Utterance types in the August database

Linda Bell and Joakim Gustafson

Centre for Speech Technology KTH
{bell, joakim_g}@speech.kth.se

Future information systems will benefit from using speech technology components, both in terms of accessibility and user-friendliness. These natural language systems should be designed so that non-specialists can operate them. The experimental multimodal August system was developed to study how non-trained users would communicate with an animated agent.

Traditionally, spoken dialogue systems have been designed with a specific and rather narrow task in mind. Many such systems are built around a structured dialogue model that specifies almost every aspect of the human-computer interaction. In a recent study, Heeman et al. suggest that this is motivated by the fact that structured systems presently are the only systems that are relatively simple to build [1]. The human-computer interactions in such systems are almost exclusively system-directed and place limitations on the users' linguistic input to the system. Allen et al. have argued that unless we have working systems, it becomes nearly impossible to make fair evaluations of models and theories on dialogue management [2] The experimental August system was exposed to the general public from the very beginning. The system had a recognition lexicon of no more than 500 words and idiomatic phrases. However, the system responses had been manually pre-processed, integrating prosodic information as well as head- and facial movements [3]. This resulted in a system that sometimes appeared to handle almost anything and which generated quite human-like dialogues, while it sometimes failed almost completely.

One of the aims of the August system was to collect spontaneous speech input from people who had little previous experience of spoken dialogue systems. The users of the system were given little or no information about the system's capacities. The corpus collected has been studied so that user reactions during error resolution have been analyzed. These studies focus on how users adapt their speech when the dialogue fails [4, 5]. Many of the people who interacted with the August system seemed to be more interested in making the system respond to their spoken input rather than searching for information. This resulted in a large number of what in our analysis is referred to as **social** utterances. In addition, the users asked factual questions that were clearly out-of-domain, commented on the system itself, referred back to previous dialogue turns and sometimes even tried to deceive the system. The August system was designed with a number of simple domains instead of a single complex one, and one of these domains handled greetings and other social utterances. Nonetheless, it is clear that the performance of the system did not always match the users' expectations. This is also reflected in the current database, which contains a number of utterances that are referred to below as **insults**. We here describe the categorization of the spoken input in the August database into utterance types. Lexical and syntactic aspects of the August corpus are also briefly discussed to see whether the utterance types are distinguishable in terms of linguistic complexity.

The material analyzed in the August database consists of 10,058 utterances. The utterances were transcribed orthographically and labeled with some basic speaker characteristics. The total number of speakers was 2685, out of which 50% were judged to be men, 26% women and 24% children. The average number of utterances per user was 4.1 for men, 3.3 for women and 3.5 for children. The number of utterances originating from a single speaker ranges from one to forty-nine. The number of words per utterance in the August corpus also varied greatly, although most utterances were relatively short. The average utterance contained four words, but this figure varied from a single word to twenty-two words. The database consists of almost 40,000 words, out of which close to 3000 are unique. Approximately half of these unique words occurred only once in the database. The 200 most frequently occurring words in the database covered about 80% of all the words in the corpus.

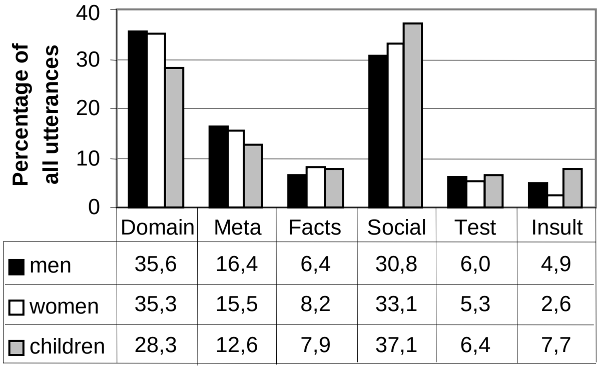
The input utterances were mostly short and their syntactic patterns were seldom complex. As reported in [4], 188 sentence types covered 80% of all the utterances in the database. The number of sentence types differed greatly between the different categories of utterances in the database. 35 sentence types covered 80% of the utterances used for seeking information from the system while 26 sentence types covered 80% of the utterances used for socializing with the animated agent. The largest number of different sentence types was used during error resolution and by users who were testing the limits of the system.

