

Adaptive Language Behavior in HCI: How Expectations and Beliefs about a System Affect Users' Word Choice

Jamie Pearson

University of Edinburgh
Edinburgh, EH8 9JZ
Jamie.Pearson@ed.ac.uk

Jiang Hu

Stanford University
Stanford, CA 94305
huj@stanford.edu

Holly P. Branigan

University of Edinburgh
Edinburgh, EH8 9JZ
Holly.Branigan@ed.ac.uk

Martin J. Pickering

University of Edinburgh
Edinburgh, EH8 9JZ
Martin.Pickering@ed.ac.uk

Clifford I. Nass

Stanford University
Stanford, CA 94305
nass@stanford.edu

ABSTRACT

People display adaptive language behaviors in face-to-face conversations, but will computer users do the same during HCI? We report an experiment ($N=20$) demonstrating that users' use of language (in terms of lexical choice) is influenced by their beliefs and expectations about a system: When users believe that the system is unsophisticated and restricted in capability, they adapt their language to match the system's language more than when they believe the system is relatively sophisticated and capable. Moreover, this tendency is based entirely on users' expectations about the system; it is unaffected by the actual behavior that the system exhibits. Our results demonstrate that interface design engenders particular beliefs in users about a system's capabilities, and that these beliefs can determine the extent to which users adapt to the system. We argue that such effects can be leveraged to improve the quality and effectiveness of human-computer interactions.

Author Keywords

HCI, language behavior, adaptation, alignment, interaction technologies, natural language

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): H.5.2 User Interfaces

INTRODUCTION

Research on human speech has demonstrated that conversation partners coordinate their contributions in dialogue [4, 6]. Through the course of a dialogue or series of dialogues, paired speakers tend to express themselves in similar ways, termed *alignment* by some psycholinguists [2, 10]. Researchers also refer to this phenomenon as "accommodation" [13], "entrainment" [6], "convergence" [4], and "adaptation" [1]. Such alignment appears to a

fundamental aspect of successful communication [11]. For example, people engaged in a conversation tend to converge on the same labels for objects [4].

With the emergence and fast development of natural language (NL) interface, dialogues between users and computer interfaces have gradually become a part of everyday life. Consequently, computational linguists and psychologists have explored the dynamics in natural language-based human-computer dialogues [1, 3, 4, 7, 10, 12, 14]. In a seminal Wizard-of-Oz study that simulated NL database inquiries, Brennan [3] compared linguistic behaviors displayed by participants in HCI to those found in computer-mediated communication (CMC). In general, participants' typed questions appeared to be shaped by the Wizard's preceding answers over the course of the conversation. More specifically, participants typed short questions after getting short answers, and long questions after long answers. However, differences between the two conditions were observed. For example, there was significantly less use of pronouns in HCI than in CMC.

More recently, Branigan et al. [2] conducted a series of lab experiments focusing on alignment of syntax and word choice in HCI vs. CMC. Participants were led to believe that they were interacting with a computer interlocutor or with a human one via a computer terminal (i.e., CMC); the actual behavior that they experienced from their interlocutor, however, was always identical. Hence, the tendency to align more with a "computer" interlocutor than with a "human" interlocutor occurred irrespective of the actual behavior that the interlocutor displayed. The researchers found that not only did participants align with the interlocutor in both CMC and HCI, mirroring the interlocutor's choice of words and sentence form, but they did so to a much greater extent in HCI than in CMC. Findings from these studies and those of Brennan's [3] suggest that alignment is sensitive to people's beliefs about what they are interacting with.

Thus one possible explanation for the different alignment behavior observed in HCI vs. CMC is that people believe that computers have a more limited language capacity than humans [2]. If so, we might then expect users to be even more likely to display this adaptive behavior when interacting with a computer that they believe to be of

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2006, April 22–27, 2006, Montréal, Québec, Canada.
Copyright 2006 ACM 1-59593-178-3/06/0004...\$5.00.

limited sophistication than with one that they believe to be of greater sophistication.

To test this prediction, we carried out an experiment using a method similar to those reported in Branigan et al. [2] to explore computer users' word choice in HCI. Because the manipulation through verbal instructions to induce different experimental conditions (i.e., HCI vs. CMC) generated strong effects and explicit labeling of software as "smart" or "dumb" is impractical in the real world, we decided to employ a more subtle way of manipulation in the present experiment, as described in the next section.

Method

In an ostensible picture-naming and -matching game on computer, participants and a computer alternated naming pictures and selecting a picture that matched the name given by their partner. On experimental trials, the computer named a picture that had two acceptable names, one of which is much more common. The computer used either the common or uncommon name.¹ Subsequently the participant had to name the same picture. We examined the extent to which the participant aligned with the computer, i.e., used the same name, even when the name was uncommon and less preferred.

