Does Social Information Influence Sentence Processing?

Laura Staum Casasanto (lstaum@stanford.edu)
Department of Linguistics, Margaret Jacks Hall
Stanford, CA 94305 USA

Abstract

Does knowledge of sociolinguistic variation influence how we perceive and understand speech coming from different kinds of people? A series of experiments investigated whether listeners have knowledge about *t/d* deletion, a sociolinguistic variable, and, if so, whether this knowledge influences their language comprehension. Experiment 1 investigated listeners' knowledge of the social correlates of *t/d* deletion. Experiment 2 investigated whether social information listeners gather from the non-linguistic context is used in formulating expectations about sentence meanings. Results indicate that listeners have implicit knowledge about *t/d* deletion, and they use this information in resolving ambiguity, suggesting that social information is a part of language understanding, and should be included in models of language processing.

Keywords: Sentence processing; sociolinguistics.

Introduction

What kinds of information do listeners use to understand language? The simplest assumption might be that they use information from the speech stream. But despite the fact that understanding spontaneous speech is something that people are remarkably good at, the information in the speech stream sometimes does not seem to be adequate for the task. Many other types of information from the non-linguistic context are predictive of linguistic information, however, and various types of non-linguistic information have been shown to influence sentence processing.

One of the earliest examples of non-linguistic information affecting language comprehension is the McGurk effect (McGurk and MacDonald 1976), in which getting articulatory information from the way a speaker's mouth moves that conflicts with the acoustic information in the speech stream causes listeners to believe they have heard an intermediate sound. For example, watching a video of a person saying the syllable [ga] while listening to the person saying the syllable [ba] results in the percept of having heard the syllable [da], which is intermediate between [ga] and [ba] in place of articulation.

The referential context of an utterance can also influence how listeners interpret it. The existence of "competitor" items in the context (for example, a frog on a napkin and a frog that is not on a napkin) changes whether syntactically ambiguous material (Put the frog on the napkin…) is interpreted as a modifier of a potential referent (telling you which frog) or as a potential goal location for an action (telling you where to put the frog) (Tanenhaus et al., 1995; Trueswell , Sekerina, Hill, and Logrup, 1999).

Affective information has also been shown to have an impact on ambiguity resolution. Emotional tone of voice influences whether listeners choose the emotional or the neutral meaning of a pair of homophones (e.g., flower/flour) (Nygaard and Lunders 2002).

ERP evidence indicates that pragmatic information is also available to listeners when they comprehend language (Van Berkum et al. 2008). Hearing a child say a phrase like *I just quit smoking* or a man saying *I'm pregnant* induces an ERP component indicating surprise, suggesting that listeners are considering the plausibility of an utterance given inferences they have made about the speaker during sentence comprehension.

People, then, use information about the scene, the speaker, and other parts of the context to understand language. But what about social information? Do listeners use their knowledge of how different types of people tend to talk to make predictions about what kinds of language a speaker will produce?

Sociolinguistic variation is the variable production of linguistic material that is conditioned by social factors, such as the age, gender, ethnicity, or social class of a speaker (among many other factors). This social information about speakers is also part of the context of any linguistic interaction, and like the other aspects of context mentioned above, it is highly predictive of linguistic behavior.

Since its inception, the study of sociolinguistic variation has been primarily the study of sociolinguistically conditioned variable *production* (Labov, 1966, 1972; Rickford, 1987). The variationist project has been documenting and describing the way different groups of people produce language, studying the factors that condition variation at all levels of linguistic description, from phonetics to discourse patterns. More recently, sociolinguists have begun to analyze the social hierarchies and relationships that underpin this variable production (Eckert, 1989). However, very little is known about the *comprehension* of this variable linguistic behavior - what do listeners do with the structure that exists in socially conditioned variable production? If this information is somehow monitored, do listeners use the knowledge they accumulate?

