

# Human-Robot Interaction in Concept Acquisition: a computational model

Joachim de Greeff, Frédéric Delaunay and Tony Belpaeme, *Member, IEEE*

**Abstract**—This paper presents a discussion and simulation results which support the case for interaction during the acquisition of conceptual knowledge. Taking a developmental perspective, we first review a number of relevant insights on word-meaning acquisition in young children and specifically focus on concept learning supported by linguistic input. We present a computational model implementing a number of acquisition strategies, which enable a learning agent to actively steer the learning process. This is contrasted to a one way learning method, where the learner does not actively influence the learning experience. We present results demonstrating how dyadic interaction between a teacher and learner may result in a better acquisition of concepts.

**Index Terms**—concept acquisition, human-robot interaction, language games, learning interaction

## I. INTRODUCTION

**A**RTIFICIAL systems that interact with humans typically need semantic knowledge. Only very simple (behaviour-based) systems can display a repertoire of interactive behaviours which do not rely on knowledge. However, for all other tasks, such as reasoning, learning and most importantly-linguistic interaction, the artificial system will need internal representations. These representations can be implicit in the system, for example in the weights and states of a neural network, but when they are explicitly present in the system they are typically pre-programmed. In natural language processing systems for example, the semantic interpretation of a word is a well-defined script of actions programmed by a skilled human programmer. The premise of the proposed study is that semantic knowledge can be acquired autonomously by artificial agents and, more importantly, that language plays a crucial role in this.

Relevant to this is that young children, in addition to learning directly from sensory exploration, rely on linguistic labels to acquire the meaning of words (for an overview of early language acquisition see [1]). Xu [2], for example, demonstrates how linguistic labels help 9-month old infants to establish a representation for different objects; learning without linguistic labels, or with the presence of tones, sounds or emotional expressions is not effective. This implies that language is crucial in acquiring novel concepts from a very early age on. Plunket et al. [3] come to the same conclusion in a tightly controlled experiment where they demonstrate how category formation in 10-month old infants is influenced by linguistic labels. Linguistic labels also have an effect on

category learning in adults; adults who learn a new category did so significantly faster and showed more robust category recall when the learning experience was accompanied by novel linguistic labels [4], [5]. Aforementioned studies show that linguistic labels facilitate category acquisition, both in pre-linguistic infants and adults. These insights tie in with linguistic relativism, which states that language and cognition influence each other. Recently, linguistic relativism gained renewed attention as a series of psychological experiments demonstrated how perception of stimuli and use of categories is influenced by the words we know; this has been notably demonstrated for categories of time, colour and space e.g. [6]–[8].

Central to this paper is the insight that concepts are shaped by language. The concept of CHAIR is of course related to the visual and tactile perception of a chair and its function, but delimiting CHAIR and distinguishing it from other concepts, such as TABLE, STOOL or RACK, can only come about through naming all objects that belong to CHAIR as “chair” and consequently, by naming all objects that do not belong to CHAIR something else. This might seem trivial, but linguistic labels, as mentioned above, play a crucial role in concept acquisition in children and adults. Linguistic labels might be thought of as facilitating supervised learning of unknown categories, this view however is too limited. Learning categories through language allows for more complex learning interactions than one-way supervised learning (such as contrastive learning or validating the learnt word-category pairs through querying the teacher). It also involves different teachers, allowing access to different hypotheses. The role of linguistic labels in category acquisition is perhaps still seen as a form of supervised learning. Indeed, the experimental paradigms of developmental psychologists constrain the experiments so that visual stimuli and linguistic labels are presented as pairs to infants, while neglecting interaction [2], [3]. However, it has been suggested that language acquisition is a process which, in addition to learning mechanisms, relies on social-cognitive skills and multi-level interaction with peers and caretakers (cf. [9], [10]). By extension, if category acquisition is under the influence of language, it must itself be sensitive to social interaction.