We manipulated the apparent sophistication of the computer by using a start-up screen that makes the computer system appear old-fashioned and unsophisticated ("basic" computer condition) or up-to-date and sophisticated ("advanced" computer condition). The start-up screen for the "basic" condition writes "Basic version," bears a 1987-dated copyright, and has a fictional computer magazine review stressing its limited features but cheap price and value for money (see Figure 1). In the meantime, the start-up screen for the "advanced" condition is marked "Advanced version: Professional edition," bears a current-year copyright, and offers a fictional computer magazine review stressing its expensiveness and its impressive range of features and sophisticated technology (see Figure 2).

However, the actual behavior of the computer was always identical. If beliefs and expectations about a computer's sophistication affect participants' behavior, we would expect greater alignment in the "basic" than in the "advanced" condition, because participants would assume that the basic computer had a limited vocabulary.

Procedure

Twenty native-English-speaking members of the University of Edinburgh community participated in the experiment as paid subjects. They were randomly allocated to the "basic" or "advanced" computer system condition.

¹ Experimental pictures had a preferred name (95% or greater agreement in a pretest) and an equally acceptable but uncommon name, e.g., "bench" vs. "seat."

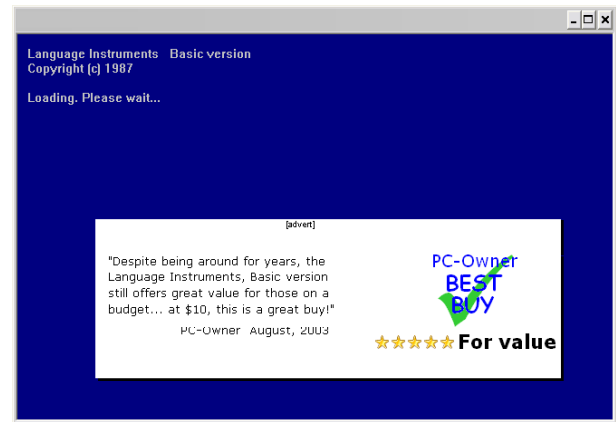


Figure 1. Start-up screen displayed in "basic" condition

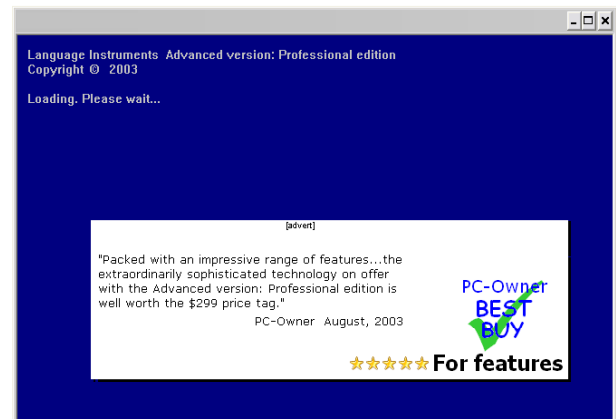


Figure 2. Start-up screen displayed in "advanced" condition

Participants were told that they were to play a picture-naming and -matching game with a computer via a network. They were seated in front of a PC that presented the pictures and names on-screen. Participants and the computer alternated between naming pictures and selecting a picture that matched the name given by their partner.

Prior to any interaction between a participant and the computer, the start-up screen was displayed for 10 seconds while the computer was ostensibly loading up. Half of the participants saw the "basic" start-up screen; the other half was presented with the "advanced" version.

During the experiment, the computer was referred to as "A," and the participant was referred to as "B." On each of the computer's naming turns, which was also the participant's matching turn, two pictures were presented side by side above a textbox marked "A" (see Figure 3). Names produced by the computer appeared after 4000ms, 4500ms, or 5000ms. The participant selected the appropriate matching picture by pressing the key "1" or "2" for the left or the right picture, respectively.

