Construction of bilingual multimodal corpora of referring expressions in collaborative problem solving

TOKUNAGA Takenobu **IIDA Ryu** **YASUHARA Masaaki** **TERAI Asuka**
{take,ryu-i,yasuhara}@cl.cs.titech.ac.jp asuka@nm.hum.titech.ac.jp
Tokyo Institute of Technology

David MORRIS **Anja BELZ**
D.Morris@brighton.ac.uk a.s.belz@itri.brighton.ac.uk
University of Brighton

Abstract

This paper presents on-going work on constructing bilingual multimodal corpora of referring expressions in collaborative problem solving for English and Japanese. The corpora were collected from dialogues in which two participants collaboratively solved Tangram puzzles with a puzzle simulator. Extra-linguistic information such as operations on puzzle pieces, mouse cursor position and piece positions were recorded in synchronisation with utterances. The speech data was transcribed and time-aligned with the extra-linguistic information. Referring expressions in utterances that refer to puzzle pieces were annotated in terms of their spans, their referents and their other attributes. The Japanese corpus has already been completed, but the English counterpart is still undergoing annotation. We have conducted a preliminary comparative analysis of both corpora, mainly with respect to task completion time, task success rates and attributes of referring expressions. These corpora showed significant differences in task completion time and success rate.

1 Introduction

A referring expression (RE) is a linguistic device that refers to a certain object of interest (e.g. used in describing where the object is located in space). REs have attracted a great deal of attention in both language analysis and language generation research. In language analysis research, reference resolution, particularly anaphora resolution (Mitkov, 2002), has a long research history as far back as the mid-1970s (Hobbs, 1978). Much research has been conducted from both theoretical and empirical perspectives, mainly concerning the identification of antecedents or entities mentioned within the same text. This trend, targeting reference resolution in written text, is still dominant in the language analysis, perhaps because such techniques are intended for use in applications such as information extraction.

In contrast, in language generation research interest has recently shifted from the generation of one-off references to entities to generation of REs in discourse context (Belz et al., 2010) and investigating human referential behaviour in real world situations, with the aim of using such techniques in applications like human-robot interaction (Piwek, 2007; Foster et al., 2008; Bard et al., 2009).

In both analysis and generation, machine-learning approaches have come to replace rule-based approaches as the predominant research trend since the 1990s. This trend has made annotated corpora an indispensable component of research for training and evaluating proposed methods. In fact, research on reference resolution has developed significantly as a result of large scale corpora, e.g. those provided by the Message Understanding Conference (MUC)[1] and the Automatic Content Extraction (ACE)[2] project. These corpora were constructed primarily for information extraction research, thus were annotated with co-reference relations within texts. Also in the language generation community, several corpora

[1] http://www.nlpir.nist.gov/related_projects/muc/

[2] http://www.itl.nist.gov/iad/tests/ace/

Proceedings of the 8th Workshop on Asian Language Resources, pages 38–46,
Beijing, China, 21-22 August 2010. ©2010 Asian Federation for Natural Language Processing

have been developed (Di Eugenio et al., 2000; Byron, 2005; van Deemter et al., 2006; Foster and Oberlander, 2007; Foster et al., 2008; Stoia et al., 2008; Spanger et al., 2009a; Belz et al., 2010). Unlike the corpora of MUC and ACE, many are collected from situated dialogues, and therefore include multimodal information (e.g. gestures and eye-gaze) other than just transcribed text (Martin et al., 2007). Foster and Oberlander (2007) emphasised that any corpus for language generation should include all possible contextual information at the appropriate granularity. Since constructing a dialogue corpus generally requires experiments for data collection, this kind of corpus tends to be small-scale compared with corpora for reference resolution.

