes, they cannot settle on a successful repertoire or at least have much greater difficulty in doing so due to unstable concepts. So there must be scenes, at least initially, which can be handled by making simple distinctions (such as between left and right). Such simplified environments can be seen as analogous to the initial environment of infants: because many sensory capabilities are not available at birth, the child learns its initial categorizations in the context of what is effectively a simplified perceptual environment (Elman et.al. 1996).

We were also able to exclude two external factors which we determined were unnecessary.

a. Global view or central control. A central puzzle in the origins of language is how a population of distributed autonomous agents can reach coherence without a central controlling organism and without giving individual agents access to a global view. A model should never introduce any central control on language or give agents a global view of the language as a whole. Our experiments have shown extensively and convincingly that self-organization is perfectly adequate to explain language coherence without this.

b. Total coherence. It is often assumed that all individuals have exactly the same linguistic competence and that deviations are only due to performance

errors. We have shown that this assumption is unnecessary. The conceptualizations and lexicons of the individual agents in the experiment were NEVER exactly the same. They had different degrees of knowledge and there were unavoidable individual differences arising from the absence of a global view. The experiment shows that communicative success can nevertheless be reached without such absolute coherence. For example, words can often be maintained in a polysemous state without causing confusion in a series of environments, while synonyms are tolerated because agents can understand words that they themselves might not necessarily choose to use.

0.4. Conclusion

This paper aims to show how experiments based on software simulations or robotic setups, like the Talking Heads experiment, can play an important role in the debate on the origin and evolution of human languages. In a field where "real" experimentation is not possible, this type of experiments allows researchers to compare hypotheses and use models to test which factors are crucial and which are contingent to achieve a communication system. Similar experiments have also studied the emergence and complexification of grammar (Steels 1998).

In summary, we have established the following internal factors for the evolution of a lexical system in a group of distributed agents without prior set of categories nor lexicon and without telepathic capability to guess meanings independently of physical action(like pointing) or language:

1. Agents must be able to engage in coordinated interactions;
2. Agents must have parallel non-verbal ways to achieve the goals of verbal interactions.
3. Agents must have ways to conceptualize reality and to form these conceptualizations, constrained by the set of categories underlying the emerging lexicon and the types of situations they encounter.

4. Agents must have ways to recognize word forms and reproduce them and
5. Agents must have the ability to discover and use the strongest associations (between words and meanings) in the group.

We also established the following external factors:

1. There must be sufficient group stability to enable a sufficient set of encounters between agents,
2. Initial group size should not be too large so that there are enough encounters between the same individuals,
3. There must be sufficient environmental stability and different degrees of complexity in the environment.

These various constraints help us to understand the conditions under which language may have emerged in human societies and what kind of minimal cognitive and sensori-motor architecture is needed to get a lexical system off the ground.

0.5. Acknowledgement

This work was carried out at the Sony Computer Science Laboratory in Paris and the VUB AI laboratory in Brussels (financed by a GOA grant). We are extremely grateful to the various sites that hosted physical installations, in particular to Barbara Vanderlinden and Hans-Ulrich Obrist who organized the LABORATORIUM exhibition in Antwerpen, to Adam Lowe who organized the NO1SE exhibition that lead to sites in Cambridge (UK) and London, to Marie Canard and Eric Emery who organized the exhibition on Animal Communication in the Palais de la Découverte in Paris, and to Ben Krose who was instrumental in getting the Amsterdam site operational. We are also indebted to the reviewers and editor of this volume for many useful comments regarding the text.

0.6. References

Bishop, C.M. (1995) *Neural Networks for Pattern Recognition.* (Oxford: Oxford University Press).

Elman, J, Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. (1996) *Rethinking Innateness.* (Cambridge, MA: MIT Press)

Fodor, J. (1998) *Concepts: Where Cognitive Science Went Wrong*. (Oxford: Clarendon Press).

Harnad, S. (1990) The Symbol Grounding Problem. *Physica* D 42: 335-346.

Hurford, J. (1989) Biological Evolution of the Saussurean Sign as a Component of the Language Acquisition Device. *Lingua* 77, 187-222.

