LEXICAL ACCOMMODATION IN MACHINE-MEDIATED INTERACTIONS

Laurel Fais
ATR Interpreting Telecommunications Research Laboratories
2-2 Hikaridai, Seika-cho Soraku-gun
Kyoto, Japan 619-02
email:fais@itl.atr.co.jp

Abstract

We report results of lexical accommodation studies involving three different interpretation settings: human-human monolingual; human-interpreted bilingual; and machine-interpreted bilingual. We found significant accommodation in all three conversational settings, with the highest rate in the human-interpreted setting. There is evidence for long-range mutual accommodation in that setting, as compared to short-range accommodation in the machine-interpreted setting. Motivations discussed in the accommodation literature, including speakers' concern for social standing and communicational efficiency, are examined in the light of the results obtained. Finally, we draw implications for the design of multimedia human-computer interfaces.

1 Introduction

For real-time, real-situation, human-computer interaction to approach reality, the burden of understanding and conveying information cannot be shared equally between the two interactors. Humans need to make allowances for features of the computer interface such as synthesized speech, limitations on range of knowledge base, and imperfect speech recognition. However, in order for naive users to accept and use computers effectively in an interactive format, the restrictions placed upon them need to be as minimal and as natural as possible. Exploring the linguistic behavior of humans interacting with computers in an unrestricted environment allows us to determine how humans are *naturally* inclined to accommodate to the current limitations of human-machine interaction. Encouraging those natural inclinations in real human-machine systems will have a better chance of success than imposing artificial restrictions. In addition, systems designers can take advantage of the accommodations that humans make naturally to improve the performance of their systems (Fais *et al.*, 1995)

Below, we will discuss a particular kind of accommodation in conversational interaction, namely lexical accommodation. In lexical accommodation, one conversant adopts the lexical items used by the other conversant. This type of accommodation is one way in which users adapt to the limitations of computer interfaces, i.e., they converge to the limited lexicon of the computer. Thus, it has important implications for the design of workable human-computer interfaces. We discuss experiment results from three conversational settings: human-human monolingual; human-interpreted bilingual, and machine-interpreted bilingual.

1.1 Background

Of course, there is no *a priori* reason why interactors could not conduct conversations in completely different styles, using different phonologies, sentence structures, vocabulary, etc. However, it has been widely demonstrated that they do not. Lexical accommodation is only one instance of a diverse range of convergence behaviors that humans display in conversation. Giles *et al.* (1987) cite studies demonstrating convergence of speech styles, dialect, non-verbal behavior, vocal intensity, prosody, speech rate and duration, and pause length. Garrod and Doherty (1994) report on a study in which conversational interactors in a language community showed a high level of convergence on a particular description language over the course of the task (a maze game). Fais (1994a) discusses both lexical and syntactic accommodation over a range of natural conversational styles.

Accommodation has also been shown to be present in human-computer interactions. Zoltan-Ford (1991) and Leiser (1989) demonstrate accommodation by users to the phrasing and vocabulary of the confirmation output of a computer in information manipulation and retrieval tasks. Speakers also unconsciously adapt their speaking behavior to the limitations of a speech recognition system, as demonstrated by Mellor and O'Connor (1995).

But while the phenomenon of accommodation, both between humans and between human and computer, is amply demonstrated, the motivations behind the phenomenon are less often discussed. Those in the field of human-computer interactions usually note simply that accommodation exists. They express relief when it is found to act as a natural constraint on the user's vocabulary or syntax (e.g., Leiser, 1989; Zoltan-Ford, 1991), and distress when accommodation to the "natural speech" style of some computer output encourages matching casual speech from the user which is difficult to process (e.g., Spitz *et al.*, 1991).

Speech Accommodation Theorists, who fall under a broad category which might be called socio-linguistics, tend to ascribe one of three motivations to speakers who accommodate: "evoking listeners' social approval, attaining communicational efficiency between interactors, and maintaining positive social identities." (Giles *et al.*, 1987) p.15). These motivations can be grouped into two major categories: concern with social standing and identity, and concern

with communicational efficiency.

1.2 Hypotheses

Specifically, then, if we base our predictions on these standard accounts in the literature, we should expect the following results concerning both level and direction of accommodation:

• In human-human interaction, we should find significant lexical accommodation. Because this is essentially an information-giving and -receiving task, we expect that the receiver of the information will accommodate to the giver, adopting the lexical items used by the speaker who imparts information.

