A Japanese corpus of referring expressions used in a situated collaboration task

Philipp Spanger Yasuhara Masaaki Iida Ryu Tokunaga Takenobu
Department of Computer Science
Tokyo Institute of Technology
{philipp, yasuhara, ryu-i, take}@cl.cs.titech.ac.jp

Abstract

In order to pursue research on generating referring expressions in a situated collaboration task, we set up a data-collection experiment based on the Tangram puzzle. For a pair of participants we recorded every utterance in synchronisation with the current state of the puzzle as well as all operations by the participants. Referring expressions were annotated with their referents in order to build a referring expression corpus in Japanese. We provide preliminary results on the analysis of the corpus from various standpoints, focussing on *action-mentioning expressions*.

1 Introduction

Referring expressions are a linguistic device to refer to a certain object, enabling smooth collaboration between humans and agents where physical operations are involved. Previous research often either selectively focussed only on a limited number of expression-types or set up overly controlled experiments. In contrast, we intend to work towards analysing the whole breadth of referring expressions in a situated domain. For this purpose we created a corpus (in Japanese) and analysed it from various standpoints.

From very early on in referring expression research, there has been some interest in the collaborative aspect of the reference process (Clark and Wilkes-Gibbs, 1986). This has more recently developed into creating situated corpora in order to analyse the referring expressions occurring in situated collaborative tasks. The *COCONUT* corpus (Di Eugenio et al., 2000) is collected from keyboard-input dialogues between two participants who are collaboratively working on a simple 2-D design task (buying and arranging furniture for two rooms). In contrast, the *QUAKE* corpus (Byron et al., 2005) – as well as the more recent *SCARE* corpus (Stoia et al., 2008), which is an extension of *QUAKE* – is based on an interaction captured in a 3-D virtual reality (VR) world where two participants collaboratively carry out a treasure hunting task. There has been ongoing work to exploit these two resources for research on different aspects of referring expressions (Pamela W. Jordan, 2005; Byron, 2005).

However, while these resources have inspired new research into different aspects of referring expressions, at the same time they have clear limitations. The *COCONUT* corpus is collected from dialogues in which participants refer to symbol-like objects in a 2-D world. It thus resembles the more recent (non-collaborative) TUNA-corpus (van Deemter, 2007) in tending to encourage very simple types of expressions. Furthermore, limiting participants' interaction to keyboard input makes the dialogue less natural. While the *QUAKE* corpus deals with a more complex domain (3-D virtual world), the participating subjects were only able to carry out limited kinds of actions (pushing buttons, picking up or dropping objects) as compared with the complexity of the three-dimensional target domain.

In contrast to these two corpora, we set up a comparatively simple collaborative task (Tangram Puzzle) allowing participants to freely communicate via speech and to perform actions various enough to accomplish the given task, e.g. picking, moving, turning and rotating pieces. All utterances by participants were recorded in synchronisation with operations on objects and the object arrangement. The utterances were transcribed and all referring expressions found were annotated together with their referents. Thus, this corpus allows us to study in detail human-human interaction, particularly referring expressions in a situated setting. In what follows, we first describe details of the building of the corpus and then provide

Proceedings of the 12th European Workshop on Natural Language Generation, pages 110–113,
Athens, Greece, 30 – 31 March 2009. ©2009 Association for Computational Linguistics

results of our preliminary analysis. This analysis reveals a novel type of referring expression mentioning an action on objects, which we call *action-mentioning expressions*.

2 Building the corpus

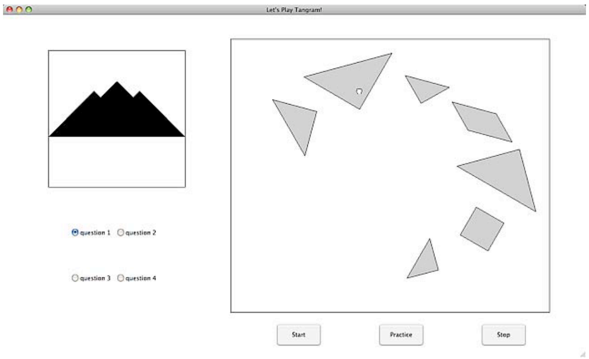

Figure 1: Screenshot of the Tangram simulator

2.1 Experimental setting

We recruited 12 Japanese graduate students (4 females, 8 males) and split them into 6 pairs. Each pair was instructed to solve the Tangram puzzle (an ancient Chinese geometrical puzzle) cooperatively. The goal of Tangram is to construct a given shape by arranging seven pieces of simple figures as shown in Figure 1.