Kaplan, F. (1998) Role de la Simulation Multi-agent pour Comprendre l'origine et l'évolution du Language. In: Barthès, J-P, V. Chevrier, and C. Brassac, (eds), *Systèmes multi-agents: de l'interaction à la socialité*, (Paris: Hermès) , 51-64

Kaplan, F. (1999) "Dynamiques de l'auto-organisation lexicale: simulations multi-agents et 'Têtes parlantes'", In *In Cognito*, 15 , 3-23.

Kaplan, F. (2001) La naissance d'une langue chez les robots (Paris: Hermès)

Lindgren, K., and M. Nordal (1991) Coooperation and Community Structure in Artificial Ecosystems. In: Langton, C. (ed.) *Artificial Life. An overview*. (Cambridge: The MIT Press), 15-37.

MacLennan, B. (1991) Synthetic Ethology: An approach to the study of communication. In: Langton, C., C. Taylor, J.D. Farmer and S. Rasmussen (eds.) (1991) *Artificial Life II*. (Reading: Addison-Wesley), 631-658.

McIntyre, A. (1998) Babel: A testbed for research in the origins of language. In *Proceedings of Coling-ACL 98*, (Montreal: ACL). 830-835.

Mitchell, T. (1997). *Machine Learning* (New York: McGraw Hill)

Noble, J. and D. Cliff (1996) On simulating the evolution of communication. In: Maes, P., M. Mataric, J.-A. Meyer, J. Pollack and S.W. Wilson (eds.) *From Animals to Animats 4: Proceedings of Fourth International Conference on Simulation of Adaptive Behavior*. (Cambridge, MA: the MIT Press), 608-617.

Oliphant (1996) The Dilemma of Saussurean Communication. *BioSystems*, 37:31-38.

Steels, L. (1997) The Synthetic Modeling of Language Origins, *Evolution of Communication Journal* 1 (1): 1-34.

Steels, L. (1998) The Origins of Syntax in Visually Grounded Robotic Agents. *Artificial Intelligence*, 103:1-24.

Steels, L. and Kaplan, F. (1998a) Stochasticity as a source of innovation in Language Games. In: Adami C., R. Belew, H. Kitano and C. Taylor (eds.), *Proceedings of Artificial Life VI*, (Cambridge, MA: MIT Press), 368-376

Steels, L. and Kaplan, F. (1998b) Spontaneous lexicon change. *Proceedings of Coling-ACL 98*, (Montreal: ACL). 1243-1249

Steels, L. and Kaplan, F. (1999) Collective Learning and Semiotic Dynamics. In: Floreano, D., J.D. Nicoud and F. Mondada (eds) (1999) *Advances in Artificial Life* (Proceedings of ECAL 99). Lecture Notes in Computer Science. (Berlin: Springer-Verlag), 679-688.

Steels, L. and McIntyre, A. (1998) Spatially Distributed Naming Games, *Advances in Complex Systems* 1 (4), 301-324.

Vogt, P. (1999) Grounding a Lexicon in a Coordination on Mobile Robots. In Postma, E. and Gyssens, M. (eds). *Proceedings of Eleventh Belgium-Netherlands Conference on Artificial Intelligence.* (University of Maastricht), 275-276.

Yanco, H. and Stein, L. (1993) An adaptive communication protocol for cooperating mobile robots. In: Meyer, J-A., Roitblat, H.L. and Wilson, S.W. (eds) (1993). *From Animals to Animats 2: Proceedings of the Second International Conference on the Simulation of Adaptive Behavior*. (Cambridge, MA: The MIT Press/Bradford Books). pp. 478-485.

