Second Language Tutoring using Social Robots

Project No. 688014

L2TOR
Second Language Tutoring using Social Robots

Grant Agreement Type: Collaborative Project
Grant Agreement Number: 688014

D4.2 Input module for the space domain

Due Date: **31/03/2018**
Submission Date: **09/04/2018**

Start date of project: **01/01/2016** Duration: **36 months**

Organisation name of lead contractor for this deliverable: **Plymouth University**

Responsible Person: **Tony Belpaeme** Revision: **1.0**

Project co-funded by the European Commission within the H2020 Framework Programme		
Dissemination Level		
PU	Public	**PU**
PP	Restricted to other programme participants (including the Commission Service)	
RE	Restricted to a group specified by the consortium (including the Commission Service)	
CO	Confidential, only for members of the consortium (including the Commission Service)	

Contents

Executive Summary

This deliverable describes work done towards developing and evaluating the input modalities to support the learning of spatial words in a target language by young children. It consists of published and submitted work on the input device (a touch-screen computer) and on studies evaluating the ability of children to understand and express spatial words in a second language through manipulating objects on the screen using tapping and dragging.

Principal Contributors

The main authors of this deliverable are as follows (in alphabetical order):

Tony Belpaeme, University of Plymouth/Ghent University
Christopher Wallbridge, University of Plymouth
Fotios Papadopoulos, University of Plymouth
Daniel Hernandez, University of Plymouth

Revision History

Version 0.1 (T.B. 27-03-2018)
First draft.

Version 1.0 (T.B. 09-04-2018)
Final version containing submitted ROMAN paper.

1 Overview of the Space Domain Input

Work Package 4, multimodal input processing, aims to leverage existing software and methods for social signal processing. The solutions devised as part of this work package will provide the input from sensing data to enable the lessons for the L2TOR evaluations. The intention is that wherever possible, the input module should utilise current state-of-the-art software following evaluation of its suitability for the specific scenarios in L2TOR. This approach encourages efficient use of resources, whilst simultaneously providing substantial information for use by the robotic platform from the world (and specifically, interacting partners) around it.

2 Tablet game

The Table Game is a HTML based 3D game (described in Deliverable 4.1) with collision detection and virtual object manipulation through touch that supports spatial reasoning through Underworlds[1] (a "behind the scenes" 3D reasoning engine for spatial relations in a Human-Robot Interaction context [1]). It has been further developed to allow the operator to design and build a variety of scenes within the user interface. An editor has been embedded in the Tablet Game and a designer can easily add or remove 3D objects and animations to the scene. An example of a scene developed with the editor can be seen in Figure. 1.

Figure 1: Bakery scene designed with the game editor.

In this example, the designer added a number of 3D objects, some of them including animations to provide a more life-like environment. An example is the oven which contains an animation of the rising of bread. The completed scenes can be saved as JSON files locally that can later be loaded on demand by the Interaction Manager. Additionally, the Tablet Game can move objects around the scene

[1]Underworlds is available as an open source project at `https://github.com/underworlds-robot/underworlds`.

when called from the Interaction Manager in various locations, for example to mimic a child running up and down the road. Various effects have been added to the game to give feedback to the children such as fireworks at the end of each scene, shaking or jumping 3D models and rolling animations. In addition, the designer or the Interaction Manager can change the perspective of the scene by changing the angle and distance of the camera in the 3 game engine. For example, the scenario might involve interaction with specific areas of the scene therefore the camera can zoom in and change the angle to allow easier object manipulation from the child. The modifiable camera perspective also has the benefit of providing a better spatial understanding of the scene as it allows the child to a view on the scene from different angles.

The collision model has been updated to allow seamless integration with Underworlds. Figure. 2 shows a Zoo scene from the spatial domain with multiple animals that the child must interact with following the instructions of the robot. Underworlds is cloning the Tablet environment constantly and is aware of the spatial relationships between the visible objects on the screen. Every time the child is dragging a 3D object in the game, Underworlds will mirror the positions and produce a spatial relationship between the objects. For example, if the child moves the giraffe on top of the lake, Underworlds will output the spatial relation on top between the giraffe and the lake. The Interaction Manager will use this output the check if the child has completed the required task and provide the appropriate feedback via the robot and the Tablet Game.

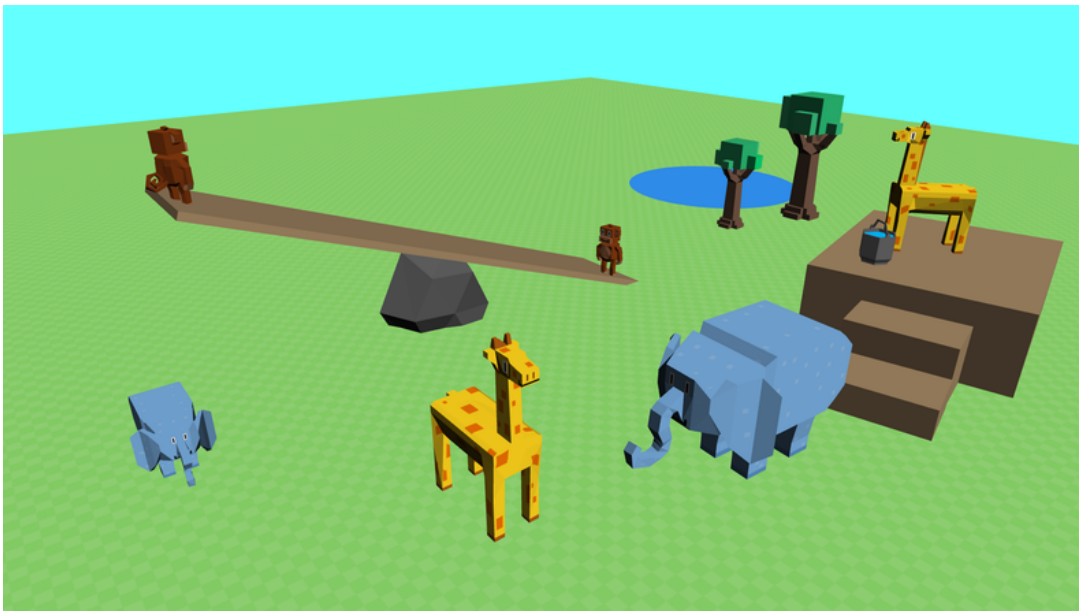

Figure 2: Zoo scene for the space domain

Finally, we added an error resuming function that allows the Tablet Game to resume the last saved state of the scene if any of the modules that comprise the system crashes and the operator needs to restart the system. In such as a case, all the modules will resume from the saved state allowing the child to continue from the same point.

2.1 Further developments to Underworlds

To enable the teaching of the spatial domain the development efforts for Underworlds focused to now include all the target spatial words required. It has also been integrated into the interaction manager to enable quicker communication between the two modules. Further developments have also been made to the way in which Underworlds loads models into a scene to reduce overall loading time and offer a smoother user experience.

3 Evaluating the comprehension and production of spatial language

We have been investigating the use of spatial language in two ways:

1. Encouraging the production of spatial language in L2
2. Looking at how children use spatial language naturally in L1

3.1 Encouraging Production

We were especially looking at the difference between receptive (what we can understand) and productive (what we can say) vocabularies. It has been established that the receptive vocabulary is often bigger than productive vocabulary [2, 3], and that often L2 learners perform much worse on productive tests than in receptive ones [4]. This has been formalised as a hierarchy [5] from bottom to top:

1. *Passive recognition* - The student is able to select the L1 word from a choice of words when provided the word in L2.
2. *Active recognition* - The student is able to select the L2 word from a choice of words when provided the word in L1.
3. *Passive recall* - The student is able to give the meaning of a word in L1 when provided the word in L2.
4. *Active recall* - The student is able to give the L2 word when provided the word in L1.