The research reported in this paper was prompted by a desire to study new modes of learning in embodied robots. Typically, machine learning starts by collecting a set of training examples, such as pairs of visual images and words, and proceeds by offering this training set to a feature detector followed by a learning algorithm. This form of machine learning, called

J. de Greeff, F. Delaunay and T. Belpaeme are with the School of Computing, Communications and Electronics, University of Plymouth, UK. e-mail: joachim.degreeff@plymouth.ac.uk

supervised learning, does not involve the learner in steering the learning experience and as such, we believe, misses out on a fundamental property of human development. We aim to study how robots and humans can engage in an interaction whereby the robot learner actively steers the learning experience. The algorithms studied in this paper form the basis of this interaction.

The paper presents a learning mechanism and three interactive variations together with experimental results drawn from simulations. The learning mechanism is based on the *language games* mechanism, where teacher and learner engage in a dyadic contextualised interaction [11], [12]. Eventually the teaching role could be assumed by a human, and the learning role by a robot. The model presented in this paper exists within simulation only, the roles of both teacher and learner are assumed by software agents. The variations on the learning mechanism are informed by recent insights in developmental psychology and the intuitions of the authors; more detail is given in the following sections.

## II. MODEL

### A. Representation

The model we use consists of a *conceptual space* which allows for a geometrical representation of conceptual knowledge along various quality dimensions [13]. In a nutshell, a conceptual space is a collection of one or more domains (like colour or tone), where a domain is postulated as a collection of inseparable quality dimensions with a metric. Examples of quality dimensions are *weight*, *temperature*, *brightness*, *pitch*, *loudness* and RGB<sup>1</sup> values. For instance, to express a colour in RGB the different quality dimensions ‘red’, ‘green’ and ‘blue’ are all necessary to express colour values and are hence inseparable. Other domains may consist of more or just one quality dimension. A concept can be represented as a point in the conceptual space, where the coordinates of the point determine the features of the concept. For example, the concept RED is represented as a point on (255, 0, 0) in the RGB colour domain and BLUE as a point on (0, 0, 255). In principle any domain may be used, although for some domains it might be easier to extract the relevant dimensions than for others. In this research the colour domain is used as a test case, but in CIE  $L^*a^*b^*$  encoding because this is more in line with how humans perceive colours [14]. Hence, a colour stimulus consists of three values, where the  $L^*$  dimension encodes for the lightness of the colour and the  $a^*$  and  $b^*$  dimensions respectively encode for a red-green and yellow-blue dimension.

A newly observed colour stimulus can be classified as belonging to a particular existing colour concept by calculating the weighted distance from the stimulus to every concept already present in the conceptual space. The observed stimulus is then assigned to the closest existing concept. Furthermore, the model allows for the representation of concepts through prototypes, which enables it to display typicality effects observed in human conceptualisation. Rosch [15] pointed out

that many everyday concepts are prototypical in nature, i.e. humans regard certain instances for a specific concept to be more typical than others. For example, for the concept BIRD, the instance ROBIN is thought to be more bird-like than the instance PENGUIN. Hence, it seems that specific instances exhibit a graded membership to an idealised prototype concept. Following Gärdenfors, in our model, a conceptual prototype is built through the addition of exemplars (colour stimuli) for the specific concept, where the mean values of all dimensions encode for the coordinates of the prototype and the variance of all exemplars determines the prototype’s size. Hence, a conceptual prototype will not be an exact point in the conceptual space, but rather define a certain convex region.

