Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Knowledge Technology
Department of Informatics
University of Hamburg

Seminar Thesis

Lingodroids: Language Learning Robots

Submitted by: Jörg Winkler
Matr. Nr. 6591861

on: December 17, 2014

Degree/Programme: Master of Science / Bioinformatics

Advisor: Dr. Sven Magg
Co-advisor: Stefan Heinrich

Contents

1 Introduction **5**

2 Foundations of language development **7**

3 Lingodroids playing language games **9**

3.1 Grounding of locations 10

3.1.1 Where are we? 10

3.1.2 Go to 11

3.2 Learning temporal concepts 11

3.3 Communication with different cognitive capabilities 12

4 Comparison of human and android language learning approaches **13**

5 Conclusion and outlook **15**

Bibliography **17**

1 Chapter 1 Introduction

The use of languages is characteristic of human communication, which provides a basis for our social life. We exchange information, we share our emotions and opinions with other people and we are able to make an appointment with someone at a given place and time. If we consider robots to operate with humans they need to communicate as well – in particular, they have to understand spatial and temporal specifications, which are the most fundamental aspects of the natural world [1].

Australian scientists at the University of Queensland [2] have developed language-learning robots called *Lingodroids*. These robots can learn simple terms for space and time and use these terms to meet other robots at a certain location. Lingodroids are equipped with three infra-red proximity sensors, a forward camera, speakers and a microphone. They have wheels for locomotion, but nevertheless look like a small rat (figure 1.1). Their task is to play language games for exchanging and improving their knowledge.

Due to the fact that human language is very complex in syntax and semantics, one might ask whether it is possible for androids to learn human language. To answer this question, let us compare how humans build and

Figure 1.1: Two iRat robots are meeting to play a language game. [2]

develop their language with the way robots do. When and how do individuals establish a new word or phrase and how is it spread among the community afterwards? First of all, we have to discuss some basic concepts of human language development in chapter 2 to get an idea of the most important points we have to compare later in chapter 4. Before that, the Lingodroids project and its methods are introduced in chapter 3 and the different type of language games are described there.

2 Chapter 2 Foundations of language development

Today there is still no consensus about the origins of human language, but it has been figured out that language is the outcome of three evolutionary processes that influence and reinforce each other: social, cultural and biological evolution. The interaction of these processes is shown in figure 2.1. Social evolution is caused by ecological pressures due to the complexity of human social structures. Cultural evolution means the development of more sophisticated language and it tells how specific features emerge and propagate in a population. With biological evolution we can explain the origins of the neural architecture necessary for language, which includes increased brain capacity. [3]

Steels [3] hypothesises that social evolution is the most important factor for language evolution to take place. For cultural and biological evolution there are two different kind of views. *Biolinguists* claim that there is a strong biological determination of the structure of language and they try to show that language features are genetically or epigenetically transmitted and subject to natural selection. On the other hand, *Evolutionary Linguistics*

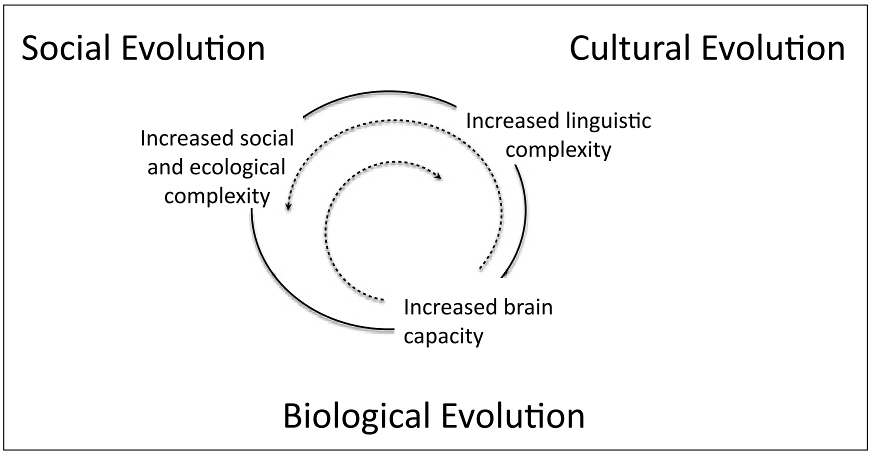


Figure 2.1: Three evolutionary processes as basis for language evolution [3]

say that language is shaped by cultural forces and they want to prove that language features are not linked to biological fitness. [3]

For forming languages, it is necessary to be able to link names to one's own internal representations, which is called *private grounding*. That means to create a personal map that is based on experience. Due to the fact that language is shared among a community, one needs to develop a standard language for shared terms, which is called *social grounding*. All in all, social grounding is the connection of the private representations to shared symbols. In this context the term *grounding* means to make symbols intrinsic to a system, instead of just linking them to other symbols. [4]

Chapter 3

3 Lingodroids playing language games

The aim of language games is the transmission of the language in a population from one generation to the next. In one single game there are two participants, the *speaker* and the *hearer*. Typically, the speaker is from the older generation and the hearer from the new generation. At the start of an agent's lifetime it has no 'knowledge' and is taught by the other agents. Later on, when the agent has gotten some experience, it is able to teach other agents. [5]

In the Lingodroids project, the scientists want to develop a practical system that can be used by real robots to form a shared spatial language, and they therefore designed the Lingodroid robots to understand the relationship between shared language and private maps [4]. They make these robots play language games, which are described in the order of development in the following sections.

