(Proceedings, 1996 International Symposium on Spoken Dialogue, ISSD-96. Philadelphia, PA, pp. 41-44. Copyright © 1996 by the Acoustical Society of Japan.)

LEXICAL ENTRAINMENT IN SPONTANEOUS DIALOG

Susan E. Brennan

Department of Psychology
State University of New York
Stony Brook, NY 11794-2500
Email: susan.brennan@sunysb.edu

ABSTRACT

It is often desirable to predict or constrain the lexical choices people make with spoken language systems. I discuss lexical variability in language use; while there is a great deal of variability across conversations, there is little within. In conversation, people achieve *conceptual pacts*, or shared conceptualizations, which they mark by using the same terms; this process of *lexical entrainment* limits and systematizes lexical variability (Brennan & Clark, 1996). In text and speech dialogs with computer systems, people often adopt the systems' terms; this process shares some aspects in common with lexical entrainment, and it differs in some ways as well. After summarizing five experiments on lexical variability with human partners and computers, I present some implications of this work for spoken language systems.

1. THE VOCABULARY PROBLEM

There is potential for enormous variability in people's lexical choices in dialog, whether with human or machine partners. Not only is the mental lexicon large–by some estimates, a high school graduate may know as many as 100,000 words–but the same message can be expressed in an indefinite number of ways. This potential for variability has been dubbed *the vocabulary problem* by Furnas, Landauer, Gomez, and Dumais (1983, 1987) in their studies of command languages. Furnas et al. found that the probability of any two people producing the same term for the same command ranged from only .07 to .18 (Furnas et al., 1987). For instance, to remove a file, people might try such terms as *remove, delete, erase, kill, omit, destroy, lose, change*, and even *trash*. Command language designers have tried to cope with the vocabulary problem by having systems accept synonyms for commands (see, for instance, Good, Whiteside, Wixon, & Jones, 1984). But as Furnas et al. (1987) discovered in their experiments, even with as many as 20 synonyms for a single command, the likelihood that two people would choose the same term for that command was only about 80% (Furnas et al. 1987). Furthermore, synonyms lead to additional problems: In an application with only 25 commands, the likelihood that two people who chose the same term meant the same command by it was only 15%.

For natural language interfaces without explicit constraints on the syntax of input utterances, there are many more choices. One group of computational linguists working on a natural language interface to a database query application listed 7000 distinct ways to ask about *programmers who work for department managers*, using common words and syntax (Brennan, 1990). And for speech interfaces, the vocabulary problem can be worse still. Speech interfaces are often designed to accept a limited vocabulary because both perplexity and processing time increase as the search space increases. Under these circumstances, what people can say at each point in the dialog is constrained. There are two common ways to do this: Some systems offer a small and explicit menu of choices, such as "yes" or "no" for people to accept or decline a collect call. Others place the burden on speakers to remember which terms are acceptable. The first of these approaches works only for highly scripted interactions under the system's control, while the second approach may work for expert users but not for casual ones. If our eventual goal is to provide people with spoken language interfaces that are both natural and robust, then we need a new approach. How do people solve the vocabulary problem when they talk with other *people*? To address this question, I will consider what we have learned about how people choose their terms while referring to objects in conversation.

2. LEXICAL ENTRAINMENT AND CONCEPTUAL PACTS IN CONVERSATION

Lexical variability can be just as high in human conversation as in human-computer dialog. In one of our referential communication studies, the likelihood that people in one conversation would choose the same terms for the same common object as people in another conversation was only 10% (Brennan & Clark, 1996). But while variability is high between conversations, it is relatively low *within* a conversation (Brennan & Clark, 1996; Garrod & Anderson, 1987). When two people repeatedly discuss the same object, they come to use the same terms. This phenomenon, called *lexical entrainment*, has several possible explanations.