One study of the effects of the sociolinguistic variable ING (e.g. *walkin'* vs. *walking*) on listeners' attitudes suggests that listeners do make use of linguistic variation to make judgments about speakers. In the matched guise paradigm (Lambert, Hodgson, Gardner and Fillenbaum, 1960), only the critical feature is manipulated between speech samples, which are then evaluated by naïve listeners. Manipulating the realization of the final nasals in ING influenced listeners' judgments about the person who used it – the alveolar nasal (n) makes speech sound more casual

and less educated/intelligent, while the velar nasal (ng) makes speakers sound more formal and more educated/intelligent (Campbell-Kibler, 2006). Listeners use the realization of ING to make attitude judgments about speakers, and they have both explicit and implicit beliefs about who uses this variable and what it means about them. But what about the other way around – do people use information about speakers to make judgments about speech?

Listeners do use information about speakers in making metalinguistic judgments. Strand (2000) showed that seeing a picture of a man or a woman affects how people categorize ambiguous stimuli between sibilants *s* and *esh* and between the back vowels in *hood* and *hud*, and more stereotypical men and women elicit stronger effects than less stereotypical men and women. Because these variables are associated with gender through vocal tract characteristics, it is unclear to what extent the effect relates to sociolinguistic knowledge. The social meaning of fricative frequency and back vowel formant frequencies is not something well-understood. - we don't know to what extent speakers use them as sociolinguistic resources or how variation in these domains is structured.

More importantly, because phone categorization is not a part of normal language comprehension, Strand's (2000) study leaves open the question of whether people use information about speakers when they're not making an overt metalinguistic judgment, but actually trying to understand a speaker in real time. Is social information one of the clues listeners use when figuring out the puzzle of spontaneous speech?

The Variable

The ideal sociolinguistic variable for a study of language comprehension is one that has been well-studied from the point of view of production. Consonant cluster reduction (a.k.a. *t/d* deletion) is a phonetic variable in English in which final coronal stops in consonant clusters may be deleted in some environments. This variable makes a good test case because "…over the past thirty-five years, this phenomenon has been studied in more detail than probably any other variable phonological phenomenon" (Coetzee, 2004).

In addition, consonant cluster reduction has a very convenient property, which is that the deletion can sometimes cause ambiguity between two words. For example, the word *mast* produced without its final consonant becomes ambiguous with the word *mass*. This situation provides a good opportunity to see the effects of contextual information on the resolution of this ambiguity.

This reduction is conditioned by several aspects of the linguistic environment, including features of the segment before the stop (as in *last night* vs. *hard night*), features of the segment following the stop (*fast car* vs. *fast action*), and the morphological status of the stop (*past resolution* vs. *passed resolution*) (Labov et al., 1968; Fasold, 1972). The possible realizations of the final consonant vary along a continuum from an aspirated *t* with a strong release burst to a completely deleted *t* which leaves few signs that it ever existed in the acoustic signal. Although fine-grained levels of *t/d* deletion have been shown to be socially meaningful (Podesva, 2006), the standard taxonomy of consonant cluster reduction distinguishes only between two variants: the deleted and non-deleted variants. It is these two categories of consonant realization that have been studied extensively from the perspective of production, and so that is the distinction I consider in the experiments presented in this paper.

Consonant cluster reduction is also conditioned by many stylistic and social factors, and this conditioning has been studied extensively in a variety of social groups. The deleted variant is less formal than the non-deleted variant, younger people use the deleted variant more than older people do, men use the deleted variant more than women do, and African Americans use the deleted variant more than Anglo Americans do (Wolfram, 1969). Because ethnicity is a very robust conditioning factor, and because it is possible to manipulate the purported ethnicity of a speaker by showing pictures, it is the relationship between ethnicity and consonant cluster reduction that I examine.

Experiment 1 investigates whether listeners have implicit knowledge of the relationship between ethnicity and *t/d* deletion in American English. Experiment 2 determines whether listeners make use of this knowledge to resolve ambiguity.

