A hybrid model for learning word-meaning mappings*

Federico Divina[1] and Paul Vogt[1 2]

[1] Induction of Linguistic Knowledge / Language and Information Science
Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands
[2] Language Evolution and Computation Research Unit, School of Philosophy,
Psychology and Language Sciences, University of Edinburgh, U.K.
f.divina@uvt.nl,paulv@ling.ed.ac.uk

Abstract. In this paper we introduce a model for the simulation of language evolution, which is incorporated in the New Ties project. The New Ties project aims at evolving a cultural society by integrating evolutionary, individual and social learning in large scale multi-agent simulations. The model presented here introduces a novel implementation of language games, which allows agents to communicate in a more natural way than with most other existing implementations of language games. In particular, we propose a hybrid mechanism that combines cross-situational learning techniques with more informed feedback mechanisms. In our study we focus our attention on dealing with referential indeterminacy after joint attention has been established and on whether the current model can deal with larger populations than previous studies involving cross-situational learning. Simulations show that the proposed model can indeed lead to coherent languages in a quasi realistic world environment with larger populations.

1 Introduction

For language to evolve, the language has to be transmitted reliably among the population, which is only possible if the individual agents can learn the language. In human societies, children have to learn for instance the sounds, words and grammar of the target language. In the current paper, we focus solely on the evolution and acquisition of word-meaning mappings. The way children acquire the meanings of words still remains an open question. Associating the correct meaning to a word is extremely complicated, as a word may potentially have an infinite number of meanings [1].

Different mechanisms that children may adopt when acquiring the meanings of words have been suggested, see, e.g., [2] for an overview. For example,

* This research and the New Ties project is supported by an EC FET grant under contract 003752. We thank all members the New Ties project for their invaluable contributions. Opinions and errors in this manuscript are the authors' responsibility, they do not necessarily reflect those of the EC or other New Ties members.

Tomasello has proposed that *joint attention* is a primary mechanism [3]. According to this mechanism, children are able to share their attention with adults on objects, e.g., through gaze following or pointing. Moreover, children can learn that adults have control over their perceptions and that they can choose to attend to particular objects or aspects of a given situation. This allows children to focus their attention on the same situation experienced by adults, thus reducing the number of possible meanings of a word.

This mechanism, however, is not sufficient, because it is still uncertain whether a word relates to the whole situation, to parts of the situation or even to a completely different situation. This is known as the *referential indeterminacy* problem illustrated by Quine [1] with the following example: Imagine an anthropologist studying a native speaker of an unfamiliar language. As a rabbit crosses their visual field, the native speaker says “gavagai” and the anthropologist infers that “gavagai” means *rabbit.* However, the anthropologist cannot be completely sure of his inference. In fact, the word “gavagai” can have an infinite number of possible meanings, including *undetached rabbit parts*, *large ears*, *it’s running*, *good food* or even *it’s going to rain.*

To overcome this problem, additional mechanisms have been proposed to reduce the referential indeterminacy. Among these is a representational bias known as the *whole object bias* [4], according to which children tend to map novel words to whole objects, rather then to parts of objects. Another mechanism that children appear to use is the *principle of contrast* [5], which is based on the assumption that if a meaning is already associated with a word, it is unlikely that it can be associated with another word.

There is also evidence that children can acquire the meanings of words more directly by reducing the number of potential meanings of words across different situations [6, 7]. This *cross-situational learning* can work statistically by maintaining the co-occurrence frequencies of words with their possible meanings [8, 9] or simply by maintaining the intersection of all situations in which a word is used [10, 11]. Crucially, cross-situational learning depends on observing a sufficient degree of one-to-one mappings between words and meanings. Although theoretically, the level of uncertainty (i.e. the number of confounding – or background – meanings) in situations may be quite large, this may have a large impact on the time required to learn a language [11].

Cross-situational learning yields poor results when the input language is less consistent regarding the one-to-one mapping. This has been found in simulation studies of language evolution with increased population sizes [9]. In such simulations, different agents create many different words expressing the same meaning when they have not yet communicated with each other. So, the more agents there are, the more words can enter a language community during the early stages of evolution. In models that use explicit meaning transfer, there are positive feedback loops that reduce the number of words sufficiently over time, allowing the language to converge properly [12]. However, when there is no positive feedback loop, as is the case with cross-situational learning, there appears to be no efficient mechanism for reducing the number of words in the language.

A possible solution to this problem could be to include an additional mechanism that imposes a bias toward one-to-one mappings between words and meanings [13].

