

Language & Cognition

2002–2003

University Seminar #681

Columbia University
New York, New York

Language & Cognition

What can the study of language contribute to our understanding of human nature? This question motivates research spanning many intellectual constituencies, for its range exceeds the scope of any one of the core disciplines. The technical study of language has developed across anthropology, electrical engineering, linguistics, neurology, philosophy, psychology, and sociology, and influential research of the recent era of cognitive science has occurred when disciplinary boundaries were transcended. The seminar is a forum for convening this research community of broadly differing expertise, within and beyond the University. As a meeting ground for regular discussion of current events and fundamental questions, the University Seminar on Language and Cognition will direct its focus to the latest breakthroughs and the developing concerns of the scientific community studying language.

University Seminar #681, Founded: 2000

SEMINAR ADMINISTRATION

CHAIR: Robert E. Remez
Ann Whitney Olin Professor
Department of Psychology, Barnard College
(212) 854-4247
remez@columbia.edu

RAPPORTEUR: Jennifer S. Pardo
Post-doctoral Research Fellow in Psychology, Columbia University
(212) 854-7033
jsp2003@columbia.edu

WEBPAGE: <http://www.columbia.edu/~remez/langcog.html>

Table of Contents

1. Toward a dual processing model of language: Normal and neurologic studies DIANA VANLANCKER-SIDTIS.....	7
2. Actions of vocal organs as phonological units LOUIS GOLDSTEIN	15
3. The Nature of social meaning in sociolinguistic variation PENELOPE ECKERT	27
4. Fearful symmetry: <i>Similar</i> and similar concepts LILA GLEITMAN	39
5. The Biological basis of speech: Talking to the animals and listening to the evidence J. D. TROUT.....	47
6. The Mapping of sound structure to the lexicon: Evidence from normal subjects and aphasic patients SHEILA BLUMSTEIN	65

3 OCTOBER 2002

**Toward a dual processing model of language:
Normal and neurologic studies**

Diana Vanlancker-Sidtis

*Department of Speech-Language Pathology & Audiology
New York University
&*

Nathan Kline Institute for Psychiatric Research

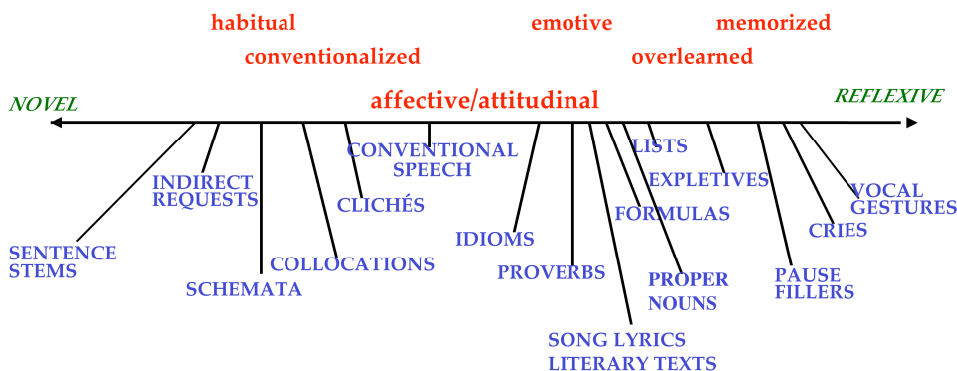
Most contemporary approaches to linguistic modeling, influenced by generative grammar theory, emphasize newly created, novel sentences. Nevertheless, there is a growing awareness of the importance of formulaic language, including formulas, idioms, expletives, memorized speech, slang, conventional expressions, sayings, clichés, and proverbs. Review of linguistic descriptions, psycholinguistic experiments, sociolinguistic studies, observations in aphasia, and functional brain imaging reveals differentiated roles for novel and nonnovel verbal functions. The generation of novel sentences from rules operating on lexical items, and the management of holistic, prefabricated expressions represent two legitimate and separable processes in language behavior. These essentially disparate processing modes each have unique attributes, achieve specialized goals in communication, and are highly integrated in normal language behavior. Interaction and synchronizing of these two processes contributes to creativity in language and is essential for normal communicative function.