Experiment 1: Attitudes

Experiment 1 was designed to determine whether listeners associate deleted final stops more with African Americans than with white speakers, as the distribution in the input would predict (if listeners are keeping track of the input in some way). If listeners have knowledge about the distribution of the variants with respect to ethnicity, then it will be possible to find out if they use this information in processing language.

Methods

Participants 111 native English-speaking Stanford University undergraduates received course credit for their participation in this study.

Materials 24 sentences were constructed so that each included a word with a consonant cluster that could be subject to *t/d* deletion (e.g. mast, least, wind). These words were all in phonological environments that promote consonant cluster reduction. In half of the sentences, this word was presented with its normal orthography (Fig. 1A). In the other half of the sentences, this word appeared with its final stop replaced by an apostrophe (Fig. 1B).

Each questionnaire contained 6 items written in normal orthography and 6 written with an apostrophe, and an equal number of fillers, half of which contained another nonstandardism (such as *coffee* spelled *cawfee*). Pictures of

potential speakers were taken from a database of university ID photos, and included the shoulders and head of college-aged individuals, on a white or neutral background (Fig. 1). Four black and four white individuals were selected.

A: The mist predicted by the weatherman surprised me.
B: The mis' predicted by the weatherman surprised me.

Figure 1: In Experiment 1, participants saw a pair of faces along with either the A or the B version of each sentence.

Procedure Participants were asked to try to imagine what these sentences would sound like in their mind's ear and then circle the picture in the pair above the sentence of the person they think is more likely to have said the sentence. Each sentence had one black and one white picture above it. Using written stimuli allows this experiment to address the influence of *t/d* deletion without the influence of auditory cues. I did not expect participants' knowledge of the relationship between *t/d* deletion and ethnicity to be conscious. Thus, participants were not made aware that consonant cluster reduction or ethnicity were of interest.

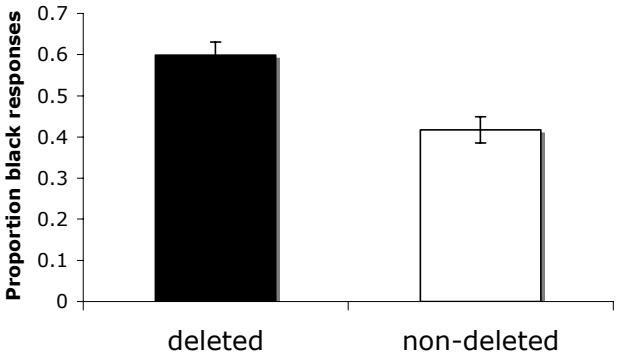

Figure 2: Experiment 1 results: proportion of black faces selected as speaker for deleted vs. non-deleted tokens.

Results and Discussion

Participants attributed 60% of the "deleted" sentences, represented with apostrophes, to the African American person pictured, but they attributed only 42% of the non-deleted sentences, with normal orthography, to the African American ($t(1,109)=4.86$, $p<.001$) (Fig. 2, above). These deleted sentences were significantly more likely to be attributed to African Americans than the other non-standardisms ($t(1,52)=1.97$, $p=.03$), indicating that participants' responses to the *t/d* deletion sentences was not simply a reflection of a general belief that nonstandardisms are more likely to be said by African Americans.

Experiment 2: Ambiguity

Experiment 1 showed that in this population, listeners have at least implicit knowledge that speaker ethnicity is correlated with *t/d* deletion. However, this experiment does not tell us whether listeners make use of this knowledge when they are engaged in the process of language comprehension. Experiment 2 investigates this question using pairs of sentences that when spoken aloud are temporarily ambiguous between two interpretations – one in which the crucial word had a final coronal stop that was deleted (e.g. *mas[t]*), and another in which it never had a final stop (e.g. *mass*):

1. The (mas[t]/mass) probably lasted…

If listeners use their knowledge of sociolinguistic variation when they understand sentences, then they will be more likely to predict a consonant cluster reduction when they believe the speaker is black than when he is white. Thus, they should reach the *mast* interpretation, which involves inferring a deleted stop, faster when they believe the speaker is black than when they believe he is white. By contrast, they will reach the *mass* interpretation, which involves rejecting the alternative with a deleted stop, faster when they believe the speaker is white than when they believe he is black.