In this paper we propose a hybrid model for the evolution of language that combines joint attention, cross-situational learning and the principle of contrast as mechanisms for reducing the referential indeterminacy. In addition, a feedback mechanism and related adaptations are used as a synonymy damping mechanism. This model is used to investigate the effect that context size has on the development of language, but more importantly it is used to investigate how this model can deal with large populations. The model is embedded in the New Ties project[3], which aims at developing a benchmark platform for studying the evolution and development of cultural societies in very large multi-agent systems [14].

The paper is organised as follows: in the next section, we provide a brief description of the proposed model (for details, consult [14, 15]). In Section 3 we present some experiments, whose aims are to show that the proposed hybrid model can lead to the evolution of a coherent lexicon in large population sizes and with varying context sizes. The results are discussed in Section 4. Finally, Section 5 concludes.

2 The Model

2.1 New Ties agent architecture

The New Ties project aims at developing a platform for studying the evolution and development of cultural societies in a very large multi-agent system. In this system, agents are inserted in an environment consisting of a grid world in which each point is a location. The world, which is inspired by Epstein & Axtell's [16] sugar scape world, is set up with tokens, edible plants, building bricks, agents, different terrains of varying roughness, etc. The aim for the agents is to evolve and learn behavioural skills in order for the society to survive over extended periods of time. As part of these skills, language and culture are to develop.

At each time step each agent receives as input a set of perceptual features and messages, which constitute the context of an agent, and outputs an action (see Fig. 1 for the basic agent architecture). These actions are collected by the environment manager, and when all agents have been processed, the collected actions are executed and the environment is updated.

The perceptual features an agent receives represent both objects and actions that occur in its visual field. These features are processed with a categorisation mechanism based on the discrimination game [17] (a detailed description of this mechanism is given in [14, 18]). Basically, each object is mapped onto a set of categories, where each category corresponds to a feature. So, if an object is described by n features, it will be categorised into n categories. Messages are

[3] New Ties stands for New Emerging World models Through Individual, Evolutionary and Social learning. See http://www.new-ties.org

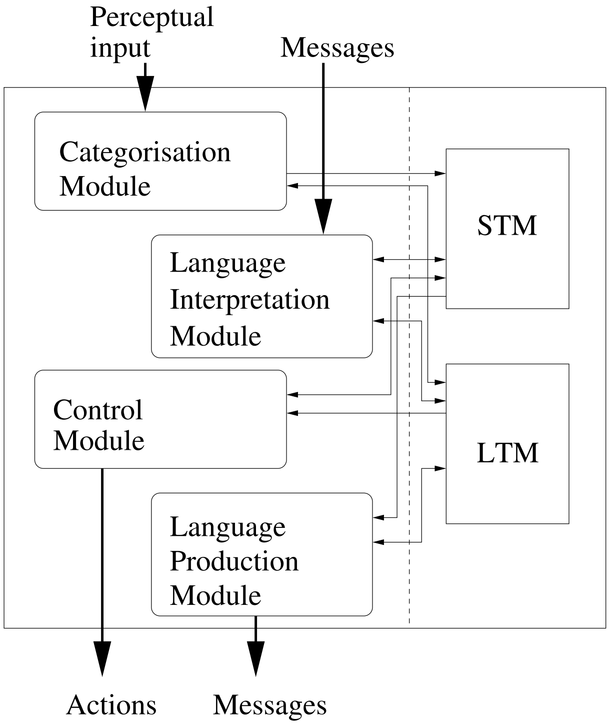

Fig. 1. The basic architecture of a New Ties agent. Perceptual features of objects and actions are processed by the categorisation module, while messages are interpreted with the language interpretation module. The control module outputs actions and the language production module produces outgoing messages. Various sources of knowledge are stored in the short- and long-term memories.

processed with a language interpretation module, described in Section 2.2, and also yield a set of categories. All these categories are stored in the short-term memory (STM), which can be accessed by the control module, as well as all other modules.

Once the perceptual features and messages have been processed, the controller is used to determine the action to perform. This controller is represented by a *decision Q-tree* (DQT), which is a decision tree that can change during an agent's lifetime using reinforcement learning [14]. The possible actions include, among others, *move, turn left, turn right, mate, talk, shout,* ... In case the output of the DQT is either the *talk* or *shout* action, the agent must produce a message, which is done by the language production module, described below. Each action performed costs a certain amount of energy, and when an agent's energy level decreases to zero or below, it dies. Energy levels can be increased by eating plants. Agents also die when they reach a predefined age.

