The Impact of Non-verbal Communication on Lexicon Formation

Paul Vogt

Universiteit Maastricht, Infonomics / IKAT
P. O. Box 616, 6200 MD Maastricht

Abstract

This paper presents a series of experiments in which two mobile robots develop a shared lexicon of which the meaning is grounded in the real world. The experiments investigate the impact of non-verbal communication on lexicon formation. Non-verbal communication is used to establish joint attention or to evaluate feedback. The experiments implement adaptive language games in which two agents try to communicate some real world object. When the agents fail, they can adapt their memory in order to improve performance on future occasions. As the experimental results show, the quality of the evolved lexicon is better when feedback is used rather than joint attention.

1 Introduction

One major issue in the study of human language development is to what extend non-verbal cues are available to language learners. The issue has become famous as part of the *poverty of the stimulus argument* [3]. This argument is based on the observation that infants are capable of learning a full-fledged language without being exposed to the complete language. A related problem is the *no negative feedback evidence* [2]. This problem is based on the observation that infants, when making mistakes during their language acquisition do not receive as much feedback as one would expect. If, however, infants receive little or no feedback, are there other non-verbal cues available to the child? One cue that has been proposed to play a role is joint attention to the referent during the presentation of a word, see e.g. [7].

Recent computational studies have shown that artificial agents can learn a shared lexicon using either joint attention or feedback[1] as a non-verbal cue [1, 4, 5, 10]. This paper investigates, using robotic experiments, the influence of these different types of non-verbal cues on the quality of the lexicon formation. The paper is based on experiments reported earlier in the author's Ph. D. thesis [9].

[1] In this paper 'feedback' means an evaluation whether communication has been performed correctly or not. I.e. whether two agents were able to communicate the same referent.

The experiments implement the model of language games that has been introduced by Luc Steels [6]. This model is used to investigate how a population of robots can develop a shared lexicon of which the meaning is grounded in reality.

The paper is organized as follows. The next section provides some psycholinguistic background on the issues about non-verbal communication in language acquisition. It also discusses related work that has been done in the adaptive behaviour community. Section 3 presents a brief description of the implemented model. Experimental results are given in section 4, followed by a discussed in section 5. Finally, section 6 concludes.

2 Non-verbal cues in language development

During their early childhood, infants need to learn how words are associated with reality. How do they know what word belongs to what referent?

Although infants are exposed to a limited part of a language, an adult language user can express an unlimited number of sentences. This observation provides the basis of *poverty of the stimulus* argument [3].

One of the observations that provide evidence for the poverty of the stimulus argument is the so-called *no negative feedback evidence* problem, see e.g. [2]. The problem is that it seems that language can only be learned when both positive and negative feedback on language is available to a language learner. Such feedback consists of (positive) rewards and (negative) corrections. However, psycholinguistic research has shown that hardly any negative feedback is provided by adult language users.

One alternative for the feedback is the establishment of joint attention *prior* to the verbal communication. Early studies of Tomasello showed that children can learn a language better when joint attention is established [8].

In a later study, Tomasello and Barton report on experiments where children learn novel words under two different conditions [7]. In one condition, children do not receive non-verbal cues when the word-form is presented. When at a later moment the corresponding referent is shown, a positive feedback is given if the child correctly relates the referent with the given word-form. Negative feedback is provided when an incorrect relation is made. In the second condition, joint attention is established simultaneous with the presentation of the word-form while no feedback is given. Tomasello and Barton showed that children could equally well learn novel word-meaning relations in both conditions.

Recent studies using computer simulations and robotic experiments have investigated the difference between feedback and joint attention too [1, 4, 5, 10]. Many of these studies use either Hebbian learning or reinforcement learning. In Hebbian learning associations are formed between two (or possibly more) simultaneously active patterns. Hence, when learning form-meaning associations, some form of joint attention mechanism is required. Reinforcement learning, on the other hand, uses feedback to learn an association correctly.

Billard and Oliphant have shown, independently of each other, that simulated and robotic agents can learn a lexicon when they simultaneously have access to both a word-form (or form for short) and a meaning without using feedback [1, 5][2]. Yanco and Stein have shown that agents could develop a shared lexicon without access to both meaning

[2]Note that Oliphant calls this type of learning *observational learning* [5], while Billard calls it *learning through imitation* [1].