--------- FIGURES ----------

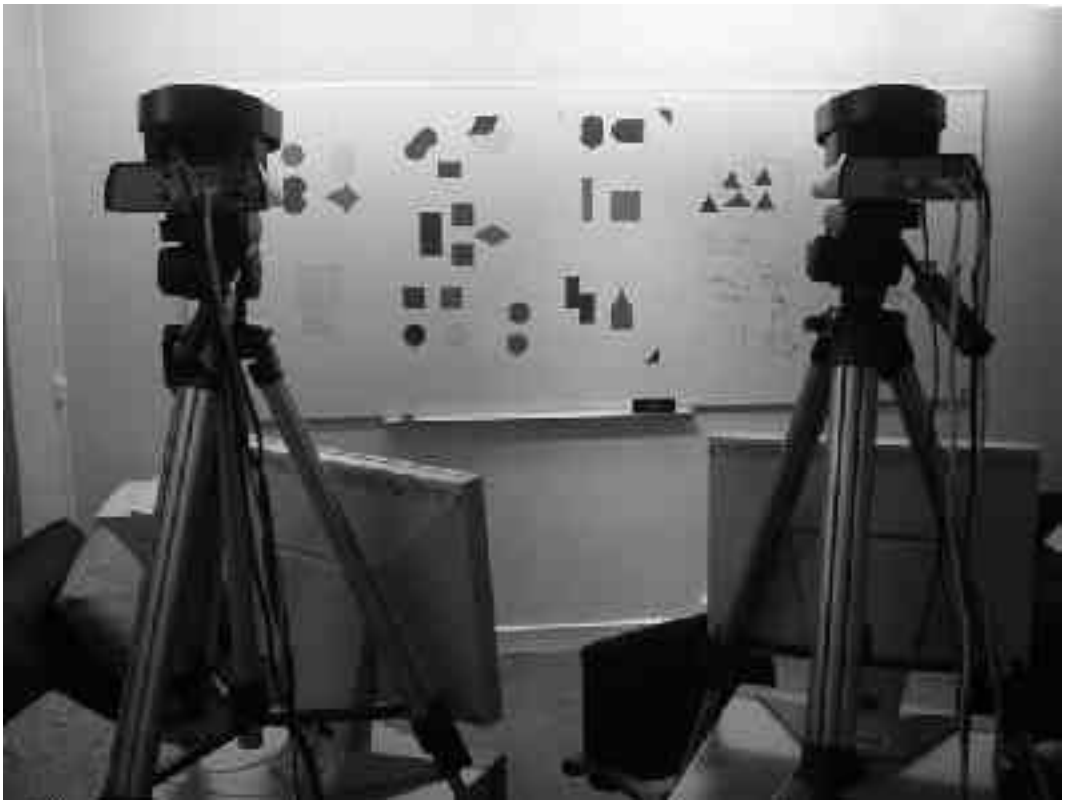

Figure 1. Typical example of a Talking Heads setup. Two steerable cameras are connected to computer equipment and oriented towards a white board on which geometric figures are pasted.