3.1 Grounding of locations

This section deals with the way Lingodroids learn spatial language. Space is one of the most fundamental things in the natural world – that is why the group who have developed the Lingodroids started with spatial concepts.

In the private grounding process, the agents build a so-called *cognitive map* of their environment. As stated in section 2, these maps are private and dependent on the agent's personal experience. A cognitive map enables a robot to pay attention to locations in space and not just the objects – this is the key innovation for the Lingodroids. The Lingodroids' mapping abilities are based on the RatSLAM platform, which provides a complete autonomous navigation system [4] and is inspired by the rodent hippocampus [2]. In the following paragraphs, the language games for forming a toponymic lexicon are described.

3.1.1 Where are we?

In this language game, the robots are continually driving around in their environment and every time they meet a colleague they talk to each other about the location where they are. A meeting means in this context, that they are close enough to communicate with each other. The communication proceeds the following way: after a handshake, which establishes shared attention, the speaker produces an utterance to name the current location. The choice of this utterance is covered in section 4. The hearer updates its lexicon and sends an acknowledgement. If the speaker receives the acknowledgement, it also updates its lexicon. After several conversations the robots have formed a toponymic lexicon that covers all locations in the environment and links shared words to private experiences. [4]

3.1.2 Go to

After a handshake (as in the paragraph before), the speaker names a goal word from its lexicon which refers to a remote location. The hearer looks after this word in his lexicon and if no match is found, the conversation stops here. Otherwise, if the hearer can plan a path to the destination, both robots start moving and independently follow their route to the goal. On arrival, both agents announce their arrival to each other. The aim of this game is to test the correlation of the languages, so the agents do not update their lexicons. [4]

3.2 Learning temporal concepts

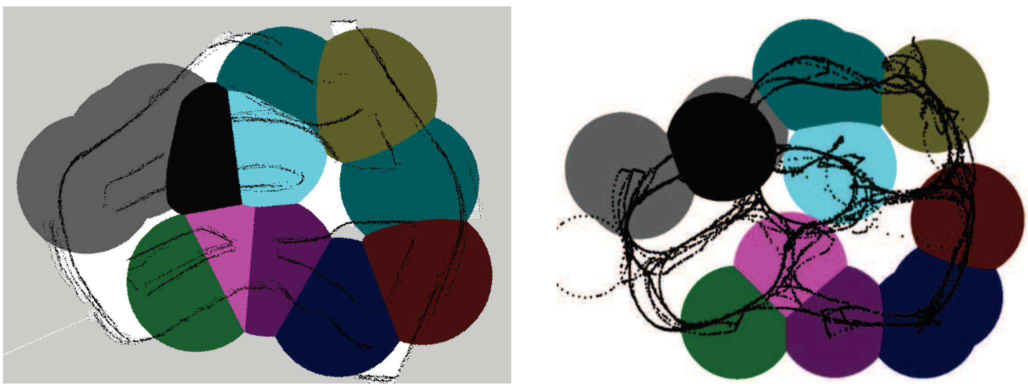
In a further step, the Lingodroids learn terms for time. In *how-long* conversations, they calculate the duration of a journey from one location to another by the distance travelled [6]. In the *how-long-since-we-met* game, the robots estimate the time since they last interacted and name this duration. The final step, which brings spatial and temporal concepts together, is *meet-at* games. One robot specifies location and time of a future meeting and the challenge for both robots is to arrive at the specified location at the specified time. [2]

These experiments show that Lingodroids are indeed able to learn terms for temporal concepts that are grounded in their own experience and that they can be used in a meeting task. The problem with time is, that humans often use inexact words like *recently* or *soon*, which is in general hard to estimate for a machine because the time scale depends on the context in which these terms are used. [2]

3.3 Communication with different cognitive capabilities

In the previous sections, the Lingodroids always have identical cognitive architectures, sensors and mapping systems. This setting is not given in real world conversations, because there may be different species or available sensors. To test how Lingodroids can cope with that, the scientists use two different kinds of robots. The first type has got a forward facing camera to estimate appearance, and for generating a topological map it uses a bioinspired continuous attractor network. The second type estimates range with a laser scanner and uses a probabilistic filter to generate an occupancy grid. [7]

The results are two different maps due to the different mapping systems. Nevertheless, they show a high degree of similarity between the locations for all of the toponyms (see figure 3.1). So there are very different spatial representations formed from the characteristic sensors and they cannot be shared by direct transfer. Instead of interpreting each other's maps, the Lingodroids are able to use the shared experience. For this it is necessary that the robots share a common process for learning a shared symbol, such as hearer/speaker roles and sharing attention. [7]



(a) Map of laser robot (b) Map of camera robot

Figure 3.1: Maps and Toponymic Lexicons for different robots [7]

Chapter 4

4 Comparison of human and android language learning approaches

As shown in the previous chapter, artificial agents are set up with some necessary functions enabling them to play successful language games. These functions, however, do not include any concrete choices about how to formulate the meanings they want to transmit.