We conducted a study to see if we could use a robot to encourage the production of spatial language. Spatial language proves an interesting challenge for production as unlike an object it can't be just be pointed to, and is harder to explain using an image. We believed that a robot may give us a better measure of a child's language production capability due to previous evidence of their ability reduce anxiety in students [6, 7].

For the study itself we used a Sandtray environment [8], with a child sitting opposite an agent(experimenter or robot) across a large touch screen device(Figure. 3). After a lesson with a French tutor in the morning, in small groups, children would come to play the production quiz game individually. The game itself involved the children describing in French the position of the teddy bear in relation to the chair in the image displayed on the Sandtray using the words they had been taught. Children would play this game either with a robot, or with one of the experimenters.

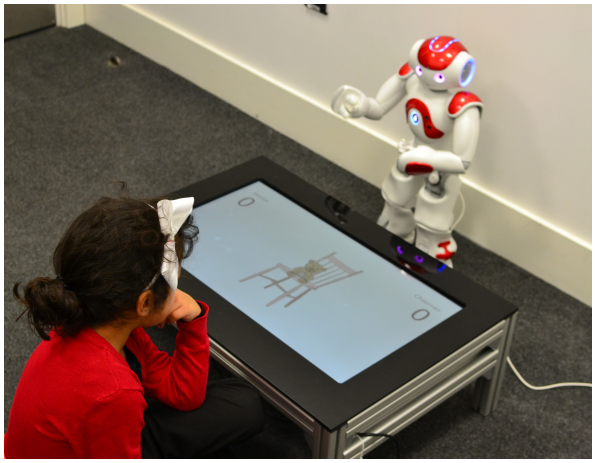

Figure 3: A child interacting with the robot in our study. The agent – in this case a robot – stands opposite from the child. An interactive table (sandtray) displays an image of a teddy bear and a chair. The child must use a word from a second language to describe the position of the bear in relation to the chair.

We found that the robot was able to match a human experimenter, despite the greater social ability that the person was able to display. This is a very encouraging result, as it demonstrates the robot is able to assess production as well as comprehension of learned words. The set of encouragement that could be given was severely limited so as not to change the nature of the task. This made the game very repetitive for an experimenter, which could have led to breaks in protocol. As expected this was a challenging task for the children, and it could be emotionally stressful for an experimenter attempt to assess the children without breaking protocol when they were struggling. Further developments in social robotics may lead to robots being able to surpass a human in the ability to encourage production.

3.2 Use of Spatial Language

We wanted to look at the natural use of spatial language in L1 for young children to establish a benchmark for how spatial language was used by children that could inform future design decisions. We also wanted to see if this was affected by the presence of a robot. For this study we used the same sandtray environment as in section 3.1. In the study children were asked to describe the location of objects –using a reference map (Figure 4– to a manipulator (either another child, an experimenter, or the robot) who would then move the objects on the sandtray to place them in the position described.

We found in contradiction to typical HRI implementations, these revolve around a single complete description that eliminates ambiguity, that the process of describing the position of objects was much more fluid and ambiguous. Rather than a ‘walkie-talkie’ like interaction the manipulator was often involved in the description process, by guessing based on limited information, enabling the describer to use more words to narrow down a position.

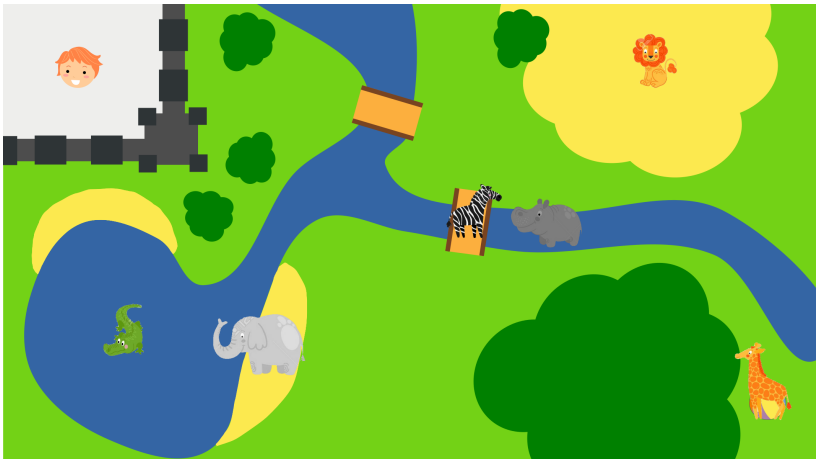

Figure 4: An example of the reference map given to a child to describe. The eight items (face, crocodile, elephant, zebra, hippo, lion, giraffe and ball) are shown in the desired location that they need to be moved to. The child describes the position on his map for an agent to manipulate into the correct position.

4 Annex descriptions

4.1 Wallbridge, C.D. (2018), Encouraging the Production of Spatial Concepts in L2 for Young Children Using a Robot Peer

Bibliography – Wallbridge, C.D., Van den Berghe, R., Hernandez Garcia, D., Kanero, J., Lemaignan, S., Belpaeme, T. (2018) Encouraging the Production of Spatial Concepts in L2 for Young Children Using a Robot Peer. Submitted to *The 27th IEEE International Conference on Robot and Human Interactive Communication (ROMAN 2018)*.
Abstract – When discussing second language learning, we must recognize the difference between the language we understand (receptive vocabulary) and the language we use (productive vocabulary). As receptive vocabulary is considered to be an easier and more sensitive measure of a student's knowledge, productive vocabulary is not often measured. At the same time, previous studies on foreign language learning have found that robots can help to reduce language anxiety, leading to improved results. We conducted a study with 25 children to measure the effectiveness of a robot measuring and encouraging production compared to a human experimenter. We found that a robot is able to match the experimenter's performance in getting children to produce, despite the person's advantages in social ability, and discuss the extent to which a robot may be suitable for this task.
Relation to WP – Increasing input to robotic systems of L2.

4.2 Wallbridge, C.D. (2018), Spatial Referring Expressions in Child-Robot Interaction: Let's Be Ambiguous!

Bibliography – Wallbridge, C.D., Lemaignan, S., Senft, E., Edmunds, C., Belpaeme, T. (2018) Spatial Referring Expressions in Child-Robot Interaction: Let's Be Ambiguous!. To be published in *Proceedings of the 4th Workshop on Robots for Learning - Inclusive Learning*.
Abstract – Establishing common ground when attempting to disambiguate spatial locations is difficult at the best of times, but is even more challenging between children and robots. Here, we present a

study that examined how 94 children (aged 5-8) communicate spatial locations to other children, adults and robots in face-to-face interactions. While standard HRI implementations focus on non-ambiguous statements, we found this only comprised about 20% of children's task based utterances. Rather, they rely on brief, iterative, repair statements to communicate about spatial locations. Our observations offer strong experimental evidence to inform future dialogue systems for robots interacting with children. **Relation to WP –** Design considerations for input on spatial language.

References

[1] Séverin Lemaignan, Mathieu Warnier, E Akin Sisbot, Aurélie Clodic, and Rachid Alami. Artificial cognition for social human–robot interaction: An implementation. *Artificial Intelligence*, 247:45–69, 2017.

[2] Batia Laufer and T. Sima Paribakht. The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning*, 48(3):365–391, 1998.

[3] Batia Laufer. The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics*, 19(2):255–271, 1998.

