

rightsretained
00002017

Who Has to Do It? The Use of Personal Pronouns in Human-Human and Human-Robot-Interaction

Brigitte Krenn
Austrian Research Institute for
Artificial Intelligence
Freyung 6/6
Vienna 1010, Austria

Stephanie Gross
Austrian Research Institute for
Artificial Intelligence
Freyung 6/6
Vienna 1010, Austria

Lisa Nußbaumer
Austrian Research Institute for
Artificial Intelligence
Freyung 6/6
Vienna 1010, Austria

INTRODUCTION AND BACKGROUND

In human communication, pronouns are an important means of perspective taking, and in particular in task-oriented communication personal pronouns are an indicator of who has to do what at a certain moment in a given task. The ability of handling task-related discourse is a factor for robots to interact with people in their homes in everyday life. Both, learning and resolution of personal pronouns pose a challenge for robot architectures as there has to be a permanent adaptation to the human interlocutor's use of personal pronouns. Especially the use of *ich*, *du*, *wir* (*I*, *you*, *we*) may be irritating for the robot's natural language processing system.

As regards related work, Roy et al. [?] present a set of representations and procedures that enable a robot to maintain a "mental model" of its physical environment by coupling active vision to physical simulation. Within this model, "imagined" views can be generated from arbitrary perspectives (e.g., *my left* versus *your left*). Gold & Scasselatti [?] developed a system that can learn the correct deictic meaning for *I* and *you* by observing interactions between other agents. It uses contextual information from already understood words and sensory information from its environment. In addition, empirical studies have shown that in many languages personal pronouns can be used to transmit structural knowledge and general truths, see [?] for an overview.

In the present contribution, we look into the use of the personal pronouns *ich*, *du*, *wir* in four different types of task descriptions where a human (H) teacher verbally explains and shows a specific task to a human or robot (R) learner. In Task 1, 22 teachers explain for a non-present human learner into a video camera how a selection of fruit should be arranged and re-arranged on a table. Task 2 is a collaborative task where in 22 HH-dyads a teacher and a learner together manipulate a board. In Task 3, 22 teachers explain to human or robot learners (16 HH-dyads, 6 HR-dyads) how to mount a tube in a box with holdings. Task 4 is a navigation task where in 16 HH-dyads and in 6 HR-dyads a teacher instructs a learner to follow some path. For a detailed description of the tasks and related data collection experiments see [?].

In the present paper, the data is analysed with respect to (i) the distribution of occurrences of *ich*, *du*, *wir* in the different tasks, and (ii) the literal or impersonal use of *du*, *wir* taking into account their multimodal contexts of use. To our best knowledge, this is the first work on this topic in situated, multimodal, task-related communication including HH- and HR-dyads. Whereby the tasks vary according to physical presence or absence (Task 1) of the learner, as well as the activity level of teacher and learner, i.e., teacher and learner conduct a task collaboratively (Task 2), only the teacher is acting (Task 3), only the learner is acting (Task 4).

DATA ANALYSIS AND RESULTS

Depending on the task-assignment a certain trend in favour of one predominantly used pronoun can be observed, see Figure 1. While in Task 1, where the learner is not physically present, *ich* is predominantly used, *wir* is the prevailing personal pronoun in the collaborative Task 2, the use of *ich*, *du*, *wir* is a bit more balanced with a surplus of *ich* in the HH-dyads of Task 3 (active teacher), and there is a clear bias to *ich* in the HR-dyads. In Task 4, where the learner has the active part, *du* is the most prominently used personal pronoun in both, HH- and HR-dyads.

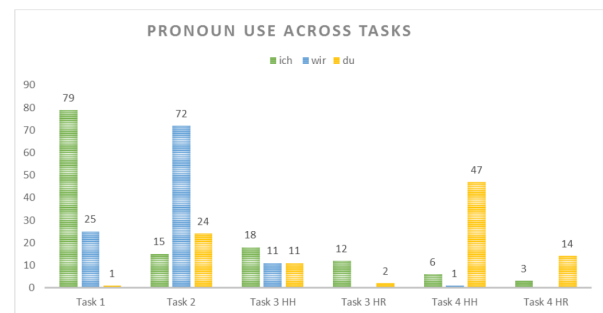


Figure 1: Pronoun use across all 4 tasks in human-human and human-robot interactions.

This flexibility of using pronouns in natural language represents a challenge for human-robot interactions. Similar to human-human interactions, 1st person *ich* principally can be interpreted literally, 2nd person *du* and *wir* are ambiguous:

- (i) literal *you*: teacher explains an action and the addressee actually performs it
- (ii) impersonal *you*: teacher explains and performs action

In these cases, the robot needs to correctly identify the human intention and decide whether it should start acting or continue listening and watching.

