Robots, Communication, and Language: An Overview of the Lingodroid Project

Ruth Schulz[1], Arren Glover[2], Gordon Wyeth[2] and Janet Wiles[1]
School of Information Technology and Electrical Engineering
The University of Queensland[1]
School of Engineering Systems
Queensland University of Technology[2]
{ruth, wiles}@itee.uq.edu.au
{aj.glover, gordon.wyeth}@qut.edu.au

Abstract

The foundation of a language for a mobile agent – robot or human – is the representation of spatial and temporal concepts. These concepts include where and when events, objects and agents are located in space and time. This paper presents an overview of the Lingodroid project, in which real and simulated robots play language games to ground concepts for effective communication about their world. A series of location language games is described, with typical lexicons presented. We present future plans to extend the abilities of the robots to ground concepts for objects attributes, actions, and time.

1 Introduction

For robots to operate effectively in unstructured environments in close proximity to ordinary people, including children, injured people and the elderly, communication is a fundamental requirement. People in proximity to one another develop natural languages (such as English, Warlpiri, Mandarin, or Swahili), which enable communication about the concepts and events that are relevant to completing daily tasks and coordinating joint actions. Human cognition includes powerful learning systems that enable a new member to learn the language of a group and also the group to continually develop new concepts and the words, phrases and cultural practices to use them effectively in communication [Keller, 1994]. When a group of people have no shared language they develop pidgin languages [Aitchison, 1991].

No robotic or computational system has yet approached the scope and flexibility of human communication despite half a century of research in linguistics and artificial intelligence. The limitations are not in the words or grammars of natural language, although these can be complex, but rather in understanding natural concepts and the contexts in which they can be used. The connection between a word and its meaning or referent in the world is called *grounding*, and it is at the heart of understanding and effectively using words [Harnad, 1990].

To communicate effectively with ordinary people, robots need to understand concepts about the real world, and have the learning abilities to learn new names and new concepts, as well as how to exchange information (that is, hold a natural conversation). Communication is also ambiguous – words are missed, misheard or misunderstood, and word meanings evolve over time.

The most effective methods for teaching robots grounded concepts have been based on robot language games – games where robots learn both concepts and names through interacting with each other. Robot language games focus on the tasks and embodiment of the robots themselves rather than attempting to directly learning human languages which are optimised for human tasks and communication.

In this paper, we present the Lingodroid project, which uses real and simulated robots playing a series of language games to form grounded concepts for places, directions, distances, and durations. The robots learn a rich set of grounded experiences in both virtual and real world environments. We also present our future plans for extending the series of games to objects, attributes of objects, types of places or objects, actions, time, matching imagined representations to real experiences, and the co-development of grammars and lexicons.

2 Representations, worlds and language games

Space and time are core concepts to be represented and used by mobile agents [Peuquet, 2002]. As such, they also form the foundation for cognition and language [Levinson, 2003]. There are two streams in the Lingodroid project. The first stream is concerned with developing the representations used by the robot to interact with the world, including experience generation and categorisation. The second stream is concerned with developing the language abilities of the robots, including lexicon formation and utilisation. This section provides a brief review of representing experiences and establishing languages.

The internal representation of the world used by the Lingodroids is inspired by the cognitive map representation found in the hippocampus [O'Keefe and Nadel, 1978, Tolman, 1948]. The maps are built by the agents individually, and are unique to each agent, formed through direct experience of and interaction with the world.

To communicate effectively with each other, the robots need to use symbols that are privately grounded in their own experience and socially grounded in shared interactions. A method for establishing a shared grounded lexicon is the use of language games [Steels, 2001]. Using the language game framework, grounding has been solved for concepts that refer to direct experiences [Steels, 2008], but there are still open challenges for grounding [Nolfi and Mirolli, 2010].