In order to record detailed information of the interaction (position of pieces, participants' actions), we implemented a Tangram simulator in which the pieces on the computer display can be moved, rotated and flipped with simple mouse operations. Figure 1 shows the simulator interface in which the left shows the goal shape area and the right the working area. We assigned two different roles to participants, a *solver* and an *operator*; the solver thinks of the arrangement of the pieces to make the goal shape and gives instructions to the operator, while the operator manipulates the pieces with the mouse according to the solver's instructions.

A solver and an operator sit side by side in front of their own computer display. Both participants share the same working area of the simulator. The operator can manipulate the pieces, but cannot see the goal shape. In contrast, the solver sees the goal shape but cannot move pieces. A shield screen was set between the participants in order to prevent them from peeking at their partner's display. In this asymmetrical interaction, we can expect many referring expressions during the interaction.

Each pair is assigned four exercises and the participants exchanged roles after two exercises. We set a time limit of 15 minutes for an exercise. Utterances by the participants are recorded separately in stereo through headset microphones in synchronisation with the position of the pieces and the mouse actions. In total, we collected 24 dialogues of about four hours. The average length of a dialogue was 10 minutes 43 seconds.

2.2 Annotation

Recorded dialogues were transcribed with a time code attached to each utterance. Since our main concern is collecting referring expressions, we defined an utterance to be a complete sentence to prevent a referring expression being split into several utterances. Referring expressions were annotated together with their referents by using the multi-purpose annotation tool SLAT (Noguchi et al., 2008). Two annotators (two of the authors) annotated four dialogue texts separately. We annotated all 24 dialogue texts and corrected discrepancies by discussion between the annotators.

3 Analysis of the corpus

We collected a total of 1,509 tokens and 449 types of referring expressions in 24 dialogues. Our asymmetric experimental setting tended to encourage referring expressions from the solver, while the operator was constrained to confirming his understanding of the solver's instructions. This is reflected in the number of referring expressions by the solver (1,287) largely outnumbering those of the operator (222). There are a number of expressions (215 expressions; 15% of the total) referring to multiple objects (referring to 2 or more pieces) and we excluded them from our current analysis. We exclusively deal here with expressions referring to a specific single piece or indefinite expressions, i.e. those that have no definite referent (in total 1,294 tokens).

We found the following syntactic/semantic features used in the expressions: i) demonstratives (adjectives and pronouns), ii) object attribute-values, iii) spatial relations, iv) actions on an object and v) others. The number of these features is summarised in Table 1. (Note that multiple features can be used in a single expression.) The right-most column shows an example with its En-

Table 1: Features of referring expressions

	Feature	*types*	*tokens*	Example
i)	*demonstrative*	118	745	
	adjective	100	196	"*ano migigawa no sankakkei* (that triangle at the right side)"
	pronoun	19	551	"*kore* (this)"
ii)	*attribute*	303	641	
	size	165	267	"*tittyai sankakkei* (the small triangle)"
	shape	271	605	"*ôkii sankakkei* (the large triangle)"
	direction	6	6	"*ano sita muiteru dekai sankakkei* (that large triangle facing to the bottom)"
iii)	*spatial relations*	129	148	
	projective	125	144	"*hidari no okkii sankakkei* (the small triangle on the left)"
	topological	2	2	"*ôkii hanareteiru yatu* (the big distant one)"
	overlapping	2	2	"*sono sita ni aru sankakkei* (the triangle underneath it)"
iv)	*action-mentioning*	78	85	"*migi ue ni doketa sankakkei* (the triangle you put away to the top right)"
v)	*others*	29	30	
	remaining	15	15	"*nokotteiru ôkii sankakkei* (the remaining large triangle)"
	similarity	14	15	"*sore to onazi katati no* (the one of the same shape as that one)"

glish translation. The identified feature in the referring expression is underlined.

We note here a tendency to employ object attributes, particularly the attribute "shape" as well as use of demonstratives, particularly demonstrative pronouns. These kinds of referring expressions are quite general and appear in a variety of other non-situated settings as well. In addition, we found another kind of expression not usually employed by humans outside of situated collaboration tasks; referring expressions mentioning an action on an object. We have 85 expressions (over 6% of the total) of this type in our corpus.

4 Action-mentioning expressions

We further analysed those expressions that mention an action on an object, which we call *action-mentioning expressions* hereafter. Although there was significant variation in usage of action-mentioning expressions among the pairs, all 6 pairs of participants used at least one action-mentioning expression, indicating that it is a fundamental type of expression for this task setting. *Action-mentioning expressions* are different from *haptic-ostensive* referring expressions (Foster et al., 2008) since *action-mentioning expressions* are not necessarily accompanied by simultaneous physical operation on an object.