Agents start their life with a small initial DQT, which, as mentioned above, can be changed by reinforcement learning. This initial DQT is the result of evolution. When two agents reproduce, they produce an offspring who inherits

$$\begin{array}{c|ccc} & m_1 & \dots & m_N \\ \hline w_1 & \sigma_{11} & \dots & \sigma_{1N} \\ \vdots & \vdots & \vdots & \vdots \\ w_M & \sigma_{M1} & \dots & \sigma_{MN} \end{array} \qquad \begin{array}{c|ccc} & m_1 & \dots & m_N \\ \hline w_1 & P_{11} & \dots & P_{1N} \\ \vdots & \vdots & \vdots & \vdots \\ w_M & P_{M1} & \dots & P_{MN} \end{array}$$

Fig. 2. A simplified illustration of the lexicon. The lexicon consists of two matrices that associate meanings m_j with words w_i. The left matrix stores association scores σ_{ij} and the right matrix stores co-occurrence probabilities P_{ij}.

its genome from its parents, subject to cross-over and mutations. This genome carries the code for producing the initial DQT and other biases, which regulate, for instance, the 'socialness' of the agent. This socialness gene is a bias for an agent to be social; the more social an agent is, the more frequently it will communicate and the more likely it is to provide more information regarding the meaning of a message. Unlike standard evolutionary algorithms, reproduction is not processed cyclical, but acyclical, i.e., two agents can reproduce when they decide to, but only if they are of different sex and in nearby locations.

2.2 Communication and learning word-meaning mappings

The language evolves in the society by agents' interacting through language games. While doing so, each individual constructs its own lexicon, which is represented in the long-term memory (LTM) by two association matrices (Fig. 2). Each matrix associates words w_i with meanings m_j. The first matrix stores association scores σ_{ij}, while the second stores co-occurrence probabilities P_{ij}. The former is updated based on feedback the agents may receive regarding the effectiveness (or success) of their interaction. However, as this feedback is not always available, the agents also maintain the co-occurrence frequencies of words and the potential meanings as they co-occur in a given situation (or context). The two matrices are coupled via the *association strength*, $strL_{ij}$, which is calculated as:

$$strL_{ij} = \sigma_{ij} + (1 - \sigma_{ij})P_{ij}. \tag{1}$$

This coupling allows the agents to infer the right word-meaning mappings across different situations using the co-occurrence probabilities when there has been little feedback. However, when there has been sufficient feedback on the language use of the agents, the association score σ_{ij} may become high enough to overrule the co-occurrence probabilities.

Both matrices are updated after each language game. If a language game is considered successful based on the feedback mechanism, the association score σ_{ij} of the used association is increased by

$$\sigma_{ij} = \eta \cdot \sigma_{ij} + 1 - \eta, \tag{2}$$

where $\eta = 0.9$ is a constant learning parameter. In addition, the scores of competing associations are laterally inhibited by

$$\sigma_{ij} = \eta \cdot \sigma_{ij}. \tag{3}$$

An association α_{nm} is competing if either the word is the same $(n = i)$ or the meaning $(m = j)$, but not both. If the game has failed according to the feedback mechanism, σ_{ij} is also decreased this way. The association score is unchanged if no feedback is processed.

In each game, irrespective of its outcome, the co-occurrence frequencies f_{ij} of words with potential meanings in that situation are increased, thus affecting the co-occurrence probabilities:

$$P_{ij} = \frac{f_{ij}}{\sum_i f_{ij}}. \tag{4}$$

The reason for adopting this dual representation is that earlier studies have indicated that using the mechanism for updating the association scores (Eqs. 2 and 3) work much better than for updating the co-occurrence probabilities (Eq. 4) if there is feedback, while the opposite is true for cross-situational learning [19].

Unlike standard implementations, such as [17, 18], a language game is initiated by an agent when its controller decides to talk or shout[4], or otherwise with a certain probability proportional to socialness gene. This agent (the speaker) then selects an arbitrary object from its context as a *target object*[5] and decides on how many words it will use to describe the object. This number, expressed in the *task complexity* T_c, is determined by generating a random number between 1 and 5 following a Gaussian distribution with the average age of the target audience in tens of 'New Ties years' (NTYrs)[6] as its mean and a standard deviation of 0.75. This way, the agent will tend to produce shorter messages when addressing a young audience and longer messages when addressing an older audience.