Against this background, we have been developing multimodal corpora of referring expressions in collaborative problem-solving settings. This paper presents on-going work of constructing bilingual (English and Japanese) comparable corpora in this domain. We achieve our goal by replicating, for the English corpus, the same process of data collection and annotation as we used for our existing Japanese corpus (Spanger et al., 2009a). Our aim is to create bilingual multimodal corpora collected from dialogues in dynamic situations. From the point of view of reference analysis, our corpora contribute to augmenting the resources of multimodal dialogue corpora annotated with reference relations which have been minor in number compared to other types of text corpora. From the point of view of reference generation, our corpora contribute to increasing the resources available that can be used to further research of this kind. In addition, our corpora contribute to comparative studies of human referential behaviour in different languages

The structure of the paper is as follows. Section 2 describes the experimental set-up for data collection which was introduced in our previous work (Spanger et al., 2009a). The setting is basically the same for the construction of the English corpus. Section 3 explains the annotation scheme adopted in our corpora, followed by a description of a preliminary analysis of the corpora in section 4. Section 5 briefly mentions related work to highlight the characteristics of our corpora. Finally, Section 6 concludes the paper and looks at possible future directions.

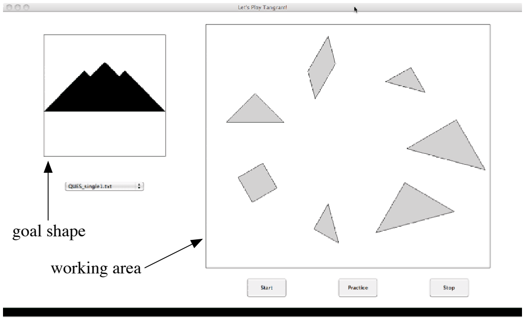

Figure 1: Screenshot of the Tangram simulator

2 Data collection

2.1 Experimental set-up

We recruited subjects in pairs of friends and colleagues. Each pair was instructed to solve Tangram puzzles collaboratively. Tangram puzzles are geometrical puzzles that originated in ancient China. The goal of a Tangram puzzle is to construct a given goal shape by arranging seven simple shapes, as shown in Figure 1. The pieces include two large triangles, a medium-sized triangle, two small triangles, a parallelogram and a square.

With the aim of recording the precise position of every piece and every action the participants made during the solving process, we implemented a Tangram simulator in which the pieces can be moved, rotated and flipped with simple mouse operations on a computer display. The simulator displays two areas: a goal shape area and a working area where the pieces can be manupulated and their movements are shown in real time.

We assigned a different role to each participant of a pair: one acted as the *solver* and the other as the *operator*. The operator has a mouse for manipulating Tangram pieces, but does not have a goal shape on the screen. The solver has a goal shape on the screen but does not have a mouse. This setting naturally leads to a situation where given a certain goal shape, the solver thinks of the necessary arrangement of the pieces and gives instructions to the operator how to move them, while the operator manipulates the pieces with the mouse according to the solver's instructions.

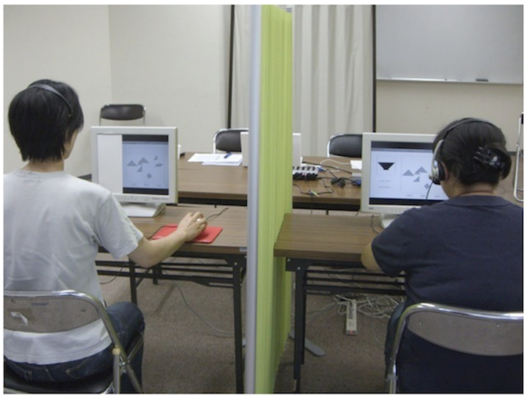

Figure 2: Picture of the experiment setting

As we mentioned in our previous study (Spanger et al., 2009a), this interaction produces frequent use of referring expressions intended to distinguish specific pieces of the puzzle. In our Tangram simulator, all pieces are of the same color, thus color is not useful in identifying a specific piece, i.e. only size and shape are discriminative object-intrinsic attributes. Instead, we can expect other attributes such as spatial relations and deictic reference to be used more often.

Each pair of participants sat side by side as shown in Figure 2. Each participant had his/her own computer display showing the shared working area. A room-divider screen was set between the solver (right side) and operator (left side) to prevent the operator from seeing the goal shape on the solver's screen, and to restrict their interaction to speech only.