An important question is the speaker's decision of whether the robot should use the most likely word it already knows or invent a new one. According to [2,4,7], the probability p_i of inventing a new word is

$$p_i = \exp\left(-\frac{h_{ij}}{T(1-h_{ij})}\right)$$

where h_{ij} is the likelihood of the current best word and T is a scaling parameter to adjust the rate of word invention. That means that if a known word fits well to the semantic the robot wants to express (high h_{ij}), the nominator increases while the denominator decreases, which causes the probability of inventing a new word to be very low.

While robots use probability formulae, human word creation is inspired by cultural evolution, as introduced in section 2. Steels [3] shows that we increase the complexity of our language by introducing new terms. For example, instead of the term *yellow-red* we started to use the word *orange* at the beginning of the 16th century. While humans often create neologisms with associations (for example the name of the fruit orange for the appropriate colour), robots create new words from a given set of simple syllables.

According to social evolution, a population chooses one language over another based on its social interactions. A study with two separate populations of Lingodroids shows that the communal language depends on the way communities interact with each other. Although there was limited interaction between populations, the acquisition of new words was substantial and the mutual understanding could be well established. [8]

5 Chapter 5
Conclusion and outlook

We have seen that the way Lingodroids learn languages is highly bio-inspired. They are able to learn and use spatial and temporal concepts, even if they have different cognitive capabilities. Future work on Lingodroids will contain temporal phrases in different contexts and scales [2]. For example the phrase "the train will come *soon*" means about 5 minutes, while the statement "Christmas is *soon*" could be half a month.

It is an interesting fact that language spreads quickly among the population and it is obvious that self-organisation of language happens with every conversation. Robots always update their lexicons such that common words are established and rare words die out. If, for instance, humans move to a different county where a certain dialect is spoken, they are very likely to adopt some of the special phrases there to reach more successful conversations.

Although spatial and temporal concepts are very important for conversations, they are still not enough to replace human dialogue partners. There is a lack of many additional concepts for real human interaction, and language games performing conversations with humans would be necessary. All in all the current learning techniques are a well-suited basis to deal with these issues.

As the studies of Steels implicitly assume the necessity of feedback on language use, the Lingodroids project assumes that coherence is accomplished without any feedback on the success of the interaction [4]. The Lingodroids language games have shown that unsupervised learning without feedback indeed leads to convergence.

The grammar of the robots' language is given by the structure of the conversation, as explicitly described in sections 3.1.1 and 3.1.2. This is a simplification, which cannot be applied in human-robot-conversations. Grammatical features of human languages could be introduced in advanced language games, but this grammar is so complex that a huge amount of interesting future work with the Lingodroids will emerge.

Bibliography

[1] Ruth Schulz, Arren Glover, Gordon Wyeth, and Janet Wiles. Robots, communication, and language: An overview of the lingodroid project. *Australasian Conference on Robotics and Automation (ACRA)*, 2010.

[2] Scott Heath, Ruth Schulz, David Ball, and Janet Wiles. Lingodroids: learning terms for time. *ICRA 2012, The International Conference on Robotics and Automation, Saint Paul, MN*, pages 1862–1867, 2012.

[3] Luc Steels. Modeling the cultural evolution of language. *Physics of Life Reviews*, 8(4):339–356, Dec 2011.

[4] Ruth Schulz, Gordon Wyeth, and Janet Wiles. Lingodroids: socially grounding place names in privately grounded cognitive maps. *Adaptive Behavior, 19 6*, pages 409–424, 2011.

[5] Paul Vogt. The emergence of compositional structures in perceptually grounded language games. *Artificial Intelligence*, 167(1-2):206–242, Sep 2005.

[6] Ruth Schulz, Gordon Wyeth, and Janet Wiles. Are we there yet? grounding temporal concepts in shared journeys. *IEEE Transactions on Autonomous Mental Development*, 3:163–175, 2011.

[7] Scott Heath, David Ball, Ruth Schulz, and Janet Wiles. Communication between lingodroids with different cognitive capabilities. *ICRA 2013, The International Conference on Robotics and Automation*, pages 490–495, 2013.

[8] Ruth Schulz, Matthew Whittington, and Janet Wiles. Language change in socially structured populations. In Thomas C. Scott-Phillips, Mónica Tamariz, Erica A. Cartmill, and James R. Hurford, editors, *9th International Conference on the Evolution of Language*, pages 312–319. World Scientific, Mar 2012.