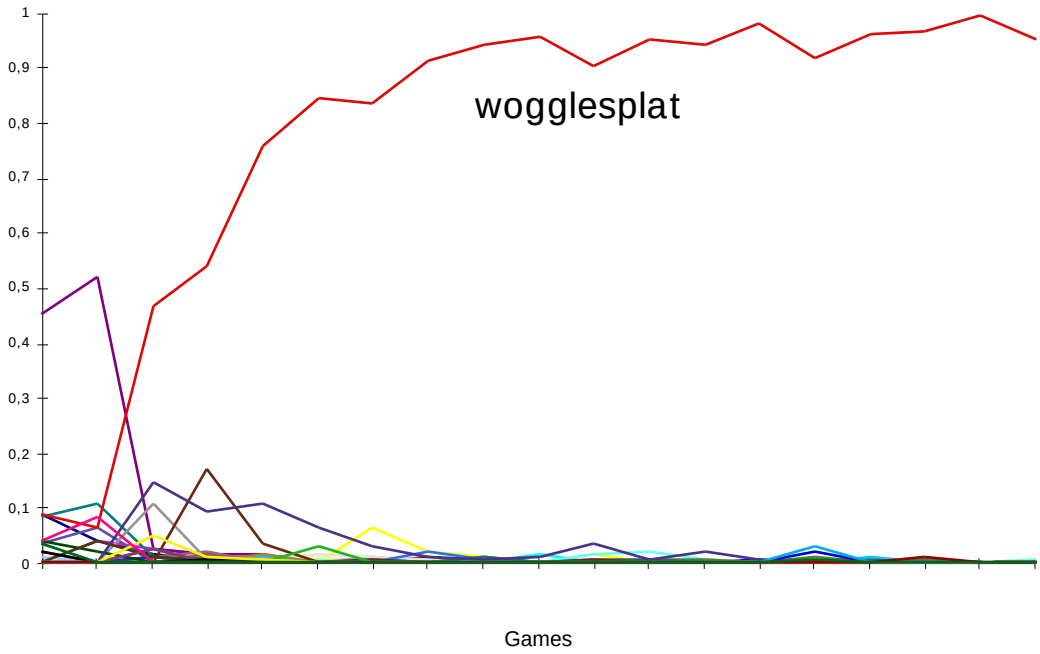

Figure 2. Evolution of word-meaning relations for the same meaning over a period of 90,000 games (as taken from the experiment). The graph shows the average use of all the words per 10 games. One word (wogglesplat) comes out as the winner.

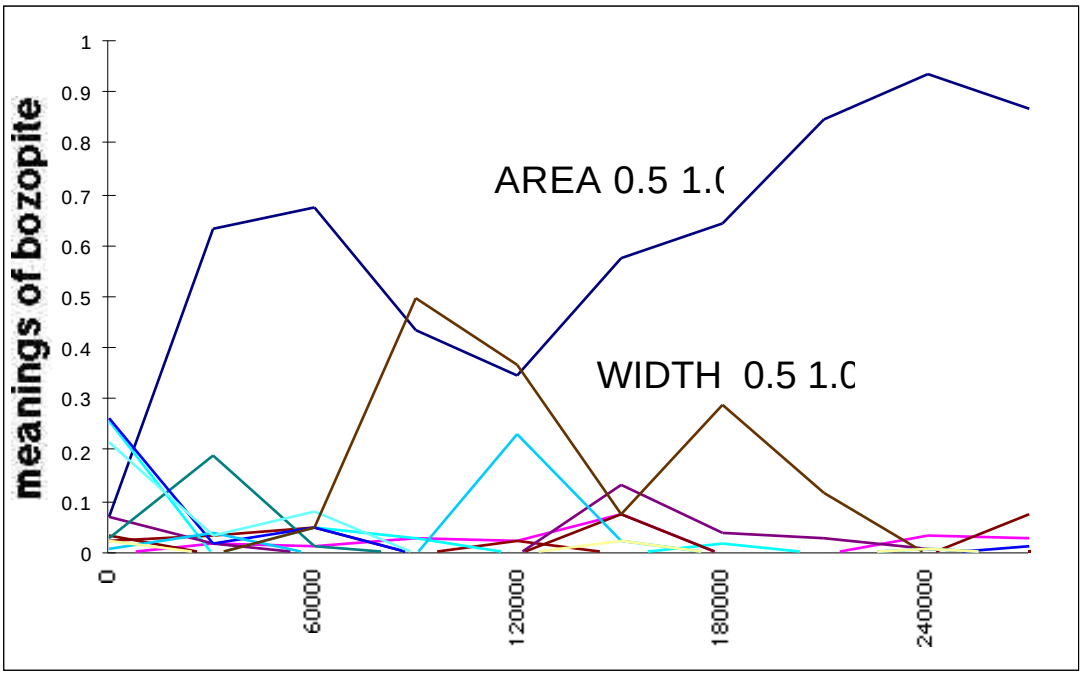

Figure 3. Evolution of word-meaning relations for the same word "bozopite" over a period of 250,000 games (as taken from the actual experiment). The graph shows the average use of all the meanings per 10 games. There is a struggle between two meanings: large area and large width, until one comes out as the winner.