My presentation today is going to be a kind of “kitchen sink” talk. The native English speakers present here today will know what I mean, and I use that utterance as an example. Non-native speakers may not know exactly what I am referring to, that this talk includes “everything but the kitchen sink.” My talk will be very heterogeneous, covering a variety of topics—many unrelated findings and phenomena thrown together in one place. However, so many of these formulaic, over-learned, or memorized expressions are difficult to describe because they have a much deeper meaning than a surface description permits. I am going to argue that we learn them in a different manner than we learn other forms of language, such as grammar and lexicon, the typical constituents of linguistic descriptions. I am going to describe evidence from both normal, natural language populations and language use, as well as the language of aphasic patients. The evidence points to the conclusion that there are two distinct modes of processing in the native competence of every speaker.

Formulaic expressions have not been very popular, as you can see from this *New Yorker* cartoon. In the cartoon, a group of people are standing at Inspiration Point, where they are supposed to be having deeply reflective thoughts, but instead, they are thinking things like “No news is good news, ” or other such clichés. This idea that these forms of expression are somehow deficient forms of language, lacking in creativity, has been prevalent in linguistics since the field’s beginning, and intensified in 1957 by Chomsky’s attack on behaviorism. This same idea has recently appeared in *The Language Instinct*, by Steven Pinker, who says, “First, virtually every sentence that a person utters or understands is a brand new combination of words, appearing for the first time in the history of the universe” (1995 p. 22). Although one could argue that he did say, “virtually,” I want to argue that formulaic language is a more prominent and functional form of language than such accounts allow. This kind of language is processed differently as shown in both psycholinguistic and neurological studies.

So what do I mean by formulaic expressions? One way to describe them is to call them non-propositional speech. However, there are around 200 terms that can be used to describe these kinds of utterances, which I will refer to collectively as formulaic or fixed expressions. What they all share is the quality that the individual words of such expressions, as defined in the lexicon, do not provide enough information about the intended meaning of the expression. These expressions must be learned as a whole to be mapped onto a complex meaning that includes specific attributes of the relationship of the items in the utterance. As opposed to the difficulty in learning about grammatical forms, such as passive constructions, the meaning of formulaic expressions is immediately clear. Moreover, literal interpretations of such expressions are unexpected. For example, a former colleague of mine used to delight in answering indirect requests such as, “Do you know what time it is?” as if he were making a literal interpretation of the utterance, by responding, “Yes.”

Continuum of Formulaic Expressions



It might be instructive to think of this collection of formulaic expressions as fitting along a continuum of reflexivity. From involuntary/automatic at the sub-cortical level to voluntary/selective at the cortical level. What makes this category fit together is that the utterances are not novel. They all have some quality that is different from a purely novel utterance such as, "The cat walked into the room and jumped up onto the table." This utterance is a completely new sentence that you probably have not heard before, and that I have never said before. There is some sense in which we can all agree that using a vocal gesture, such as saying "Ouch" when you are hurt, is more automatic than using a sentence stem like, "I'd like you to meet..." One question I have is, how do these familiar utterances differ both psycho-linguistically in the mind and neurologically in the brain?

The unique properties of formulaic expressions are also potentially possible in novel utterances, but are more central for this category of expression. They have a stereotyped form, conventional meaning, and emotional-attitudinal content. As far as the meaning is concerned, the idiomatic literature points out that they are non-compositional, with some degree of flexibility. Also, they are context-sensitive in the sense that there are certain points in a given

Figure 1. A continuum of formulaic expressions arrayed from novel to reflexive.

conversation where they are more or less appropriate, perhaps to serve pragmatic goals. And, above all, they constitute a special knowledge or competence in the native speaker.

