Paralinguistic correlates of conceptual structure

DALE J. BARR
University of California, Riverside, California

How is conceptual knowledge transmitted during conversation? When a speaker refers to an object, the name that the speaker chooses conveys information about category *identity*. In addition, I propose that a speaker's confidence in a classification can convey information about category *structure*. Because atypical instances of a category are more difficult to classify than typical instances, when speakers refer to these instances their lack of confidence will manifest itself "paralinguistically"—that is, in the form of hesitations, filled pauses, or rising prosody. These features can help listeners learn by enabling them to differentiate good from bad examples of a category. So that this hypothesis could be evaluated, in a category learning experiment participants learned a set of novel colors from a speaker. When the speaker's paralinguistically expressed confidence was consistent with the underlying category structure, learners acquired the categories more rapidly and showed better category differentiation from the earliest moments of learning. These findings have important implications for theories of conversational coordination and language learning.

When we talk, we reveal something about ourselves not only by *what* we say, but also by *how* we say it. The idea that our linguistic behavior, especially our errors and disfluencies, can illuminate the structure of the mind goes back to Freud (1914/1966). In modern psycholinguistic investigation, researchers have found speech errors, disfluencies, and hesitations to be informative about the basic units of speech production (Butterworth, 1980; Dell & Reich, 1981; Goldman-Eisler, 1968; Levelt, 1983; Maclay & Osgood, 1959). The idea that these acoustic features of spontaneous speech have theoretical value presupposes a systematic correspondence with the underlying structure of thought.

Although the problems that speakers experience in producing fluent speech typically have some acoustic manifestation, this need not perturb the linguistic content of speech. A speaker might hesitate for long periods, produce "filled pauses" such as *um* or *uh*, or vary the pitch or the loudness of his or her speech, while still producing a linguistically grammatical and coherent utterance. In other words, acoustic features such as these lie "around the edge of language" (Bolinger, 1964) and therefore might be characterized as *paralinguistic*.

The fact that theorists have exploited paralinguistic dimensions of spontaneous speech to inform theories of language production suggests that perhaps listeners could also find them useful in solving the more everyday problems of language understanding. Recently, research has begun to address this question. Such studies have revealed that speakers paralinguistically signal imminent delays or disfluencies (Fox Tree, 2001; Fox Tree & Clark, 1997) and that the form or timing of such disfluencies can guide listeners in how they integrate the disfluent speech with surrounding speech (Brennan & Schober, 2001; Fox Tree, 2001).

Although paralinguistic information also correlates with epistemic dimensions of thought, comparatively little research has addressed the sensitivity of listeners to these dimensions. Smith and Clark (1993) observed that paralinguistic features correlate strongly with a speaker's "feeling of knowing" or degree of certainty in a proposition. The less confident the speaker was, the longer the length of the pre-speech delay, the higher the probability of a filled pause, and the higher the probability of rising versus neutral intonation. Brennan and Williams (1995) showed that listeners use these cues to make judgments about a speaker's knowledgeability. Similarly, Sabbagh and Baldwin (2001) found effects of speaker confidence on word learning. Children were more likely to learn novel words from speakers whose behavior suggested that they were knowledgeable rather than ignorant, though in their study, degree of certainty was conveyed verbally as well as paralinguistically.

Given that speakers provide paralinguistic information about certainty and listeners are sensitive to it, it is important to understand how this sensitivity might benefit listeners in conversation. One domain where this information would be especially useful is referential communication. Reference is fundamentally an act of classification (Brown, 1958). Because of the graded structure of most everyday categories, ease of classification depends on an instance's typicality (Rosch & Mervis, 1975). For instance, if someone pointed to a patch of color and asked, *What color is*

this? one would reply more quickly and with a more confident intonation if the color in question were focal blue than if it were some less typical shade. While the label *blue* would inform the listener about the classification of the referent, the paralinguistic information would potentially reveal how typical this particular instance was of the corresponding category.