2.1 Worlds

The Lingodroid project explores language formation by agents in three types of world: Simulated agents in a grid world, simulated agents in a virtual reality world constructed using high fidelity visual images, and Pioneer DX robots in a real office environment (see Figure 1). The use of different levels of real-world fidelity for the Lingodroid agents allows us to explore different features of the system. The grid world enables a focus on the representations and algorithms specifically for the language abilities of the agents, disregarding any real world ambiguity about shared attention or location of the agents in the world. The simulation world allows the addition of some of the ambiguity for the representations of the simulated robots, but allows access to actual location of the robots within the simulated world. The real world provides a fully embodied system that enables testing of the representations and algorithms in a real environment.

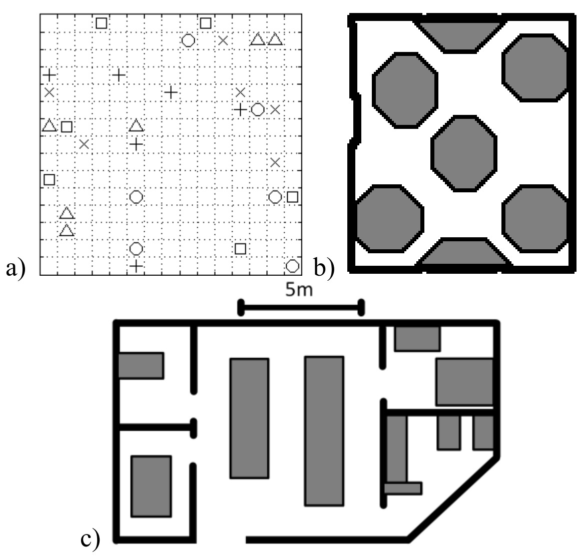

Figure 1 Worlds for the Lingodroids: a) a grid world, b) simulated office, and c) a real office

2.2 RatSLAM

The spatial representation used by the Lingodroids is provided by the RatSLAM system [Milford and Wyeth, 2010], a robot SLAM system based on models of the mapping and navigation processes in the rodent brain. The system consists of several components, of which the *experience map* is directly used by the Lingodroid project (see Figure 2). The experience map provides a semi-metric topological representation of the world that can be used for navigation, and is used by the Lingodroids as the underlying representation for building a lexicon. The experience map is a graph consisting of nodes (experiences) and links that encode the physical and temporal distance between experiences.

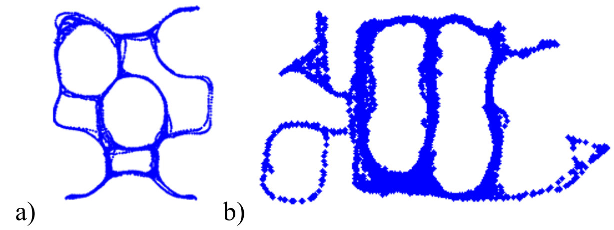

Figure 2 Experience maps in a) the simulated world and b) the real world

2.3 Location Language Games

Language games provide a framework by which agents can build a shared lexicon to describe a set of experienced concepts. In a game, shared attention is used to provide a common topic to two agents. One agent describes the topic with an utterance, and the other agent listens to the utterance. Both agents may update their lexicons based on the utterance. Over time, as many games are played, a range of concepts and utterances are formed to describe the topics experienced by the two agents.

The set of games that have been implemented to date in the Lingodroid project is a series of location language games, in which the focus of the shared attention of the agents is different locations in the world. The games are named for the simple questions location-based that two robots can ask, such as “Where are we?” in which they select a name for their current location, “Go to [X]” in which one robot sets a target location for a meeting. The games implemented so far are described in the next section, together with representative lexicons. In the lexicons presented in the next section, playing 100 language games of a particular type, with a variety of locations specified in each game, generally allows the agents to form a usable shared lexicon. More games are required when the concept space is large, for example in larger worlds.

2.4 Lexicon Representations

Over the course of the project, a variety of lexicon representations have been tested. The most effective has proven to be a system that is similar to an exemplar view of concepts [Medin and Schaffer, 1978, Murphy, 2002]. Each time a word is used in a language game, the word is associated with the concept that is currently the focus of the agent’s shared attention. The agent’s shared attention may be focused through co-location of the agents, or through specifying other concepts through existing combinations of words.