• The human-interpreted setting constitutes both a human-human interaction and a more stressful communication environment, one in which communicational efficiency is a concern. For that reason, we expect an even greater level of accommodation in the human-interpreted setting than in the human-human setting. In the human-interpreted setting, we examine the accommodation between client and interpreter. The client is information receiver, and the interpreter is the imparter of information, not the originator; thus, neither client nor interpreter is in a dominant role. For this reason, we cannot predict whether the client will accommodate to the interpreter or *vice versa*.

• The machine-interpreted setting only indirectly involves human-human interaction; all dialogue is mediated by the "machine" interpreter. Therefore, we conjecture that interactors in this setting will not be concerned with social standing. On the other hand, this is the most difficult communication environment of the three, involving as it does not only the limited understanding of the machine translator but also limited speech recognition, a difficult-to-understand modulated speech signal, and rigid turn-taking constraints. For this reason, communicational efficiency *will* be a concern. However, whether this will generate more or less accommodation than concern for social standing generates in the human-human case is an open question.

Since users in the machine-interpreted setting should not be concerned with social standing, we might predict a lower rate of accommodation than in the human-interpreted setting. However, again, the greater difficulties in communication in the machine-interpreted setting might make up for a lack of concern with social standing, resulting in a rate of accommodation comparable to that in the human-interpreted setting. Results discussed below will shed some light on the interaction of these two factors.

• We expect that clients will accommodate to the machine to some extent, that clients' word choice will be affected by their perception of "what works," or "what the machine knows." Given this possibility, then, we predict that the results will show client accommodation to the machine-interpreter, but at a lower level than in the human-interpreted setting.

2 Methods

We conducted three experiments to evaluate lexical accommodation in various interpretation settings. In the first experiment, native English-speaking subjects acting as "clients" were instructed that their task was to get directions to the site of a conference they were "attending" by engaging in a cooperative dialogue with a native English-speaking "conference agents." Their interaction was human-human. In two other, similar experiments, English-speaking clients interacted with Japanese-speaking agents, both to get directions and to make hotel reservations. In one of these experiments, speech was translated by human interpreters; in the other, by a simulated automatic machine translation system ("Wizard of Oz" style; see (Fais, 1994b) and (Fais *et al.*, 1995) for further details of these two experiments). We will refer to this latter setting as the "machine-interpreted" setting; keep in mind, however, that translation was actually done by trained interpreters mimicking a computer-based system.[1] The experimental configurations are shown in Figure 1.
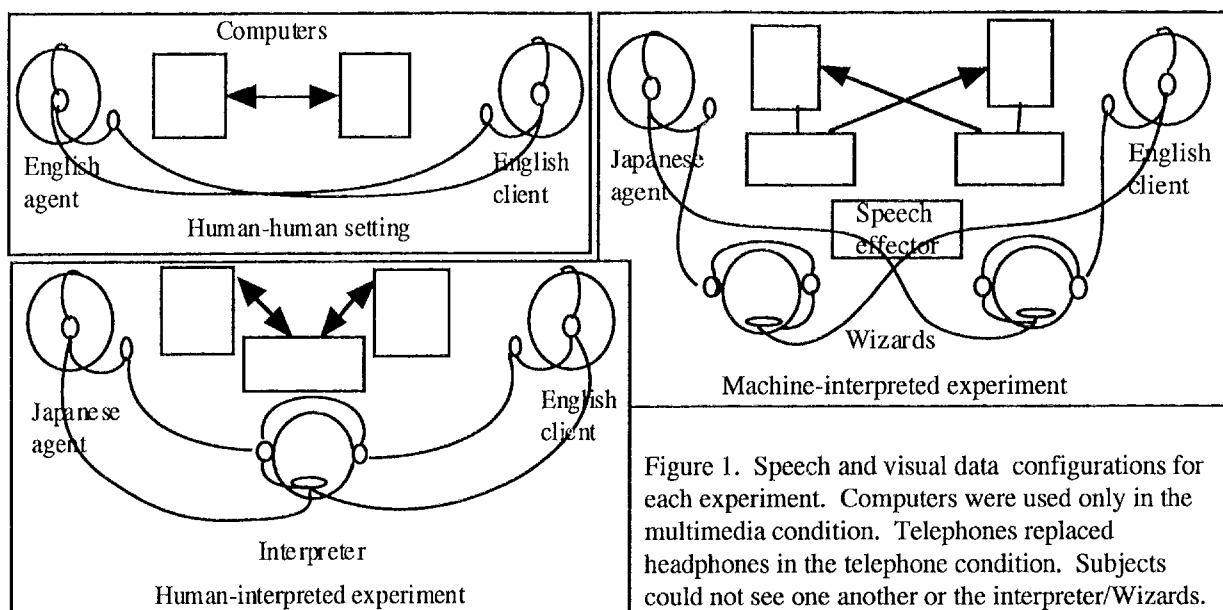
2.1 Measures

Lexical Accommodation: We measured lexical accommodation by examining the number of lexical items which were used by *both* interactors in the course of a conversation. The accommodation rate for each conversation in the three experiments was calculated by dividing the number of (different) lexical items the two speakers had in common by the total number of (different) lexical items in the conversation.