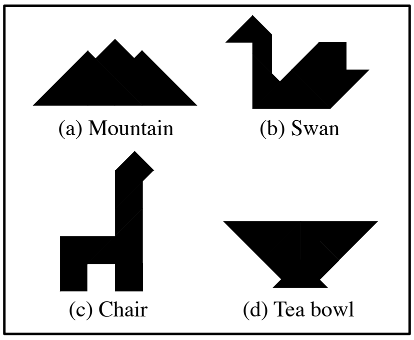

Figure 3: The goal shapes given to the subjects

Each participant pair was assigned 4 trials consisting of two symmetric and two asymmetric goal shapes as shown in Figure 3. In Cognitive Science, a wide variety of different kinds of puzzles have been employed extensively in the field of Insight Problem solving. This has been termed the "puzzle-problem approach" (Sternberg and Davidson, 1996; Suzuki et al., 2001) and in the case of physical puzzles has relatively often involved puzzle tasks of symmetric shapes like the so-called T-puzzle, e.g. (Kiyokawa and Nakazawa, 2006). In more recent work Tangram puzzles have been used as a means to study various new aspects of human problem solving approaches, including collection of of eye-gaze information (Baran et al., 2007). In order to collect data as broadly as possible in this context, we set up puzzle-problems including both symmetrical as well as asymmetrical ones as shown in Figure 3.

The participants exchanged their roles after two trials, i.e. a participant first solves a symmetric and then an asymmetric puzzle as the solver and then does the same as the operator, and vice versa. The order of the puzzle trials is the same for all pairs.

Before starting the first trial as the operator, each participant had a short training exercise in order to learn how to manipulate pieces with the mouse. The initial arrangement of the pieces was randomised every time. We set a time limit of 15 minutes for the completion of each trial (i.e. construction of the goal shape). In order to prevent the solver from getting into deep thought and keeping silent, the simulator is designed to give a hint every five minutes by showing a correct piece position in the goal shape area. After 10 minutes have passed, a second hint is provided, while the previous hint disappears. A trial ends when the goal shape is complete or the time is up. Utterances by the participants are recorded separately in stereo through headset microphones in synchronisation with the position of the pieces and the mouse operations. Piece positions and mouse actions were automatically recorded by the simulator at intervals of 10 msec.

Table 1: The ELAN Tiers of the corpus

Tier	meaning
OP-UT	utterances by the operator
SV-UT	utterances by the solver
OP-REX	referring expressions by the operator
OP-Ref	referents of OP-REX
OP-Attr	attributes of OP-REX
SV-REX	referring expressions by the solver
SV-Ref	referents of SV-REX
SV-Attr	attributes of SV-REX
Action	action on a piece
Target	the target piece of Action
Mouse	the piece on which the mouse is hovering

∗ Indentation of Tier denotes parent-child relations.

Table 2: Attributes of referring expressions

dpr	: demonstrative pronoun, e.g. “the same one”, “this”, “that”, “it”
dad	: demonstrative adjective, e.g. “that triangle”
siz	: size, e.g. “the large triangle”
typ	: type, e.g. “the square”
dir	: direction of a piece, e.g. “the triangle facing the left”.
prj	: projective spatial relation (including directional prepositions or nouns such as “right”, “left”, “above”. . .) e.g. “the triangle to the left of the square”
tpl	: topological spatial relation (including non-directional prepositions or nouns such as “near”, “middle”. . .), e.g. “the triangle near the square”
ovl	: overlap, e.g. “the small triangle under the large one”
act	: action on pieces, e.g “the triangle that you are holding now”, “the triangle that you just rotated”
cmp	: complement, e.g. “the other one”
sim	: similarity, e.g. “the same one”
num	: number, e.g. “the two triangle”
rpr	: repair, e.g. “the big, no, small triangle”
err	: obvious erroneous expression, e.g. “the square” referring to a triangle
nest	: nested expression; when a referring expression includes another referring expression, only the outermost expression is annotated with this attribute, e.g. “(the triangle to the left of (the small triangle))”
meta	: metaphorical expression, e.g. “the leg”, “the head”

2.2 Subjects and collected data

For our Japanese corpus, we recruited 12 Japanese graduate students of the Cognitive Science department, 4 females and 8 males, and split them into 6 pairs. All pairs knew each other previously and were of the same sex and approximately same age[3]. We collected 24 dialogues (4 trials by 6 pairs) of about 4 hours and 16 minutes. The average length of a dialogue was 10 minutes 40 seconds (SD = 3 minutes 18 seconds).