[4] Jan-Arjen Mondria and Boukje Wiersma. Receptive, productive, and receptive + productive l2 vocabulary learning: What difference does it make? In Paul Bogaards and Batia Laufer, editors, *Vocabulary in a Second Language: Selection, Acquisition and Testing*, pages 79–100. John Benjamins Publishers, 2004.

[5] Batia Laufer and Zahava Goldstein. Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54(3):399–436, 2004.

[6] Sungjin Lee, Hyungjong Noh, Jonghoon Lee, Kyusong Lee, Gary Geunbae Lee, Seongdae Sagong, and Munsang Kim. On the effectiveness of robot-assisted language learning. *ReCALL*, 23(01):25–58, 2011.

[7] M Alemi, A Meghdari, and M Ghazisaedy. The impact of social robotics on l2 learners' anxiety and attitude in English vocabulary acquisition. *International Journal of Social Robotics*, pages 1–13, 2015.

[8] Paul Baxter, Rachel Wood, and Tony Belpaeme. A touchscreen-based'sandtray'to facilitate, mediate and contextualise human-robot social interaction. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 105–106. ACM, 2012.

A Annexes

Encouraging the Production of Spatial Concepts in a Second Language with a Robot Peer

Christopher D. Wallbridge[1], Rianne van den Berghe[2], Daniel Hernandez Garcia[1], Junko Kanero[3], Séverin Lemaignan[1], Charlotte Edmunds[1], Tony Belpaeme[1,4]

***Abstract*— When discussing second language learning, we must recognize the difference between the language we understand (receptive vocabulary) and the language we use (productive vocabulary). As receptive vocabulary is considered to be an easier and more sensitive measure of a student's knowledge, productive vocabulary is not often measured. At the same time, previous studies on foreign language learning have found that robots can help to reduce language anxiety, leading to improved results. We conducted a study with 25 children to measure the effectiveness of a robot measuring and encouraging production compared to a human experimenter. We found that a robot is able to match the experimenter's performance in getting children to produce, despite the person's advantages in social ability, and discuss the extent to which a robot may be suitable for this task.**

I. Introduction

Learning the language of a new home region is vital for migrant children. It is required for them to integrate with their peers, and necessary to prevent them from falling behind in school. Children need the opportunity to practice their language skills, but it may be difficult if no one at home is able to speak the language of the host region. Finding qualified teachers or tutors that know both the new language and the language of children's old homeland can also be challenging. With robots we may be able to support children's language learning needs.

When learning a second language (L2), it is difficult to master vocabulary both receptively and productively. L2 learners may find themselves capable of understanding the L2, while still struggling to produce L2 words. Indeed, previous research has shown that receptive vocabulary tends to be bigger than productive vocabulary in first language (L1) [1], [2], and that L2 learners obtain lower scores on productive tests as compared to receptive tests [3]. Thus, people are able to recognize more words than they can produce, both in their L1 and L2. This has been formalised into a hierarchy for word knowledge by Laufer et al. [4], based on knowing the words passively or actively and in being able to recognize them or recall them. The hierarchy is as follows, from easiest to most difficult: passive recognition → active recognition → passive recall → active recall. These are defined as follows:

- *Passive recognition* - The student is able to select the L1 word from a choice of words when provided the word in L2.
- *Active recognition* - The student is able to select the L2 word from a choice of words when provided the word in L1.
- *Passive recall* - The student is able to give the meaning of a word in L1 when provided the word in L2.
- *Active recall* - The student is able to give the L2 word when provided the word in L1.

This poses a challenge for L2 vocabulary interventions in which the trainer wants to assess the trainee's learning gains: L2 learners have difficulty learning the words productively (i.e. learning to produce foreign words), and will struggle to actively recall newly learned L2 words. There are several tests to assess an L2 learner's productive vocabulary, including assessments in which the participant has to describe pictures (e.g., the Expressive Vocabulary Test [5], the Expressive One-Word Picture Vocabulary Test [6], or the Clinical Evaluation of Language Fundamentals Test [7], writing tests in which the learner has to fill in the blank (e.g., the Productive Vocabulary Levels Test [8]), or, for very

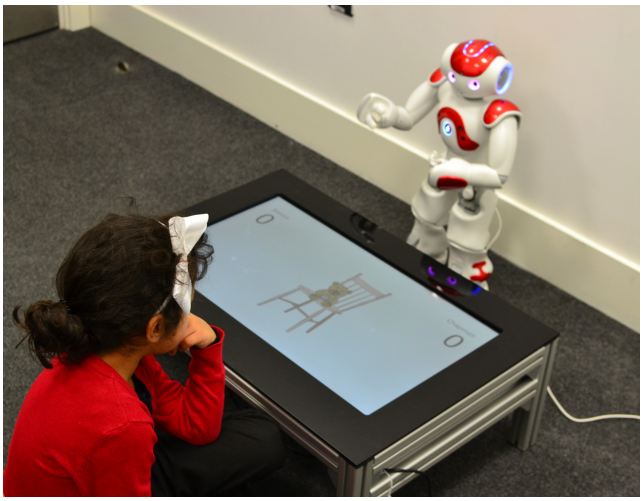

Fig. 1: A child interacting with the robot in our study. The agent – in this case a robot – stands opposite from the child. An interactive table displays an image of a teddy bear and a chair. The child must use a word from a second language to describe the position of the bear in relation to the chair.

[1]With the Centre for Robotics and Neural Systems, University of Plymouth Drake Circus, Plymouth, Devon, PL4 8AA, UK <firstname.lastname>@plymouth.ac.uk

[2]With the Department of Special Education: Cognitive and Motor Disabilities, Utrecht University, Heidelberglaan 1, 3584 CS Utrecht, Netherlands m.a.j.vandenberghe@uu.nl

[3]With the Department of Psychology, Koç University, Rumelifeneri Yolu, Sarıyer 34450, Istanbul, Turkey jkanero@ku.edu.tr

[4]With the IDLab Imec, Ghent University, iGent Toren, Technologiepark-Zwijnaarde 15 B-9052 Gent, Belgium tony.belpaeme@ugent.be

young children, parental or teacher reports [9].

In many situations, it may not be possible to use one of these tests. For example, when the words learned concern abstract concepts, which cannot be easily depicted, it is not possible to use a picture test. If the learner is illiterate, one cannot use a writing test. Parents or teachers may struggle to report the childs L2 if they do not speak that language themselves. To further complicate the issue, producing L2 words may be intimidating for L2 learners. Even if the learner is able to produce the word, they may not produce it due to anxiety of pronouncing the word incorrectly [10].

A social robot may help overcome some of the issues described above in assessing L2 learners vocabulary. While not being able to solve by itself the issue of vocabulary being more difficult to learn productively than receptively, a social robot may help in innovating novel ways to assess L2 vocabulary, or in reducing L2 anxiety in L2 vocabulary test settings. A robot may be less intimidating than an adult assessor, especially for young children, encouraging more speech production. This study evaluates whether school children may produce more L2 words in a productive L2 vocabulary test when playing with a social robot than with an adult. Below, we discuss relevant robot-assisted language learning (RALL) studies before detailing our study.

II. Previous Work

RALL has been found to be effective in reducing foreign language anxiety (FLA), and teaching robots are able to improve oral skills of young students learning English as a foreign language [11]. Alemi et al. [12] performed a study using a robot teaching assistant. In the study, Persian-speaking students in Iran were taught English. A survey of the students showed that those who learned from the robot were significantly less anxious compared to the control group that did not have the robot. While a number of factors were thought to contribute to this reduction in anxiety, the authors claimed a major reason to be intentional mistakes the robot made. The mistakes not only gave the students a chance to correct the robot, but also made them less afraid of making errors of their own.