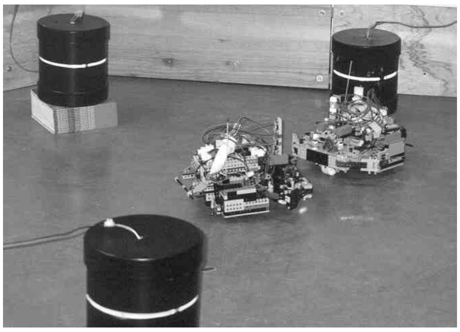


Figure 1: The robots situated in their environment as used in the experiments.

and form, but with using feedback [10]. Where the first studies use Hebbian learning, the latter study uses reinforcement learning. Finally, De Jong showed how agents can learn under both conditions [4].

So, in both approaches a lexicon can be learned [1, 10] or even evolve from scratch [4, 5]. But how do the two approaches influence the quality of the lexicon? The experiments discussed in the remainder of this paper investigates this question.

3 Adaptive Language Games

The experiments are done using two LEGO robots that are situated in a small environment consisting of four light sources, see figure 1. Each light source is placed at a different height and the robots are equipped with four light sensors, each mounted at a corresponding height. The aim of the experiment is that the robots develop a shared lexicon of which the meaning is grounded in their world. The lexicon is a set of form-meaning associations, each relating ideally to some light source. Prior to the experiments, the robots have no categories that constitute part of the meaning and their lexicon is empty.

The lexicon formation is guided by a series of *guessing games*. A guessing game is a variant of a language game in which the hearer tries to guess what light source the speaker is naming. The guessing game makes no use of joint attention, but of feedback. After the guessing game is explained, the observational game that does use joint attention, but no feedback is explained. As the language game model is meanwhile well known, it is described very briefly. The interested reader is referred to [9] for the details.

1. The guessing game starts when both robots are close to each other with their backs oriented towards each other.

2. One by one, the robots acquire a spatial view of 360^o by means of rotating. This is called *sensing*, which results in a set of raw sensory data which is sent to a PC for further processing.

3. The view is preprocessed by segmentation and feature extraction processes. It results in a description of the sensing, called the *context*, which is a set of feature vectors. A feature vector is a 4 dimensional vector describing some properties of

the sensing. Each feature vector is supposed to relate to the detection of one light source. It is important to note that not always all four light sources are detected, nor is the mapping always correct.

4. After the context is constructed, the speaker chooses one arbitrary feature vector as the topic. The hearer considers all feature vectors as a potential topic; it has to guess which feature vector is the real topic.

5. Both robots individually try to find a distinctive categorization for the (potential) topic(s). This is modeled by so-called *discrimination games*. The aim of a discrimination game is to find a categorization of a topic that distinguishes the topic from all other feature vectors in the context. When no distinctive category can be found by the discrimination game, new categories can be created for which the feature vector of the topic is used as an exemplar. If the discrimination game succeeds, the distinctive categories are forwarded to the naming part of the guessing game. If in turn the guessing game succeeds, the prototypical category is moved towards the feature vector of the topic.

6. After both robots acquired distinctive categories that relate to the (potential) topic(s), the speaker tries to produce an utterance that names the distinctive category of its topic. It does so by searching its lexicon for elements of which the meaning matches the distinctive category. If there are more than one, it selects the one that has the highest association score and the associated form is uttered. The association score indicates the element's past effectiveness in the communication. If the speaker does not find such an element, it creates a new form with a certain probability, associates this with the distinctive category and adds the new element to its private lexicon.

7. When the hearer receives an utterance, it tries to interpret it. It searches its own lexicon for elements of which the form matches the utterance, *and* of which the meaning matches a distinctive category of one of the potential topics. If there is more than one such element, the hearer selects the one that has the highest association score. If there are none the lexicon is adapted as explained below. The feature vector to which the matching distinctive category relates is then chosen as the hearer's topic of the game. I.e. this feature vector is what the hearer guessed the speaker has named.

8. Feedback is provided on whether the hearer found a lexical element and if so, whether both robots communicated the same topic. The outcome of the feedback is known to both robots.

9. Depending on the outcome (provided by the feedback) the lexicon is adapted. Three possible outcomes / adaptations remain:

   (a) The hearer has not found an element in its lexicon that matches the received utterance and of which the meaning is consistent in the game's context. In this case, the hearer adopts the uttered form and associates it with the distinctive categorization(s) of one arbitrarily selected feature vector. The speaker decreases the association score of the used form-meaning association. The guessing game fails.

(b) The hearer has found a matching element, but the selected topic is not consistent with the speaker's topic; there is a *mismatch in referent*. In this case, both robots decrease the association score of the used form-meaning association. The game fails.

(c) Both robots have selected a lexical element in relation to a consistent topic. The hearer guessed right and the game succeeds. Both robots increase the association score of the used element and competing elements are *laterally inhibited*. An element is competing when the form matches the communicated form, but not its meaning or when the meaning matches the meaning of the used element, but not its form.