We calculated lexical accommodation rates for client and agent in the same-language, human-human experiment setting; for client and (Japanese-to-English) interpreter in the human-interpreted experiment setting; and for client and (Japanese-to-English) "Wizard" interpreter in the machine-interpreted experiment setting. Although the actual measurement of lexical items was done for the English speech of the Japanese-to-English *interpreters*, these interactors in the conversation will be referred to below as "agents," to conform to the human-human setting in which we assessed the lexical accommodation of the agents directly.

Direction of Accommodation: Accommodation is not necessarily a mutual phenomenon (Giles, 1987). In order to determine if one conversant was accommodating more than the other, we examined the number of words used first by the client and the number used first by the agent. We reasoned that the subject who used a particular lexical

Figure 1. Speech and visual data configurations for each experiment. Computers were used only in the multimedia condition. Telephones replaced headphones in the telephone condition. Subjects could not see one another or the interpreter/Wizards.

item first was *not* accommodating, and, by extension, then, the subject who did not use an item first, *was* accommodating.

The following objection to this definition might be made: the fact that interactors use words that their partners have used does not necessarily mean that they are accommodating to the other's prior use of that word. But what other justification could there be for saying that accommodation *is* taking place? Because even lexical accommodation is rarely a conscious act, speakers' intuitive judgments are not helpful. On the other hand, there is no other external evidence available. Thus, we will use the quantitative criterion described above. We will argue that accommodation is a real phenomenon in dialogue; it follows, then, that at least some of the instances in which conversants use lexical items previously used by their partners are instances of accommodation.

Relative Importance of Words-in-Common: We also looked beyond the initial use of lexical items to determine what role was played in subsequent conversation by the words that agent and client both used. That is, once one conversant accommodates to the other by adopting a lexical item, does that conversant continue to use that lexical item in a significant way in the remainder of the conversation? In order to explore this question, we estimated the percent of usage for each word-in-common, for both client and agent. That is, for each word-in-common, we divided the number of times each subject used the word by the total number of words uttered by that subject in order to determine what proportion of the subject's conversation consisted of the uses of that word. By comparing these proportions for the roles of client and agent, we ascertained the relative frequency of the word for each role, giving us some idea of the "importance" of the word for that role.

Coincidental Overlap: Of course, a certain amount of lexical overlap is inevitable as a simple artifact of cooperative conversation. In order to

determine the extent of coincidental overlap, we measured the lexical overlap in the speech of clients and agents from the first experiment who had not participated in the experiment together. That is, the speech of clients who had participated in the experiment with Agent A was compared with the speech of Agent B. Likewise, the speech of clients who had participated in the experiment with Agent B was compared with the speech of Agent A. Because these conversants were not talking to one another, the lexical overlap in their speech could not be a result of accommodation to one another. However, because the overlap was calculated for speakers engaging in cooperative dialogues concerned with the same task and via the same media , it reflects the extent to which overlap occurs simply because these are speakers in similar situations talking about similar topics.

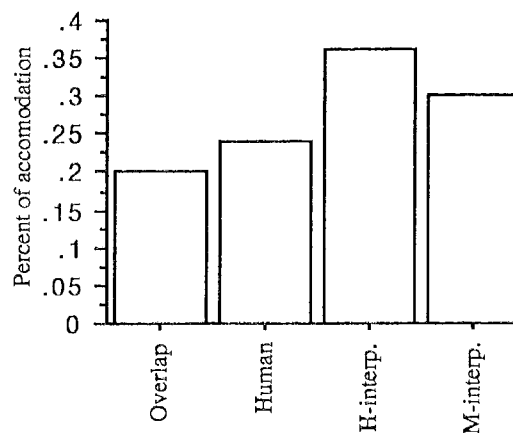# 3 Results

3.1 Lexical Accommodation

Figure 2. Rates of accommodation for coincidental overlap and all three settings.