In category research, it has been established that people learn categories better when given typical rather than atypical instances (Mervis & Pani, 1980; Posner & Keele, 1968). I am proposing that paralinguistic information can help listeners differentiate good from bad examples so that they can rely more on the good ones as they learn. The main question is, Do listeners take advantage of this information in order to learn concepts?

In the following experiment, listeners learned a set of novel color categories from an expert teacher. For half of the categories, the paralinguistic characteristics of the teacher's utterances were consistent with the category structure: difficult-to-classify instances (i.e., atypical) were marked by unconfident-sounding utterances, and easy-to-classify instances (i.e., typical), by confident-sounding ones. For the other half, the characteristics were inconsistent with the category structure. If listeners take advantage of paralinguistic cues, they should acquire the categories faster when the information is consistent than when it is inconsistent.

METHOD

Participants

Forty-four undergraduates participated in the study for course credit or payment. Eight were excluded because they had guessed the purpose of the experiment, leaving a total of 36 (22 females and 14 males).

Procedure

Participants were asked to imagine that they were new employees in the stockroom of an Internet clothing store, where their job was to fill orders. For instance, if they received an order for a "scarf in *riallo*," they would need to know how to pick the correctly colored scarf. A male "expert teacher" would train them on the store's color scheme. They were told that he was a previous participant in the experiment who had been taught the color categories, and that he would produce no erroneous classifications.

There were six *color sets* to be learned. Each set consisted of 3 color categories, for a total of 18 categories in all. The 3 categories in each set were all from the same range of the color spectrum (e.g.,

one set contained 3 reddish colors, another 3 bluish colors, etc.). Each category, in turn, contained five instances that varied in shade (though all of them were referred to by the same name, much as there can be many different shades of the color blue).

The participants learned these color sets one at a time. For each set, the learning procedure consisted of a study phase and a test phase. In the study phase, the 15 instances were presented 1 at a time in a random order. The participants saw a circular patch approximately 7 cm in diameter appear on a computer screen and heard the teacher's pre-recorded classification. Then the participants rated the typicality of the instance on a seven-point scale: (1) *very bad*, (2) *bad*, (3) *somewhat bad*, (4) *fair*, (5) *somewhat good*, (6) *good*, (7) *extremely good*.

After the participants had studied all 15 instances, the test phase began. In this phase, they saw the instances one at a time in a random order and attempted to classify each one by selecting the color name from a pop-up window. They did not receive feedback. Before they could continue on to the study phase for the next color set, they had to meet a performance criterion, which required them to classify at least 4 out of the first 6 test instances correctly. If they failed, the test phase was ended immediately and they had to repeat the study phase for the current set. If they succeeded, they finished classifying all of the instances for that set and then continued on to the study phase for the next set. For each color set, the computer recorded the number of study phases, or "blocks" required for the participant to meet the criterion. This was the principal measure of interest.

A two-factor mixed design was used. The main independent variable was *consistency*, which was a two-level, within-participants variable. Three color sets appeared in the *consistent* condition; the other three appeared in the *inconsistent* condition (details below). A second, between-participants factor, *block order*, corresponded to whether participants completed the three consistent blocks first, followed by the inconsistent sets, or vice versa.

Materials

Eighteen novel color names were invented that had no clear preexisting associations with specific colors (e.g., *Murray*, *Alba*, *Riallo*). Each name defined a category that included five instances varying in shade. None of these categories directly corresponded to typical colors such as blue, green, yellow, and so forth. Though each category spanned a unique portion of color space, the three colors in each color set were similar in that they were adjacent in the space. Table 1 provides an example of a single color set, with the color vector provided for each instance. This vector corresponds to standard RGB values (which range from 0 to 255). Note that Murray-5 is similar in shade to Alba-1, and Alba-5 is similar in shade to Riallo-1. Because of this intercategory similarity, these instances would be most difficult to classify. In contrast, Murray-1, Alba-3, and Riallo-5 should be easiest, because they are farthest away from boundaries with adjacent categories.