To provide some examples of the stereotypic form of formulaic expressions, Terbeek, Canter, and I have examined some ditropic utterances spoken with either literal or idiomatic meaning. For example, one could say the expression, "It broke the ice," with either a literal or idiomatic meaning. Listeners were able to categorize such utterances with about 80% accuracy, and acoustic analyses revealed that formulaic versions were shorter overall, had reduced pitch variation, and shorter pauses. We then tested these items with both native and non-native English speakers, among these either very fluent speakers or ESL students, and found that accuracy varied with degree of language experience. The ESL students were at chance in their ability to distinguish a literal from non-literal production of a formulaic expression, even though they understood the meaning of the expression, the fluent non-native speakers were a little better, and the native speakers were around 90% correct. We concluded that prosodic information is very important for signaling a formulaic meaning, and that non-native speakers have greater difficulty in interpreting such information. Furthermore, we examined our native speakers, and found that they split into two different groups, and that the members of the group with lower performance all had non-native parents. This implies that the ability to use such prosodic information may develop relatively early in a native speaker's competence.

Historically, there have been numerous approaches to idiom comprehension that have not had much success. There is the literal processing notion, in which you decompose a literal interpretation, then use a list to look-

up the real interpretation. An idiom could be treated as the same thing as a very long word, memorized as a chunk linked to its meaning. The problem is that there are too many such expressions for such listing strategies. Other models are likewise deficient, for example, the direct access model, configurational model, and idiom decomposition hypothesis. Finally, it is even suggested that idioms and literals are processed in the same way. I would argue that they can be processed the same way, but that is not typically the case.

The reality is that such "frozen expressions," even the most frozen one, "Kick the bucket," can be varied in intonation, or grammatical construction, for example in the right context, I could say, "The bucket sure was kicked here." The truth is that we need some way to account for both the inherent difference between formulaic expressions and novel utterances, and the array of variation in formulaic expressions. We could begin to understand this phenomenon by proposing that there are two separate modes of processing, one that handles holistic canonical forms, the other one generates novel sentences from a lexicon using rules; and, the second one can work on the first one.

To establish a dual mode system, we can examine maturational schedules to demonstrate distinctive acquisition contingencies. We can show that there are specialized uses of memory, different communicative functions, and separate neurological systems. The point is that these two modes are smoothly integrated in language use: There is a huge repertoire of familiar expressions with canonical forms, and generative rules can operate on them in any legal way. This idea is not entirely new, other proponents of a dual process model are D. Bolinger, D. Cannon, J. Sinclair, A. Wray, F. Lounsbury, P. Hopper, C. Fillmore, and P. Lieberman.

Numerous speech corpora studies have found that formulaic expressions are relatively widespread in mainstream language use. There is a high incidence of fixed expressions, with fixed expressions being used more often than completely novel utterances. Among the fixed expressions, a large number are used, so that use of any individual fixed expression is relatively rare. Many propose that frequency of occurrence explains knowledge of fixed expressions, but that explanation is not supported by the finding that any given expression is not used very frequently. So, how do we know them? Well, perhaps there is something about the context of the first occurrence that allows the fixed expression to get shunted off into a different knowledge system. One problem with these earlier studies is that they lack outside confirmation of the meaning of the expressions, so it is not clear that their measures of usage were accurate.

In order to have outside confirmation of fixed expressions, we studied the screenplay of the film, *Some Like it Hot*. We identified possible candidates for formulaic expressions, including one-word utterances like "Right," "Okay," or "Thanks;" as well as more standard forms like "It's Goodbye, Charlie," "I've been on the wagon," or "Refresh my memory." We found that they constituted 25% of the utterances in the screenplay. We then tried to determine whether our classification of these utterances as formulaic had any outside confirmation, so

we created a survey using 75 formulaic and 25 novel expressions from the screenplay. We had native speakers fill in missing words in a recall task, and then, we had the same informants indicate whether each expression was novel or familiar. Roughly 70% of the words written in for the Familiar Non-literal Expressions (FNE) were the same, compared to only 30% of the words for the novel expressions. Identification of these expressions as novel or familiar was comparable at around 85% agreement. So, we confirmed that a large proportion of utterances are formulaic expressions, and that there is some agreement on the content and identification of formulaic expressions.