### B. Lexicalisation

While the conceptual space incorporating conceptual prototypes may stand in itself, in this work it is tightly linked to a lexicon of linguistic labels used to describe the concepts. The linguistic labels can potentially be stored as a string of characters, a visual icon or an acoustic sample. Moreover, these labels play a crucial role in the formation of new concepts. This approach is based on the idea that language is a prime forming factor for the acquisition of new concepts as described in the introduction. Typically, a learning agent is confronted with a stimulus accompanied by a linguistic label. The stimulus is then integrated as exemplar data into the existing conceptual knowledge represented in the conceptual space of the agent. The accompanying label may influence to which existing concept the new data should be assigned. And the other way around, the meaning of the linguistic labels can be perceptually grounded through the values of the associated concepts. Concepts are linked with labels through an association matrix which determines the strength of the connection between every known concept in the conceptual space and every label in the lexicon. Hence, when the agent needs a label to express a specific concept or vice versa, this can be found through consulting the association matrix. When a new concept-label connection is added to the association matrix, it is initiated with a default strength of 0.5. When a connection needs to be increased or decreased (based on success or failure of language game interaction), this is done by increasing/decreasing the strength with 0.01 to 0.0 (minimum) or 1.0 (maximum). Multiple concepts may be associated with the same label.

### C. Knowledge acquisition

The basic learning mechanism for concept acquisition is a language game. This is then augmented with three different interactive features, resulting in four different learning regimes. A description of this is given:

A language game (LG) is implemented as a combination of a *discrimination game* and a *guessing game* [12], [16]. A language game typically consist of two agents, where one agent takes the role of teacher and the other agent acts as learner. Different from the original language game, in our implementation the teacher and learner do not switch roles because we view the teacher as “all knowing”. The teacher

<sup>1</sup>RGB is a technical colour representation suited for reproducing colour on display devices. Red, Green and Blue values here range from 0 to 255.

has a fixed concept-label mapping and uses 11 concepts (basic English colour names)<sup>2</sup>. Both agents are given a set of training stimuli, called the context, from which one specific stimulus is assigned as topic which is known by the teacher only. The teacher then communicates the topic to the learner by stating its associated linguistic label. If the learner is able to correctly identify the topic from the context through the perceived label, the game succeeds. If the learner is not able to correctly identify the topic, either because the label is not known by the learner or because the learner points to the wrong item from the context, the game fails. A failed game provides an opportunity for the learner to improve its conceptual and lexical knowledge. By employing a sequence of language games, the learner is able to build a body of knowledge containing conceptual prototypes and associated labels.

The language game is modified to allow for more interaction by adding features that enable the learner to actively steer learning experience towards gaps in its knowledge and may stimulate the formation of more robust knowledge. These features are:

- 1) *Active learning* (AL). During a guessing game, instead of using a randomly picked topic from the context, the learner actively chooses the topic. This is done by picking the stimulus from the context for which the distance to the most nearby already learnt concepts is the greatest. That is, the most unfamiliar stimulus is chosen as topic<sup>3</sup>. The idea behind AL is that selecting the most unfamiliar stimulus as the topic enables the agent to reach far corners of the conceptual space more quickly. By selecting the stimulus which bears the least resemblance to already known concepts, the agent should be able to achieve a more distributed conceptual knowledge structure. AL could be viewed as a way of modeling novelty preference which is typically observed in young children.
- 2) *Knowledge querying* (KQ). After a specified number of guessing games, the learner queries some of its knowledge built up so far with the teacher. This is

<sup>2</sup>This is the case because we are not only interested in the dynamics of learning concepts through language games as such, but also wish to study mechanism of how an artificial system may learn such knowledge from a human teacher. Indeed, the aim is to eventually implement this mechanism onto robotic hardware which will learn concepts through interaction with humans.

<sup>3</sup>Inspiration has been drawn from Oudeyer and Delaunay [17], which also featured a mechanism called Active learning. The difference with our implementation of AL and that of Oudeyer and Delaunay consists in the fact that we aim to actively explore the far corners of the conceptual space quickly. Hence, the aim is to enable the agent to experience unknown stimuli and build concepts for this. Instead, in Oudeyer and Delaunay the active selection of meaning by the agent serves as a mechanism to gradually control the growth of different meanings and thus strive for a more robust shared lexicon. Because the agent considers introducing a new meaning based on certain criteria (for instance, the average success of the meanings already in use), this active selection can be seen as a method to consolidate the knowledge already learnt, leading to faster convergence among the population. This form of AL is essentially aimed at employment within a community of agents which all interact with one another, while ours is aimed at the learning agent only. In summary, although the term “active learning” is the same, the actual implementation functions differently. It is called “active” because in both cases agents are actively engaged in the dynamics that govern the acquisition of meaning.

