Referring under Restricted Interactivity Conditions

Raquel Fernández, Tatjana Lucht, David Schlangen
Department of Linguistics
University of Potsdam, Germany
`{raquel|lucht|das}@ling.uni-potsdam.de`

Abstract

We report results on how the collaborative process of referring in task-oriented dialogue is affected by the restrictive interactivity of a turn-taking policy commonly used in dialogue systems, namely *push-to-talk*. Our findings show that the restriction did not have a negative effect. Instead, the stricter control imposed at the interaction level favoured longer, more effective referring expressions, and induced a stricter and more structured performance at the level of the task.

1 Introduction

The collaborative process by means of which people coordinate in identifying referents in dialogue has motivated a fair amount of psycholinguistic studies. While most of them experiment with natural, fully interactive conditions (Clark and Wilkes-Gibbs, 1986; Garrod and Anderson, 1987) some, like e.g. (Krauss and Weinheimer, 1966; Clark and Krych, 2004), have investigated how the referring process is affected by non-interactive settings that lack cotemporality (speakers do not receive messages in real time) and simultaneity (speakers cannot communicate at once). This is done by letting speakers talk to a tape recorder for future addressees, which fully precludes any form of interaction.

In the work we report here, we wanted to investigate a condition with *restricted* interactivity, in which cotemporality is allowed but si-

multaneity is inhibited. This is a setting commonly found in spoken dialogue systems that use a *push-to-talk* turn-taking strategy. To investigate the effects of this restriction in isolation, we conducted an experiment where we let subjects do a referring task either with free turn-taking or with turn-taking controlled by a half-duplex channel managed by *push-to-talk*.

Such restrictions of interactivity are often seen as having negative impact on the efficiency of the dialogue (Whittaker, 2003) as they affect the ability to give immediate and concurrent feedback and hence disturb the grounding process (Clark and Schaefer, 1989). As we have reported in recent work (Fernández et al., 2007), however, we found that subjects in the restricted condition were able to solve the task in roughly the same time, with no loss of efficiency. We hypothesised that one of the reasons behind this was a more cautious strategy whereby subjects proceed by more firmly grounding each step in the task, which was favoured by the turn-taking restriction, In this short paper we extend the analysis to investigate in more detail the effect of the *push-to-talk* restriction on the shape of the referring process. As we shall see, our findings support our previous conclusion that, for some tasks, higher interactivity is not necessarily advantageous.

After briefly describing the experimental procedure in the next section, in Section 3 we summarise the global patterns observed in the dialogues and then focus on the referring process in Section 4. We close with some conclusions and pointers for further work.

2 Experimental Setup

2.1 Task & conditions

In the task to be carried out in our experiment an *instruction giver* (IG) tells an *instruction follower* (IF) how to build up a *Pentomino* puzzle (see Figure 1). The IG has the solution of the puzzle, while the IF has the puzzle outline and the set of loose pieces.
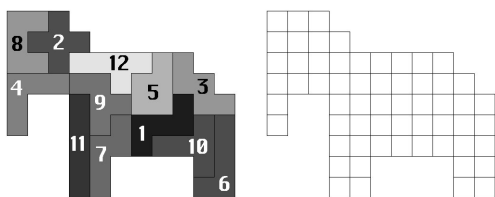
Figure 1: Puzzle and Outline

The IG is asked to tell the IF how to assemble the puzzle following the numbers shown in Figure 1. The pieces that the IF has at her disposal are however not numbered and are all the same colour.

We experiment with two different conditions: a fully interactive *free turn-taking* (FTT) condition, and a *push-to-talk* (PTT) condition where interactivity is restricted. In both conditions, subjects are in different sound-proof rooms and communication is only verbal. In FTT participants communicate by means of headsets with a continuously open audio channel. In PTT subjects use walkie-talkies that only offer a half-duplex channel that precludes simultaneous communication. Speakers have to press a button to get the turn, hold it to keep it, and release it again to yield it.

Twenty German native speakers, 11 females and 9 males between 20 and 40 years old, participated in the experiment. They were grouped in 10 IG-IF pairs and 5 pairs were assigned to each of the two conditions.