The voice used for the teacher's sound recordings was that of a male confederate who was instructed to speak each color name five times with confidence that varied from very unsure to very sure. The

Table 1
Paralinguistic Features of Acoustic Stimuli (Consistent Condition)

	Outer Category 1 Murray					Inner Category Alba					Outer Category 2 Riallo				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Red (0–255)	125	131	136	142	147	147	162	176	191	205	215	224	233	241	250
Green (0–255)	45	34	23	11	0	0	0	0	0	0	10	27	44	61	78
Blue (0–255)	61	66	71	76	81	99	101	103	104	106	106	111	115	120	124
Onset (msec)	335	782	924	1,696	3,001	2,671	1,194	632	984	2,770	3,076	1,416	988	917	229
Filled pauses	0%	17%	33%	83%	100%	100%	33%	0%	17%	100%	83%	50%	33%	0%	0%
Rising intonation	0%	33%	17%	83%	100%	100%	83%	0%	50%	67%	100%	83%	50%	33%	0%

Note—Murray, Alba, and Riallo are sample names. 1–5 are instances of the category represented by the sample name.
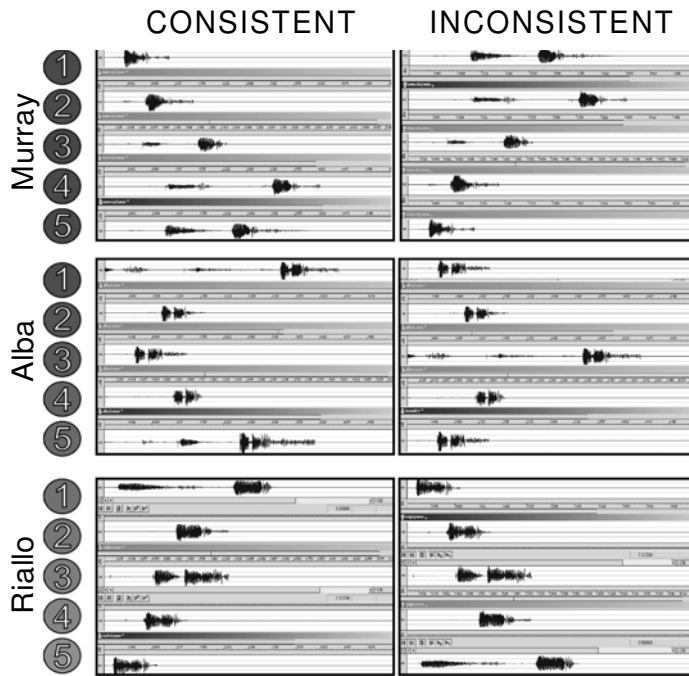
Figure 1. Example of consistent/inconsistent mapping of audio stimuli to category instances.

paralinguistic characteristics are given in Table 1 (for the consistent condition) averaged over color set. On average, the onset of the category name, measured from the beginning of the sound file, occurred at 400 msec for the most confident utterances and 2,879 msec for the least confident ones. In addition, 95% of the least confident utterances began with a filled pause (e.g., *uh*, *um*, *mm*), whereas none of the most confident ones did. Ninety-two percent of the least confident utterances also included rising intonation, whereas none of the most confident utterances did.

The teacher sound files were matched up with category instances in two different manners to form the consistent and inconsistent conditions (see Figure 1 for an example of the pairings for one color set). In the consistent condition, the acoustic characteristics were matched to correspond with the assumed ease of classification. As can be seen in the figure, the onset of the category name correlated with the instance's distance from the boundary with a neighboring category. Thus, Murray-1 was paired with a confident-sounding utterance that contained a word onset latency of approximately 370 msec and descending intonation, whereas Murray-5 was paired with an unconfident-sounding utterance that had a latency of 2,375 msec and included a filled pause. In the inconsistent condition, this pattern was reversed. Note that the materials used in the inconsistent condition are nearly identical; the main difference is the mapping between sounds and instances.[1] Thus, any effect would most likely be due to the difference in the mapping rather than the stimuli themselves.