The lexicon information is stored in a distributed lexicon table, which stores counts of the number of times a concept element and word have been used together. Word production and comprehension algorithms allow the agents to generalise from the exemplars experienced to similar concepts.

Words are invented probabilistically, depending on how well the best word matches the current concept. Words will be used in situations that match the previous situation that they have been used in, as well as situations that are close to the previous situations. For example, a place name may be re-used for a location that is a meter away, but a new word will be invented for a location that is five meters away.

3 Lingodroids

Location language games have been played in each world type over a variety of studies [Schulz, et al., 2008, Schulz, et al., 2010]. Presented in this section are examples of each lexicon type formed through the sequence of games.

3.1 Where are we?

There *where-are-we* game involves the agents creating a language to describe specific locations in the world, called *toponyms*, similar to the names of towns and states on a map. To agree upon a location, both agents need to attend to the same aspects of the world. In the *where-are-we* games, they share attention through co-location where co-location is established by one robot hearing an audible signal from the other robot. The robots estimate position in their individual RatSLAM maps, and name the current, shared location (see Figure 3). At the end of a *where-are-we* game, both agents update their toponymic lexicons, increasing the association count between their current location and the word used.

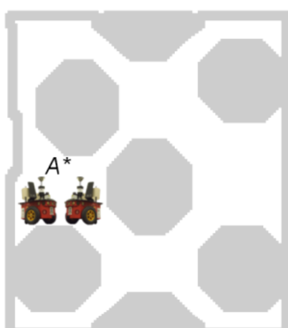

Figure 3 The *where-are-we* game allows the agents to develop a shared toponymic lexicon by updating the associations for the word used (A) and the current location

The *where-are-we* game results in a shared toponymic lexicon in which the agents can refer to several locations in the world through a set of shared toponyms (see Figure 4). The specificity of the toponymic language depends on the hearing distance between the agents, and on the neighbourhood size used by the agents for word production and comprehension.

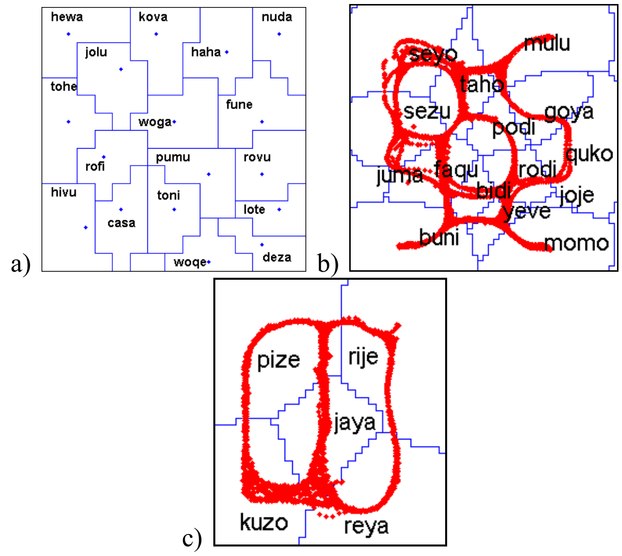

Figure 4 Toponym lexicons in a) the grid world, b) the simulation world, and c) the real world.

3.2 Go to

Go-to games provide a test for whether the shared toponymic lexicon developed by the agents is actually useful (see Figure 5). By testing whether the agents can meet each other at a distant location, we are testing how similar the two lexicons are for comprehension. In a test of the real world language from the previous section, at least one of the agents heard the other agent in 38 of the 50 *go-to* games played.