When looking at speaking skills, the focus can not just be on vocabulary gains, but pronunciation as well. Lee et al. [13] conducted a series of lessons to help Korean children from grades 3 to 5 (roughly 8 to 10 years old) learn English. In South Korea children start learning English from grade 3. As part of a lesson series they were given a pronunciation training with a robot, that used a lexicon that included often confused phonemes, so that the robot could correct the child's pronunciation. It was reported that the children's speaking skills improved significantly with a large effect size when measured by a teacher. As well as the improvement in speaking skills all three affective factors – interest, confidence and motivation – all improved significantly.

Instances of robots acting as care-receivers also occur in RALL. In a study by Tanaka and Matsuzoe [14], Japanese children were given the role of teaching English verbs to a NAO robot. The children had to guide the robot's arm to act out the target verbs, e.g. brushing teeth. In a comprehension post-test the children answered correctly more often with words they had taught the robot than those learnt during a regular verb-learning game. While the robot only learned from 'Direct' teaching, where the child was guiding the motion of the robot, there was a high frequency of verbal teaching using English.

We can see that there are many instances where RALL is able to assist in teaching an L2 to students. Many of these show a reduction in FLA and increase in confidence and willingness to learn in the students. In all these cases, however, they use the robot to teach, whether directly in the role or acting as a care receiver or assistant. Robots were not used in assessment, and in most cases the tests performed were aimed at measuring the comprehension of the L2 words that were being taught. We want to explore the possibility of using a robot to assess the L2 production of children. Due to the reported reductions in anxiety and increase in confidence when using a robot, we may see an increase in the amount of production.

III. Study Design

This study was conducted at a local school with English-speaking 5- to 6-year-old children. We decided to teach spatial language, more specifically spatial prepositions, because while those concepts are more abstract than physical objects, we can still represent them using images. Spatial language itself is also particularly challenging to L2 learners as the meaning can often differ depending on context and the referent. Every morning, five children were randomly selected to participate in the study for that day and assigned a condition, balanced across gender. These five children were first given a French lesson before playing our production quiz game on an interactive table [15] individually throughout the rest of the day (Figure 1). An agent (robot or experimenter depending on our condition) is placed opposite to the child and gives instructions and encouragement to the children. The interactive table displays an image of a teddy bear and a chair. The child would have to use one of the French words taught to describe the position of the bear relative to the chair.

As well as the teacher three experimenters were involved in the study:

1) *Lead Experimenter* - The lead experimenter acted as the interaction point for the children outside of the one to one sessions. Either the lead experimenter or the wizard was required to be in the presence of the child while outside their classroom. The lead experimenter was certified in the children's health and well being, and was there to ensure the health and safety of the children as required by the school.
2) *Wizard Experimenter* - The wizard experimenter controlled the robot remotely via a laptop interface. The wizard experimenter was also certified in the children's health and well being, but had minimal interaction with the children so as to minimise interference during the study.

3) *Blind Experimenter* - The blind experimenter facilitated the interactions before the main study began, provided the comprehension test and acted as the agent in the child-human condition. The blind experimenter was unaware of the purpose of the study to reduce influencing the outcome.

A. Hypotheses

With our study we wanted to test the following hypothesis:

H1 The presence of a robot will allow children to produce more spatial words verbally in an L2 than when working with a human experimenter.

We expected to see with our data similar findings to those of Laufer et al. [4]:

H2 Comprehension (passive recognition) is easier than production (active recall), and a hierarchy between the two can be shown.

B. Teaching

The children were taught five French words: *Nounours* (Teddy Bear), *chaise* (chair), *devant* (in front of), *sur* (on), *sous* (under). Of these, the first two were supporting words and the last three were the target words for the study. The content of the lesson was created and taught by a professional French teacher, with a goal of enabling the children to produce these words after one lesson. We decided to use a professional teacher as we did not want a robot teacher that would also influence our results. It has also been shown that human teachers can still outperform a robot teacher. [16]. The lead experimenter acted as a teacher's assistant. The children were taught in groups of five. The lesson was designed to last 30 minutes.

The teacher started the lesson by introducing the children to the support words. At all stages the children were encouraged to repeat any French words they heard. The children were taught a song that used the three target words and hand gestures to go along with them. After singing, the children would position themselves relative to the chair based on the words announced by the teacher. The children were then each given a teddy bear and repeated the process with the bear. The children then played a game of 'Telephone'. In this game one child was first given one of the target words, and each child would whisper the word to the next child down the line until the last child. The last child would announce to the rest of the group the word they heard. The game was repeated several times with the children re-organised into a different order so that the announcing child changed each time. This was followed by a game of 'Corners'. In each corner of the lesson area, a teddy was placed in a position relative to a chair that referred to one of the target words. The children were then encouraged to sing and move around until the teacher would stop them, and say one of the target words. The children then had to move to the relevant corner and say the word three times. Variants of this game were then played in teams with the chairs lined up, and then individually. Finally each child was told to say one of the target words and then go stand by the correct chair. The lesson wrapped up with one more repetition of the song they had been taught near the beginning.

Fig. 2: A child being administered the comprehension test before moving onto the main production quiz.

During the interaction we also established any prior Knowledge in the target language. They were split into the following categories:

1) *No Exposure* - The children have not been exposed to any French, other than potentially those used in popular culture e.g. *C'est la vie*.
2) *Beginner* - The child has potentially received some lessons in French and knows simple phrases that do not include our target words e.g. *Je m'appelle John*.
3) *Intermediate* - The child has knowledge of French, including our target words.
4) *Advanced* - The child has an intricate knowledge of French, and is able to produce words with a high capability or are fluent.

Children of intermediate or advanced knowledge were excluded from the data analysis. 25 children took part in our study of which three were excluded from the analysis of results, leaving 22 children.

C. Individual Interactions

Upon completing another familiarity task and a 10 minute activity with the robot–that required the child to describe the position of objects to the robot in English–a comprehension test was administered by a blind experimenter who was unaware of the purpose of the study (Figure 2). This served as a small refresher of what the children had learned earlier in the day, as well as allows us to establish a baseline for the efficacy of the lesson. For the comprehension test there were 6 sheets with 3 images each (representing the 3 target words), placed on the left, in the centre or on the right. Together, the 6 sheets covered all possible permutations of the 3 target words (*devant*, *sur*, *sous*) with each of the 3 positions. The images were similar but not the same as the ones used for the production quiz questions. For each sheet the experimenter asked the child to point at the picture that

Fig. 3: The 'wizard' experimenter was positioned behind the child to minimise interaction between them.

matches the statement (see below). If the child pointed to the wrong picture they were allowed to try again until they pointed to the correct image. We repeated each target word twice to account for guessing and to ensure they weren't just picking based on location on the question sheet. The statements and their order were the same for every child:

1) Le nounours est sous la chaise.
2) Le nounours est devant la chaise.
3) Le nounours est sur la chaise.
4) Le nounours est devant la chaise.
5) Le nounours est sur la chaise.
6) Le nounours est sous la chaise.

The child then played the production quiz with either the robot or the blind experimenter based on the group they were in (child-robot or child-human). In both conditions, the production quiz was displayed on the sandtray. The robot was controlled through a Wizard-of-Oz interface, with the 'wizard' sat behind the child, out of sight, so as to minimise effects on the child (Figure 3). The rules of the game were explained by the agent (blind experimenter or robot). The child was sat in front of the sandtray upon which the production quiz game was displayed. The agent sat opposite the child. The sandtray displayed an image of the teddy bear in a position relative to the chair, and the agent or child must answer "Où est le nounours?" (Where is the teddy bear?). The agent was to give the answer in the form "sur/sous/devant la chaise", but any answer given by the child that included one of the target words 'sur', 'sous' or 'devant' was accepted. Each correct answer scored a point. If either the question was answered correctly or both the child and the agent answered incorrectly then the production quiz moved onto the next question. If the child did not answer after a short period then the agent would give encouragement in proceeding levels:

1) Encourage the child to guess e.g. "Just have a guess".
2) Targeted encouragement, such as asking them to remember the lesson from the morning.