An *ahistorical* explanation considers each choice of terms as an independent event, affected by informativeness, the availability of the terms, and the salience of the referent's features. The simplest ahistorical explanation is that speakers design their referring expressions to distinguish a referent from a set of alternatives (Brown, 1958; Olson, 1970) and that they choose the most concise expression that will enable their addressees to do so (Grice, 1975). So when speakers want to refer to an Irish Setter in a set that also contains a beagle and a poodle, they shouldn't use "dog," which isn't informative enough, or "the big red Irish Setter with the tongue hanging out," which is over-informative, but rather "the Irish Setter." In addition, availability (or ease of lexical access) plays a role; the most available labels are usually basic level nouns–"dog," for example, as opposed to "animal" or "Irish Setter" (Cruse, 1977; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). So when the set consists of the Irish Setter, a bicycle, and a hammer, then speakers should be likely to use "dog," a common basic level term (as opposed to "animal," which is informative enough). This account of lexical choice was presented by Cruse (1977), who proposed that the most common term that is informative enough is considered unmarked; all other terms are marked. An ahistorical, frequency-based explanation underlies the expectation that people in dialogs with computers will limit their choices to the most common lexical items (e.g., Guindon, 1991). But Furnas et al's. (1987) results suggest this expectation is unjustified.

Historical explanations of lexical entrainment, on the other hand, appeal to past references within a conversation and to partner-specific conceptualizations that two people achieve interactively. We take the view that referring is a collaborative process (Clark & Wilkes-Gibbs, 1986). When a speaker labels an object, she is proposing a conceptualization of it, a proposal her addressee may or may not agree to. So her referring expression is provisional until accepted or modified by her partner. This view leads to several predictions: (1) once people establish a shared conceptualization or *conceptual pact*, they appeal to it in later references even when they could use simpler or less informative references; (2) speakers rely on conceptual pacts more often the more firmly they have been established, and (3) conceptual pacts are based not only on a speaker's own history, but on history *with a particular addressee*.

In a series of experiments, we tested these predictions. Pairs of people who were visually separated conversed in order to line up identical sets of picture cards in the same order. The target cards (the items of interest to us) were a particular shoe, dog, car, and fish. In the *A* trials, the targets were all unique in their basic level categories, and so we expected that many people would choose basic level terms. In the *B* or non-unique trials, people matched a different set of cards that included the same four targets, but also additional shoes, dogs, cars, and fish; we expected that they would modify their labels to be more informative. In the *C* trials, they returned to matching the same set of (unique) cards as in the A trials. The trial of particular interest was the first C trial, in which speakers *could* return to using basic level terms (which were informative enough, as well as shorter than the more informative terms from the B trials). This is what an ahistorical view of lexical choice predicts. A historical view, on the other hand, predicts that in the first C trial, people should continue to use the same or similar terms as in the B trials (and not revert to basic level terms). This is what we found in two experiments; 46-52% of the time, people used the same terms in the first C trial as in the B trials, and when terms were not precisely the same, they were strongly related (our coding system used strict criteria for equivalence and counted "the big red dog with the tongue hanging out" as not the same as "the big red dog with the tongue"). On average there were 11 references to other objects between each reference to a particular target item (including some very complex references to abstract geometric objects that served as distracters) and so 46-52% represents a substantial degree of lexical entrainment.

As for our second prediction, the likelihood of lexical entrainment in the first C trial was strongly affected by number of B trials (either one or four) that people had experienced; people were less likely to revert to basic level terms like "dog" and "fish," and more likely to rely on conceptual pacts, the more firmly these pacts had been established. In our third experiment, people matched four B trials (with the non-unique cards) and then did four C trials (with the same targets, but unique in their basic level categories) either with the same partner or with an entirely new partner (who had never matched the cards before). If people were really forming conceptual pacts with specific partners, they should use the same terms in the first C trial as in the last B trial more often when continuing to match cards with the same partner than when matching cards with a new partner. As predicted, the results showed a partner-specific effect; speakers with continuing addressees appealed to conceptual pacts they had already established, even when these were over-informative, while speakers with new addressees were more likely to switch to basic level terms in the first or subsequent C trials.