done by selecting the concept which has been the least successful during previous language game interaction. This concept is stated to the teacher, along with the associated label. If the teacher confirms the query, i.e. if the label of the teacher for the queried concept is equal to the label of the learner, the strength of the association between the label and the concept of the learner is increased. If the query is not confirmed, this association is weakened. With KQ we aim to implement a common sense intuition, namely that it makes sense to check learned information from time to time and adapt if necessary.

- 3) *Contrastive learning* (CL). During a guessing game, after the learner has successfully identified the topic through the label uttered by the teacher, not only the association between label and topic is increased, but contrastive information is utilised as well. For each stimulus in the context which is not the topic, the learner finds the concept in its knowledge body which is closest, and weakens the association between this concept and the label that the teacher used to describe the topic. This is supported by experimental results from developmental psychology [18] and bears resemblance to lateral inhibition [19] and lexical contrast [20].

#### D. Description of algorithms

Discrimination games are used by the agent to build up a conceptual knowledge body, and guessing games are used to assign the proper labels to the learnt concepts, with help of the teacher.

The discrimination game is as follows:

- 1) Agent  $A$  is confronted with context  $O = \{o_1, \dots, o_N\}$  containing  $N$  stimuli and an index  $i$ , specifying the topic  $o_t$ .
- 2)  $A$  finds the best matching concept  $c$  from its knowledge body  $K_A$  for each stimulus in the context:  $\{o_1, \dots, o_N\} \rightarrow C = \{c_1, \dots, c_N\}$ .
- 3) If the best matching concept on  $i$  is unique in  $C$  the game succeeds, otherwise it fails.

The discrimination game can fail in several ways: this is an opportunity to improve the agents knowledge body. When  $K_A$  is empty, a new category is created on the coordinates of  $o_t$ . When no unique discriminating concepts can be found, there are two possible actions: (1) a new concept is created on  $o_t$ , or (2) the best matching concept  $c$  is adapted to better represent  $o_t$ . This is done by shifting  $c$  towards  $o_t$ . Action (1) is taken when the discriminative success<sup>4</sup> of the agent is below a threshold  $adapt = 0.9$ , otherwise action (2) is taken. In all cases the discrimination game results in  $A$  stating a concept from  $K_A$ .

<sup>4</sup>The discriminative success of an agent is the global success of the agent of all discrimination games it has engaged in. It is measured by dividing the number of times the agent has successfully discriminated the topic from the context by the total number of discrimination games the agent has played.

The guessing game is as follows:

- 1) Teacher  $A^T$  and learner  $A^L$  are confronted with context  $O = \{o_1, \dots, o_N\}$  containing  $N$  objects and the index of the topic, specifying  $o_t$ .
- 2)  $A^T$  plays a discrimination game for  $o_t$ , the discrimination game succeeds and returns the concept  $c^T$ .
- 3)  $A^T$  finds the associated label  $l^T$  and communicates this to  $A^L$ .
- 4)  $A^L$  hears  $l^T = l^L$  and finds the associated concept  $c^L$ .
- 5)  $A^L$  points to  $o^L$  closest to  $c^L$ .
- 6) if  $o^L = o_t$ , the guessing game succeeds; if not, it fails.

When the guessing game is successful, the connection strength between  $l^L$  and  $c^L$  is increased and  $o_t$  is added as an exemplar of  $c^L$ , effectively shifting the coordinates of  $c^L$  a bit towards  $o_t$ .