Depending on this task complexity, the agent selects arbitrarily T_c different categories that represent the object. Recall that each category relates to one perceptual feature of an object, such as the object's colour, shape, distance or weight. For each category, the speaker then searches its lexicon for associations that have the highest strength $strL_{ij}$. If no such association is found, a new word is invented as an arbitrary string and added to the lexicon. Each word thus found is then appended to the message which is distributed to the agent(s) in the speaker's vicinity.

[4] The 'talk' action is directed to only one visible agent, while 'shout' is directed to all agents in the audible vicinity of the initiator.

[5] In later studies we intend to make this selection depending on the decision making mechanism determined by the DQT, so the communication will be more functional with respect to the agent's behaviour.

[6] In the current paper, a year in 'New Ties time' equals to an unrealistic 365 time steps.

On certain occasions, for instance, when the hearer had signalled that it did not understand the speaker, the speaker may accompany the message with a pointing gesture to draw the attention to the target (such a gesture is only produced with a probability proportional to the socialness gene mentioned earlier). This way, the agents establish joint attention, but still the hearer does not necessarily know exactly what feature of the object is signalled (cf. Quine's problem).

When an agent receives a message, its language interpretation module tries to interpret each word in the message by searching its lexicon for associations with the highest strength $strL_{ij}$. If the association score σ_{ij} of this element exceeds a certain threshold (i.e., $\sigma_{ij} > \Theta$, where $\Theta = 0.8$), then the hearer assumes the interpretation to be correct. If not, the hearer may – with a certain probability proportional to the socialness gene – consider the interpretation to be incorrect and signal a 'did not understand' message, thus soliciting a pointing gesture; otherwise, the hearer will assume the interpretation was correct.

In case the interpretation was correct, the hearer may – again with a probability proportional to its socialness gene – signal the speaker that it understood the message, thus providing feedback so that both agents increase the association score of used lexical entries and inhibit competing elements as explained above. In all cases, the co-occurrence probability P_{ij} is increased for all categories in the context that have an association with the expressed words. In case the speaker had pointed to the object, this context is reduced to the perceptual features of this object. Otherwise, the context contains all categories of all visible objects, which may differ from those the speaker sees – including the target object. All interpretations are added to the STM, which the controller uses to decide on the agent's next action.

When no interpretation could be found in the lexicon, the agent adds the novel word to its lexicon in association with all categories valid in the current context (i.e., either all objects and events perceived or the object that was pointed to). The frequency counters of these associations are set to 1 and the association scores σ_{Nj} are initialised with:

$$\sigma_{Nj} = (1 - \max_i(\sigma_{ij}))\sigma_0, \tag{5}$$

where $\max_i(\sigma_{ij})$ is the maximum association score that meaning m_j has with other words w_i, $\sigma_0 = 0.1$ is a constant, and $i \neq N$. This way, if the agent has already associated the meaning (or category) m_j with another word w_i, the agent is biased to prefer another meaning with this novel word. Hence, this implements a notion of the principle of contrast [5]. Note again that the hearer may not have seen the target object and thus may fail to acquire the proper meaning.

3 Experiments

In the experiments we test the effectiveness of the model described in the previous section. In particular, we are interested to see whether reasonable levels of communicative accuracy can be reached with relatively large populations. In addition, we investigate the influence of considering a different number of perceptual features that agents have at their disposal for inferring word-meaning mappings. In order to focus on these questions, the evolutionary and reinforcement learning mechanisms were switched off. So, although agents could reproduce, each agent has exactly the same hand-crafted controller that did not change during their lifetimes. As a result, in the simulations reported here, agents only move, eat, reproduce (with no evolutionary computation involved) and communicate with each other. When an agent's energy level decreased below zero, they died. The same happens when agents reach a certain age (set at 80 'New Ties years', i.e. 29,200 time steps).

We performed a set of experiments in which we varied the number of features considered for each object, from a minimum of 2 to a maximum of 10 features. Varying the number of features has an influence on the number of possible meanings in the language. The following table indicates how many meanings there are for the different number of features available:

No. of features	2	3	4	5	6	7	8	9	10
No. of meanings	10	16	19	23	26	35	40	45	48

Remember that a category relates to one feature, so the more features are used to describe an object, the more possible meanings can be associated to a word. Effectively, increasing the number of features increases the context sizes. A recent mathematical model describing cross-situational learning [11] shows that learning word-meaning mappings is harder when the context size is larger. So, we expect that considering a higher number of features will lead to the evolution of a lower level in communicative accuracy, or to a slower learning rate.