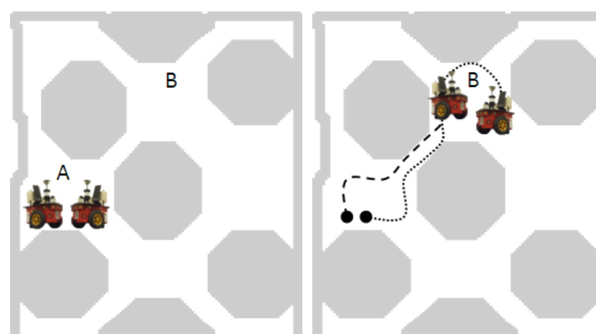

Figure 5 The *go-to* game allows the agents to test their toponymic lexicon by specifying a target location (B) to meet at

3.3 How far?

Once the agents have developed a shared toponymic language, they can use that language to direct the attention of the other agent to remote locations in the world. This ability allows the agents to refer to relationships between these locations. The *how-far* game uses this ability, with the agents referring to the distance between two locations in the world (see Figure 6). The result of the game is a distance lexicon, with which the agents can refer to distances (see Figure 7).

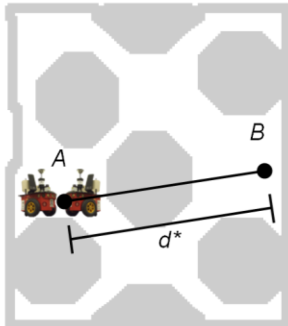

Figure 6 The *how-far* game allows the agents to develop a shared lexicon for distances by finding the distance between two toponyms

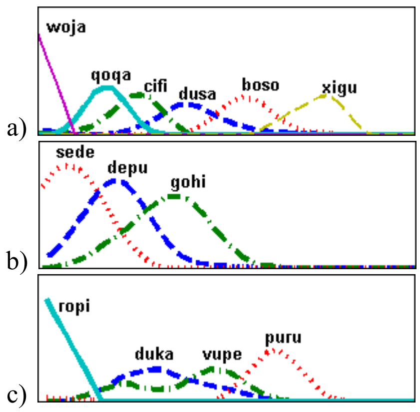

Figure 7 Distance lexicons in a) the grid world, b) the simulation world, and c) the real world.

3.4 What direction?

The *what-direction* games also build on the agents ability to refer to remote locations and the relationships between them, with the agents referring to the angle between two locations when situated at a third location (see Figure 8). The result of the game is a direction lexicon, with which the agents can refer to directions (see Figure 9).

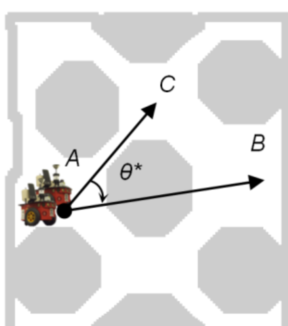

Figure 8 The *what-direction* game allows the agents to develop a shared lexicon for directions by finding the angle between three toponyms

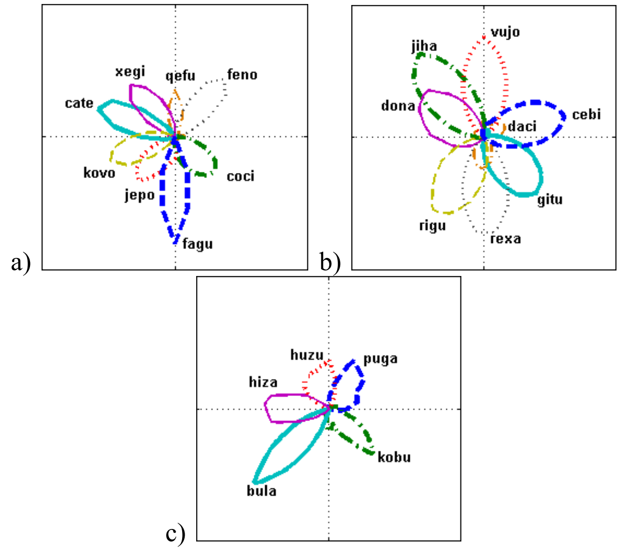

Figure 9 Direction lexicons in a) the grid world, b) the simulation world, and c) the real world

3.5 Where is there?

The *where-is-there* games build on all of the previously formed lexicons: toponyms, distances, and directions. In a *where-is-there* game, the agents use toponyms to specify a current and orientation location together with a distance and direction to refer to a target location (see Figure 10). A toponym for the target location may be invented if there is not already a word in use for that location.