On each of the participant's naming turns, which was also the computer's matching turn, two pictures were presented side by side above a textbox marked "B," one of which was

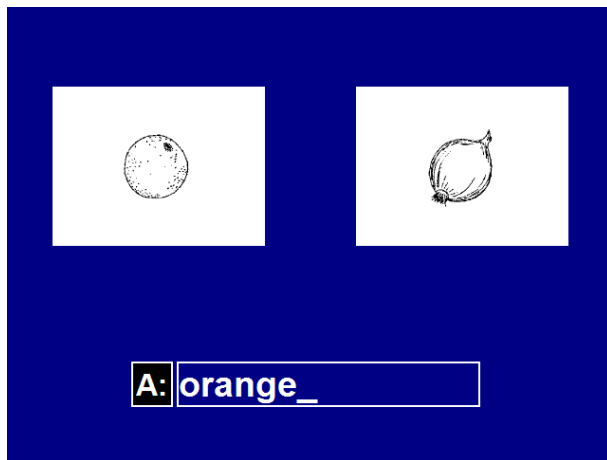


Figure 3. User to select the picture that matches the name provided by computer

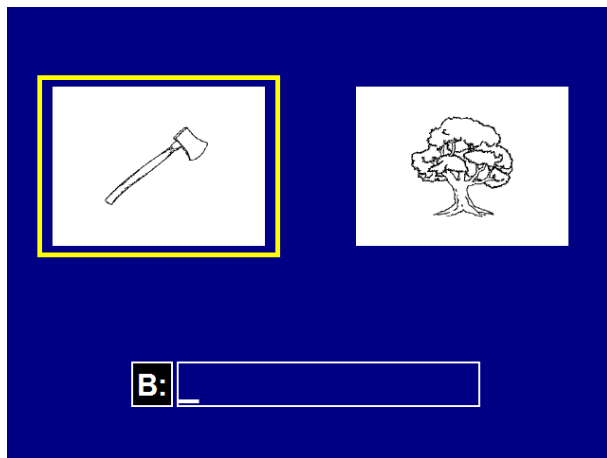


Figure 4. User to type a name of a picture for the computer to match

highlighted after 2000ms with a yellow box surrounding the picture. The highlighted picture could be the left or the right one (see Figure 4). The participant typed in the name of the highlighted picture, and pressed “Enter” to send the name to the computer. Typos could be corrected using “Backspace,” but all keys other than the alphanumeric keys, “Enter” and “Backspace,” were inactive. The participant was told that the computer would use the name given by the participant to select one of two pictures. The picture that the computer had “selected” to match the name given by the participant was indicated to the participant after 1500ms or 2000ms with a red box surrounding the picture. The selected picture was always appropriate.

There were 18 experimental pictures and 268 filler pictures. All experimental pictures have a preferred and a dispreferred names. For example, name pairs included (preferred name first) “bench/seat,” “glasses/spectacles,” “axe/hatchet,” etc. On each experimental trial, the computer used the preferred or dispreferred name; on a subsequent turn, the participant named the same object. We measured

whether the participant aligned with the name previously used by the computer. In both the “basic” and “advanced” conditions, the computer behaved identically.

RESULTS

Responses were coded as *aligned* (i.e., the participant used the same term as the computer had used) or *misaligned* (i.e., the participant used a different term than the computer had used). Cohen’s Kappa was used to calculate alignment. Kappa controls for random agreements, such that kappa is 1 for complete agreement, 0 for identical-to-chance agreement. Kappa was calculated for each participant across all 18 experimental items.

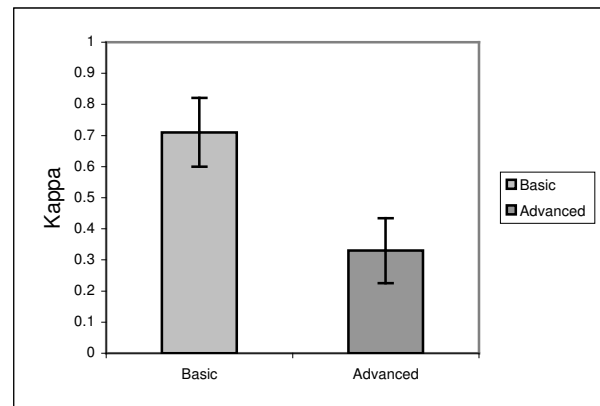


Figure 5. Comparison of alignment

Figure 5 shows the results from the experiment. When chance was controlled for, participants clearly aligned overall with the computer’s choice of name: Alignment was significantly greater than chance in both the “basic” condition, $K(144) = .711$, $ASE = .051$, $p < .001$, and in the “advanced” condition, $K(144) = .333$, $ASE = .061$, $p < .001$.

Consistent with the idea that participants take the computer’s competence into account when selecting words, alignment was significantly greater in the “basic” condition than in the “advanced” condition, $t(18) = 2.33$, $p < .05$. Following the Landis and Koch criteria [8], $Kappa = .711$ (“basic” condition, $SD = 0.35$) is “substantial,” whereas $Kappa = .333$ (“advanced” condition, $SD = 0.37$) is only “fair.”