Methods

Participants 40 American native English speakers from the Stanford University community participated in this study in exchange for payment.

Materials 24 pairs of sentences were constructed which were identical for the first few words (the section underlined in 2a and 2b below) except for a critical word (italicized), which was identical except for the presence or absence of a stop at the end of a final consonant cluster:

2a. The *mast* probably lasted through the storm.
2b. The *mass* probably lasted an hour on Sunday.

These nearly identical sections would be ambiguous when spoken aloud if a speaker used the deleted variant of a word like *mast*. The pairs of sentences, however, are all disambiguated by the endings of the sentences, which

are much more consistent with one of the interpretations of the beginning than the other. For example, *through the storm* is much more consistent with the *mast* interpretation of the beginning, and *an hour on Sunday* is much more consistent with the *mass* interpretation of the beginning.

24 filler pairs were created that also contained an ambiguity that was resolved later in the sentences:

3a. They saw her duck under the fence.
3b. They saw her duck swimming away with the ducklings.

None of these ambiguities were related to *t/d* deletion. In addition, 48 unambiguous filler sentences were constructed of similar length and complexity. Each item was recorded once by a black speaker, and once by a white speaker. The actual ethnicity of the speaker was counterbalanced across items, so that the pairing of voice and face was equally felicitous across conditions, on average.

The speaker pictures from Experiment 1 were used for the crucial trials and one third of the fillers (4 black males and 4 white males), while 8 females of various ethnicities were displayed with the other two thirds of the fillers. Within each subject, each face was paired with only one voice, to maximize the plausibility of the premise that the pictures represented the speakers.

Procedure Participants were instructed to listen to a short sound clip while looking at a picture of a face, which they were told represented the speaker of the clip. They heard the ambiguous portion of one of the sentence pairs, which contained no final stops. Sound files were excerpted from a recording of the version of the sentence that never contained *t/d* (e.g. the *mass* version), so that there are no cues in the speech stream indicating the presence of a deleted stop.

Participants then saw one of the sentence endings appear below the picture of the speaker. For example, in one trial, participants heard:

4. The [mas] probably lasted

After this clip, one of the following endings appeared on the screen:

5a. …through the storm.
5b. …an hour on Sunday

In half the cases participants saw a continuation that made sense if the ambiguous word had no final stop (e.g. an hour on Sunday, which makes sense if the word was *mass*), and in the other half of cases they saw the other continuation, which made sense if the ambiguous word did have a final stop that had been deleted (e.g. through the storm, which makes sense if the word was *mast*). Participants' job was to assess whether the ending created a sensible sentence in combination with the beginning they had heard, and response times were measured from the time the continuation appeared on the screen.

Results and Discussion

As predicted, participants responded faster to the continuation that was compatible with the word that has an underlying *t* (the *mast* interpretation) when they saw a black face, but they responded faster to the continuation that was compatible with the word that had no underlying *t* (the *mass* interpretation) when they saw a white face ($F1(1,39)=5.64$, $p=.02$, $F2(1,23)=9.23$, $p=.006$) (Fig. 3, below).

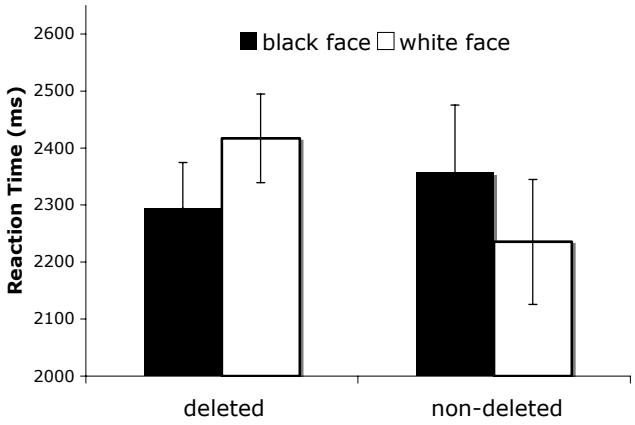

Figure 3: Experiment 2 results, reaction time in milliseconds.