372

Table 1. Significance levels for differences in lexical accommodation (two-way ANOVA).

	human-human	human-interpreted	machine-interpreted
Coincidental overlap	p<.03	p<.0001	p<.001
Human-human		p<.0001	p<.02
Human-interpreted			p<.01

The rates for lexical accommodation for all three experiments were significantly different from the level established for coincidental lexical overlap (Figure 2, Table 1; data were subjected to two-way analyses of variance). In addition, the lexical accommodation rates for each experiment also differed significantly from those for each of the other two experiments. Human-human accommodation was higher than coincidental overlap, but lower than both of the interpreted settings. The human-interpreted setting had the highest rate of accommodation.

3.2 Direction of Accommodation

When we examined the percentage of words-in-common used first by each role (agent or client), the following patterns emerged (Figure 3; data were subjected to three-way analyses of variance). In the human-human setting the agent used a significantly higher percentage of words first (p<.03); the client accommodated to the agent. In the human-interpreted setting, there was no difference. It is not possible to say that one or the other interlocutor was responsible for the accommodation found. In the machine-interpreted setting, the agent used a significantly higher percentage of words first (p<.005); the client accommodated to the agent.
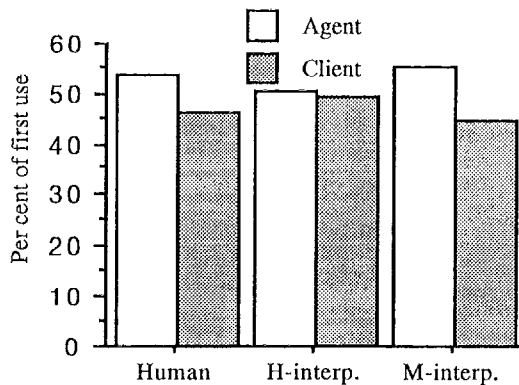

Figure 3. Percent of first use of words-in-common for agent and client in each setting.

3.3 Relative Importance of Words-in-common

An examination of the use of each word-in-common with respect to overall word use for client and agent, i.e., the word's "importance," showed the following results (Figure 4; three-way ANOVA). In the human-human setting, client use of words-in-common made up a significantly greater percent of total word use than agent use of these words (p<.0003). There was no significant difference between client and agent in

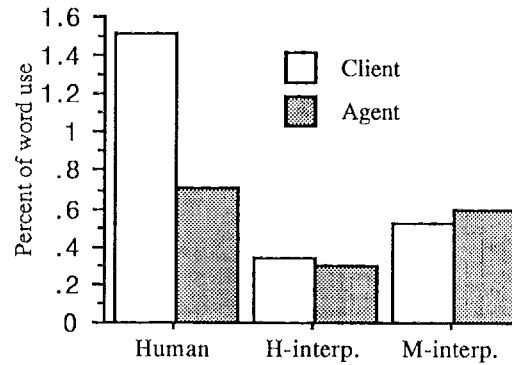the human-interpreted setting or in the machine-interpreted setting.

Figure 4. Frequency of use of words-in-common for agent and client in each setting.

4 Discussion

We analyzed lexical accommodation in a variety of interactions in order to determine how accommodation can be expected to operate in a machine-interpreted context, and learn ways in which to support lexical accommodation in the design of human-machine interfaces. It is encouraging that lexical accommodation happens spontaneously. As our initial results show, it is not simply a coincidental byproduct of conversing about common topics. There is a significant difference between that case (what we have called "coincidental overlap") and the case of two people talking about the same topic *to one another*. While lexical accommodation has been shown for typed human-computer interactions (Leiser, 1989), and lexical and structural accommodation has been demonstrated for typed and spoken human-computer interactions in constrained contexts (Zoltan-Ford, 1991), these results demonstrate lexical accommodation for unconstrained, spoken, machine-mediated human-human interactions. Given that lexical accommodation is a "real" phenomenon, then, how can we characterize the patterns of accommodation for each experimental setting, and what can we learn from them?

In the human-human setting, there was a non-trivial, but low level of accommodation. The client accommodated to the agent, using the words-in-common more frequently in subsequent conversation than did the agent. This is consistent with the interpretation that the agent acted as information-provider and the client acted as information-receiver in a non-stressful communication environment, as our initial hypothesis stated. The interactors maintained a level of accommodation high enough to satisfy their

concern for social standing, but since they were native speakers of the same language and the communication channel was clear and direct, they had minimal concern for communicational efficiency. There was no incentive to extend lexical accommodation.