2.2 Coding

The 10 dialogues collected make up a total of 194.54 minutes of recorded conversation (in German). The recordings were transcribed and segmented into a total of 2,262 turns, 4,300 utterances and 28,969 words using the software Praat (Boersma, 2001).

Using MMAX (Müller and Strube, 2001), we annotated the dialogues at three different levels:

Dialogue acts (DAs). We distinguish between task acts (including a tag for *description acts* where a piece or a location are described) and grounding acts (including different types of feedback acts and clarification requests). More details on the scheme used can be found in (Fernández et al., 2007).

Moves. The task can be divided into 12 *moves* or cycles, one for each piece. A move covers all speech that deals with a particular piece, from the point when the IG starts to describe it (*"The next piece looks like Oklahoma"*) to the point when the subjects move on to the next item. Moves are sometimes closed with errors, which may lead to later repairs. All speech that deals with the repair of a previously closed move is annotated as a *repair sequence*.

Referential expressions. We annotated the referential expressions used by the subjects distinguishing between those that referred to a piece (*"the Swiss cross"*), those that referred to part of a piece (*"a square sticking up"*), and those that referred to a location on the board (*"between the legs of the elephant"*). Note that referential expressions and description acts are different kinds of units, with the former typically being part of the latter.

3 Global Patterns

All pairs of subjects were able to finish the task and in both conditions they did so in roughly the same time (18.7 min in PTT and 19.8 min in FTT on average; no significant difference). The PTT condition thus did not have any significant impact on task efficiency, although it did have an effect on the shape of the dialogues. PTT pairs were able to finish the task using significantly fewer words than FTT pairs. The structural patterns observed were also highly different across conditions: FTT dialogues contain roughly twice as many turns and utterance as PTT dialogues, with turns and utterances in PTT being much longer than in FTT. Table 1 gives an overview of these results.

average # of	FTT	PTT	t-test, df=8
words/dialogue	3540	2254	$p < 0.05$
turns/dialogue	328	115	$p < 0.005$
utts/dialogue	596	264	$p < 0.005$
words/utt	6	8.6	$p < 0.01$
words/turn	11.3	20.2	$p < 0.05$

Table 1: Summary of structural patterns

We also found that there were significant differences in the distribution of DAs. In particular, the proportion of positive feedback acts, like backchannels and acknowledgements, was consistently higher in FTT (33.8% vs 25.7% on average; χ^2 test, $p < 0.01$), while PTT dialogues contained a higher proportion of task-related acts (45.4% vs 36.7% on average; χ^2 test, $p < 0.01$). The reader will find an extensive discussion of these results in (Fernández et al., 2007).

4 Analysing the Referring Process

In this section, we report some results of our analysis of the referring process and of how this is affected by the global patterns brought about by our two experimental conditions.

4.1 Internal structure of moves

The moves that deal with the different pieces of the puzzle include several sub-tasks: (i) identifying the piece in question, (ii) optionally describing its orientation, and (iii) establishing its location on the board. The latter is the most challenging of the three and the one on which subjects spend most of the effort: in both FTT and PTT, slightly over 60% of the referring expressions used deal with the identification of board locations. For each move, there is minimally one change in sub-tasks, typically with a transition from (i) to (ii). These sub-tasks, however, are not always addressed in the canonical order and often subjects go back and forth between them during a single move.

To measure the orderliness of the referential process, we counted the number of times subjects changed to a different sub-task within a move and found significant differences between conditions. We observed that, on average, subjects in PTT dialogues change to a different referential sub-task 1.2 times per move, while in FTT dialogues the average number of changes

per move is 2. These differences are statistically significant (t=3.18, df=8, $p < 0.02$). Thus, participants in PTT dialogues tend to follow a more structured strategy where they first deal with the description of a piece and then with its location, making sure that each of these phases is grounded. This suggests that the stricter control imposed by the turn-taking restriction on the interaction level leads to a stricter and better structured performance at the task level.

4.2 Referential expressions

In the collaborative model put forward by (Clark and Wilkes-Gibbs, 1986), the referential process is divided into three phases: an *initiating* phase, a *refashioning* phase, and a *concluding* phase. In a basic exchange, like e.g. (1), only the first and last phases occur, which correspond to the presentation and acceptance phases of grounding any dialogue contribution.