The guessing game can fail in several ways. (1) The discrimination game of  $A^T$  fails; in this case the guessing game fails as well. (2)  $A^L$  does not know  $l^T$ .  $A^L$  then plays a discrimination game for  $o_t$ , finds  $c^L$  and adds  $l^T$  to its lexicon with a default connection 0.5 to  $c^L$ . (3)  $A^L$  knows  $l^T$ , but points to the wrong topic.  $A^L$  then decreases the connection strength between  $l^L$  and  $c^L$ , plays a discrimination game for  $o_t$ , finds  $c^L$  and adds  $l^T$  to its lexicon with a default connection 0.5 to  $c^L$ .

When interactive learning is used, the language game is augmented with interactive features. These features are:

- Active learning. During the guessing game, when  $A^L$  is confronted with context  $O$ : (1)  $A^L$  finds best matching concept  $c$  in  $K_{A^L}$  for each stimulus in  $O$ :  $\{o_1, \dots, o_N\} \rightarrow C = \{c_1, \dots, c_N\}$ . (2) The distance between every  $o_i$  and  $c_i$  is calculated and stored in  $D = \{d_1, \dots, d_N\}$ . (3) The  $o_i$  with the highest  $d_i$  is chosen as topic for the guessing game by  $A^L$ .
- Knowledge querying. After each language game the success of the concept  $c^L$  used by  $A^L$  is recorded. After a specified number of language games  $A^L$  initiates a knowledge query: (1)  $A^L$  finds the concept in  $K_{A^L}$  with the lowest success rate  $c_{low}$  and the associated label  $l_{low}$  and communicates this to  $A^T$ . (2)  $A^T$  finds the closest concept in  $K_{A^T}$  and the associated label  $l_{match}$ . (3) If  $l_{low} = l_{match}$ ,  $A^T$  answers positive, and otherwise negative. (4) Based on the answer from  $A^T$ ,  $A^L$  increases or decreases the connection strength between  $c_{low}$  and  $l_{low}$ .
- Contrastive learning. (1) After a successful guessing game  $A^L$  examines all objects  $\neg o_t$  in the context and finds  $C = \{c_1, \dots, c_N\}$  in  $K_{A^L}$ . (2)  $A^L$  decreases the connection between  $l^L$  and all objects in  $C$ .

### III. RESULTS

#### A. Experimental setup

In each language game the context consisted of 4 stimuli, including the topic. This context was generated by randomly picking 4 samples from a dataset containing 25,000 pixels drawn with uniform probability from the RGB space and converted into CIE  $L^*a^*b^*$  space. Between all stimuli in the context there was a minimum distance of 50 (to give the reader

an idea of the CIE  $L^*a^*b^*$  distance between typical colours: green-blue is 258, red-blue is 177, yellow-blue is 232 and yellow-green is 70). The teacher and learner engaged in 2000 language games. For all learning regimes (LG, AL, KQ and CL) 300 replicas were run and the average correctness score was calculated.

#### B. Evaluation

To evaluate the performance of the different learning regimes, the conceptual knowledge held by the learner after learning sessions is compared to that of the teacher. This is done by employing a test scenario in which teacher and learner are shown a set of 100 random stimuli<sup>5</sup>. Both teacher and learner then state their associated label for each stimulus in the set. If the two labels are equal, the learner has learnt the label correctly. In this way the learner is assigned score  $S$  which reflects the percentage of correctly learnt labels.  $S$  is calculated as the number of stimuli correctly named by the learner divided by the total number of stimuli in the given set.

#### C. Result

To compare the results of the various learning regimes the LG learning method was used as a baseline performance. Then the interactive features AL, KQ and CL were compared to the baseline LG, which is shown in Figure 1, 2 and 3. Figure 4 displays the performance of the base condition against interactive learning with all three features enabled. As can be observed from the graphs, performance of interactive learning is very close to that of LG learning in all cases. CL performs slightly less, while the other learning regimes perform somewhat better than LG. With all three interactive features concurrently enabled, performance gain is the highest. To determine whether or not the differences were significant, a Welsch two sample t-test was performed (see Table I). It turns out that in all cases except for LG vs CL the difference in performance is small but significant. Overall it can be observed that the learning of concepts through language game interaction may improve somewhat when interactive features are added; both in terms of learning speed and in terms of final performance after learning. The performance of any learning regime never really converges to 100% accuracy. This is due to the fact that although the learner is generally able to form concepts that functionally resemble the concepts of the teacher, the boundaries of the learners' concepts are never exactly the same. So while testing, there will always be border cases which the teacher may call 'concept 1', but the learner 'concept 2'.