For the comparable English corpus, we recruited 12 native English speakers of various occupations, 6 males and 6 females. Their average age was 30. There were 6 pairs all of whom knew each other beforehand except for one pair. Whereas during the creation of the Japanese corpus we had to give extra attention to ensuring that social relationships did not have an impact on how the subjects communicated with one another, for the English corpus there was no such concern. We collected 24 dialogues (4 trials by 6 pairs) of 5 hours and 7 minutes total length. The average length of a dialogue was 12 minutes 47 seconds (SD = 3 minutes 34 seconds).

3 Annotation

The recorded speech data was transcribed and the referring expressions were annotated with the Web-based multi-purpose annotation tool SLAT (Noguchi et al., 2008)[4]. Our target expressions in this corpus are referring expressions referring to a puzzle piece or a set of puzzle pieces. We do not deal with expressions referring to a location, a part of a piece or a constructed shape. These expressions are put aside for future work. The annotation of referring expressions is threefold: (1) identification of the span of expressions, (2) identification of their referents, and (3) assignment of a set of attributes to each referring expression.

Using the multimodal annotation tool ELAN,[5] the annotations of referring expressions were then merged with extra-linguistic data recorded by the Tangram simulator. The available extra-linguistic information from the simulator consists of (1) the action on a piece, (2) the coordinates of the mouse cursor and (3) the position of each piece in the

[3] In Japan, the relationship of senior to junior or socially higher to lower placed might affect the language use. We carefully recruited pairs to avoid the effects of this social relationship such as the possible use of overly polite and indirect language, reluctance to correct mistakes etc.

[4] We did not use SLAT for English corpus annotation. Instead, ELAN was directly used for annotating referring expressions.

[5] http://www.lat-mpi.eu/tools/elan/

Table 3: Summary of trials

ID	time	success	OP-REX	SV-REX	ID	time	success	OP-REX	SV-REX
E01	15:00				J01	8:40	o	10	48
E02	15:00				J02	11:49	o	7	55
E03	15:00				J03	11:36	o	5	26
E04	15:00				J04	7:31	o	2	21
E05	15:00				J05	15:00		23	78
E06	15:00				J06	11:12	o	5	60
E07	15:00				J07	12:11	o	3	59
E08	15:00				J08	11:20	o	4	61
E09	10:39	o			J09	14:59	o	36	84
E10	15:00				J10	6:20	o	3	47
E11	15:00				J11	5:21	o	2	14
E12	8:30	o			J12	13:40	o	37	77
E13	14:33	o	8	95	J13	15:00		8	56
E14	7:27	o	1	62	J14	4:48	o	1	29
E15	14:02	o	16	127	J15	9:30	o	20	39
E16	3:57	o	1	31	J16	5:07	o	3	17
E17	13:00	o			J17	13:37	o	10	46
E18	6:40	o			J18	8:57	o	4	51
E19	15:00				J19	8:02	o	0	37
E20	12:32	o			J20	11:23	o	1	59
E21	15:00				J21	10:12	o	7	71
E22	15:00				J22	10:24	o	9	64
E23	15:00				J23	15:00		0	69
E24	5:36	o			J24	14:22	o	0	76
Ave.	12:47		6.5	78.8	Ave.	10:40		8.3	51.8
SD	3:34		7.14	41.4	SD	3:18		10.4	20.1
Total	5:06:56	10	26	315	Total	4:16:01	21	200	1,244

working area. Actions and mouse cursor positions are recorded at intervals of 10 msec, and are abstracted into (1) a time span labeled with an action symbol ("move", "rotate" or "flip") and its target piece number (1–7), and (2) a time span labeled with a piece number which is under the mouse cursor during that span. The position of pieces is updated and recorded with a timestamp when the position of any piece changes. Information about piece positions is not merged into the ELAN files and is kept in separate files. As a result, we have 11 time-aligned ELAN Tiers as shown in Table 1.