(1) A: Number 2 is a cross
    B: OK, I have that one.

Refashioning may take place because the initial reference is not properly understood or not accepted, or simply because the speaker considers it insufficiently adequate.

According to this model, *"there is a trade-off between initiating the noun phrase and refashioning it. The more effort a speaker puts into the initial noun phrase, in general, the less refashioning it is likely to need."* As the authors point out, however, due to constraints like time pressure and the possible complexity of the referring task, speakers do not always put in enough effort to avoid refashioning, which—in conditions with full interactivity—leads to a more collaborative and interactive process. Indeed, our FTT dialogues are full of installments, provisional references and descriptions presented by proxy, as in the following example (translated from German), all of which are rare in PTT dialogues.

(2) IG: It looks kind of a bit like...
    IF: Like an inverted L with an extra bit.
    IG: Yeah, could be. Basically like a duck.

Although the referential expressions used in PTT tend to be longer, overall their average length

is not significantly different across conditions (12.6 vs 9.7 words on average; not significant). Averaging over all referential expressions, however, conceals important differences in the way in which the referential process unfolds. The differences are to be found in those expressions used in the initiating phase of the referring process. In particular, we observed that the average length of the initial descriptions used to refer to locations (which, as mentioned above, is the most prominent sub-task) is significantly higher in PTT dialogues (13.6 vs 8.7 words on average; t=2.30, df=8, $p < 0.05$). More generally, the average length of the referential expressions used in initial moves (as opposed to repair sequences; see Section 2.2) is also higher in the restrictive interactivity condition (12.6 vs 9.26; t=2.34, df=8, $p < 0.05$).

This underlines the aforementioned trade-off between the cost of producing detailed initial descriptions and the cost of interactively designing the referential expressions. The turn-taking restriction favours longer initial descriptions, which turns out to be advantageous since interactive refashioning, as in (2) above, is harder in this condition.

5 Conclusions and Further Work

We have reported some first results of our ongoing investigation of the referring process in restricted task-oriented dialogue. We have seen that there is a correspondence between the interaction and the task levels, with restricted interactivity leading to more orderly task performance. We have also observed that our PTT condition favours a strategy whereby participants put more effort in the initiating stages of the referring task, which seems to be advantageous for the task at hand.

Our findings so far support the idea that tasks that require complex, spontaneously generated contributions may not be adversely affected or even be supported by interactivity restrictions. Although understanding such complex descriptions as occur in our corpus is of course way beyond the current state of spoken dialogue systems, our results should be of more immediate significance for designing computer-mediated interaction systems.

We are currently developing a classification scheme for locative referring expressions in the lines of the taxonomy used in (Clark and Wilkes-Gibbs, 1986) for noun phrases. This will allow us to analyse the referring process further and investigate how phenomena like e.g. lexical entrainment and the increasing simplification of referring expressions (which, as shown by (Krauss and Weinheimer, 1966; Clark and Krych, 2004) are severely affected in non-interactive setting) are altered under the restricted interactivity imposed by a PTT policy.

References

P. Boersma. 2001. Praat, a system for doing phonetics by computer. *Glot International*, 5(9–10).

H. Clark and M. Krych. 2004. Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, 50:62–81.

H. Clark and E. Schaefer. 1989. Contributing to discourse. *Cognitive Science*, 13:259–294.

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

R. Fernández, D. Schlangen, and T. Lucht. 2007. Push-to-talk ain't always bad! comparing different interactivity settings in task-oriented dialogue. In *Proceedings of DECALOG*, Trento, Italy.

S. Garrod and A. Anderson. 1987. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27:181–218.

R. Krauss and S. Weinheimer. 1966. Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, 4:343–346.

C. Müller and M. Strube. 2001. MMAX: A tool for the annotation of multi-modal corpora. In *Proceedings of the 2nd IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*.

S. Whittaker. 2003. Theories and methods in mediated communication. In *The Handbook of Discourse Processes*, pages 243–286. Lawrence Erlbaum Associates.