We also computed the performance of agents performing randomly, i.e. the agent picks a random label to respond to a stimulus. Typically, the performance of the agent is around 10% (this is proportional to inverse number of labels of the teacher). As can be seen from the graphs, the learners quickly perform better than random and achieve a performance of over 60% after about 100 interactions. Quite a number of concepts

<sup>5</sup>Because of running time considerations we did not use the full set of 25,000 colour samples for evaluation after each training interaction.

the agent holds at that point will in fact be learnt after one or two interactions only. Hence, the learning process bears a resemblance to fast mapping in young children [21]. The fact that the agent scores only 60% is because (1) through random selecting the stimuli not all colour categories may be encountered already at this point, and (2) the discrimination game of the teacher may fail sometimes (depending on the distance between all the stimuli in the context), rendering some of the 100 interactions not suitable for learning.

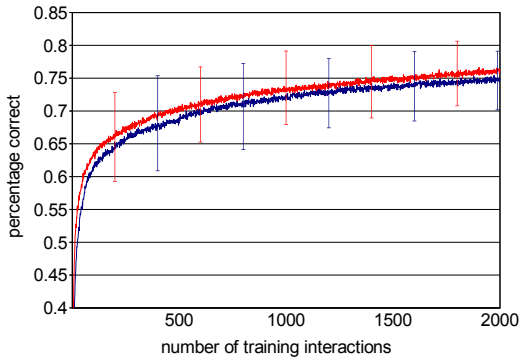


Fig. 1. Performance of LG vs AL. The darker (blue) line indicates LG, the lighter (red) line indicates AL.

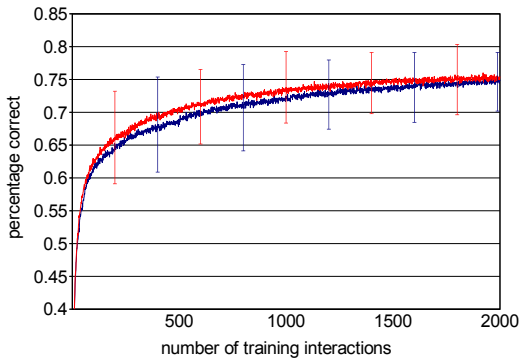


Fig. 2. Performance of LG vs KQ. The darker (blue) line indicates LG, the lighter (red) line indicates KQ.

TABLE I  
TEST RESULTS

	LG vs AL	LG vs KQ	LG vs CL	LG vs all
t-test	t = -2.6114	t = -2.4928	t = 1.5676	t = -7.1749
	df = 597.9	df = 597.3	df = 596.9	df = 590.9
	p = 0.0092	p = 0.013	p = 0.12	p < 0.0001

#### IV. DISCUSSION

The results show that the language game [11], [12] already provides a good base model for acquiring categories in a continuous semantic space. This is perhaps surprising as the language game algorithm was designed to study diachronic

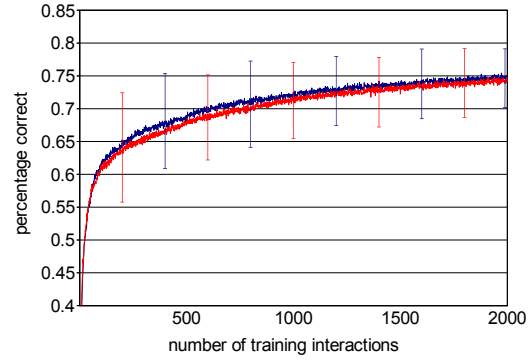


Fig. 3. Performance of LG vs CL. The darker (blue) line indicates LG, the lighter (red) line indicates CL.