Six different color set orders were crossed with the two block orders to generate 12 lists. The materials were counterbalanced across participants so that each color set appeared an equal number of times in both conditions.

RESULTS

If people learn categories only through the verbal labels that communicate category *identity*, there should have been no difference between the consistent and inconsistent conditions. On the other hand, if people pay attention to the paralinguistic cues that correlate with internal category *structure*, this should have enhanced learning in the consistent condition. The results, presented in Table 2, support the latter view. Overall, learners acquired the color categories 16% faster in the consistent condition. A paired samples *t* test on the overall means revealed a reliable difference [$t(35) = 2.00$, $p = .05$]. These findings provide clear support for the hypothesis that listeners exploit certainty cues when learning concepts.

It is possible to rule out certain artifactual explanations. In a postexperiment questionnaire, 32 out of the 36 participants agreed that the descriptions sounded natural, often citing the filled pauses and hesitations as justification for their belief. Furthermore, it is not possible that listeners learned the relation between certainty and category structure over the course of the experiment. If this had been the case, the half of the participants who completed the first three color sets in the inconsistent condition should have learned the wrong contingencies, leading to worse

Table 2
Blocks to Learn by Block Order and Consistency

Block Order	Consistency	M	SD	Median	Mode
Overall (collapsed)	Consistent	1.62	.61	1.33	1.00
	Inconsistent	1.92	.71	1.67	1.67
Consistent first	Consistent	1.74	.65	1.67	1.33
	Inconsistent	1.98	.78	1.83	1.83
Inconsistent first	Consistent	1.50	.56	1.33	1.00
	Inconsistent	1.85	.65	1.67	1.67

performance in the consistent condition. However, learners who completed the inconsistent block before the consistent block acquired the categories 19% faster in the consistent condition (see Table 2). There was no main effect of block order [$F(1,34) = 1.29, p > .2$], nor did that factor interact with consistency [$F(1,34) < 1$].

Was the observed learning effect general or localized to atypical instances? Perhaps hearing an unconfident classification led listeners to pay more attention while encoding that instance. This extra attention might have produced better learning in the consistent condition, where the unconfident-sounding utterances were paired with the harder-to-classify atypical instances. If atypical instances provided the greatest opportunity for learning, the difference in accuracy between the two conditions should have been greater for these than for typical instances.

To test this possibility, the proportion of correct classifications was computed for the most typical and atypical instances of each color set. The typical instances included those that were the farthest away from a category boundary (e.g., Murray-1, Alba-3, and Riallo-5), whereas atypical instances included those that were on the boundary (e.g., Murray-5, Alba-1, Alba-3, and Riallo-1). The arcsine transformed proportions were submitted to an analysis of variance with two within-participants factors, consistency and typicality. There was a marginally reliable difference in accuracy between the two conditions [mean accuracy, consistent vs. inconsistent: .73 vs. .69; $F(1,35) = 3.60, p < .07$] and a main effect of typicality [typical vs. atypical: .87 vs. .55; $F(1,35) = 209.46, p < .01$]. However, the factor of consistency did not interact with typicality [$F(1,35) < 1$]. In short, paralinguistic information did not

simply call attention to atypical instances, but yielded better overall accuracy in classification.

Further insight into learning can be gained by examining listeners' typicality judgments during the study trials. Learners could rely on two sources of information when making their judgments. As in standard category learning tasks, learners could determine typicality by comparing the perceptual features of the current instance to stored category information. In this experiment, however, they could also rely on the speaker's level of certainty about the classification. How might listeners combine these two sources of information?

In considering the typicality judgments, only the ratings collected during the first study phase of each color set are included, since learners acquired the categories at different speeds in the consistent and inconsistent conditions. Thus, these data represent learners' typicality judgments from their very first exposure to the instances. To reduce variability caused by differences in how individual learners anchored their responses to the response scale, the ratings for each learner were converted into z scores, using the mean and standard deviation for that learner.[2] Figure 2 presents the results, with a sample color set on the x-axis (although the data are averaged over all six color sets). The overall similarity of the two curves suggests that learners very strongly weighted stored category information. However, certain differences, such as the flatter contours of the outer categories in the inconsistent condition, suggest that paralinguistic information may have modulated the learners' responses.