In addition to reducing the number of features, referential indeterminacy can be reduced by means of pointing. As mentioned, the probability with which agents point is proportional to the socialness gene. As the evolutionary mechanisms are switched off in these experiments, the socialness gene is now initialised individually with a random value.

The initial population size is set to 100 agents. When the agents reach the age of 10 NTYrs (3,650 time steps), they start to reproduce. So, from then onward the population size can grow, though this may not happen if the agents tend to die faster than they reproduce.

Recall that all agents are evaluated once during each time step. So, during one time step, multiple language games can be played by different agents. Moreover, different agents can speak to one another simultaneously, as they do not wait for their turn. Playing one language game takes 2-3 time steps: (1) sending a message, (2) receiving a message, occasionally, (3) signalling feedback and (4) receiving feedback.

The simulations are evaluated based on *communicative accuracy.* Communicative accuracy is calculated each 30 time steps by dividing the total number of successful language games by the total number of language games played

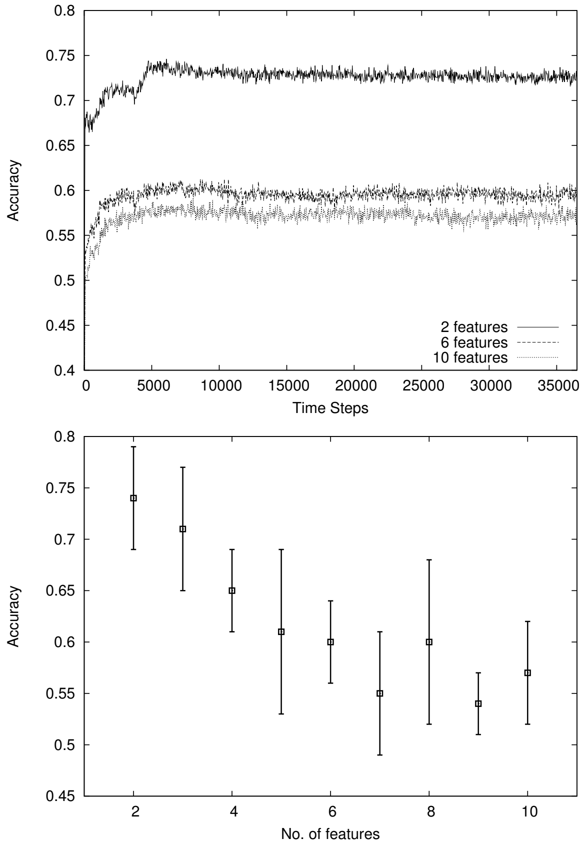

Fig. 3. (Top) The evolution of communicative accuracy (y-axis) over time (x-axis) for the conditions with 2, 6 and 10 features. Notice the odd scale on the y-axis. (Bottom) Communicative accuracy measured at the end of each simulation, averaged over the 5 trials with their standard deviation. The results relate to the number of perceptual features varied from 2 to 10 with incremental steps (x-axis).

during this period. A language game is considered successful if the hearer interpreted the message from the speaker such that the interpreted category exactly matched the intended category (so not the object). Simulations were repeated 5 times with different random seeds for each condition and the results reported are averages over these 5 trials.

Figure 3 (top) shows communicative accuracy for the cases with 2, 6 and 10 features. In all cases, accuracy increased to a level between 0.50 (10 features) and 0.68 (2 features) during the first 30 time steps. After this, accuracy first increased quite rapidly and then stagnated more or less around 0.57 (10 features), 0.60 (6 features) and 0.73 (2 features). Although the language is not learnt perfectly in any condition, accuracy is reasonable and much better than chance. For instance, in the case where there are 6 features, chance is between 1/26 (if all possible meanings are in the context – cf. above mentioned table) and 1/6 (if the target object was pointed to).

For comparison, we tested the model in a simulation where pointing was used to *explicitly* transfer the intended *meaning* (i.e. categories) – at least in those interactions where pointing was used. Under this condition, communicative accuracy yielded on average 0.97±0.02 at the end of the simulations.

It is clear that the levels of communicative accuracy decreased when the number of features increased up to 6 or 7 features, after which there is no more significant change (Fig. 3 bottom). Although differences between subsequent numbers of features are not significant, the difference between using 2 features and 7 features is. This is consistent with our prediction mentioned earlier and also with the findings from the mathematical model [11]. However, in the mathematical model all word-meaning mappings could be learnt perfectly, but at the expense of longer learning periods for larger context sizes (i.e. more features).