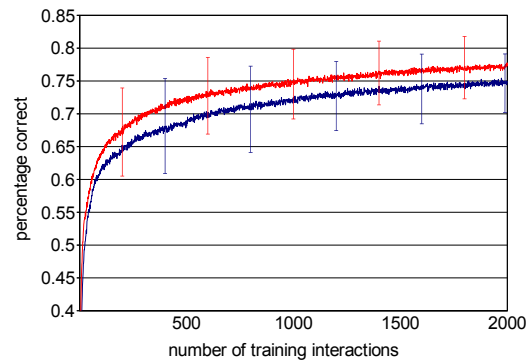


Fig. 4. Performance of learning with all interactive features enabled. The darker (blue) line indicates LG, the lighter (red) line indicates learning with interactive features.

language evolution and not to serve as a machine learning algorithm to acquire human-like knowledge. Variations on the language game, whereby the learner is allowed to actively steer the learning experience, add a relatively small but nevertheless significant improvement to the learners performance, both in terms of speed of learning as in terms of accuracy.

When teaching categories and associated labels to an autonomous agent, such as a robot or software agent, one of the most efficient learning methods remains direct instruction. In this form of instruction prototypes and associated labels are offered to the learner. Several learning algorithms are able to absorb these training examples immediately and the learner will as such, after a limited number of training examples, faithfully reproduce the teacher's categories and labels. While this might be desirable, this learning method is unrealistic for a number of reasons. One is that children and caretakers do not generally seem to use direct instruction, even though children can learn from direct instruction as evidenced by [3]. Their experimental setup however, was not exactly a natural learning setting. Also, learning proper names is usually a one-shot learning experience. However, generally acquiring a category and label requires repeated exposure to different examples of the category together with its label. In some cases learning categories and labels seems inexplicably hard

for children: colour category and colour word learning evolve slowly; even though children are exposed to unambiguous learning experiences on a daily and frequent basis, it takes on average two to three years before basic colour categories and words are correctly learnt [22], [23]. A second reason for not only implementing direct instruction is that caretakers rather use a blend of instruction methods and young children subsequently use a range of learning methods to handle this mixed mode of instruction. A third reason is that learning examples are not always available: when children learn about trees, a prototypical tree is not always readily available. Rather, the concept of tree is gradually constructed from numerous examples of tree seen from different angles, at different stages of growth and in different seasons. Children already use incomplete concepts, both linguistically and non-linguistically, and use the implicit and explicit feedback they receive to refine, limit or extend concepts. A fourth and final reason is related to human-machine interaction: the intent of this research is to explore novel learning methods to allow artificial agents to acquire a repertoire of concepts and language through long-term natural interaction with people. Direct instruction is not compatible with this goal.

This research ties in with a larger project, in which it is studied how robots can learn concepts from humans through a natural interaction scheme<sup>6</sup>. Emphasis is both on how human-robot interaction can effectively facilitate the learning process, and on the actual learning algorithms that need to be deployed by the robot in order to properly learn and represent concepts. Language is considered to be crucial for this. Future research will look at how hierarchical structures can be supported within the conceptual space of a robot and explore methods to transfer conceptual knowledge from one robot to another.

#### ACKNOWLEDGMENT

This work is supported by the CONCEPT project (EPSRC EP/G008353/1) and the EU FP7 ITALK project.