The second type of language game that is investigated is called the *observational game*. In this game, the robots achieve joint attention prior to the verbal communication. This means that in step 4 above, the attention of the hearer is drawn to the topic. In addition, no feedback mechanism evaluates a game's success. Association scores are updated as under point 9 (c) whenever a robot selected a form-meaning association. In this case the game is considered successful. A mismatch in referent will not occur because the topic is selected by both robots in advance. When the hearer has no association between the uttered form and the distinctive category of the referent, it will adopt the form and associate it with this category.

4 The Experimental Results

With the above models, several experiments have been done, most of which are reported in [9]. In this paper two of these experiments are discussed.

A data set of raw sensory data has been acquired for the experiments. This data set consists of the sensing in approximately 1,000 situations. Each situation bears the context of one language game (guessing game or observational game) for the two robots. With these 1,000 situations 10 runs of 10,000 language games have been played. In each game, one robot is arbitrarily selected to play the role of the speaker, the other is the hearer. The speaker in each game randomly selects one feature vector as the topic of the game.

Figure 2 shows the communicative success of the two experiments. The communicative success is the average number of successful language games of the past 100 games averaged over the 10 different runs. The results do not exceed 80%, because sometimes the hearer has not detected the topic. As the figure shows, the experiments do much better than chance (23%)[3]. The results confirm the results of [1, 4, 5, 10] that a lexicon can be developed by using either joint attention or feedback.

The observational game outperforms the guessing game in that the lexicon develops faster and the overall communicative success is higher. Although in the end the guessing game approaches the success-rate of the observational game.

What are the qualitative differences of the lexicons developed by the two approaches? To investigate this, it is useful to look at the lexicons that developed in a typical run of the experiments, figure 3. The graphs show how the robots use referents and forms in relation

[3] The a priori chance for success is 23% rather than 25%. This is because the robots detect on the average 3.5 feature vectors during a game and then there is 20% chance that the hearer has not detected the topic.

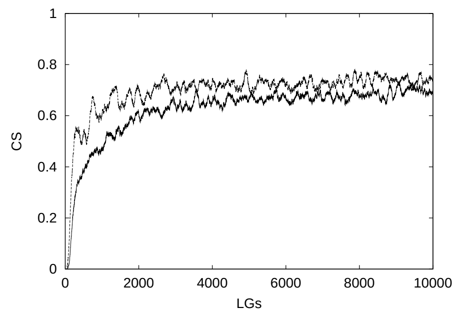


Figure 2: The communicative success of the observational games (upper line) and the guessing games (lower line).

to the meanings they constructed. The connections indicate the co-occurrence frequencies of the meaning relative to the occurrence frequencies of the referents the forms.

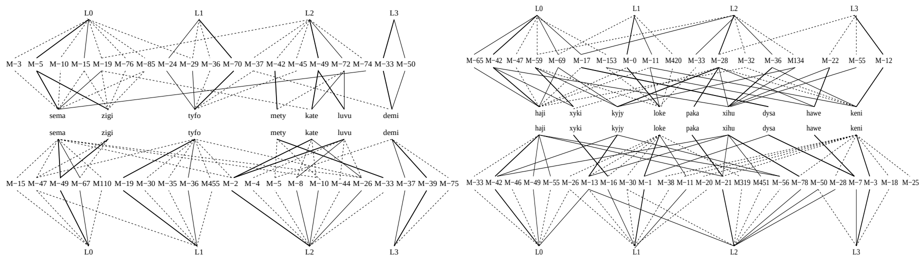
In both graphs the most frequently used associations are displayed with bold lines. It is clear that in both experiments these associations connect the referent with forms one-to-one for both robots coherently. Ideally, the connections between referents and forms for both robots should not cross-connect, otherwise there appears too much synonymy and polysemy in naming the referents consistently. This is achieved to a large degree in the guessing game,but the lexicon of the observational game reveals more cross-connections (figure 3). Hence the observational game bears more synonymy and polysemy. As a result, the lexicon developed in the guessing games is qualitatively more effective.

5 Discussion

So, although the observational game quantitatively outperforms the guessing game, qualitatively the lexicon developed by the observational game is worse. The communicative success of the observational game is higher and the lexicon is learned faster as figure 2 shows. On the other hand the observational game develops more polysemy and synonymy, which makes the lexicon less effective (figure 3).