Two annotators (two of the authors) first annotated four Japanese dialogues separately and based on a discussion of discrepancies, decided on the following criteria to identify a referring expression.

- The minimum span of a noun phrase including necessary information to identify a referent is annotated. The span might include repairs with their reparandum and disfluency (Nakatani and Hirschberg, 1993) if needed.
- Demonstrative adjectives are included in expressions.
- Erroneous expressions are annotated with a special attribute.
- An expression without a definite referent (i.e. a group of possible referents or none) is assigned a referent number sequence consisting of a prefix, followed by the sequence of possible referents as its referent, if any are present.
- All expressions appearing in muttering to oneself are excluded.

Table 2 shows a list of attributes of referring expressions used in annotating the corpus.

The rest of the 20 Japanese dialogues were annotated by two of the authors and discrepancies were resolved by discussion. Four English dialogues have been annotated so far by one of the authors.

4 Preliminary corpus analysis

We have already completed the Japanese corpus, which is named REX-J (2008-08), but only 4 out of 24 dialogues have been annotated for the English counterpart (REX-E (2010-03)). Table 3 shows a summary of the trials. The horizontal lines divide the trials by pairs, "o" in the "success" column denotes that the trial was successfully completed in the time limit (15 minutes), and the "OP-REX" and "SV-REX" columns show the number of referring expressions used by the operator and the solver respectively. The following subsections describe a preliminary comparison of the English and Japanese corpora.

Table 4: Task completion time

Lang.\Shape	(a)	(b)	(c)	(d)
English	832.0	741.2	890.3	605.8
	(105.4)	(246.5)	(23.7)	(287.2)
Japanese	774.7	535.0	571.7	633.8
	(167.3)	(168.5)	(242.2)	(215.2)

* Average (SD)

4.1 Task performance

We conducted a two-way ANOVA with the task completion time as the dependent variable, and the goal shape and the language as the independent variables. Only the main effect of the language was significant ($F(1, 40) = 5.82,\ p < 0.05$). Table 4 shows the average and the standard deviation of the completion time. Note that we set a time limit (15 minutes) for solving the puzzle. We considered the completion time as 15 minutes even when a puzzle was not actually solved in the time limit. We also conducted a two-way ANOVA using only the successful cases. Both main effects and their interaction were not significant.

We then conducted an ANOVA with the number of successfully solved puzzles by each pair as the dependent variable and the language as the independent variable. The main effect was significant ($F(1, 10) = 6.79,\ p < 0.05$). Table 5 shows the average number of success goals per pair and the success rate with their standard deviations in parentheses.

Finally, we conducted an ANOVA with the number of pairs who succeeded in solving a goal shape as the dependent variable and the goal shape as the independent variable. The main effect was not significant.

Table 5: The number of solved trials and success rates

Lang.	solved trials	success rate [%]
Japanese	3.50 (0.55)	87.5 (13.7)
English	1.67 (1.63)	41.7 (40.8)

* Average (SD)

In summary, we found a difference in the task performance between the languages in terms of the task completion time and the success rate, but no difference among the goal shapes. This difference could be explained by the diversity of the subjects rather than the difference of languages. The Japanese subject group consisted of university graduate students from the same department (Cognitive Science) and roughly of the same age (Average = 23.3, SD = 1.5). In contrast, the English subjects have diverse backgrounds (e.g. high school students, university faculty, writer, programmer, etc.) and age (Average = 30.8, SD = 11.7). In addition, a familiarity with this kind of geometric puzzle might have some effect. However, we collected a familiarity with the puzzle only from the English subjects, we could not conduct further analysis on this viewpoint. Anyhow, in this respect, the independent variable should have been named "subject group" instead of "language".

4.2 Referring expressions

It is important to note that since we have only completed the annotation of four dialogs, all by one pair of subjects, our analyses of referring expressions are tentative and pending further analysis.