One striking aspect of formulaic speech is that it is relatively preserved in aphasic populations. Aphasic patients can produce so-called automatic speech or non-propositional speech. They can count, produce serial speech, formulas, sentence stems, pause fillers, and recited speech. Where in the damaged brain is residual aphasic speech represented? Randy McIntosh, Scott Grafton, and I sought to determine this in a PET activation study comparing counting and naming in normal and aphasic subjects. First, we established that our aphasics differed from the normal subjects in their ability to produce lexical items in a naming task, but they did not differ in their ability to count or to produce non-lexical vocalizations. We used a Partial Least Squares analysis instead of a more typical subtraction method to identify three significant latent variables that are associated with particular active brain regions identified in PET scans. The first latent variable identified naming and vocalization, which activates the left frontal areas. The second latent variable identified naming only, which activates Broca's area and the anterior left frontal hemisphere. The third latent variable identified counting, and there is no left hemisphere activation, but there are a lot of disparate sites, including some sub-cortical areas. So, the more traditional language areas in the left hemisphere were associated with naming and vocalization, whereas more right and sub-cortical activation was found for counting.

Additional evidence for the preservation of formulaic expressions comes from a patient called A. C., who is a transcortical sensory aphasic. This patient had fluent, well-articulated speech, but very poor comprehension and naming. It occurred to me that what was preserved was mostly formulaic expressions that he would string together. We analyzed a sample of his speech, and found that an overwhelming proportion (>80%) of his utterances were formulaic expressions. When we compared the proportion of formulaic to novel utterances for our aphasic talker with the screenplay data and two telephone conversation transcriptions, it is clear that there is a huge deficit in producing novel utterances. Others who study severe aphasia in British English and German note that the only preserved forms are swear words, interjections and greetings, numbers, sentence stems, proper nouns, and other miscellaneous non-propositional utterances.

To study comprehension of formulaic expressions more systematically, Dan Kempler and I created a Formulaic and Novel Language Comprehension Test. (FANL-C) containing 20 familiar non-literal expressions and 20 literal expressions matched for surface structure and length. We used a set of line

drawings for response choices in a 4AFC design. None of the drawings depicted the literal meaning of the familiar expressions. We presented these to 250 normal English speakers ranging in age from 2 to 20 years old, and found that although comprehension of literal phrases reached adult levels by age 7, comprehension of non-literal phrases lagged until around age 11. It appears that there is a maturational component to comprehension of formulaic expressions. We also presented this task to left and right hemisphere damaged patients, and performance for literal forms was worse for those with damage in the left hemisphere, but performance for non-literal forms was worse for those with right hemisphere damage. So, you better have a right hemisphere to handle these non-literal formulaic expressions.

Geschwind at Harvard studied an individual, E. C., who had his entire left hemisphere removed as an adult because of an infiltrating tumor. He was interviewed about 4 months after his surgery, and in a period of a 5-minute interview, we have "goddamit" 6 times—perfectly articulated and emphatically pronounced, exactly as one would say it in context. Another expletive and pause fillers were produced as well, which are very much neglected in considerations of speech production, but are very much a part of natural speech. In many accounts of these kinds of phenomena with aphasics, there is always speculation that it must be supported by remnants of some part of the left hemisphere, but here that cannot be maintained, as the entire left hemisphere has been removed.

Although expletives are a hallmark of aphasic speech, most discussions of expletives bring up Tourette's syndrome, because most of these individuals go through a period where they manifest *coprolalia*: They use expletives in their typical speech production. There was a neurologist at UCLA who posited that it was just confined to what he termed the Anglo-Saxon gutteral, where a talker would just find the mouth in a certain position, and could not help but say, "Shit, shit shit!" In response to this preposterous suggestion, I surveyed cross-language reports of coprolalic utterances. If it is the case that coprolalia is a purely motoric phenomenon, then Tourette patients in Italy should also say "Shit!" What I found is that such patients explore a diverse phonetic repertoire of expletives in their own language. So, it is not purely a motor phenomenon, but some sort of semantic or lexical expression. These kinds of utterances in general are preserved in aphasia, as well as in left hemispherectomy, and hyper-activated cross-culturally in Tourette's syndrome, which is associated with basal ganglia disease.