In contrast to the *where-are-we* games which require robots to move throughout their environment and meet at each named location, the *where-is-there* games can be played through communication alone. The games are called "generative" because they enable locations to be specified using combinations of existing terms in the robots' lexicons.

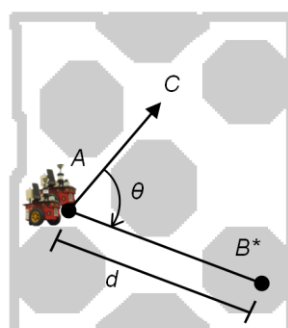

Figure 10 The *where-is-there* game allows the agents to expand their toponymic lexicon generatively by combining concepts from their toponymic, distance, and direction lexicons

The *where-is-there* game allows the agents to expand their toponym lexicon generatively by linking toponyms to locations beyond their existing map of the world through the use of pseudo-experiences (see Figure 11). Pseudo-experiences are only formed generatively (using distance and direction terms combined with already named locations), and do not have any direct experience associated with them. However, if the format of the world changes, and robots are allowed to explore areas of the world that were previously not available, then toponyms formed for those areas become potential targets for the *go-to* game.

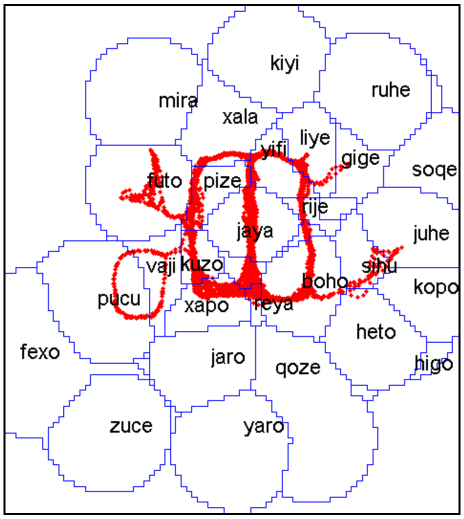

Figure 11 Expanded toponym lexicon in the real world.

An interesting aspect of the toponymic language is revealed through extended periods of *where-is-there* games. As the robots continue to play these communication-only games, their distance and toponym lexicons expand to increasingly larger distances and correspondingly more distance locations as the focus of the agent's shared attention expands with the lexicon (see Figure 12).

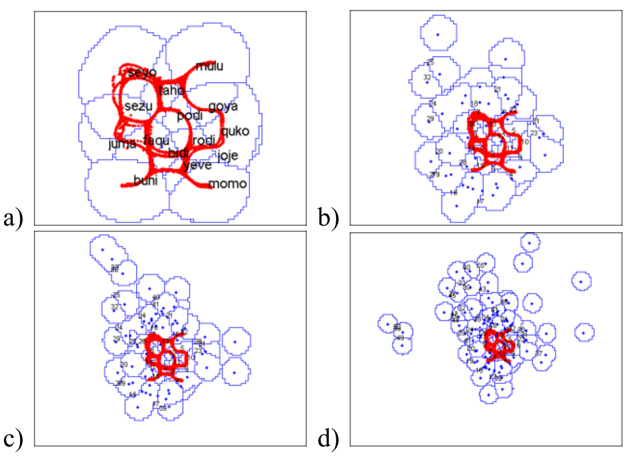

Figure 12 The extended toponym lexicon developing over time in the simulation world a) after *where-are-we* games, b) after sequential *how-far*, *what-direction*, and *where-is-there* games, and c) and d) after concurrent *how-far*, *what-direction*, and *where-is-there* games.

3.6 How long?

The preceding sequence of games constructs a set of spatial concepts. Temporal concepts are also core concepts to be represented by mobile agents. A metaphor that is common to many languages maps time into space [Lakoff and Johnson, 1980]. In the final game described in this section, the *how-long* game (see Figure 13), the concepts formed are the temporal concepts of duration. The duration concepts are constructed by using the existing toponyms as referents. The *how-long* game can be played together with the *go-to* game, with shared attention being the path between two locations in the world, specified using two toponyms. There are a variety of ways in which duration concepts may be formed, including the predicted distance or time taken for a path, and the actual distance travelled or time taken to complete a *go-to* game (see Figure 14).