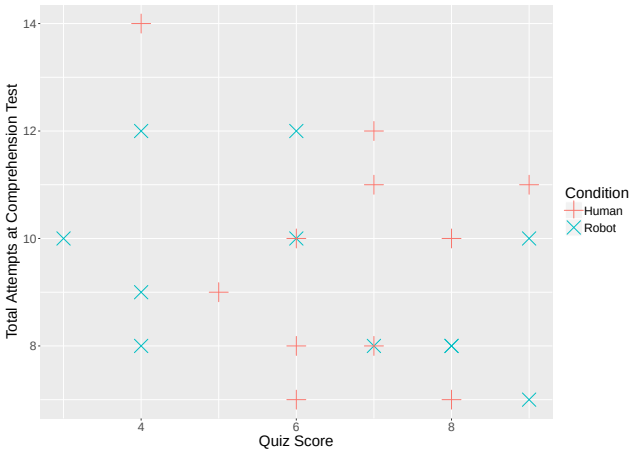

Fig. 4: A comparison between the score in the production quiz and the number of attempts required to complete the comprehension test. No significant correlation was found.

3) The agent will attempt the question.
   - If the child was ahead on points then the agent (adult/robot) would answer correctly so as to keep up an appearance of a challenging opponent in the game.
   - If the child was level or behind the agent (adult/robot) then the agent would answer incorrectly to demonstrate a willingness to answer even if wrong.

If the child still did not have a guess after all stages then the game proceeded as if they had answered incorrectly. The agent began the production quiz after explaining how to play by answering the first question correctly. There were nine subsequent questions which we expected the child to answer, three for each target word.

IV. Results

A. Participants

25 children took part in our study of which three were excluded from our analysis of results leaving us with 22 children. 11 Children were in the Human Condition (4 Female) and 11 in the Robot Condition (6 Female). There were 11 5 year olds (6 Female) and 11 6 year olds (4 Female). Of these children two had an L1 other than English (1 Female), but their English level was high enough to still participate.

B. Comprehension

We measured the answers on the comprehension test as the number of attempts to find the correct answer. The mean total number of attempts for the comprehension test was 9.5 (SD=1.92), with 6 being the highest possible score and 18 being the lowest. In the Human condition the children averaged 9.73 (SD=2.20) attempts at the comprehension test while in the Robot condition the children averaged 9.27 (SD=1.68). Using a Welch Two Sample t-test, no significant difference between the two conditions was found (t= -0.55, df =18.72 p=0.59). This shows that the groups between our two

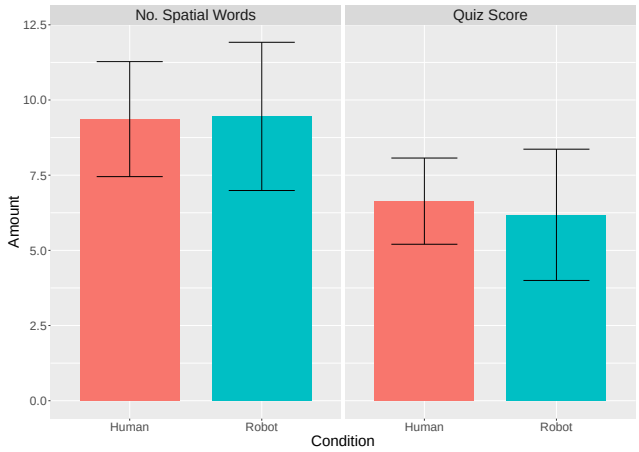

Fig. 5: Analysis of L2 spatial words used during the production quiz. Left: spatial words used without additional prompting to attempt the question; right: number of correct words said by the children during the production quiz. In both cases no significant difference was found between the robot and adult conditions. Error bars are showing the standard deviation.

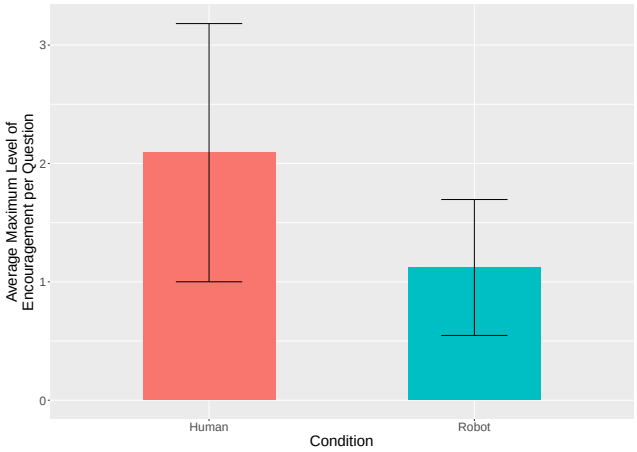

Fig. 6: Analysis of the average maximum level of encouragement reached across conditions. A significant difference is seen between the two conditions, Human and Robot. Error bars are showing the standard deviation.

conditions were roughly equal in ability before beginning the production quiz. The number of attempts was similar throughout the task, and no learning effect was seen when the first half and the second half of the comprehension test were compared (first half: mean=4.5, SD=1.26; second half: mean=5 SD=0.93; t=-1.50, df = 38.51, p=0.14).

C. Production

Children in the child-human condition scored M=6.64 (SD=1.43) out of 9 on the production quiz and M=6.18 (SD=2.18) in the child-robot condition. Using a Welch Two Sample t-test no significant difference between the two conditions was found (t=-0.58, df =17.27, p=0.57).

We also analysed the total number of spatial vocabulary used in L2 (Figure 5). Due to a break in protocol, children were sometimes prompted to attempt a question again instead of moving on in the production quiz. As such our analysis is on words used without being prompted for an additional attempt. In the Robot condition, the children averaged M=9.45 (SD=2.46) spatial words, compared to M=9.36 (SD=1.91) in the Human condition. Using a Welch Two Sample t-test no significant difference was found (t=0.10, df=18.4, p=0.92).

Finally we analysed the amount and level of encouragement given (see levels in Section III-C). While encoding encouragement given to the children we added a fourth level for analysis of the results:

4) Encouragement is given that changes or disrupts the task, e.g. telling the child that the current question is the same as a previous one.

The mean amount of encouragement given was M=12.36 (SD=7.46) in the Human condition and M=13.09 (SD=7.78) in the Robot condition. No significant difference was found between the conditions (p=0.83). However we see a significant difference in the average maximum level of encouragement per question across the two conditions (Robot: M=1.12, SD=0.57. Adult: M=2.09, SD=1.09, p=0.02). This is strongly influenced by the amount of level 4 encouragement given by the adult, of which we see 33 instances across 10 children. We see a significant difference between the average amount of level 4 encouragement given per child between the amount given in the first half of the study compared to the second showing an increase in deviation from the protocol over time (First Half: M=1.25, SD=.0.88. Second Half: M=4.25, SD=2.64, p=0.04).

D. Comprehension and Production

Across both conditions the children had an average score on the production quiz of 6.41 (SD=1.82) out of 9 and is significantly above chance (p=0.03). A negative but non-significant correlation was found between attempts at the comprehension test and their production quiz score (Pearson's r=-0.29, p=0.19). The lack of correlation suggests that abilities in comprehension and production are not directly related.