Partner-specific effects have two possible causes. First, speakers may tag a conceptual pact such as *the fish with different colors* as shared with a particular partner (Clark & Marshall, 1978). So when speakers continue with the same partner, the identity of the partner serves as an additional memory cue for the old conceptualization. When people meet a new partner, that cue is missing, so they should be more likely to abandon the precedent and offer simpler, basic level terms. A second possibility is that new partners who haven't yet formed a conceptual pact with anyone should expect the basic level term "fish" on the first C trial. If the more experienced partner doesn't use "fish," but proposes a more informative conceptualization, then the new partner may encourage her to use the basic level term, as in this example (in which pairs of asterisks denote overlapping speech):

Partner 1: it's a fish with *different colors*
Partner 2 *yeah* okay

Partner 1 offered an expression that was over-informative, while Partner 2 interrupted with *yeah*, indicating that the basic level term was sufficient. Thereafter, this pair used "fish." Examples like this show that a term can emerge from the conceptual coordination of two people interacting.

Conceptual pacts are flexible and temporary agreements to conceptualize an object in a particular way. Over repeated referring, the expressions that mark a pact are often simplified: e.g., "the rainbow fish with the curved back" may be shortened to "the rainbow fish." It appears that when speakers shorten referring expressions in this way, they are marking the same or related conceptualizations; in our experiments, pairs who had included the basic level term in their more specific B terms (e.g., "the rainbow fish") were significantly more likely to revert to the unadorned basic level term ("fish") at some point during the C (unique) trials than were pairs who had not included the basic level terms in their B terms (e.g., "the rainbow trout"). Finally, conceptual pacts are flexible in the additional sense that speakers can easily abandon them when necessary. A full report of our findings appears in Brennan & Clark (1996).

3. LEXICAL CONVERGENCE WITH COMPUTERS

Do people using speech or natural language interfaces adopt the terms used by the computer systems they interact with? We have found that people are at least as likely to adopt the terms of their computer partners as those of their human partners (Brennan, Ries, Rubman, & Lee, 1996. In two Wizard-of-Oz experiments using a database query task, one simulating a text-based natural language interface and the other simulating a speech recognition interface (with synthesized speech output), the system sometimes responded with different terms than those proposed by users, as here:

User:	what college does Aida attend?
System:	the school Aida attends is Williams.

Users then had three more opportunities to refer to the same items, and we coded whether they adopted the system's terms. Conversation analysts have proposed that when two people in conversation use different terms to refer to the same item, either explicitly or implicitly, one is attempting to correct the other's perspective (Jefferson, 1982). The previous example shows what Jefferson would label an *embedded correction*, while the next example shows an *exposed correction* (within a helpful error message):

User:	what college does Aida attend?
System:	by college, do you mean school?
User:	yes
System	the school Aida attends is Williams.

We hypothesized that people would be more likely to adopt the system's term after an exposed correction than after an embedded one. We also expected that people would often adopt the system's term after an embedded correction as well (even though the system had "understood" their original term and answered their question). In both experiments (text and speech), the system's answer contained the user's own terms one third of the time (control items), different terms without comment one third of the time (embedded corrections), and different terms within a helpful error message one third of the time (exposed corrections).

A second variable we investigated was the extent to which memory for the partner's contrasting term would play a role in lexical convergence. We controlled whether the opportunity for re-referring took place immediately or after a delay consisting of references to several other objects. This we did by having users ask questions about the missing information in a small database depicted by a spreadsheet, following either a vertical or a horizontal order over the empty cells. The database concerned some hypothetical people and their attributes; the attributes were unlabeled, since we wanted users to infer them from the values present in the spreadsheet and then generate their own terms. If people adopted the system's embedded term only in the immediate memory condition, this would be evidence that lexical convergence is an automatic process probably driven by priming. If they frequently adopted the system's embedded term in the delayed memory condition as well, this would be evidence that lexical convergence is a more strategic process.