These results seem to support the hypothesis that listeners make use of information about sociolinguistic variation in language processing. However, the two sentences in each pair differ in a lot more ways than just whether or not they contained an instance of *t/d* deletion. For these results to indicate that ethnicity information is used to make predictions about *t/d* deletion, it's important to rule out the possibility that the effect is due to other differences in the continuations that make them more associated with one ethnicity or the other.

To rule out this alternative interpretation, I conducted a norming experiment on written versions of the sentence pairs used in Experiment 2.

In 58 subjects, there was no significant difference between the two lists ($t(1,57)=1.05$, $p=.29$) (Fig. 3, below). In addition, the slight numerical difference between them is in the opposite direction from the difference that would be needed to account for the results of Experiment 2 without appealing to pronunciation differences. Thus, general differences in meaning, lexical items, etc. cannot account for these results.

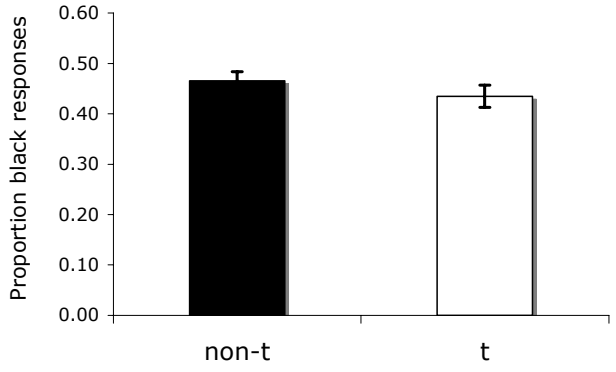

Figure 4: Proportion of black faces selected as the speaker for non-t words (e.g. *mass*) and t words (e.g. *mast*).

General Discussion and Conclusions

The field of psycholinguistics has been discovering more and more kinds of information that people use during sentence processing. Some of these types of information are clearly internal to language and others external to it. Others, however, are less clearly definable: for example, referential context can be interpreted as a property of the world or as a property of the discourse, making it difficult to classify as linguistic knowledge or extralinguistic knowledge.

Knowledge of sociolinguistic variation is similarly difficult to classify, because it is both social and linguistic in nature. The age, gender, social class, or ethnicity of a speaker is not traditionally considered to be part of the linguistic system, which makes it tempting to locate sociolinguistic knowledge outside of language. Yet, knowledge of social characteristics alone does not give rise to the effect observed in the above experiments. The social characteristics are only meaningful and predictive in combination with their relationships to linguistic behavior. While these characteristics do constitute extralinguistic *information* in the input, the stored *knowledge* listeners must draw on, accumulated over a lifetime of linguistic experience, is best described as linguistic. Just as listeners might predict that *t/d* deletion is more likely before a consonant than before a vowel, they are predicting that it is more likely from a black speaker than from a white speaker. The similarity of these predictive processes makes it unsatisfying to classify the socially based phenomenon as stemming from outside the language system. The reasoning behind both these observations is parallel, and it is possible that the way listeners acquire the knowledge about phonological and social conditioning of this variable could be very similar – both generalizations, that following consonants make *t/d* deletion more likely and that black speakers delete more often, could be learned via tracking statistics about co-occurrence.