We also looked at the hierarchy between comprehension and production. We marked a child as having achieved comprehension on a particular word if they required less than four attempts across the two relevant questions in the comprehension test. For example if we were looking at whether a child could comprehend the word 'sur' we would look at the number of attempts they took for questions three and five. If a child takes two attempts on question three and one attempt on question five their total number of attempts for 'sur' would be three. We would mark this child as being able to comprehend 'sur'. We marked a child as being able to produce a word if they scored at least two points in the production quiz on the three relevant questions. Using Guttman's Coefficient of Reproducibility (reported in Table I), we were unable to find a hierarchy. A hierarchy would show that comprehension is needed for production. Guttman's Coefficient measures whether such a hierarchy

exists based on the number of deviations from that hierarchy. A coefficient of over 0.9 is expected to display such a hierarchy.

	Sur	Sous	Devant
No. Deviations	5	3	4
Guttman's Coefficient λ_4	0.11	0.57	0.56

TABLE I: Table detailing the number of deviations from the expected hierarchy and the Guttman's Coefficient of reproducibility. In the case of all three words, we fail to meet the reliability expectation of 0.9

V. Discussion

A. Effectiveness of the robot to support L2 production

The scores from the production quiz are higher than we expected. From the literature we expected L2 production to be difficult for the children, and our expert tutor believed that it would take two to three sessions for most children to produce at all. The observed prowess of the children may be partially explained by the design of the lessons, directly aimed at encouraging the children to produce the target words for this study. It should be noted that most productions were only single words. Only two children produced any of the support words (*nounours* – teddy bear, and *chaise* – chair).

While this study does not show statistical improvement to a child's ability to produce by using a robot over a person, it does show equal performance in this task. It may still be desirable to use a robot to allow standardization and automation of assessment. With a minimal amount of support being provided by an agent, only a narrow set of phrases can be given – otherwise the nature of the task could be changed from production. This can make interactions very repetitive for the assessor. Though the scores were higher than expected it still proved to be a challenging task for the children. With the minimal amount of support available to an experimenter it could be emotionally stressful to be unable to intervene when a child is finding the task difficult.

Several factors may contribute to the high performance of the experimenter. Even within the context of a limited set of responses a person is able to provide much better cues and encouragement based on reading the child. These kind of social skills are still a gold standard to which robotics researchers strive. Though this experiment was conducted using a 'wizard', their position and the time delay in actions for the robot prevented this fine grained social interaction. Some of the cues provided by the experimenter were not programmed into the robot but should be added into its repertoire

1) *Direct phonetic cues* - Giving part of the word e.g. the starting s.
2) *Indirect phonetic cues* - Giving clues to the word about how it sounds e.g. "It's the one with a strange sound in it"
3) *Rhythmic cues* - Giving the syllables of the word e.g. "Duh-dum". This may work well for the small target vocabulary, like ours, where this could refer to a single word, but may be less effective in larger vocabularies.
4) *Gestural cues* - Movements with the hands that mimic gestures used by the teacher in the lesson.

Despite the more limited social skills of the robot, it was still able to match the performance of a person. This may be the expected reduction of anxiety balancing the limited social behaviours.

However we also saw a large amount of encouragement given to the children by the blind experimenter that was outside of the original protocol, that could be deemed to have affected the scores of the children in an undesirable way. While in the first half the amount of these encouragements by the experimenter remained low, there was a sharp increase in the latter half. This could be caused by forgetting the protocol over the days of the study or just growing more lax in its use, or even the emotional stress that is put on a person by the children's difficulties.

The presence of a wizard in the room may also have been a contributing factor. The presence of a person, even when not in view, may have prevented the robot from reducing anxiety as much as it could have done, as the child might be aware someone else is listening in. However the majority of children did appear to forget that he was there, and focused on the robot.

Finally, it must be noted that the school where we performed the study cultivated a much friendlier relationship between adults in the school and the students than is typically seen. This may have made the children feel more comfortable and confident in the presence of our experimenter, reducing anxiety. Future work will focus on broadening this study to multiple schools to see whether our results can be replicated in different settings.

B. Relative difficulty of comprehension versus production

The lack of correlation shown between the production quiz score and the number of attempts on the comprehension test (Figure 4) shows that there was no direct relation between comprehension and production vocabularies. However when we look at the possibility of a hierarchy from comprehension to production we do not find evidence to support a hierarchy. This could have had several causes. While we were hoping to find support within our data, we were not directly testing for this hierarchy. Laufer et al. [4] looked at students 16 years and older at high school and university who had been studying their L2 as part of a national curriculum between 6 and 9 years. Ours is based on a single lesson focused entirely on being able to say the target words. The younger children in our study may also have been more receptive to learning words productively, as they are still increasing their phonological vocabulary. These skills have been shown to have a correlation with word vocabulary [17]. These factors could account for an increase in deviations from the previously established hierarchy.

VI. Conclusion

We hypothesized that a robot could surpass human performance in encouraging the production of spatial language: this hypothesis is not supported by our study; however, the robot nevertheless matches the performance of a human facilitator. This was despite the greater social ability of the human experimenter, and is suggestive that the robot does make the children less anxious. Future work expanding the robot's social ability may improve the robot's ability to assess and support a student's learning. Measuring the production skills of a child at this level is a repetitive and lengthy task. An autonomous robot that is able to measure the production level of a child may alleviate these factors, enabling more accurate data collection for both research and assessment purposes.

Currently we are planning on expanding this work to more schools while increasing the social skills of the robot.

VII. Acknowledgements

This work was supported by the EU H2020 L2TOR project (grant 688014). The authors would also like to thank the teacher, who wished to remain anonymous, who provided the French lessons for the children. All statistics and graphs were obtained using R [18].

References

[1] B. Laufer and T. S. Paribakht, "The relationship between passive and active vocabularies: Effects of language learning context," *Language Learning*, vol. 48, no. 3, pp. 365–391, 1998.

[2] B. Laufer, "The development of passive and active vocabulary in a second language: Same or different?" *Applied Linguistics*, vol. 19, no. 2, pp. 255–271, 1998.

[3] J.-A. Mondria and B. Wiersma, "Receptive, productive, and receptive + productive l2 vocabulary learning: What difference does it make?" in *Vocabulary in a Second Language: Selection, Acquisition and Testing*, P. Bogaards and B. Laufer, Eds. John Benjamins Publishers, 2004, pp. 79–100.

[4] B. Laufer and Z. Goldstein, "Testing vocabulary knowledge: Size, strength, and computer adaptiveness," *Language Learning*, vol. 54, no. 3, pp. 399–436, 2004.

[5] K. Williams, *Expressive Vocabulary Test*. Minnesota: American Guidance Service, 1997.

[6] M. F. Gardner, *Expressive One-Word Picture Vocabulary Test - Revised*. Novato, CA: Academic Therapy, 1990.

[7] E. H. Wiig, W. Secord, and E. Semel, *CELF-Preschool: Clinical Evaluation of Language Fundamentals - Preschool*. New York: Psychological Corp, 1992.

[8] B. Laufer and P. Nation, "A vocabulary-size test of controlled productive ability," *Language Testing*, vol. 16, no. 1, pp. 33–51, 1999.

[9] L. Fenson, P. S. Dale, S. Reznick, D. J. Thal, E. Bates, J. Hartung, S. J. Pethick, and J. Reilly, *The MacArthur Communicative Development Inventories: Users guide and technical manual*. San Diego, CA: Singular Publishing, 1993.

[10] D. Maillat, "The pragmatics of l2 in clil. language use and language learning in clil classrooms," *Language Use and Language Learning in CLIL Classrooms*, pp. 39–58, 2010.