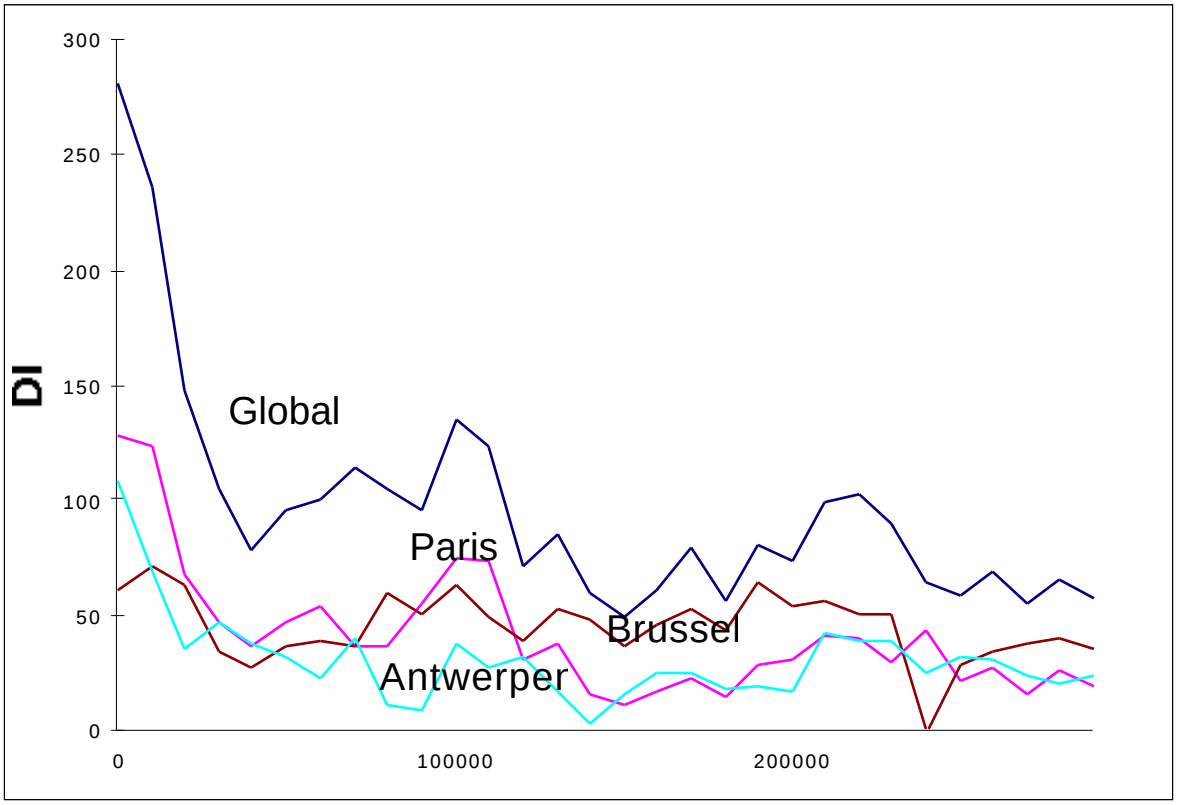

Figure 4. Evolution of word diversity. After an initial period in which many words were used, the system stabilizes around a kernel of about a hundred words which are sufficient to deal with the situations encountered.