Where does this difference in quality come from? To understand this, it is good to realise that both robots start the experiments with an empty lexicon. Furthermore, a referent may be categorised differently under different circumstances. As a result, the robots may start to name a referent differently because their adaptations do not directly associate forms with referents, but forms with meanings. When the hearer in a game receives a synonymous or polysemous utterance, it may easier succeed in the observational game, because it already knows the topic. In the guessing game the hearer has to guess the topic, for which it may need to select between several possible solutions. Naturally, this is more prone to errors, because synonymy and polysemy are important sources for misinterpretation. However, this puts more pressure on the feedback and reinforcement style of learning to disambiguate the synonymy and polysemy.

The verbal communication becomes redundant when joint attention is established beforehand. Hence the effectiveness of a lexicon is not vital for the communicative success;



(a) Guessing (b) Observational

Figure 3: The semiotic landscape shows the lexicon developed by the two robots for the (a) guessing and (b) observational game. It displays the associations between referents (L), meanings (M) and forms, given with their co-occurrence frequencies relative to the occurrence frequency of either referent or form as measured over one entire run. The winning connections are given in bold lines, the other solid lines have frequencies larger than 0.05, the dashed lines have frequencies between 0.005 and 0.05. Associations with lower frequencies have been left out for clarity.

synonymy may be present and the pressure to disambiguate the synonymy is low. The verbal communication in the guessing game bears more information and is vital for the success of a game. So, feedback is necessary in this game to develop an effective lexicon.

6 Conclusions

This paper presented a series of robotic experiments to investigate the impact of non-verbal communication on lexicon formation. This is done by comparing an experiment where the robots establish joint attention to the topic of the game and one where the robots did not. In the first experiment no feedback on the effect of the language game is used as a source of non-verbal communication, whereas in the latter it is.

The experiments confirm the results of a psycholinguistic study [7] and several computational studies [1, 4, 5, 10]. In both types of experiments the robots were able to develop a lexicon with which they could communicate rather well. Objectively, one could conclude that although the observational game yields higher quantitative results, the guessing game reveals a qualitative better lexicon.

As the no negative feedback evidence indicates, joint attention will be a more likely strategy to learn form-meaning associations. However, the results indicate that, although joint attention benefits fast learning, using feedback without joint attention might allow an infant to learn a more effective lexicon. As there may be more strategies available to a child it is likely that the child uses these different strategies, sometimes guided by its parents. Currently more research is done to investigate other strategies to associate forms with referents.

Acknowledgements

The research has been carried out at the Vrije Universiteit Brussels. The author wishes to thank Luc Steels and other colleagues of the VUB for their indispensable contributions. The author also wishes to thank Ida Sprinkhuizen-Kuyper and Nico Roos for proofreading earlier versions of the paper.

References

[1] A. Billard and K. Dautenhahn. Experiments in social robotics: grounding and use of communication in autonomous agents. *Adaptive Behavior*, 7(3-4), 1999.

[2] M. Bowerman. The 'no negative feedback evidence' problem: How do children avoid constructing an overly general grammer? In J. A. Hawkins, editor, *Explaining Language Universals*, chapter 4, pages 73–101. Basil Blackwell, 1988.

[3] N. Chomsky. Three models for the description of language. *IRE Transactions on Information Theory IT*, 2(3):13–54, 1956.

[4] E. D. De Jong. *The Development of Communication*. PhD thesis, Vrije Universiteit Brussel, 2000.

[5] M. Oliphant. The learning barrier: Moving from innate to learned systems of communication. *Adaptive Behavior*, 7 (3-4):371–384, 1999.

[6] L. Steels. Emergent adaptive lexicons. In P. Maes, editor, *From Animals to Animats 4: Proceedings of the Fourth International Conference On Simulating Adaptive Behavior*, Cambridge Ma., 1996. The MIT Press.

[7] M. Tomasello and M. Barton. Learning words in nonostensive contexts. *Developmental Psychology*, 30(5):639–650, 1994.

[8] M. Tomasello, S. Mannle, and A. Kruger. The linguistic environment of one to two year old twins. *Developmental Psychology*, 22:169–176, 1986.

[9] P. Vogt. *Lexicon Grounding on Mobile Robots*. PhD thesis, Vrije Universiteit Brussel, 2000. http://www.cs.unimaas.nl/p.vogt/thesis.html.

[10] H. Yanco and L. Stein. An adaptive communication protocol for cooperating mobile robots. In J-A. Meyer, H.L. Roitblat, and S. Wilson, editors, *From Animals to Animats 2. Proceedings of the Second International Conference on Simulation of Adaptive Behavior*, pages 478–485, Cambridge Ma., 1993. The MIT Press.