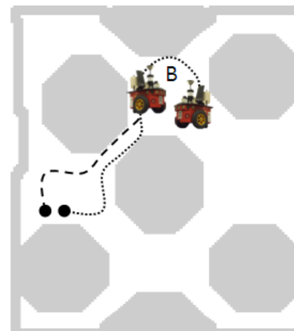

Figure 13 The *how-long* game allows the agents to develop a lexicon for durations referring to how long it takes to travel from one toponym to another

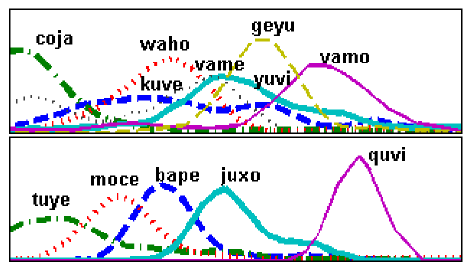

Figure 14 Duration lexicons in the simulation world using estimated (top) and actual (bottom) time taken to move from one toponym to another

4 Extended language development

A target of our current research is to teach a robot 1000 words. In order to do this, we will need to expand on our current toponym, distance, direction, and duration lexicons to new types of concepts. Space and time provide a good base of knowledge for the robots, in a manner similar to human languages in which spatial and temporal concepts are fundamental [Levinson, 2003, Peuquet, 2002].

Evolution of language studies acknowledge that culture and learning play a large role in the nature of the languages that can evolve [Brighton, et al., 2005]. In our studies to date, we have provided the robots with a sequence of language games that constrains the order in which concepts are acquired by the agents. The order in which concepts are acquired by children is constrained by the experiences and interactions the children has with the world [Roy, 2008].

The Lingodroid project is different to previous spatial language [Levit and Roy, 2007, Loetzsch, et al., 2008, Moratz and Tenbrink, 2006] and robot language [Cangelosi and Riga, 2006, Mavridis and Roy, 2006, Steels and Baillie, 2003, Vogt, 2000] models in the combination of exemplar representations and language game framework.

In the following sections, we discuss the extensions planned based on extending our current implementation to different concept types.

4.1 Space

The extensions to Lingodroids planned for the spatial domain include forming type nouns rather than the proper nouns of toponyms, and names for actions involving moving through space. The type nouns will include names for types of rooms such as kitchen, office, and corridor, and will likely involve combinations of features such as sensory-motor actions that can be performed in the area and visual features that are present.

Spatial actions will be bootstrapped from the shared attention achieved through *follow-me* games, in which one robot follows the other robot as it explores the world. The initial game will build concepts for the actions currently being performed by the robots, with a test game describing how to get from one location in the world to another.

For the further development of spatial concepts, we propose three new language games: *what-sort-is-this* for forming type nouns for rooms, *what-did-we-just-do* for developing spatial action concepts, and *how-to* for testing spatial action concepts.

4.2 Time

The temporal concepts addressed to date have been duration concepts, indicating how long it took to complete a *go-to* game. Other temporal concepts include *temponyms*, which can be used as reference points in time, and cyclic concepts such as days and seasons. With further temporal concepts, the robots will be able to situate events in time relative to each other and to the present time, enabling the planning of synchronous behaviours.

For the further development of temporal concepts, we propose a series of temporal language games, including *what-time-is-it*, *how-long-since*, and *how-long-until*.

4.3 Objects

In a parallel stream of the Lingodroid project, the robots are starting to form representations for interacting with objects in their world, and learning how to perform goal actions with those objects. Extending from this work, the robots will be able to discuss names of objects, types of objects, and how to interact with various objects. In the grid world, a *what-animal* game has been developed to name 'animals' that are located in squares in the grid. Talking about objects enables agents to talk about various features of the world that could be recognised when visited.