We have 200 and 1,243 referring expressions by the operator and the solver respectively, 1,444 in total in the 24 Japanese dialogues. On the other hand we have 26 (operator) and 315 (solver) referring expressions in 4 English dialogues. The average number of referring expressions per dialogue in Table 3 suggests that English subjects use more referring expressions than Japanese subjects. Since we have only the data from a single pair, we cannot say whether this tendency applies to the other pairs. We cannot draw a decisive conclusion until we complete the annotation of the English corpus.

Table 6 shows the total frequencies of the attributes and their frequencies per dialogue. The table gives us an impression of significantly frequent use of demonstrative pronouns (dpr) by the as limited annotation.

Table 6: Comparison of attribute distribution

	English		Japanese	
	(4 dialogues)		(24 dialogues)	
attribute	frq	frq/dlg	frq	frq/dlg
dpr	226	56.5	678	28.3
dad	29	7.3	178	7.4
siz	68	17.0	288	12.0
typ	103	25.8	655	27.3
dir	0	0	7	0.3
prj	10	2.5	141	5.9
tpl	4	1	9	0.4
ovl	0	0	2	0.1
act	5	1.3	103	4.3
cmp	17	4.3	33	1.4
sim	0	0	7	0.3
num	22	5.5	35	1.5
rpr	0	0	1	0
err	0	0	1	0
nest	1	0.3	31	1.3
meta	1	0.3	6	0.3

English subjects. The Japanese subjects use more attributes of projective spatial relations (prj) and actions on the referent (act).[6] The English subjects use more complement attributes (cmp) as well as more number attributes (num).

5 Related work

Over the last decade, with a growing recognition that referring expressions frequently appear in collaborative task dialogues (Clark and Wilkes-Gibbs, 1986; Heeman and Hirst, 1995), a number of corpora have been constructed to study the nature of their use. This tendency also reflects the recognition that this area yields both challenging research topics as well as promising applications such as human-robot interaction (Foster et al., 2008; Kruijff et al., 2010).

The COCONUT corpus (Di Eugenio et al., 2000) was collected from keyboard-dialogs between two participants, who worked together on a simple 2-D design task, buying and arranging furniture for two rooms. The COCONUT corpus is limited in annotations which describe symbolic object information such as object intrinsic attributes and location in discrete co-ordinates. As an initial work of constructing a corpus for collaborative tasks, the COCONUT corpus can be characterised as having a rather simple domain as well as limited annotation.

The QUAKE corpus (Byron, 2005) and its successor, the SCARE corpus (Stoia et al., 2008) deal with a more complex domain, where two participants collaboratively play a treasure hunting game in a 3-D virtual world. Despite the complexity of the domain, the participants were only allowed limited actions, e.g. moving step forward, pushing a button etc.

As a part of the JAST project, the Joint Construction Task (JCT) corpus was created based on dialogues in which two participants constructed a puzzle (Foster et al., 2008). The setting of the experiment is quite similar to ours except that both participants have even roles. Since our main concern is referring expressions, we believe our asymmetric setting elicits more referring expressions than the symmetric setting of the JCT corpus.

In contrast to these previous corpora, our corpora record a wide range of information useful for analysis of human reference behaviour in situated dialogue. While the domain of our corpora is simple compared to the QUAKE and SCARE corpora, we allowed a comparatively large flexibility in the actions necessary for achieving the goal shape (i.e. flipping, turning and moving of puzzle pieces at different degrees), relative to the complexity of the domain. Providing this relatively larger freedom of actions to the participants together with the recording of detailed information allows for research into new aspects of referring expressions.

As for a multilingual aspect, all the above corpora are English. There have been several recent attempts at collecting multilingual corpora in situated domains. For instance, (Gargett et al., 2010) collected German and English corpora in the same setting. Their domain is similar to the QUAKE corpus. Van der Sluis et al. (2009) aim at a comparative study of referring expressions between English and Japanese. Their domain is still static at the moment. Our corpora aim at dealing with the dynamic nature of situated dialogues between very different languages, English and Japanese.

[6] We called such expressions as *action-mentioning expressions* (AME) in our previous work.