There does not seem to be any reason, then, to posit a different representational scheme for the two types of knowledge. However, just as the social information alone is not sufficient to produce the effects discussed above, knowledge of co-occurrence between social characteristics and linguistic behavior is also not sufficient – both probabilistic knowledge about co-occurrences and social information about the speaker must be present for listeners to make inferences about speaker behavior. The characteristics of this information, rather than knowledge about how it might relate to language, differ between the case of phonemes and the case of ethnicity.

The results of Experiment 1 addressed the issue of knowledge – listeners have acquired, though through what means is still unknown, implicit knowledge of the relationship between ethnicity and likelihood of *t/d* deletion. This, however, is not sufficient to determine that the social characteristics of speakers will influence language comprehension. Listeners would also need to have access to and interpret social information about the speaker in real time, while they are using many other kinds of information, in order for this knowledge to be of any use to them in comprehension.

The results of Experiment 2 indicate that they in fact do this. When social information about the speaker is available to listeners, they integrate this information with information from the speech stream and use their sociolinguistic knowledge about probabilistic relationships between social information and language to understand language.

This suggests that information that is not represented linguistically can be integrated into language comprehension, as long as listeners have probabilistic knowledge that coordinates this type of information with linguistic representations.

Acknowledgments

Thanks to Daniel Casasanto, Herb Clark, Penelope Eckert, John Rickford, Meghan Sumner, and Arnold Zwicky for helpful discussion. This research was supported in part by NSF Grant #BCS-0720054 and by a Mellon Dissertation Year Fellowship to the author.

References

Campbell-Kibler, K. (2005). *Listener perceptions of sociolinguistic variables: The case of (ING)*. Doctoral dissertation, Department of Linguistics, Stanford University.

Coetzee, A. (2004). *What it Means to be a Loser: Non-Optimal Candidates in Optimality Theory*. Doctoral dissertation, Department of Linguistics, University of Massachusetts, Amherst.

Eckert, P. 1989. *Jocks and burnouts: Social categories and identity in the high school*. New York: Teachers College Press.

Fasold, R. (1972). *Tense marking in Black English: A linguistic and social analysis.* Arlington, Va: Center for Applied Linguistics.

Labov, W., Cohen, P., Robins, C., and Lewis, J. (1968). *A study of the non-standard English fo Negro and Puerto Rican speakers in New York City*. Final report, Cooperative Research Project 3288. Vols. I and II.

Labov, W. (1966). *The Social Stratification of English in New York City.* Washington, D.C.: Center for Applied Linguistics.

Labov, W. (1972). *Language in the Inner City*. Philadelphia: U. of Pennsylvania Press.

Lambert, W.E., Hodgson, R. C., Gardner, R. C., and Fillenbaum, S. (1960). Evaluational reactions to spoken language. *Journal of Abnormal and Social Psychology, 60*. 44-51.

McGurk, H. and McDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*. 746-748.

Nygaard, L and Lunders, E. (2002). Resolution of lexical ambiguity by emotional tone of voice. *Memory and Cognition*, *30*, 583-593.

Podesva, R. (2006). *Phonetic Detail in Sociolinguistic Variation: Its Linguistic Significance and Role in the Construction of Social Meaning.* Doctoral Dissertation, Department of Linguistics, Stanford University.

Rickford, J. R. (1987). *Dimensions of a Creole Continuum*. Stanford: Stanford University Press.

Strand, E. (1999). Uncovering the roles of gender stereotypes in speech perception. *Journal of Language and Social Psychology* 18. Special Issue: Attitudes, Perception, and Linguistic Features. Lesley Milroy and Dennis R. Preston, Guest Eds.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., and Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*. 1632-1634.

Trueswell, J. C., Sekerina, I., Hill, N. M., and Logrip, M. L. (1999). The kindergarten-path effect: studying on-line sentence processing in young children. *Cognition, 73*. 89-134.

Van Berkum, J.J.A., Van den Brink, D., Tesink, C., Kos, M. and Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*.

Wolfram, W. (1969). *A sociolinguistic description of Detroit Negro speech*. Washington, DC: Center for Applied Linguistics.