A closer look on the usage of *du* and *wir*

As there was a learner physically present in Tasks 2, 3 and 4, these three tasks were examined to find a pattern whether *du* and *wir* indicate an active learner involvement or not. In particular, we looked for each HH- and HR-dyad into the first occurrences of *du*, *wir* and their multimodal contexts.

Task 2 – object manipulation; collaborative task; active role of instructor & learner. *Du* and *wir* is always used literally. 13 times *du* is the first uttered personal pronoun, *wir* seven times and one instructor started with an imperative. In 59% of the first occurrences, the pronouns are uttered in combination with pointing at a (relevant) object. In addition, seven instructors utter *jetzt* (now) and four utter *bitte* (please) indicating immediate learner involvement.

Task 3 – object manipulation; active role of instructor; passive role of learner. Six out of 16 participants uttered *du* and six *wir*. Only one occurrence of *du* could be interpreted literally and two instructors interrupted and corrected themselves changing from *wir* to either *ich* or *man* (one indicating impersonal use). For the remaining five first occurrences of *du* the instructor was always holding a task-relevant object in his/her hand. For the remaining five first occurrences of *wir* three were uttered during a summary at the beginning (impersonal use). One of these three instructors continued uttering *wir* after the summary and was holding a task-relevant object, indicating that the teacher keeps the active role and *wir* must be interpreted as impersonal. Another one corrected immediately after the summary from *wir* to *ich*. The third one did not utter any personal pronoun at all, however, held relevant objects throughout the whole task description. In the human-robot interactions only one instructor changed from *ich* to *du* after more than half of the task description. All other instructors exclusively used *ich*.

Task 4 – navigation; passive role of instructor; active role of learner. 14 out of 16 instructors in the HH-dyads uttered the first *du* in combination with a deictic gesture indicating where the learner should move. One instructor used the imperative and one summarised the task at the beginning and only included deictic gestures when he uttered *du* in the first actual instruction. Additionally, seven instructors uttered *bitte* (please) and three *jetzt* (now) close to *du* as indication for the learner to get active. In the HR-dyads, three instructors uttered *du* when first using a personal pronoun and three started with an imperative. All six six accompanied these utterances with a deictic gesture, two uttered *bitte* close to first addressing the learner and one *bitte* and *jetzt*, indicating that the learner got to get active.

CONCLUSION AND FUTURE WORK

General results. Depending on the task, the probability is high that personal pronouns cannot always be interpreted literally. For example, in Task 3, up to three different personal pronouns were uttered, all referring to the instructor, and only one occurrence of *du* could be interpreted literally. Moreover, the analysis of meta-communication needs to be separated from task-related communication. When occurring in meta-communication such as initial summaries, the interpretation of *wir* follows different criteria than in the task-accompanying talk, e.g., there are seldom non-verbal cues which allow the listener to decide upon their involvement.

Differences between tasks. When the learner was actively involved in conducting the task (Tasks 2 & 4), all occurrences of *ich*, *du*, *wir* were to be interpreted literally, in contrast to Task 3 where only the instructor conducted the task, the personal pronouns had to be interpreted via visual cues, to identify who is intended to

conduct the task. The results show that in the majority of cases where *du* or *wir* had to be interpreted literally when uttered for the first time in the task, the instructor pointed at a (relevant) object while speaking. Also occurrences of *bitte* and *jetzt* were strong indicators for literal interpretation. *Du* and *wir*, however, could not be interpreted literally, when the instructor was grasping or holding an object within the same utterance. Two human learners misinterpreted occurrences of *du* as literal, indicated movement towards the objects and were immediately corrected by their instructor. This shows that even for humans it is not always clear how to interpret personal pronouns. However, human instructors employ communicative means to immediately correct this misinterpretation, which the robot as communication partner must be able to deal with.

Differences between HH and HR interaction. In the human-robot setting, the use of *wir* was almost in-existent. However, there were parallels in the HH- and HR-dyads regarding the prevalence of *ich* in Task 3 (active teacher) and of *du* in Task 4 (active learner). In this respect, the human instructors were more explicit about who is supposed to be the active part when interacting with the robot. These findings are a first explorative indicator, but need to be treated with caution because of the small number of HR-dyads and the rather passive appearance of the robot.

Future work. More data will be collected, especially in human-robot settings but also for the human-human settings. Moreover, to further harden/challenge the evidence, data from different task settings of the same task types need to be collected, whereby task type means manipulation versus navigation task, co-presence of teacher and learner yes/no, active involvement of the instructor or the learner or both.

ACKNOWLEDGMENTS

This research is supported by the Vienna Science and Technology Fund (WWTF, project ICT15-045) and the EU CHIST-ERA project ATLANTIS.