DISCUSSION

Previous research has demonstrated that people’s beliefs about their interlocutor influence their behavior during interactions: People adapt their use of language more towards their interlocutor’s use of language when they believe that they are interacting with a computer than with another person [2]. The present experiment extends this research in important ways, showing that not all computer systems are treated equally. That is, users make distinctions between systems, based on their expectations about the system’s capabilities. These expectations are in turn based on non-functional aspects of the interface design. When the interface design sets up an expectation that the system is

relatively unsophisticated and limited in capability, users tend to adjust their language behavior to a large degree to adapt to the system. When the interface design suggests that the system is relatively sophisticated and more powerful, in contrast, users also exhibit a willingness to adapt their language behavior to the system, but this tendency is significantly smaller.

Wang and colleagues [15] reported that when a search interface seemed to exhibit capabilities that it did not have (e.g., using full sentences and idioms to produce clarifying questions) when handling recognition and understanding errors, users were more likely to feel frustrated and to rate the interface as inefficient. The present study goes one step further to show that non-behavioral features (i.e., start-up screens) may alter users' expectations and beliefs, and subsequently affect their language use in HCI.

Some researchers have pointed out that high degree of alignment may help reduce recognition errors of NL-based systems, because these systems should be able to recognize user input that is identical or similar to the output generated by the systems [2, 3, 7]. The present study not only provides further support to that design implication, but also suggests that a natural language recognizer with low capability should never be labeled or marketed as one with high capability. Otherwise, unrealistic and high expectations of a low-capability system will significantly lower the degree of alignment, and consequently generate more recognition errors to ruin the user experience.

Beyond natural language, the current research provides systematic evidence that expectations of a system's competence can have broad effects on user responses [10]. This may explain the "Uncanny Valley" [9] and other cases in which systems that appear too human are frustrating. For example, a robot with the head of Einstein (named "Albert Hubo") [5] may lead the user to have overly high expectations of the robot's competence and try to do much more with the robot than it is capable of doing.

CONCLUSION

Users adapt their behavior to fit the system with which they are interacting. This adaptation reflects beliefs about the system's capability, and not its actual behavior. Such beliefs can be established on the basis of superficial and easily-implementable features of the interface. Natural language systems, and systems in general, should not seem smarter than they actually are.

ACKNOWLEDGMENTS

This project is supported by Scottish Enterprise.

REFERENCES

- Bell, L. *Linguistic adaptations in spoken human-computer dialogues: Empirical studies of user behavior*. Doctoral dissertation. Department of Speech, Music and Hearing, KTH, Stockholm. 2003.
- Branigan, H.P., Pickering, M.J., Pearson, J., McLean, J.F., Nass, C.I., and Hu, J. Beliefs about mental states in lexical and syntactic alignment: Evidence from human-computer dialogs. *Proc. CUNY Conference on Human Sentence Processing 2004*.
- Brennan, S.E. Conversation with and through computers. *User Modeling and User-Adapted Interaction, 1* (1991), 67-86.
- Brennan, S.E., and Clark, H.H. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition 22* (1996), 1482-1493.
- Einstein robot the star of high-tech show. <http://www.msnbc.msn.com/id/10085890/from/RL.3/>
- Garrod, S.C. and Anderson, A. Saying what you mean in dialogue: A study in conceptual and semantic coordination. *Cognition, 27* (1987), 181-218.
- Guindon, R., Shuldberg, K. and Conner, J. Grammatical and ungrammatical structures in user-adviser dialogues: Evidence for sufficiency of restricted languages in natural language interfaces to advisory systems. *Proc. 25th Annual Meeting of the ACL, Association of Computational Linguistics* (1987), 41-44.
- Landis, J.R., Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics 33* (1977), 159-174.
- Mori, M. *The Buddha in the robot*. Boston: Tuttle Publishing, 1982.
- Nass, C.I., and Brave, S.B. *Wired for speech: How voice activates and enhances the human-computer relationship*. MIT Press, Cambridge, MA, 2005.
- Pickering, M.J., and Garrod, S. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences 27* (2004), 169-226.
- Richards, M.A. and Underwood, K.M. How should people and computers speak to one another? *Proc. Interact 1984, IFIP* (1984), 33-36.
- Street, R.L. and Brady, R.M. Speech rate acceptance ranges as a function of evaluative domain, listener speech rate, and communication context. *Communication Monographs, 49* (1982), 290-308.
- Walker, M., Fromer, J., Di Frabrizio, G., Mestel, C. and Hindle, D. What can I say?: Evaluating a spoken language interface to email. *Proc. CHI 1998, ACM Press* (1998), 582-589.
- Wang, Q.Y., Nass, C., and Hu, J. Natural language query vs. keyword search: Effects of task complexity on search performance, user perceptions and preferences. *Proc. Interact 2005, IPIF* (2005), 106-116.