Results. The general pattern of results was the same for both text and speech interfaces (see Figures 1 and 2). As we expected, people were more likely to adopt the system's term at the first opportunity after an exposed correction than after an embedded correction (respectively, 94 to 37% for text and 88 to 58% for speech). That they adopted the system's term as often as they did in the embedded condition is noteworthy, since the system had interpreted their original term without providing an error message, and so they were under no obligation to abandon their own term.

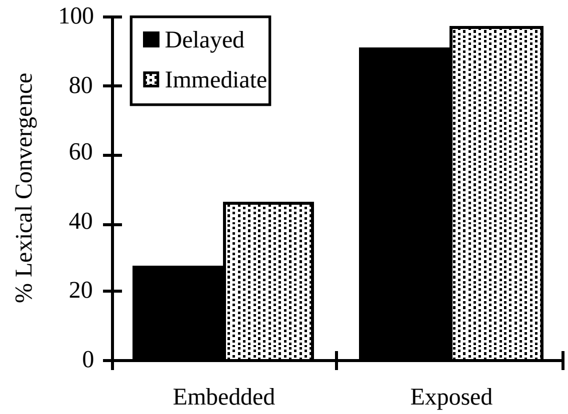

Figure 1: Lexical convergence with a text interface. *Delayed* and *Immediate* are the two memory conditions; *Embedded* and *Exposed* are the ways in which the system introduced a term that differed from the user's term.

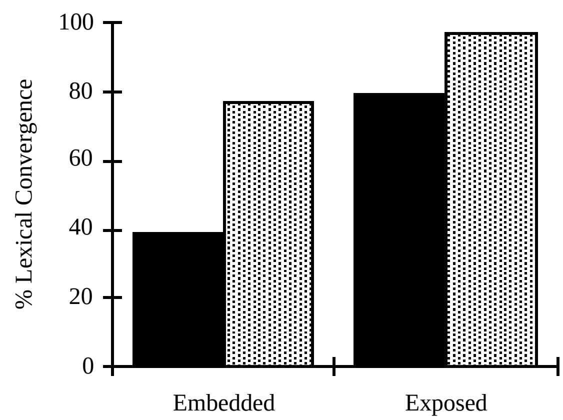

Figure 2: Lexical convergence with a speech interface. The memory conditions are the same as for Figure 1.

Memory played a role in lexical convergence as well; people were more likely to adopt the system's term when they re-referred to the same item immediately than when they did so later (72 to 59% for text and 87 to 59% for speech). While these overall means are similar for text and speech interfaces, visual comparison of Figures 1 and 2 shows that when the embedded condition is considered alone, memory made twice as much of a difference for speech as for text (38 to 19%). We must be cautious in comparing the text and speech results, since they come from two different experiments; however, the database query tasks were nearly identical and the participants were drawn from the same population of students at the State University of New York at Stony Brook. A full report of these two experiments is available in Brennan et al. (1996).

The embedded condition with a speech interface is comparable to the human conversation experiments from the previous section, and the 58% likelihood of lexical convergence with a computer partner in this condition is comparable to the 46-52% likelihood of lexical entrainment with a human partner. People appear to be at least as likely to adopt the terms of computer partners as the terms of human partners. The question is whether they do so for the same reasons. I have used "lexical convergence" as analogous to, but still distinct from, "lexical entrainment." These different terms allow for the possibility that adopting a person's term may not be a product of the same process as adopting a system's term; most systems are not in any position to negotiate, and most users are probably aware of this. Ohaeri (1995) has proposed that when people adopt a system's terms, they do so to avoid errors, expecting the system to be inflexible.