If this is correct, one would expect greater intracategory differentiation of instances in the consistent condition. In
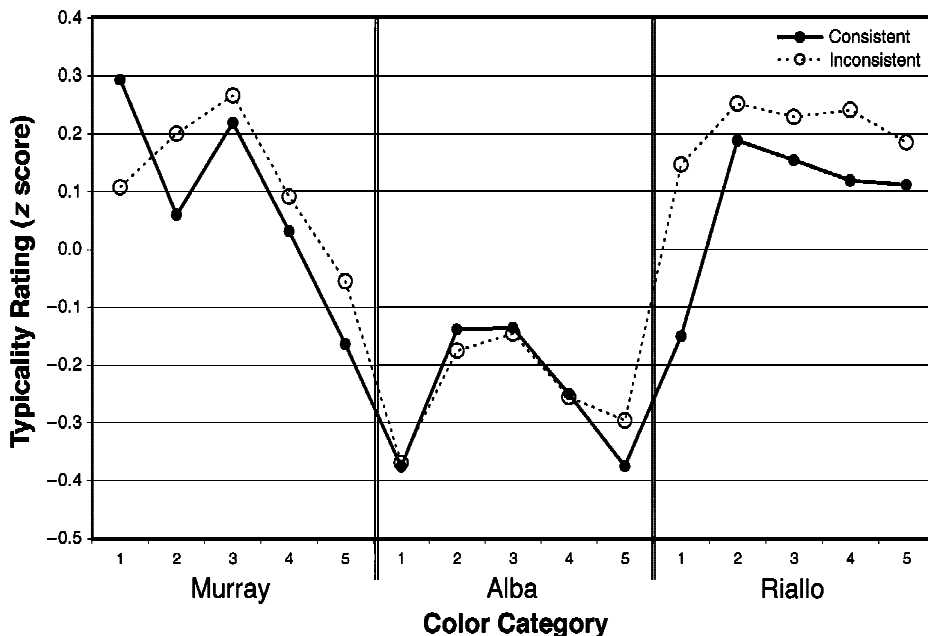

Figure 2. Typicality ratings in study phase, first block only.

other words, the difference between the most typical and least typical instances of each category should be larger in the consistent than in the inconsistent condition. This is because in the latter condition, the inconsistent paralinguistic information should modulate the rating for the most typical instance downward and the rating for the least typical instance upward. To test this hypothesis, a category differentiation score was computed on the standardized typicality ratings. This score represents the difference in rating between the most typical and the least typical instances for each category, defined in the same manner as for the accuracy analysis. The mean differentiation scores were .32 ($SD = .38$) for the consistent condition and .13 ($SD = .43$) for the inconsistent condition. This difference was reliable [$t(35) = 2.24, p < .05$]. If one includes only the ratings for the outer category, where the stimuli were identical across conditions, and excludes the inner category, where the stimuli were slightly different (see note 1), the effect is still reliable [$t(35) = 2.58, p < .05$], ruling out this artifactual explanation for the difference between the conditions. The finding that paralinguistic information modulated learners' typicality judgments in the predicted direction provides further support for the main hypothesis that learners take speakers' paralinguistically expressed confidence into account when learning concepts.

DISCUSSION

The results of this experiment support the view that language users can transmit conceptual knowledge paralinguistically. When listeners hear a speaker classify a referent, they pay attention not only to what is said, but to *how* it is said. An experiment supported the view that listeners can exploit paralinguistic cues to learn about category structure. Learners who heard cues that were consistent with the structure of the underlying categories learned faster, and showed better within-category differentiation of instances, than did learners who heard inconsistent cues. Because the materials used in the two conditions were nearly identical, it is very unlikely that these effects could have been produced by anything other than the difference in how the spoken tokens were mapped onto instances.