#### REFERENCES

- [1] P. Bloom, *How children learn the meanings of words*. Cambridge, MA: the MIT Press, 2000.
- [2] F. Xu, "The role of language in acquiring object kind concepts in infancy," *Cognition*, vol. 85, no. 3, pp. 223 – 250, 2002.
- [3] K. Plunkett, J. Hu, and L. B. Cohen, "Labels can override perceptual categories in early infancy," *Cognition*, vol. 106, no. 2, pp. 665–681, February 2008.
- [4] G. Lupyan, "Labels facilitate learning of novel categories," in *Proceedings of the 6th International Conference on the Evolution of Language*, 2006, pp. 190–197.
- [5] G. Lupyan, D. H. Rakison, and J. L. McClelland, "Language is not just for talking - redundant labels facilitate learning of novel categories," *Psychological Science*, vol. 18, no. 12, pp. 1077–1083, 2007.
- [6] A. Gilbert, T. Regier, P. Kay, and R. Ivry, "Whorf hypothesis is supported in the right visual field but not the left," in *Proceedings of the National Academy of Sciences*, 103(2) 2006, pp. 489–494.
- [7] A. Majid, M. Bowerman, S. Kita, D. B. M. Haun, and S. C. Levinson, "Can language restructure cognition? the case for space," *Trends in Cognitive Sciences*, vol. 8, no. 3, pp. 108 – 114, 2004.
- [8] D. Roberson, H. Pak, and J. R. Hanley, "Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean," *Cognition*, vol. 107, no. 2, pp. 752 – 762, 2008. [Online]. Available: <http://www.sciencedirect.com/science/article/B6T24-4PVY2XR-1/2/3b1d0525225794ef4c5d19956a236564>
- [9] M. Tomasello, *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press, 1999.
- [10] —, *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press, 2003.
- [11] L. Steels, "The synthetic modeling of language origins," *Evolution of Communication*, vol. 1, no. 1, pp. 1–34, 1997.
- [12] L. Steels and T. Belpaeme, "Coordinating perceptually grounded categories through language: A case study for colour," *Behavioral and Brain Sciences*, vol. 28, no. 4, pp. 469–89, August 2005, target Paper, discussion 489-529.
- [13] P. Gärdenfors, *Conceptual Spaces: The Geometry of Thought*. Cambridge, MA: the MIT Press, 2000.
- [14] M. Fairchild, *Color Appearance Models*. Reading, MA: Addison-Wesley, 1998.
- [15] E. H. Rosch, "Natural categories," *Cognitive Psychology*, vol. 4, no. 3, pp. 328 – 350, 1973.
- [16] T. Belpaeme and J. Bleys, "Explaining universal colour categories through a constrained acquisition process," *Adaptive Behavior*, vol. 13, no. 4, pp. 293–310, 2005.
- [17] P.-Y. Oudeyer and F. Delaunay, "Developmental exploration in the cultural evolution of lexical conventions," in *Proceedings of the 8th International Conference on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*. Lund, Brighton: Lund University Cognitive Studies, 2008.
- [18] C. O'Hanlon and D. Roberson, "What constrains children's learning of novel shape terms?" *Journal of Experimental Child Psychology*, vol. 97, pp. 138–148, 2007.
- [19] M. Oliphant, "The learning barrier: Moving from innate to learned systems of communication," *Adaptive Behavior*, vol. 7, no. 3/4, 1999. [Online]. Available: <http://www.isrl.uiuc.edu/~amag/langev/paper/oliphant99theLearning.html>
- [20] E. V. Clark, *The lexicon in acquisition*. Cambridge: Cambridge University Press, 1993.
- [21] S. Carey, "The child as word learner," in *Linguistic theory and psychological reality*, M. Halle, J. Bresnan, and G. A. Miller, Eds. Cambridge, MA: The MIT Press, 1978, pp. 305–315.
- [22] N. J. Pitchford and K. T. Mullen, "Conceptualization of perceptual attributes: A special case for color?" *Journal of Experimental Child Psychology*, vol. 80, pp. 289–314, 2001.
- [23] C. M. Sandhofer and L. B. Smith, "Learning color words involves learning a system of mappings," *Developmental Psychology*, vol. 35, no. 3, pp. 668–679, 1999.

<sup>6</sup>the CONCEPT project, see <http://www.tech.plym.ac.uk/SoCCE/CONCEPT/> for more details