[11] M. Alemi, "General impacts of integrating advanced and modern technologies on teaching english as a foreign language," *International Journal on Integrating Technology in Education*, vol. 5, no. 1, pp. 13–26, 2016.

[12] M. Alemi, A. Meghdari, and M. Ghazisaedy, "The impact of social robotics on l2 learners' anxiety and attitude in English vocabulary acquisition," *International Journal of Social Robotics*, pp. 1–13, 2015.

[13] S. Lee, H. Noh, J. Lee, K. Lee, G. G. Lee, S. Sagong, and M. Kim, "On the effectiveness of robot-assisted language learning," *ReCALL*, vol. 23, no. 01, pp. 25–58, 2011.

[14] F. Tanaka and S. Matsuzoe, "Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning," *Journal of Human-Robot Interaction*, vol. 1, no. 1, 2012.

[15] P. Baxter, R. Wood, and T. Belpaeme, "A touchscreen-based'sandtray'to facilitate, mediate and contextualise human-robot social interaction," in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. ACM, 2012, pp. 105–106.

[16] J. Kennedy, P. Baxter, E. Senft, and T. Belpaeme, "Heart vs hard drive: children learn more from a human tutor than a social robot," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*. IEEE Press, 2016, pp. 451–452.

[17] S. E. Gathercole and A. D. Baddeley, "Evaluation of the role of phonological stm in the development of vocabulary in children: A longitudinal study," *Journal of memory and language*, vol. 28, no. 2, pp. 200–213, 1989.

[18] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017. [Online]. Available: https://www.R-project.org/

Spatial Referring Expressions in Child-Robot Interaction: Let's Be Ambiguous!

Christopher D. Wallbridge[1], Séverin Lemaignan[1], Emmanuel Senft[1], Charlotte Edmunds[1], Tony Belpaeme[1,2]

1 University of Plymouth
2 University of Ghent

*** christopher.wallbridge@plymouth.ac.uk**

Abstract

Establishing common ground when attempting to disambiguate spatial locations is difficult at the best of times, but is even more challenging between children and robots. Here, we present a study that examined how 94 children (aged 5-8) communicate spatial locations to other children, adults and robots in face-to-face interactions. While standard HRI implementations focus on non-ambiguous statements, we found this only comprised about 20% of children's task based utterances. Rather, they rely on brief, iterative, repair statements to communicate about spatial locations. Our observations offer strong experimental evidence to inform future dialogue systems for robots interacting with children.

1 Introduction

For children arriving in a new country, learning the language of their new home is an important part of their integration. Proficiency in the language of the host country is a vital condition for success at school. Even for children of migrants born in the host country, this may be an issue if the language used at school cannot be reinforced in the home. As tailored language classes are expensive and limited in time, we wish to explore if robot tutors can be used to complement language tutoring. This is encouraged by robots having been shown to be able to reduce anxiety in a second language learning when acting as a peer [1]. However there is still much to be considered when designing a robotic language tutor [5].

Figure 1. A child interacting with the robot in our study.

While most language tutoring systems focus on the learning of nouns and verbs, we wish to study the learning of spatial language instead: the vocabulary and grammatical constructions serving the communication of spatial relations. Spatial language is particularly challenging, as the semantics are often vague, context dependant and referent dependant. For example, in "the apple next to the bowl" the spatial referent "next" does not have boolean membership, but rather has a graded membership depending on the distance between objects and the size of the objects. A typical assumption in Natural Language Interaction Systems (NLIS) is that referring expressions (RE) are unambiguous descriptions of object locations and that a linguistic interaction between a user and a computer system follows a quite structured and clear interaction flow using unambiguous utterances [8]. This might be the case for spoken interfaces in banking systems or telephone ordering, but the

literature in socio-linguistics and dialogue systems show that language is much more dynamic than NLIS typically allows for, and this is specifically prominent in spatial RE.

Socio-linguistics suggests that people do not tend to use fully specified RE. Instead, they reduce the cognitive load by under-specifying the description and then rely on a strategy of repair to correct misunderstanding if necessary [7]. Rather than this being a one-way communication, it is a fundamentally social process. The person being addressed is expected to be an active contributor to the process of reaching *common ground*. Each participant in the conversation will contribute until a *grounding criterion* is met [6], i.e. when each contributor to the communication believes that they have understood enough for their current purpose. Pickering and Garrod [11] describe this partial alignment of common ground as the natural way in which we communicate. Full common ground is only necessary when there is difficulty reaching alignment.

Dialogue management systems have to take into consideration these under specified statements. One assumption that often made in interaction between two agents is that what is said by one, is how the other understands it. However this is not always true, even in human-human interaction [10]. Instead, continuous communication can allow a system to re-evaluate its belief state of the current environment, and the belief state of other communicative agents. For spatial tasks they are able to use contextual language to help with the positioning of an item [2]. Instead of complex statements that try to pinpoint the exact location in one sentence, a series of much simpler statements is used.

By contrast, implementations of RE generation and understanding for use in robotics often follow Gricean Maxims [9], such as the Incremental Algorithm [8]. These algorithms focus on a single statement that eliminates ambiguity. While communicating clearly and unambiguously about spatial references is one solution to the problem of communicating about space, more recent systems also incorporate perspective taking [12], which may alleviate the need for precise but verbose REs. With perspective taking we do see a more interactive approach. But this process still relies on reaching full alignment by eliminating ambiguity.

Our present study provides real-world data of children establishing common ground in the natural course of playing a game. We observed them either interacting with other children, with adults or with a robot using a Wizard of Oz setup. The study provides opportunities for the children to use a large set of spatial language, perspective taking and establishing a common point of reference, whilst being easy to replicate.

2 Study Design

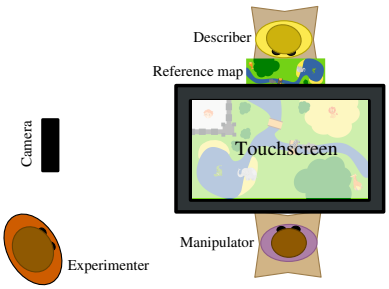

Figure 2.The experimental setup. A top down view showing the position of the manipulator and describer sitting opposite each other with the "Sandtray" screen in the middle. The experimenter is sitting to the side with a camera recording the participants.

We collected data from 94 children between the ages of 5 and 8. They were assigned to one of three conditions: child-child, child-adult or child-robot. For the child-child and child-adult conditions children from two different schools were used. They participated during the day at their school in a room for individual teaching. In the child-child condition two children from the same class participated together. In the child-adult condition a child participated with an experimenter. Those in the child-robot condition were recruited from register held by the Babylab at the University of Plymouth.

Following a sandbox paradigm [3], one child and a partner (child, robot or adult) are sitting on opposite sides of a large touchscreen (Fig. 2). The screen presents a background with different areas: a castle, a desert, two rivers with bridges, a lake, two beaches and many bushes or trees.

One agent, hereafter called the *describer*, has to guide the other agent, called the *manipulator*, to move items on the touchscreen to a desired location. The describer is provided with a reference map, which is kept hidden from the manipulator, with the desired position of eight items (Fig. 3).

While it has been shown that pointing can influence the words used [13], the task could be easily completed without words if gestures were allowed. As we were focused on the language being used, the describer was instructed not to use pointing gestures. If children attempted to use pointing they were reminded that this was not allowed.

Figure 3.An example of the reference map given to a child to describe. The eight items (face, crocodile, elephant, zebra, hippo, lion, giraffe and ball) are shown in the desired location that they need to be moved to. The child describes the position on his map for an agent to manipulate into the correct position.