Table 7: The REX-J corpus family

name	puzzle	#pairs	#dialg.	#valid	status
T2008-08	Tangram	6	24	24	completed
T2009-03	Tangram	10	40	16	completed
T2009-11	Tangram	10	36	27	validating
N2009-11	Tangram	5	20	8	validating
P2009-11	Polyomino	7	28	24	annotating
D2009-11	2-Tangram	7	42	24	annotating

6 Conclusion and future work

This paper presented an overview of our English-Japanese bilingual multimodal corpora of referring expressions in a collaborative problem solving setting. The Japanese corpus was completed and has already been used for research (Spanger et al., 2009b; Spanger et al., 2010; Iida et al., 2010), but the English counterpart is still undergoing annotation. We have also presented a preliminary comparative analysis of these corpora in terms of the task performance and usage of referring expressions. We found a significant difference of the task performance, which could be attributed to the difference in diversity of subjects. We have tentative results on the usage of referring expressions, since only four English dialogues are available at the moment.

The data collection experiments were conducted in August 2008 for Japanese and in March 2010 for English. Between these periods, we conducted various data collections to build different types of Japanese corpora (March, 2009 and November 2009). These experiments involve capturing eye-gaze information of participants during problem solving, and introducing variants of puzzles (Polyomino, Double Tangram and Tangram without any hints[7]). They are also under preparation for publication. Table 7 gives an overview of the REX-J corpus family, where "#valid" denotes the number of dialogues with valid eye-gaze data. Eye-gaze data is difficult to capture cleanly throughout a dialogue. We discarded dialogues in which eye-gaze was captured successfully less than 70% of the total time of the dialogue. Namely, we annotated or will annotate dialogues with validated eye-gaze data only.

These corpora enable research on utilising eye-gaze information in reference resolution and generation, and evaluation in different tasks (puzzles) as well. We are planning to distribute the REX-J corpus family through GSK (Language Resources Association in Japan)[8], and the REX-E corpus from both University of Brighton and GSK.

References

Baran, Bahar, Berrin Dogusoy, and Kursat Cagiltay. 2007. How do adults solve digital tangram problems? Analyzing cognitive strategies through eye tracking approach. In *HCI International 2007 - 12th International Conference - Part III*, pages 555–563.

Bard, Ellen Gurman, Robin Hill, Manabu Arai, and Mary Ellen Foster. 2009. Accessibility and attention in situated dialogue: Roles and regulations. In *Proceedings of the Workshop on Production of Referring Expressions Pre-CogSci 2009*.

Belz, Anja, Eric Kow, Jette Viethen, and Albert Gatt. 2010. Referring expression generation in context: The GREC shared task evaluation challenges. In Krahmer, Emiel and Mariët Theune, editors, *Empirical Methods in Natural Language Generation*, volume 5980 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin/Heidelberg.

Byron, Donna K. 2005. The OSU Quake 2004 corpus of two-party situated problem-solving dialogs. Technical report, Department of Computer Science and Enginerring, The Ohio State University.

Clark, H. Herbert. and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22:1–39.

Di Eugenio, Barbara, Pamela W. Jordan, Richmond H. Thomason, and Johanna. D. Moore. 2000. The agreement process: An empirical investigation of human-human computer-mediated collaborative dialogues. *International Journal of Human-Computer Studies*, 53(6):1017–1076.

Foster, Mary Ellen and Jon Oberlander. 2007. Corpus-based generation of head and eyebrow motion for an embodied conversational agent. *Language Resources and Evaluation*, 41(3–4):305–323, December.

Foster, Mary Ellen, Ellen Gurman Bard, Markus Guhe, Robin L. Hill, Jon Oberlander, and Alois Knoll. 2008. The roles of haptic-ostensive referring expressions in cooperative, task-based human-robot dialogue. In *Proceedings of 3rd Human-Robot Interaction*, pages 295–302.

[7] N2009-11 in Table 7

[8] http://www.gsk.or.jp/index_e.html

Gargett, Andrew, Konstantina Garoufi, Alexander Koller, and Kristina Striegnitz. 2010. The give-2 corpus of giving instructions in virtual environments. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC 2010)*, pages 2401–2406.

Heeman, Peter A. and Graeme Hirst. 1995. Collaborating on referring expressions. *Computational Linguistics*, 21:351–382.