In the analysis of the August database, the utterances were also labeled according to the presumed intentions of the users. The purpose of this categorization was to get a better picture of the kinds of things the users wanted to convey when interacting with the system. The concept of communicative intention is a difficult one, both in human-human and human-computer interaction [6]. The categorization of utterances always involves an arbitrary element, as one and the same utterance may express different communicative intentions depending on the context. Nevertheless, the utterances in the database were categorized in accordance with a simplified pragmatic model containing six major categories.

Table 1. The utterance types in August database.

Socializing	**Examples**
Social	*Hello August!* *That's a nice mustache!*
Insult	*You are stupid!* *Is your brain too small*
Test	*What is my name?* *I want to rent a refrigerator*
Info-seeking	**Examples**
Domain	*How many books did Strindberg write?* *Where are the restaurants on Kungsgatan*?
Meta	*What can I ask you?* *I told you that already!*
Facts	*What's the capital of Finland?* *What is two times two?*

Table 1 is an overview of the utterance types in the August database. The **social** category consists of greetings and remarks of a personal kind, while expletive expressions and swear words were placed in the category of **insults**. The category called **test** contains utterances that were spoken with what appeared to be the purpose of deceiving the system. The **domain** category includes utterances in one of the established domains which the users had been given some information about. Questions about the system itself and comments about the actual dialogue are grouped in the **meta** category. Factual questions outside the domains mostly turned out to be of an encyclopedic nature and sometimes dealt with things people would expect a computer to be good at, such as calculus. These utterances are categorized as **facts**. Figure 1 shows the distribution of these utterance types in the August database. Some differences between how men, women and children communicated with the system can be observed. For example, children in our study made use of social utterances to a greater extent than adults did. One possible explanation might be that the other domains did not particularly appeal to children. Women rarely used insults, while the children in the present study used them rather frequently. In order to be able to get a better overview of the utterance categories in the database, two main groups were created from the above mentioned six. The first one, *socializing,* includes the categories **social**, **test** and **insults** while the second one, *information-seeking,* includes the categories **domain**, **facts** and **meta**. The socializing category constituted **44%** of all utterances in the corpus. Our data suggest that men more often began by socializing with the system and then turned the dialogue to the area of information-seeking, while women more often focussed on more domain- and fact-oriented questions from the beginning. In contrast, many of the children used only social utterances over the first six turns. Very few users alternated between information-seeking and socializing during their first six turns. There seemed to be four distinguishable groups of users: firstly, those who only wanted to socialize, secondly, those who only wanted to seek for information, thirdly, those who began by exchanging greetings and then turned to information-seeking and remained in that area. The final group was a small one, and it consisted of users who tried to communicate with the system, but failed, and therefore alternated between information-seeking and socializing, trying to get the system to understand.



	Domain	Meta	Facts	Social	Test	Insult
■ men	35,6	16,4	6,4	30,8	6,0	4,9
□ women	35,3	15,5	8,2	33,1	5,3	2,6
▨ children	28,3	12,6	7,9	37,1	6,4	7,7

Figure 1. Distribution of the utterance types.

To sum up, some users of the August system started by looking for information immediately, while others preferred to socialize with the system before going into this mode. Almost half of all the utterances in the August database were categorized as socializing. Children appeared to be especially inclined to socialize with the system, perhaps because their interest in the established domains was low. The animated agent's human-like appearance probably made this social interaction make sense.

REFERENCES

1. Heeman, P. A., Johnston M. Denney J., Kaiser, E. (1998) Beyond structured dialogues: Factoring out grounding. In *Proceedings of ICSLP '98*

2. Allen J F et al. (1996) A Robust System for Natural Spoken Dialogue In *Proc.of 34th meeting of the Association for Computational Linguistics*

3. Gustafson, J., Lindberg, N. and Lundeberg, M. (1999) The August Spoken Dialogue System, Forthcoming in Proceedings of Eurospeech ‘99

4. Bell, L. and Gustafson, J (1999) Interaction with an animated agent in a spoken dialogue system. Forthcoming in Proceedings of Eurospeech ‘99.

5. Bell, L. and Gustafson, J. (1999) Repetition and its phonetic realizations: Investigating a Swedish database of spontaneous computer-directed speech. In *Proceedings of ICPhS ‘99*

6. Cohen P., Morgan, J. and Pollack, M. E. (eds.) (1990) *Intentions in Communication*. Cambridge: MIT Press