The touchscreen presents a background with different areas (Fig. 3). Eight movable items have to be moved to specified locations on the map. The reference maps were designed to elicit a number of different ways to describe the position of objects. Some objects were facing a particular direction, to encourage locutions like 'in front of' or 'behind'. Features, such as the bridges and bushes, were repeated so as to require disambiguation. Verbal disambiguation was also elicited by the relatively small size of the screen, which limits the effectiveness of joint gaze to identify the correct location for an object.

In the case of the child-child and child-adult conditions, after the first map was completed, the role of manipulator and describer would be swapped. In the case of the child-robot condition the child would be invited to describe the second map. The robot itself would appear to move objects around the touchscreen via the use of a Wizard of Oz control interface, held by an experimenter. The experimenter is able to move an object on their interface, the robot would then move its hand to point at the object and then move its hand to point at the target location, with the object moving with it.

3 Results

For statistical power reasons, we focused our current observation of results on the child-child interaction (Child-Child=60, Child-Adult=26, Child-Robot=8), while providing more qualitative observations of the other conditions in the discussion.

We observed an average of 7.12 (SD=7.50) repair statements used per round (one round consisted of one map with eight objects to be moved). The SD shows large inter-personal variations. There were comparatively few cases of repair statements requiring spatial perspective taking (M=0.56 per round). Despite being told not to use them, there was an average of 2.43 (SD=3.03) pointing gestures used per round.

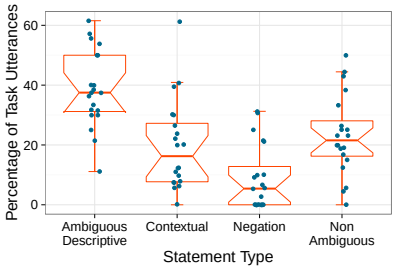

Figure 4. Break down of on-task statements. Ambiguous descriptive statements were a significantly higher proportion than the other statement types.

We took all the on-task statements from a sample of 10 child-child sessions, giving us data from 20 children. The statements were divided into the following categories: Ambiguous-Descriptive (statement refers to more than one location e.g.'the zebra is on a bridge'), Contextual (statement following from previous statements, that would make no sense to a third person entering the conversation e.g. 'the other one'), Negation(statement indicating that it is an incorrect location with no further description e.g. 'no'), Non-Ambiguous (statement that describes only one possible location e.g. 'the crocodile is in the big lake') and Pointing.

On average Ambiguous-Descriptive statements were used 38.6% of the time, Contextual in 13.1%, Negation in 9% and Non-Ambiguous in 23.2%. Using a Welch

two-sample t-test we find that the Ambiguous-Descriptive statements are used significantly more than any other type of statements, and Cohen's d test shows a large effect size in each case (Contextual: $t(38) = 4.2$, $p < .001$, $d = 1.34$; Negation: $t(38) = 7.8$, $p < .001$, $d = 2.48$; Non-Ambiguous: $t(38) = 3.7$, $p < .001$, $d = 1.17$).

4 Discussion

Our observations show that interactions between children (and between children and robots) are highly dynamic, fast-paced and relying on the situatedness and embodiment of the conversation partners [4], very unlike the "walkie-talkie exchanges" typically used in Human-Robot Interaction. Between children, as soon as the manipulator has enough information to make a guess they will often start moving the objects, without waiting until enough information is given as to be non-ambiguous. This has two possible outcomes: either they guess right, or it causes the describer to generate a repair statement. It also appears that typically it is easier for the describer to let the manipulator start moving the objects – knowing that the position they described is ambiguous – so that they may then generate a short, easily understood, repair, reducing the cognitive load. In fact we see that the robot's inability to change course after it has started moving an object caused frustration to the child describing.

In the child-robot condition there appeared to be a reduction of the repair statements when the robot moved items incorrectly. This could be caused by many factors, such as the children feeling more nervous with the robot, the expectations they have of its abilities and the absence of some basic social cues, such as back channelling and lack of eye contact, all of which made the interaction laborious.

Pointing was still prevalent, despite it being disallowed and discouraged (even the experimenter was found pointing or indicating directions). Future work could look at a different methodology to encourage the combination of gestures and language.

5 Conclusion

Counter to many implementations that seek to eliminate ambiguity entirely, we find that children tend to use many ambiguous statements when describing the location of objects. As such *the robot, when being given RE, must expect ambiguous statements.* It should not wait for further information, but rather start acting on the information it has, as this will also assist in the process of description. This also means that the robot should be prepared to react quickly to repair statements by enabling it to diverge from its current action to take into account the new information.

This also means *the robot should be allowed to be ambiguous in its descriptions.* This may be beneficial to reduce processing requirements for the robot itself, but also may help reduce the cognitive load for its conversational partner. When doing so, the robot should monitor closely the reaction of its partner, and be prepared to provide timely repairs to lead the implicit, interactive disambiguation process.

Our next steps are to implement a more interactive robot to collect more data with children interacting with the robot. Using this data we will be able to build an effective framework for natural spatial communication between children and robots.

6 Acknowledgements

This work was supported by the EU H2020 L2TOR project (grant 688014), the EU H2020 Marie Sklodowska-Curie Actions project DoRoThy (grant 657227) and the EU FP7 DREAM project (grant 611391).

References

1. M. Alemi, A. Meghdari, and M. Ghazisaedy. The impact of social robotics on L2 learners' anxiety and attitude in English vocabulary acquisition. *International Journal of Social Robotics*, pages 1–13, 2015.

2. T. Baumann, M. Paetzel, P. Schlesinger, and W. Menzel. Using Affordances to Shape the Interaction in a Hybrid Spoken Dialog System. In *Proceedings of ESSV*, Bielefeld, Germany, Mar. 2013.

3. P. Baxter, R. Wood, and T. Belpaeme. A touchscreen-based'sandtray'to facilitate, mediate and contextualise human-robot social interaction. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 105–106. ACM, 2012.

4. T. Belpaeme, S. J. Cowley, and K. F. MacDorman. *Symbol grounding*, volume 21. John Benjamins Publishing, 2009.

5. T. Belpaeme, P. Vogt, R. van den Berghe, K. Bergmann, T. Göksun, M. de Haas, J. Kanero, J. Kennedy, A. C. Küntay, O. Oudgenoeg-Paz, et al. Guidelines for designing social robots as second language tutors. *International Journal of Social Robotics*, 2017.

6. H. H. Clark and E. F. Schaefer. Contributing to discourse. *Cognitive science*, 13(2):259–294, 1989.

7. H. H. Clark and D. Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22(1):1–39, 1986.

8. R. Dale and E. Reiter. Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive science*, 19(2):233–263, 1995.

9. H. P. Grice, P. Cole, J. Morgan, et al. Logic and conversation. *1975*, pages 41–58, 1975.

10. G.-J. M. Kruijff, M. Janíček, and P. Lison. Continual processing of situated dialogue in human-robot collaborative activities. In *RO-MAN, 2010 IEEE*, pages 594–599. IEEE, 2010.

11. M. J. Pickering and S. Garrod. Alignment as the basis for successful communication. *Research on Language & Computation*, 4(2):203–228, 2006.

12. R. Ros, S. Lemaignan, E. A. Sisbot, R. Alami, J. Steinwender, K. Hamann, and F. Warneken. Which one? grounding the referent based on efficient human-robot interaction. In *RO-MAN, 2010 IEEE*, pages 570–575. IEEE, 2010.

13. A. Sauppé and B. Mutlu. Robot deictics: How gesture and context shape referential communication. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 342–349. ACM, 2014.