Action-mentioning expressions can be again divided into three categories: i) combination of a temporal adverbial with a verb indicating an action ("turned", "put", "moved", etc) (55 tokens or about 65% of action-mentioning expression), ii) use of temporal adverbials without a verb, i.e. verb ellipsis (22 tokens or about 26%) and iii) expressions with a verb without temporal adverbials (8 tokens or about 9%). The second category including verb ellipsis would be rare in English, but it is quite natural in Japanese.

Only less than 10% of this kind of expression did not include any temporal adverbial, indicating that humans tend to describe the temporal aspect of an action. This needs to be integrated into any generation algorithm for this task domain. The temporal adverbials used by the participants were the Japanese "*sakki no NP* (the NP [*verb*-ed] just before)" or "*ima no NP* (the current NP/the NP [you are *verb*-ing] now/the NP [*verb*-ed] just before)". "*Ima*" generally refers to the current time point ("now"). It can, however, refer to a past time point as well, thus it is ambiguous.

Participants tended to use "*ima*" largely in its perfect meaning (completed action). The frequency of use of "*ima*" in its perfect meaning in comparison to its progressive meaning was about 2:1. The distribution of the two types of temporal adverbials "*sakki*" and "*ima*" was about 2:3. The slight preference here for "*ima*" might be explained by its dual meanings (progressive and perfect) in contrast to the exclusive use of "*sakki*" for past actions.

Figure 2 shows the distribution of "*sakki* (just before)" and past-cases of "*ima* (now)" dependent on the time-distance to the action they refer to. For actions occurring within a timeframe of about 10 seconds previous to uttering an expression, participants had an overwhelming preference for "*ima*". The frequency of "*ima*" decreases quickly for actions that occurred 10-20 seconds prior to the utterance. In contrast, after 20 seconds from the ac-

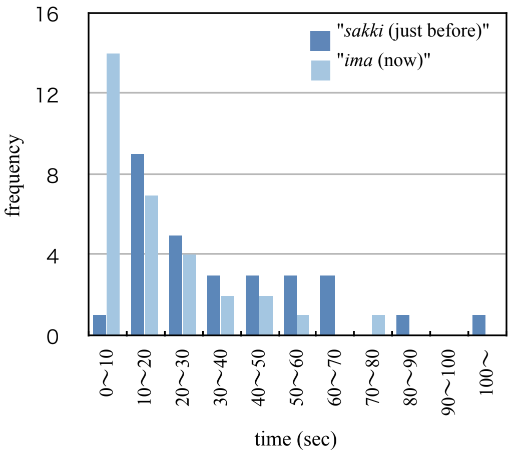

Figure 2: Frequency of “*sakki*” and “*ima*” over the time-distance to referenced action

tion, participants prefered “*sakki*”.

In addition, we investigated what actions occurred in between the utterance and the action mentioned. The actions we take into account here are basic manipulations of an object like “move”, “flip”, “click” and so on. Referring to an immediately preceding action, participants had a strong preference for using “*ima*”. Interestingly, with only one other action in between, the participants’ preference becomes opposite (i.e. “*sakki*” is preferred.). For referring to actions further in the past (i.e. more actions in between), there was a continous preference for “*sakki*” over “*ima*”. Further analysis should also investigate the phenomenon of the difference in use of temporal adverbials for other languages and whether this is related to characteristics of the Japanese language or rather an inherent property of the use of temporal adverbials in natural language.

5 Conclusion and future work

We collected a corpus of Japanese referring expressions as a first step towards developing algorithms for generating referring expressions in a situated collaboration. We carried out an initial analysis of the collected expressions, focussing on expressions that include a mention of an action on an object. We noted that they are often combined with temporal adverbials with participants seeking to make a temporal ordering of actions. In addition, we intend to further analyse other types of expressisons (demonstratives, etc) and work on developing generation algorithms for this domain.

In future work, we intend to generalise this experiment in the Tangram-domain to other domains. Furthermore, information such as gestures and eye movements should be incorporated in data collection. This will lay the basis for the development of more general models for the generation of referring expressions in a situated collaborative task.

References

Donna Byron, Thomas Mampilly, Vinay Sharma, and Tianfang Xu. 2005. Utilizing visual attention for cross-modal coreference interpretation. In *CONTEXT 2005*, pages 83–96.

Donna K. Byron. 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.

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

B. Di Eugenio, P. W. Jordan, R. H. Thomason, and J. 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.

Mary Ellen Foster, 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.

Masaki Noguchi, 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.

Marilyn A. Walker Pamela W. Jordan. 2005. Learning content selection rules for generating object descriptions in dialogue. *Journal of Artificial Intelligence Research*, 24:157–194.

Laura Stoia, 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).*

Kees van Deemter. 2007. TUNA: Towards a UNified Algorithm for the generation of referring expressions. Technical report, Aberdeen University. www.csd.abdn.ac.uk/research/tuna/pubs/TUNA-final-report.pdf.