It is not yet fully understood why there is no more significant change for variation from 6 to 10 features. One explanation could be that when there are more than 6 perceptual features, it no longer holds that all objects are described by every feature, because some features (e.g., shape and colour) are shared by all objects, while others (e.g., sex) only by some objects.

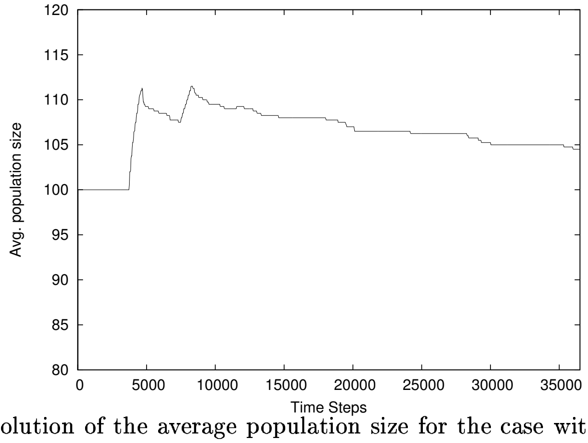

Fig. 4. The evolution of the average population size for the case with 6 features. All other simulations revealed a similar evolution.

Figure 4 shows the evolution of the average population size in the simulations with 6 features. We see that the first 3,650 time steps (10 NTYrs), the population size remains constant at 100 agents. This is because during this period, agents only start reproducing when they reached an age of 10 NTYrs. We then see a rapid increase of the population size to 110 agent, after which the population size somewhat fluctuates until it eventually slowly decreases, though the total number remains larger than 100. The decrease is due to the fact that giving birth costs a large amount of energy, which is passed on to the offspring. So agents who are less fit will have a large chance of dying after giving birth. The issue here is that these changes in the population do not seem to alter the evolution of accuracy a lot, though around the points where there is a large inflow or outflow of agents, this does seem to have some effect. This is consistent with findings from earlier simulations on language evolution, e.g., [20].

It is important to stress that these experiments are different from those focusing only on cross-situational learning as in [8, 9, 11]. In those experiments, cross-situational learning was the only learning mechanism. In these experiments, feedback regarding a game's success is provided in approximately 12% of the language games, while messages were accompanied with a pointing gesture in about 42% of all games. Note that one game can have both a pointing gesture and feedback, so none were used in an estimated 50%. Per time step, approximately 27% of all agents initiated a language game, so assuming that the population size was on average 105 over the entire period of the experiment, a total of approximately 1 million language games were played at the end of the experiments.

4 Discussion

In this paper we investigate some aspects of learning word-meaning mappings regarding Quine's problem of referential indeterminacy. In particular, we are interested in how agents can evolve a shared lexicon regarding various characteristics of objects without using explicit meaning transfer. Although agents do not always point to target objects, but when it happens, hearers still cannot determine exactly what characteristics (or features) of objects are intended by the speaker. Our proposed solution is to use cross-situational learning for such instances. However, as this learning mechanism has proved to be relatively slow and difficult to scale up in terms of population size [9], we combined this method with learning techniques based on positive feedback and the principle of contrast.

The results achieved with this model are reasonable. The population can develop a communication system with an accuracy of about 50-70% quite rapidly, while further improvement on accuracy is somewhat slower yielding levels of accuracy between 60-75% at the end of the simulations. The initial speed of learning seems very fast, but one has to realise that the agents do not communicate with all other agents. Instead, the only communicate with agents within their vicinity. In the current setting, there were groups of around 3-4 agents quite near to each other. So, although the population is larger than in any previous study using cross-situational learning, it will take a long time before all agents would have communicated with many different agents. It is unclear in the current simulations what the reach of an agent was (i.e. the number of different agents it communicated with).

The stagnation of communicative accuracy is thought to be caused by – at least – three aspects: 1) the influx of new agents, 2) the increase of task complexity and 3) mismatches in perceived contexts by different agents participating in a language game. The first two aspects start to have an influence at time step 3,650 – the time that the first agents reach an age of 10 NTYrs. This is around the same period where the stagnation starts to occur. The third aspect is caused by the 'situatedness' of the agents in their environment, because two agents cannot be at the same location simultaneously, and also because their orientation can be quite different (see [21] for a discussion). Furthermore, if an object is obscured by another one for a particular agent, this need not be the case for another agent. If the other agent already learnt the meaning of this word reliably, there is no problem, but otherwise the hearer will assume the word means something that he sees. This can be problematic for cross-situational learning, which heavily depends on consistent and reliable input [9]. Despite all this, the agents perform well beyond chance. In the future, we will assess in more detail what the exact effects of these aspects are.