Given that these effects were found in an experimental setting, it seems important to ask whether listeners would have opportunities to deploy this knowledge in more natural conversational settings. In conversation, one likely situation occurs when speakers introduce referents into discourse. When introducing a referent, speakers must not only determine the proper classification of the object in order to choose a referring expression, but also make sure that the resulting utterance conforms to the conventions of cooperative communication (Grice, 1975). When a referent is atypical of some category, speakers will experience a clash between the desire to be informative and the desire to avoid asserting things for which they lack sufficient evidence. Under such circumstances they will perhaps inevitably emit some of the side effects or "symptoms" of uncertainty, such as producing a long hesitation, and thereby implicate the atypicality of the referent.

Not only will speakers indicate their level of certainty symptomatically but they often will explicitly communicate this information to the listener during the process of *grounding* (Brennan & Clark, 1996; Clark & Brennan, 1991; Clark & Wilkes-Gibbs, 1986). In grounding, speakers and listeners collaborate to determine mutually acceptable ways of describing referents. A speaker in doubt about whether a particular expression is appropriate will seek to explicitly signal this doubt to the listener. One may do so verbally, using "hedges" (e.g., *it's sort of like an X* or *I think it's an X*) or paralinguistically, using, for instance, rising intonation (Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Sacks & Schelgloff, 1979). One function of these signals may be to invite the listener to participate in the naming process. Under this view many of the paralinguistic cues investigated in this study, such as filled pauses or rising intonation, are really metacommunicative signals that interlocutors use to manage the process of achieving mutual understanding in conversation (Clark, 1996; Clark & Fox Tree, 2002). It is therefore somewhat surprising that listeners can make inferences about a referent's typicality on the basis of paralinguistic cues, since speakers do not use these signals to communicate about typicality but simply to manage conversation.

It is clear that the relation between certainty and typicality was not learned over the course of the present experiment, but rather was part of the general knowledge that learners had acquired from conversational experience. Why would listeners attend to this relation? One possibility is that it can enhance communication between speakers and listeners. Recently, some researchers have proposed that conversation is a key site of concept learning, because interlocutors strive to develop converging conceptual representations that govern how they speak and understand (Garrod & Anderson, 1987; Garrod & Doherty, 1994; Markman & Makin, 1998). Currently, little is known about the mechanisms by which these shared representations come about. The present study suggests that paralinguistic cues might facilitate this process.

Paralinguistic cues can enhance coordination in other ways, too. A speaker's certainty, conveyed paralinguistically, can enable listeners to make predictions about whether an upcoming referent will be new or old (Barr, 2001, 2003). Speakers produce different "hesitation signatures" corresponding to two distinct processes: (1) preparing a complex verbal plan in order to describe a new referent; or (2) retrieving an existing verbal plan from memory in order to re-refer to an old referent. The "new referent" signature contains longer pauses and is more likely to include the filler *um* at the beginning of the utterance. Listeners appear to be at least somewhat sensitive to these differences and can generate predictions about the identity of the upcoming referent, even before they have heard the first word of the resulting utterance.

However, many important questions remain regarding the nature of listeners' sensitivity to paralinguistic information. First, there are many different ways in which speakers can convey their levels of uncertainty: through hesitations, filled pauses, rising intonation, even verbal hedges.

Which cues in the present experiment produced the learning effects, or were the effects driven by the combination of cues? Furthermore, because speaking is a highly complex activity, there are many reasons why a speaker may be uncertain. Is the speaker pausing to gain time to plan a complex phrase or to find a low-frequency word that is on the tip of the tongue? Has the speaker simply lost the thread of the conversation? These other disfluency-producing factors preclude a perfect correlation between typicality and certainty cues. Thus, for listeners to make use of them, it would seem that they must "interpret" the cues relative to the speaker's circumstances. A listener may be assisted in this process if the various cues cluster in a manner that provides information about the locus of the speaker's trouble, as occurs in the different "hesitation signatures" that speakers produce when re-referring versus introducing new referents.