Hobbs, Jerry R. 1978. Resolving pronoun references. *Lingua*, 44:311–338.

Iida, Ryu, Shumpei Kobayashi, and Takenobu Tokunaga. 2010. Incorporating extra-linguistic information into reference resolution in collaborative task dialogue. In *Proceedings of 48th Annual Meeting of the Association for Computational Linguistics*, pages 1259–1267.

Kiyokawa, Sachiko and Midori Nakazawa. 2006. Effects of reflective verbalization on insight problem solving. In *Proceedings of 5th International Conference of the Cognitive Science*, pages 137–139.

Kruijff, Geert-Jan M., Pierre Lison, Trevor Benjamin, Henrik Jacobsson, Hendrik Zender, and Ivana Kruijff-Korbayova. 2010. Situated dialogue processing for human-robot interaction. In *Cognitive Systems: Final report of the CoSy project*, pages 311–364. Springer-Verlag.

Martin, Jean-Claude, Patrizia Paggio, Peter Kuehnlein, Rainer Stiefelhagen, and Fabio Pianesi. 2007. Special issue on Mulitmodal corpora for modeling human multimodal behaviour. *Language Resources and Evaluation*, 41(3-4).

Mitkov, Ruslan. 2002. *Anaphora Resolution*. Longman.

Nakatani, Christine and Julia Hirschberg. 1993. A speech-first model for repair identification and correction. In *Proceedings of 31th Annual Meeting of ACL*, pages 200–207.

Noguchi, Masaki, Kenta Miyoshi, Takenobu Tokunaga, Ryu Iida, Mamoru Komachi, and Kentaro Inui. 2008. Multiple purpose annotation using SLAT – Segment and link-based annotation tool. In *Proceedings of 2nd Linguistic Annotation Workshop*, pages 61–64.

Piwek, Paul L. A. 2007. Modality choise for generation of referring acts. In *Proceedings of the Workshop on Multimodal Output Generation (MOG 2007)*, pages 129–139.

Spanger, Philipp, Masaaki Yasuhara, Ryu Iida, and Takenobu Tokunaga. 2009a. A Japanese corpus of referring expressions used in a situated collaboration task. In *Proceedings of the 12th European Workshop on Natural Language Generation (ENLG 2009)*, pages 110 – 113.

Spanger, Philipp, Masaaki Yasuhara, Ryu Iida, and Takenobu Tokunaga. 2009b. Using extra linguistic information for generating demonstrative pronouns in a situated collaboration task. In *Proceedings of PreCogSci 2009: Production of Referring Expressions: Bridging the gap between computational and empirical approaches to reference*.

Spanger, Philipp, Ryu Iida, Takenobu Tokunaga, Asuka Teri, and Naoko Kuriyama. 2010. Towards an extrinsic evaluation of referring expressions in situated dialogs. In Kelleher, John, Brian Mac Namee, and Ielka van der Sluis, editors, *Proceedings of the Sixth International Natural Language Generation Conference (INGL 2010)*, pages 135–144.

Sternberg, Robert J. and Janet E. Davidson, editors. 1996. *The Nature of Insight*. The MIT Press.

Stoia, Laura, Darla Magdalene Shockley, Donna K. Byron, and Eric Fosler-Lussier. 2008. SCARE: A situated corpus with annotated referring expressions. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*, pages 28–30.

Suzuki, Hiroaki, Keiga Abe, Kazuo Hiraki, and Michiko Miyazaki. 2001. Cue-readiness in insight problem-solving. In *Proceedings of the 23rd Annual Meeting of the Cognitive Science Society*, pages 1012 – 1017.

van Deemter, Kees, Ielka van der Sluis, and Albert Gatt. 2006. Building a semantically transparent corpus for the generation of referring expressions. In *Proceedings of the Fourth International Natural Language Generation Conference*, pages 130–132.

van der Sluis, Ielka, Junko Nagai, and Saturnino Luz. 2009. Producing referring expressions in dialogue: Insights from a translation exercise. In *Proceedings of PreCogSci 2009: Production of Referring Expressions: Bridging the gap between computational and empirical approaches to reference*.