On the other hand, the human-interpreted setting presented a more difficult communication environment in which concern for communicational efficiency was present. Since the interaction was also human-human, social standing continued to be a concern. We expected that the addition of the concern for efficiency to that for social position would result in a higher level of accommodation; in fact, the level of accommodation observed in the human-interpreted setting was the highest in all three experimental settings.

More specifically, the human-interpreted setting involved speakers from two different language backgrounds, both of whom were capable of recognizing the differences in their linguistic behaviors, and of reducing those differences to facilitate communication. Lexical choice is a surface level phenomenon, open to manipulation. Thus, lexical accommodation is an important conversational strategy for speakers who do not share linguistic conventions. The interpreters in the human-interpreted setting were native speakers of Japanese, and, though they were fluent in English, the range of overlap between their English linguistic habits and those of the native English-speaking clients was much smaller than that between two native speakers. Lexical choice was an obvious strategy for establishing shared linguistic behavior, and thus promoting effective communication. So, concern for social standing and communicative efficiency combined to generate a high rate of mutual accommodation.

Because neither client nor interpreter had a dominant role in the conversation, we could not predict the direction of accommodation. In fact, our results show that it was not possible to single out a primary accommodator in the human-interpreted setting, either in terms of proportion of words used first or the frequency with which words-in-common were used. Does this mean that both speakers accommodated or that neither did? Considering the high accommodation rates for this setting (Figure 1), we conclude that, in fact, *both* client and agent were accommodating to one another.

In the machine-interpreted setting, we saw a rate of accommodation higher than that of the human-human setting, but lower than that of the human-interpreted setting, as expected. The machine-interpreted setting is an even more stressful communication environment than the human-interpreted setting; concern for communicational efficiency resulted in a higher level of accommodation than concern for social standing did in the human-human setting. However, we do not expect humans to be concerned about their social standing with a machine, unlike in the human-interpreted setting. This explains why the rate for lexical accommodation in the machine-interpreted setting is lower than that of the human-interpreted

setting. The greater concern for communicational efficiency in the machine-interpreted setting was not enough to generate as high a level of accommodation as that found in the human-interpreted setting, where there was the additional factor of concern for social position, though it did generate a higher level of accommodation than did concern for social standing alone (the human-human setting).

As we conjectured above, clients accommodated to the machine as part of a strategy for effective communication. However, given the fact that there was a lower rate of accommodation than in the human-interpreted setting, coupled with the strong directionality observed, we conclude that this is not a case of mutual accommodation. Instead, as in the human-human setting, clients were the primary accommodators. Clients in the machine-interpreted setting may have perceived the machine to be in the dominant role, just as the agent played the dominant role in the human-human setting.

Recall that clients in the machine-interpreted setting, unlike those in the human-interpreted setting, did not use words-in-common more than the agent did in subsequent conversation. Since clients were not concerned with social standing, including establishing mutual linguistic conventions, accommodation in the machine-interpreted setting was a local phenomenon which did not extend throughout the conversation.

5 Conclusions and Future Directions

What does this tell us about the design of human-computer interfaces? Recall that these conversations were unconstrained; neither agents, clients nor interpreters, whether human or "machine," were under instructions to limit or modify their speech in any way. Thus, what we see in these results is the natural tendency of humans to accommodate to their interlocutors in a variety of communication environments. This tendency resulted in the highest level of accommodation in the human-interpreted setting. That level was achieved as a result of mutual accommodation between the two humans involved, both of whom felt a concern for both social standing and communicational efficiency. The level of accommodation observed in the machine-interpreted setting was both lower and less extensive, i.e., it did not persist throughout the conversation.

We can take advantage of even the moderately high level of accommodation found in the machine-mediated setting by building into a language processing system a preference for the lexical items used by the machine. Coupled with accommodations in other aspects of language such as discourse and syntactic complexity, fluency, and speaking rate (Fais *et al.*, in press), lexical accommodation can inform a language model to improve language processing performance by exploiting the relationships between human speech and the speech of the machine interface. Specifically, candidates for speech recognition or parsing could receive higher preference scores if they include lexical items or structures previously encountered in the discourse. Alternatively,

preferences based on accommodation could be built directly into a language model for speech recognition or decision tree for parsing.