Last, the finding that learners can exploit paralinguistic cues to learn categories has implications for the study of semantic development. One of the enigmas of development is how children attain a rich knowledge of conventional semantic categories on the basis of a seemingly small number of learning opportunities. Perhaps children, like adults, can extract valuable information about category representation from a speaker's level of certainty. Though we know that children attend to certainty cues during word learning, at least when these cues are verbalized (Sabbagh & Baldwin, 2001), we do not know whether children can exploit the relation between certainty and typicality.

In closing, the "acoustic packaging" of utterances provides insight into the mind of the speaker not just for the theorist but also for the listener (and possibly the language learner too). Listeners' ability to exploit paralinguistic cues of uncertainty in learning novel concepts has important implications for theories of language use, categorization, and language learning. In the final analysis, theorists may find it necessary to consider not only what is said but *how* it is said, in order to account for the efficient transmission of knowledge from expert to novice as well as the smooth coordination between speaker and listener.

REFERENCES

Barr, D. J. (2001). Trouble in mind: Paralinguistic indices of effort and uncertainty in communication. In C. Cavé, I. Guaïtella, & S. Santi (Eds.), *Oralité et gestualité: Interactions et comportements multimodaux dans la communication* (pp. 597-600). Paris: L'Harmattan.

Barr, D. J. (2003). *Paralinguistic correlates of discourse structure.*

Bolinger, D. L. (1964). Intonation: Around the edge of language. *Harvard Educational Review*, **34**, 282-296.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 1482-1493.

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory & Language*, **44**, 274-296.

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory & Language*, **34**, 383-398.

Brown, R. (1958). How shall a thing be called? *Psychological Review*, **65**, 14-21.

Butterworth, B. (1980). Evidence from pauses in speech. In B. Butterworth (Ed.), *Language production: Vol. 1. Speech and talk* (pp. 155-176). London: Academic Press.

Clark, H. H. (1996). *Using language.* Cambridge: Cambridge University Press.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-149). Washington, DC: American Psychological Association.

Clark, H. H., & Fox Tree, J. E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition*, **84**, 73-111.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, **22**, 1-39.

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning & Verbal Behavior*, **20**, 611-629.

Fox Tree, J. E. (2001). Listeners' uses of *um* and *uh* in speech comprehension. *Memory & Cognition*, **29**, 320-326.

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, **62**, 151-167.

Freud, S. (1966). *The psychopathology of everyday life* (A. Tyson, Trans.). New York: Norton. (Original work published 1914)

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, **27**, 181-218.

Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, **53**, 181-215.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech.* New York: Academic Press.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41-58). New York: Academic Press.

Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition*, **14**, 41-104.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous speech. *Word*, **15**, 19-44.

Markman, A. B., & Makin, V. S. (1998). Referential communication and category acquisition. *Journal of Experimental Psychology: General*, **127**, 331-354.

Mervis, C. B., & Pani, J. R. (1980). Acquisition of basic object categories. *Cognitive Psychology*, **12**, 496-522.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, **77**, 353-363.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, **7**, 573-605.

Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between theory of mind and semantic development. *Child Development*, **72**, 1054-1070.

Sacks, H., & Schelgloff, E. A. (1979). Two preferences in the organization of reference to persons in conversation and their interaction. In G. Psathas (Ed.), *Everyday language: Studies in ethnomethodology* (pp. 15-21). New York: Irvington.

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory & Language*, **32**, 25-38.

NOTES

1. To create the inconsistent stimuli for the inner categories of each color set (e.g., *Alba*) it was not possible to reverse the mapping order of the consistent stimuli, because this would have simply recreated the same typicality–certainty relationship. Therefore, in the inconsistent condition, the most certain utterance was paired with both of the two border instances. Likewise, one of the two least certain utterances was paired with the most typical instance. An analysis of learners' typicality responses in which the inner category was omitted revealed a reliable difference between the conditions. Therefore, the effects were not due to this slight difference in stimuli.

2. Incidentally, there was no main effect of consistency on typicality ratings ($F < 1$).