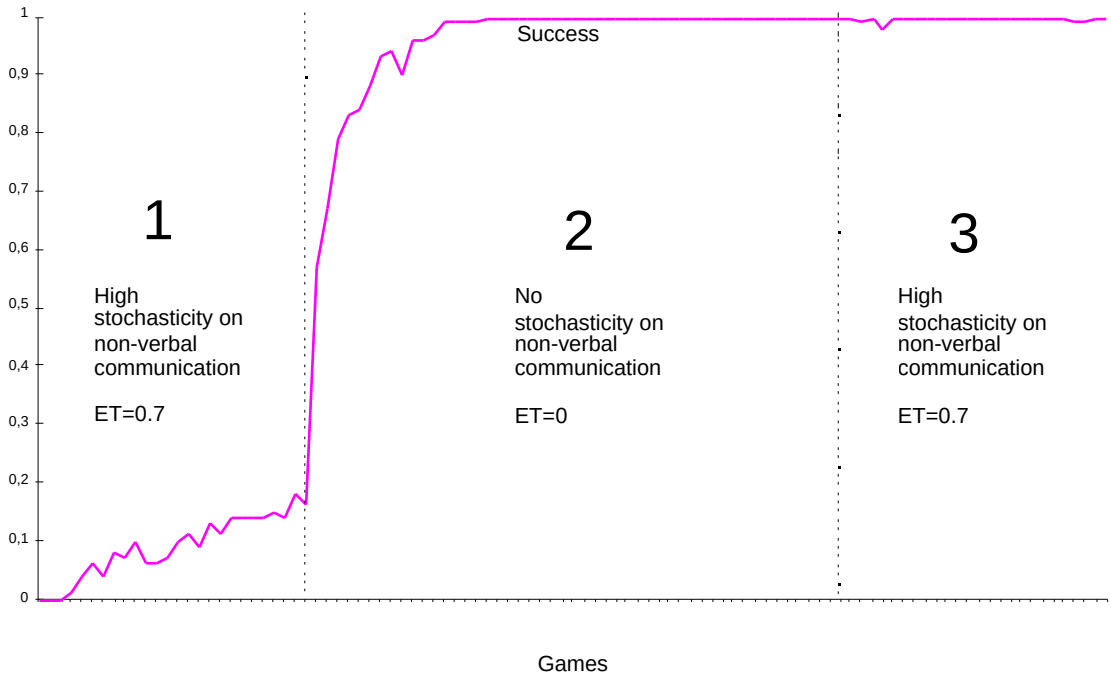

Figure 5. Simulation experiments with 20 agents and 10 meanings for 20,000 games. The x-axis plots consecutive games and the y-axis average success per 10 games. We see three phases depending on a randomness factor.

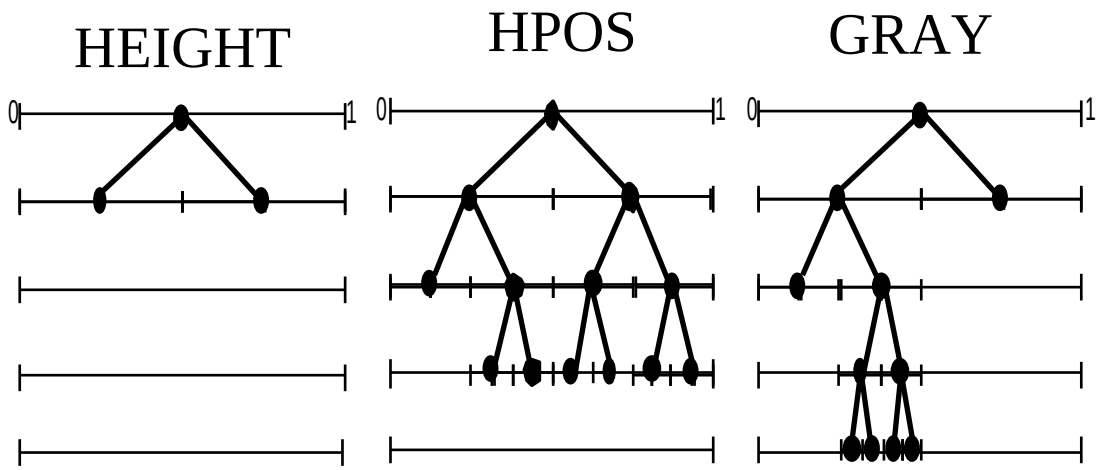

Figure 6. Example of a part of the conceptual repertoires of a single agent. Each tree divides a sensory channel into subregions, associating a concept with each region.