For the development of object concepts, we plan to add to the grid world *what-animal* game with *what-thing-is-this* in the simulation and real world. The *what-did-we-just-do* game for spatial action concepts could also be applied to object actions.

4.4 Stories

Generative language enables the formation of pseudo experiences, and once the robots have concepts for space and features of the world such as objects, they can start telling stories about what the world looks like in different locations. However, there is still an open question about how a robot could recognise and hence directly ground a location which had previously only been described to it. In current work in the grid world, an agent explores a 'zoo' filled with animals. The agent then describes this zoo to another agent by telling a story of a path through the zoo and describing what animals can be seen at every point along the path. The second agent then visits the zoo, and attempts to follow the same path as described by the first agent, adapting its imagined map of the zoo as it experiences the real zoo. The moment of recognition when an embodied experience matches a verbal description is the universal phenomena of the 'aha' moment [Kounios and Beeman, 2009]. Throughout the exploration of the zoo, the second agent may have several 'aha' moments, as it recognises places from its imagined map in the real zoo.

We propose a storytelling framework, involving one agent telling a story, another agent forming an imagined map, and the ability to use the imagined map for real exploration.

4.5 Grammars

All of the games played by the robots to date have been constructed for the particular studies, and all robots have known what types of concepts are the topics of each game. Rather than defining the grammar of each new game as new representations are available to the robots, we plan to have them develop their own series of games for forming lexicons.

For the co-development of grammars and lexicons, we propose a *what-game-is-this* game, enabling the agents to determine both the type and the content of the games played.

4.6 Summary

With grounding for effective communication and language games for robots, we have a methodology that enables robots to form lexicons for different concept types piece by piece. We began by looking at space and time, with the results presented in this paper demonstrating the power of the framework for these concepts. We are now considering the extensions that can be made to the current system, including the development of interactions and underlying representations for new concepts of actions, time, and objects. A limitation of the current framework is that the interactions and underlying representations of the robots are designed by the experimenter rather than by the robots. However, once the interactions and representations are set, the robots have a way to develop an effective shared lexicon. A challenge for the Lingodroid project is to design a more flexible framework to allow the robots to both design and form their own shared lexicons, and to use their shared language to alter behaviour and to update internal representations such as their cognitive maps.

References

[Aitchison, 1991] J. Aitchison, *Language change: progress or decay?*, 2nd ed. Cambridge: Cambridge University Press, 1991.

[Brighton, et al., 2005] H. Brighton, K. Smith, and S. Kirby, "Language as an evolutionary system," *Physics of Life Reviews*, vol. 2, pp. 177-226, 2005.

[Cangelosi and Riga, 2006] A. Cangelosi and T. Riga, "An embodied model for sensorimotor grounding and grounding transfer: Experiments with epigenetic robots," *Cognitive Science*, vol. 30, pp. 673-689, 2006.

[Harnad, 1990] S. Harnad, "The symbol grounding problem," *Physica D: Nonlinear Phenomena*, vol. 42, pp. 335-346, 1990.

[Keller, 1994] R. Keller, *On language change: the invisible hand in language*. London: Routledge, 1994.

[Kounios and Beeman, 2009] J. Kounios and M. Beeman, "The Aha! Moment: The Cognitive Neuroscience of Insight," *Current Directions in Psychological Science*, vol. 18, pp. 210-216, 2009.

[Lakoff and Johnson, 1980] G. Lakoff and M. Johnson, *Metaphors we live by*. Chicago: The University of Chicago Press, 1980.

[Levinson, 2003] S. C. Levinson, *Space in language and cognition: Explorations in cognitive diversity*. Cambridge, UK: Cambridge University Press, 2003.

[Levit and Roy, 2007] M. Levit and D. Roy, "Interpretation of Spatial Language in a Map Navigation Task," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 37, pp. 667-679, 2007.

[Loetzsch, et al., 2008] M. Loetzsch, R. van Trijp, and L. Steels, "Typological and computational investigations of spatial perspective," in *Modeling Communication with Robots and Virtual Humans, LNCS*, vol. 4930, I. Wachsmuth and G. Knoblich, Eds. Berlin: Springer, 2008, pp. 125-142.