The latter aspect can partly be solved using pointing, though – as mentioned – this only occurred on average in about 42% of all interactions. Pointing gestures can be initiated spontaneously by the speaker with a certain probability, but can also be solicited by the hearers when they send a negative feedback signal. In such cases, the context is reduced to the number of perceptual features of one object, which equals 2 in the simplest case investigated and 10 in the most difficult case. Since the language games will fail frequently early on, many negative feedback signals are sent, in which case the speaker is likely to repeat the message, but now accompanied by a pointing gesture. This way, agents can engage in a sort of 'dialogue', where the speaker repeats himself to make himself understood if requested by the hearer.

It must be stressed that the success is probably only partly due to the cross-situational learning. It is, to some extent, also due to the positive feedback that is provided when the hearer considers the language game to be successful. Recall that feedback is provided when the association score σ_{ij} exceeds a certain threshold Θ or – if this is not the case – with a probability that is inversely proportional to the value of socialness gene (which was assigned randomly in the current simulations). During the early stages of word learning, we can only expect the latter case to hold, so when through cross-situational learning a hearer has selected one possible interpretation, the association score σ_{ij} is reinforced occasionally. This association needs to be reinforced 16 times before the association score exceeds the threshold, which is set to $\Theta = 0.8$. Until then, the agents rely on cross-situational learning, accompanied by occasional 'blind' adaptations of the association scores σ_{ij}. This is, then, similar to the synonymy damping mechanism proposed in [13], which has a positive effect on disambiguating the language during cross-situational learning.

In [22], we investigated the role of feedback in a related model simulating the Talking Heads experiment. There it was found that only when feedback was used frequently enough, the results were better than when feedback was not used at all (i.e. when the learners could only rely on a variant of cross-situational learning). However, in those simulations feedback forced the speaker to point at the object and, since in those simulations objects were represented by only one category, pointing identified the target meaning more precisely. We are currently investigating more thoroughly what the role of feedback is in this model.

It is also important to realise that the language is relatively small. In case there are 2 features, an agent has only 10 categories, but in case of 10 features an agent has a total of 48 categories. Although learning individual words can take longer when there are less meanings (because it can take longer before distracting meanings no longer compete), this does not hold for the entire language, provided the context size is substantially smaller than the total number of meanings [11]. So, the smaller the language, the easier it should be learnt.

It is yet unclear what the influence of the principle of contrast is in this model, because we did not compare these results with a simulation where the principle of contrast was switched off. This will be carried out in future experiments. It is interesting to note, however, that we implemented the principle of contrast as a loose bias, rather than as a strong principle that would rule out competing word-meaning mappings entirely.

One may wonder why this particular study is carried out in a complex environment as the current one, while a similar study could have been carried out in a much more simpler simulation setting. We agree this is true, but it is important to realise that this is the first in a series of experiments being set up in the New Ties project. There are many more planned; some of which may indeed be done using a simpler set up (e.g., for investigating the effect of the principle of contrast), but most will relate to the evolution of more complex behaviours that would allow the population to remain viable over extended periods of time. Such experiments will involve various combinations of learning mechanisms to allow the population to evolve and learn how to behave properly in their complex environment. These learning mechanisms include evolutionary learning, individual (reinforcement) learning and social learning. Especially the latter is of interest, because we intend to set up experiments in which the language that evolves will be used to share information concerning the way the controller is structured, thus allowing agents to copy such structures in order to acquire more similar controllers.

5 Conclusions

In this paper we have presented a new hybrid model for the simulation of language evolution, and in particular the evolution of shared lexicons. This model is incorporated in the New Ties project, whose aim is to set up large scale simulations to study the evolution of cultural societies by combining evolutionary, individual and social learning techniques.

Using the model we show how a combination of different learning mechanisms, which include pointing as a means of establishing joint attention, the principle of contrast, a positive feedback mechanism and cross-situation learning allow agents to infer the meaning of words. In particular, we show that this model can – in contrast to previous studies [9] – deal well with relatively large populations. One reason for this ability is that the feedback mechanism acts as a synonymy damping mechanism, similar to a recent study by De Beule et al. [13].