4. IMPLICATIONS

The studies described here have several implications for modeling and constraining lexical variability in spontaneous human-computer dialogs (discussed in more detail in Brennan et al., 1996). First, a spoken language system should present as output only those terms that it can process as input. People are likely to model their lexical choices on the utterances a system presents (and on its syntax as well–see Brennan, 1991), so systems' choices will act to constrain users' choices. Second, the terms a system uses should be consistent; this includes not only terms in its output messages, but also those in its documentation. Third, repairs will be necessary in any dialog system that has real users. Spoken language systems should be supported by an architecture that enables feedback and negotiation with users. Finally, the results from the studies in Section 2 suggest several strategies that could be tried–on an experimental basis–to help cope with the vocabulary problem. A speech recognizer could begin a dialog with a large vocabulary, in order to allow a speaker to propose terms. The system would need to be able to clarify and ground these terms with the user, to ensure that they were both mapping them onto compatible conceptualizations. Then a discourse model of currently active conceptualizations and terms could be used to rapidly narrow down the vocabulary the system expected as the dialog proceeded. This model should allow for the shortening of expressions, as this may happen upon re-referring. The system should maintain a record of conceptualizations and terms used with a particular person, to evoke when that person begins a new dialog. These strategies are based on human conversation, and so there is some risk to applying them to human-computer dialog; they may succeed, or not.

ACKNOWLEDGEMENTS

I thank my collaborators: Herbert Clark, Eric Hulteen, Greg Lee, Justina Ohaeri, Pamela Reis, Claire Rubman, and Michael Schober. The research reported here was supported by National Science Foundation Grant IRI9202458 from the Interactive Systems Program and by Apple Computer, Inc.

REFERENCES

1. Brennan, S. E. Conversation as direct manipulation: An iconoclastic view. In B. K. Laurel (Ed.), *The art of human-computer interface design*, Addison-Wesley, Reading, MA, 1990.
2. Brennan, S. E. Conversation with and through computers, *User Modeling and User-Adapted Interaction*, **1**: 67-86, 1991.
3. Brennan, S. E., and Clark, H. H. Lexical choice and conceptual pacts in conversation, *Journal of Experimental Psychology: Learning, Memory, and Cognition* (in press), 1996.
4. Brennan, S. E., Ries, P. S., Rubman, C., and Lee, G. The vocabulary problem in speech and language interfaces**,** manuscript, 1996.
5. Brown, R. *Words and things*. The Free Press, Glencoe, IL, 1958.
6. Clark, H. H., and Marshall, C. Reference Diaries. In D. L. Waltz (Ed.), *Theoretical issues in natural language processing, TINLAP-2* (pp. 57-63), ACM: New York, 1978.
7. Clark, H. H., and Wilkes-Gibbs, D. (1986). Referring as a collaborative process, *Cognition*, **22**: 1-39.
8. Cruse, D. A. The pragmatics of lexical specificity, *Journal of Linguistics*, **13**: 153-164, 1977.
9. Furnas, G. W., Landauer, T. K., Gomez, L. M., and Dumais, S. T. Statistical semantics: Analysis of the potential performance of keyword information systems, *Bell Systems Technical Journal*, **62**: 1753-1806, 1983.
10. Furnas, G. W., Landauer, T. K., Gomez, L. M., and Dumais, S. T. The vocabulary problem in human-system communications, *Communications of the ACM*, **30**: 964-971, 1987.
11. Garrod, S. and Anderson, A. Saying what you mean in dialog: A study in conceptual and semantic co-ordination, *Cognition*, **27**: 181-218, 1987.
12. Grice, H. P. Logic and conversation (from the William James lectures, Harvard University, 1967). In P. Cole, & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41-58), Academic Press, New York, 1975.
13. Guindon, R.. Users request help from advisory systems with simple and restricted language: Effects of real-time constraints and limited shared context, *Human-Computer Interaction*, **6**: 47-75, 1991.
14. Jefferson, G. On exposed and embedded correction in conversation, *Studium Linguistik,* **14**: 58-68, 1982.
15. Ohaeri, J. O. Lexical convergence with human and computer Partners: Same cognitive process? Master's thesis, Department of Psychology, SUNY, Stony Brook, NY, 1995.
16. Olson, D. R. Language and thought: Aspects of a cognitive theory of semantics, *Psychological Review*, **77**, 257-273, 1970.
17. Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. Basic objects in natural categories, *Cognitive Psychology*, **8**: 382-439, 1976.