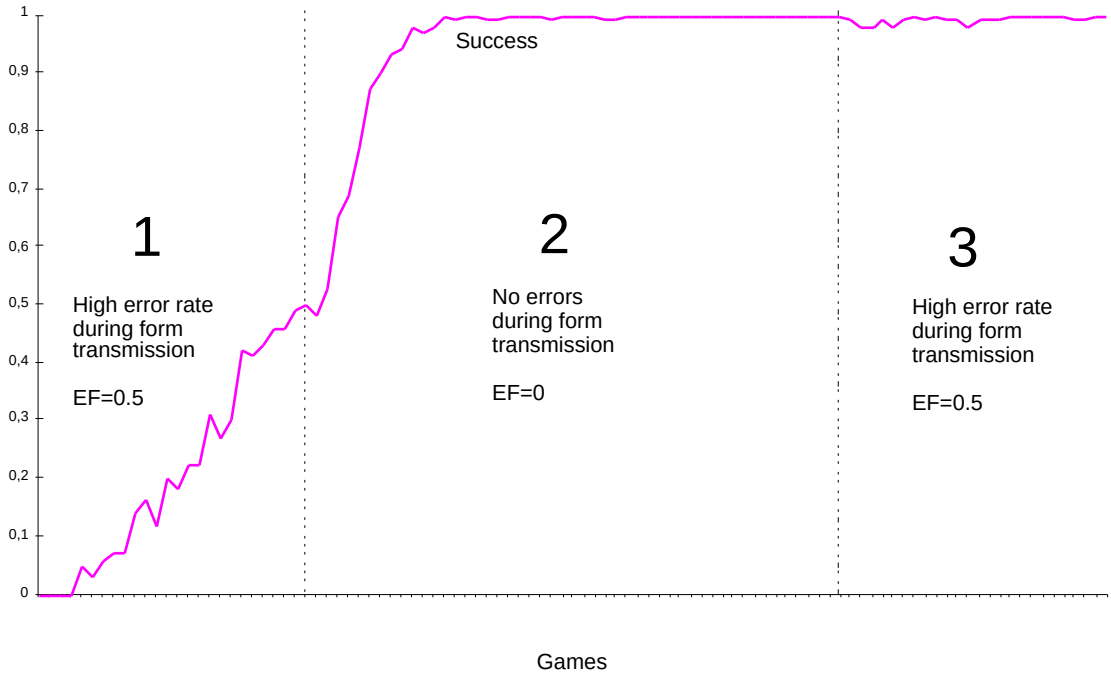

Figure 7. The x-axis plots for the number of games and the y-axis the communicative success. A parameter has been introduced to vary the accuracy with which signals are recognized and reproduced by agents.

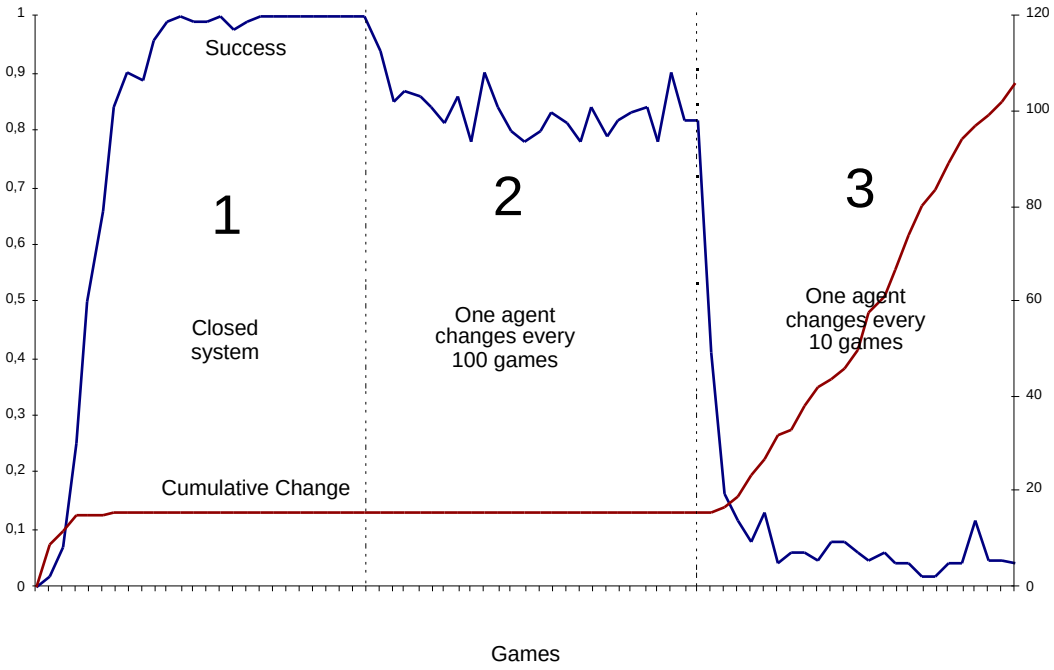

Figure 8. The graph shows the communicative success of the emergent lexicon and the lexicon change over time in a population of 20 agents over 15,000 games with different population renewal rates.