[Mavridis and Roy, 2006] N. Mavridis and D. Roy, "Grounded Situation Models for Robots: Where words and percepts meet," presented at International Conference on Intelligent Robots and Systems, 2006.

[Medin and Schaffer, 1978] D. L. Medin and M. M. Schaffer, "Context theory of classification learning," *Psychological Review*, vol. 85, pp. 207-238, 1978.

[Milford and Wyeth, 2010] M. Milford and G. Wyeth, "Persistent Navigation and Mapping using a Biologically Inspired SLAM System," *The International Journal of Robotics Research*, vol. 29, pp. 1131-1153, 2010.

[Moratz and Tenbrink, 2006] R. Moratz and T. Tenbrink, "Spatial Reference in Linguistic Human-Robot Interaction: Iterative, Empirically Supported Development of a Model of Projective Relations," *Spatial Cognition & Computation*, vol. 6, pp. 63-107, 2006.

[Murphy, 2002] G. L. Murphy, *The big book of concepts*. Cambridge, Massachusetts: The MIT Press, 2002.

[Nolfi and Mirolli, 2010] S. Nolfi and M. Mirolli, "Evolving Communication in Embodied Agents: Assessment and Open Challenges," in *Evolution of Communication and Language in Embodied Agents*, S. Nolfi and M. Mirolli, Eds. Berlin: Springer-Verlag, 2010, pp. 215-220.

[O'Keefe and Nadel, 1978] J. O'Keefe and L. Nadel, *The hippocampus as a cognitive map*. New York: Oxford University Press Inc, 1978.

[Peuquet, 2002] D. J. Peuquet, *Representations of Space and Time*. New York: The Guilford Press, 2002.

[Roy, 2008] D. Roy, "A mechanistic model of three facets of meaning," in *Symbols and Embodiment: Debates on meaning and cognition*, M. De Vega, A. Glenberg, and A. Graesser, Eds. New York: Oxford University Press, 2008, pp. 195-222.

[Schulz, et al., 2008] R. Schulz, D. Prasser, P. Stockwell, G. Wyeth, and J. Wiles, "The formation, generative power, and evolution of toponyms: Grounding a spatial vocabulary in a cognitive map," in *The Evolution of Language: Proceedings of the 7th International Conference (EVOLANG7)*, A. D. M. Smith, K. Smith, and R. Ferrer i Cancho, Eds. Singapore: World Scientific Publishing, 2008, pp. 267-274.

[Schulz, et al., 2010] R. Schulz, G. Wyeth, and J. Wiles, "Language change across generations for robots using cognitive maps," in *12th International Conference on the Synthesis and Simulation of Living Systems (Artificial Life XII)*. Odense, Denmark, 2010.

[Steels, 2001] L. Steels, "Language games for autonomous robots," *IEEE Intelligent systems*, vol. 16, pp. 16-22, 2001.

[Steels and Baillie, 2003] L. Steels and J.-C. Baillie, "Shared grounding of event descriptions by autonomous robots," *Robotics and Autonomous Systems*, vol. 43, pp. 163-173, 2003.

[Steels, 2008] L. Steels, "The symbol grounding problem has been solved. So what's next?," in *Symbols and Embodiment: Debates on meaning and cognition*, M. De Vega, A. M. Glenberg, and A. C. Graesser, Eds. New York: Oxford University Press, 2008, pp. 223-244.

[Tolman, 1948] E. C. Tolman, "Cognitive maps in rats and men," *The Psychological Review*, vol. 55, pp. 189-208, 1948.

[Vogt, 2000] P. Vogt, "Grounding language about actions: Mobile robots playing follow me games," in *From Animals to Animats 6: Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior (SAB00)*, J. A. Meyer, A. Berthoz, D. Floreano, H. Roitblat, and S. W. Wilson, Eds. Cambridge, Massachusetts: The MIT Press, 2000.