We would like to investigate the possibility of increasing the level of accommodation in the machine-mediated setting. Ideally, we would like users' accommodation to a machine interface to be as high as possible so that the lexical variability of users' speech can be as constrained as possible. One way is to use the resources of a multimedia environment to replicate the effect of the human-interpreted setting by providing the machine interface with a human-like persona. A number of human-computer systems already include such a feature (e.g., Ball and Ling, 1995; Bertenstam et al., 1995; Webber, 1995); it remains to be seen if it will have the desired effect on lexical accommodation. On the other hand, there is evidence that encouraging users to interact with machines as if they were humans may actually undermine the quality of the users' speech from the point of view of language processing. Work in the area of disfluencies in human-to-machine speech suggests that humans do, in fact, "clean up" their speech for machines (Suhm et al., 1994). These advantages may be lost if humans are encouraged to treat a machine interface as if it were human. Empirical investigation is required to determine if an optimal balance can be reached.

We have suggested that the design of speech recognition and language processing systems can take advantage of users' lexical accommodation to machine interfaces to improve system performance. This, in turn, would allow the construction of systems which make fewer demands on the willingness of users to adjust to misrecognitions and misunderstandings, and which encourage users to interact with computer interpreters as if they were interacting with human interpreters. This result would also have the effect of further increasing lexical accommodation from users.

Acknowledgments

References

Ball, J. Eugene, and Daniel T. Ling. 1995. Spoken language processing in the Persona Conversational Assistant. In *Proceedings of the ESCA Workshop on Spoken Dialogue Systems*, pages 109-112, Vigsø, Denmark, May-June.

Bertenstam, Johan, et al. 1995. The Waxholm system: A progress report. In *Proceedings of the ESCA Workshop on Spoken Dialogue Systems*, pages 81-84, Vigsø, Denmark, May-June.

Fais, Laurel. 1994a. Conversation as collaboration: Some syntactic evidence. *Speech Communication*, 15: 231-242.

Fais, Laurel. 1994b. Effects of communicative mode on spontaneous English speech. Technical Report of the Institute of Electronics, Information and Communication Engineers, NLC94-22(1994-10):1-8.

Fais, Laurel, Kyung-ho Loken-Kim, and Tsuyoshi Morimoto. (in press) Linguistic and para-linguistic differences between multimodal and telephone-only dialogues in English and Japanese. *J. Acoustical Society of Japan (E)*, (in press).

Fais, Laurel, Kyung-ho Loken-Kim, and Young-Duk Park. 1995. Speakers' responses to requests for repetition in a multimedia cooperative dialogue. In *Proceedings of the International Conference on Cooperative Multimodal Communication*, pages 129-144, Eindhoven, The Netherlands, May.

Garrod, Simon, and Gwyneth Doherty. 1994. Conversation, coordination, and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, 53:181-215.

Giles, Howard, Anthony Mulac, James J. Bradac, and Patricia Johnson. 1987. Speech accommodation theory: The first decade and beyond, in M.L. McLaughlin, editor, *Communication Yearbook 10*, pages 13-48. Sage Publications, London, UK.

Leiser, R.G. 1989. Exploiting convergence to improve natural language understanding. *Interacting with Computers: The Interdisciplinary Journal of Human-Computer Interaction*, 1(3):284-298.

Loken-Kim, Kyung-ho, Fumihiro Yato, Kazuhiko Kurihara, Laurel Fais, and Ryo Furukawa. 1993. ATR Technical Report TR-IT-0018. ATR Interpreting Telecommunications Research Laboratories, Kyoto, Japan.

Mellor, Brian, and Cian O'Connor. 1995. User adaptation to voice input interfaces. In *Proceedings of the ESCA Workshop on Spoken Dialogue Systems*, pages 97-100, Vigsø, Denmark, May-June.

Spitz, Judith. 1991. Collection and analysis of data from real users: Implications for speech recognition/understanding systems. In *Proceedings of the 4th Darpa Workshop on Speech and Natural Language*, pages 164-169, Asilomar, California, USA, February.

Suhm, B., L. Levin, N. Coccaro, J. Carbonell, K. Horiguchi, R. Isotani, A. Lavie, L. Mayfield, C. P. Rosé, C. Van Ess-Dykema, and A. Waibel. 1994. Speech-language integration in a multi-lingual speech translation system. In *Proceedings of the AAAI Workshop on Speech and Language Processing*, Seattle, Washington, USA.

Webber, Bonnie. 1995. Instructing animated agents: Viewing language in behavioral terms. In *Proceedings of the International Conference on Cooperative Multimodal Communication*, pages 5-15, Eindhoven, The Netherlands, May.

Zoltan-Ford, Elizabeth. 1991. How to get people to say and type what computers can understand. *International Journal of Man-Machine Studies*, 34:527-547.