The study further shows that the model is quite robust (but definitely not perfect) when agents need to infer the meaning when there is more referential indeterminacy, though learning is somewhat hampered in terms of communicative accuracy. Indirectly, this confirms another recent study by Smith et al. [11], who mathematically proved that cross-situational learning can work well with different levels of referential indeterminacy, though the learning speed is affected such that higher levels of indeterminacy require longer learning periods. The difference with the current study is that in the mathematical study language can be learnt with 100% accuracy, but under the assumption that an ideal language exists which needs to be learnt by one individual who receives consistent input. In the current simulation, such assumptions do not hold.

As one of the objectives of the New Ties project is to set up a benchmark platform for studying the evolution of cultural societies, which includes the evolution of language, we believe this study is a first promising step showing what sort of studies can be carried out with this platform.

References

1. Quine, W.V.O.: Word and object. Cambridge University Press (1960)
2. Bloom, P.: How Children Learn the Meanings of Words. The MIT Press, Cambridge, MA. and London, UK. (2000)
3. Tomasello, M.: The cultural origins of human cognition. Harvard University Press (1999)
4. Macnamara, J.: Names for things: a study of human learning. MIT Press, Cambridge, MA (1982)
5. Clark, E.: The principle of contrast: A constraint on language acquisition. In MacWhinney, B., ed.: Mechanisms of language acquisition. Lawrence Erlbaum Assoc., Hillsdale, NJ (1987) 1–33
6. Akhtar, N., Montague, L.: Early lexical acquisition: the role of cross-situational learning. First Language (1999) 347–358
7. Houston-Price, C., Plunkett, K., Harris, P.: 'word-learning wizardry' at 1;6. Journal of Child Language **32** (2005) 175–190
8. Smith, A.D.M.: Intelligent meaning creation in a clumpy world helps communication. Artificial Life **9(2)** (2003) 559–574
9. Vogt, P., Coumans, H.: Investigating social interaction strategies for bootstrapping lexicon development. Journal of Artificial Societies and Social Simulation **6** (2003)
10. Siskind, J.M.: A computational study of cross-situational techniques for learning word-to-meaning mappings. Cognition **61** (1996) 39–91
11. Smith, K., Smith, A., Blythe, R., Vogt, P.: Cross-situational learning: a mathematical approach. In Vogt, P., Sugita, Y., Tuci, E., Nehaniv, C., eds.: Proceedings of the Emergence and Evolution of Linguistic Communication (EELCIII), Springer (2006)
12. Baronchelli, A., Loreto, V., Dall'Asta, L., Barrat, A.: Bootstrapping communication in language games: Strategy, topology and all that. In Cangelosi, A., Smith, A., Smith, K., eds.: The evolution of language; Proceedings of Evolang 6, World Scientific Publishing (2006)
13. De Beule, J., De Vylder, B., Belpaeme, T.: A cross-situational learning algorithm for damping homonymy in the guessing game. In: ALIFE X. Tenth International Conference on the Simulation and Synthesis of Living Systems. (2006) to appear.

14. Gilbert, N., den Besten, M., Bontovics, A., Craenen, B.G., Divina, F., Eiben, A., et. al.: Emerging artificial societies through learning. Journal of Artificial Societies and Social Simulation **9** (2006)
15. Divina, F., Vogt, P.: Modelling language evolution in a complex ecological environment. ILK Research Group Technical Report Series no. 06-01 (2006)
16. Epstein, J.M., Axtell, R.: Growing artificial societies: social science from the bottom up. MIT Press, Cambridge, MA. (1996)
17. Steels, L.: The synthetic modeling of language origins. Evolution of Communication **1(1)** (1997) 1–34
18. Vogt, P.: The emergence of compositional structures in perceptually grounded language games. Artificial Intelligence **167** (2005) 206–242
19. Vogt, P.: Lexicon Grounding on Mobile Robots. PhD thesis, Vrije Universiteit Brussel (2000)
20. Steels, L., Kaplan, F.: Situated grounded word semantics. In: Proceedings of IJCAI 99, Morgan Kaufmann (1999)
21. Vogt, P., Divina, F.: Language evolution in large populations of autonomous agents: issues in scaling. In: Proceedings of AISB 2005: Socially inspired computing joint symposium. (2005) 80–87
22. Divina, F., Vogt, P.: Perceptually grounded lexicon formation using inconsistent knowledge. In: Proceedings of the VIIIth European Conference on Artificial Life (ECAL2005), Springer-Verlag (2005) 644–654