To wrap this up now, we can think of non-propositional utterances on two dimensions, one that corresponds to the left versus right hemisphere, where they are mostly sub-served by the right hemisphere. Then, there is the vertical dimension, with a division between cortical and sub-cortical structures. Most linguistic models emphasize novel expressions, as in Pinker's recent exposition, but there is recent awareness of the importance of formulaic expressions. There are profound implications of such an awareness for clinical populations. If these are an island of preserved ability in brain damage, it would be a beachhead on which to build some kind of communicative competence. If one

takes a dual-process approach, rather than trying to restore grammatical functioning in various patient populations, it would be possible to realize that there is another kind of competence that could be explored. So, to point to future directions, we can look more closely at different clinical populations, including Parkinson's patients, and at different kinds of formulaic utterances to examine how these two processes work together. Thank you.

APPLAUSE

Questions

Professor James Magnuson: I am having some trouble with the idea that these two modes are interacting all the time. This seems like a rather broad way to characterize what is going on. For example, if the basal ganglia is involved with things like swearing, and also other sorts of emotional expressions, perhaps you would predict that someone who did not swear a lot prior to brain injury would generate more such terms because it is not necessarily tied to perception and production.

Professor Vanlancker-Sidtis I think that is a great question, whether these forms are activated more after an injury because a patient is not able to say other things? Unfortunately, it is very difficult to get pre-morbid data on these patient populations, but I have been working on that, and I think it is a great question.

Professor Michele Miozzo: It seems that one of the most important aspects is to demonstrate that if formulaic expressions are a kind of hodge-podge category, then that there are different mechanisms that underlie each kind of expression. For example, I know of left brain damaged patients for whom some sub-categories are preserved. Are these preserved because they are the most frequent words, or because they are the first to be relearned by the right hemisphere?

Professor Vanlancker-Sidtis Frequency as an explanation falls apart really fast. If I were to show you all of the corpora from different patients, you would see that there is little commonality there. You could say that they are doing the same thing, but now you have to talk about frequency in their own idiolect. If frequency were a compelling cause, we would have to explain why they do not say other high frequency forms, such as the individual words that are lost in aphasia.

Professor Ernst Rothkopf: Let me make you an offer that you can't refuse...What about emerging fixed expressions? Before *The Godfather*, I would have never used such an expression.

Professor Vanlancker-Sidtis There is a high turnover of these expressions, and that is another reason why I would say that frequency is not a compelling factor. When a new form in the culture emerges, it captures attention, and these holistic expressions cycle into play, without much exposure. So it is the impact that enables them to be learned in a different way.

Professor Remez: Let us thank Professor Vanlancker-Sidtis and adjourn.

3 OCTOBER 2002

APPLAUSE

Place: Kellogg Center, Room 1512
School of International and Public Affairs
Columbia University
420 West 118th Street

Time: 4:00 PM

Chair: Prof. Robert E. Remez, Barnard College, Columbia University.

Rapporteur: Jennifer Pardo

Attendees: Abigail Batchelder, Gina Cardillo, Boris Gasparov, Peter Gordon, Joyce Kim, Jackson Liscombe, Jim Magnuson, Michele Miozzo, Ezequiel Morsella, Ernst Rothkopf, John Saxman, John J. Sidtis

Questions pertaining to this transcript should be sent to the rapporteur via email:

jsp2003@columbia.edu



7 NOVEMBER 2002

Actions of vocal organs as phonological units

Louis Goldstein

Department of Linguistics

Yale University

&

Haskins Laboratories

Phonology is a system for assembling a small number of discrete units into a large number of distinct combinations that form the words of a language. Most physical descriptions of speech, however, fail to reveal such discrete units, and instead show continuous change along many dimensions. In this talk, I will present evidence from speech errors that the act of speaking is composed of discrete constriction actions of the vocal organs. I will also consider how such discrete actions could emerge in the vocal behavior of infants developing language.

Let me begin by acknowledging the contributions of various people to this talk. Marianne Pouplier, who is a graduate student of mine, has done most of the work on speech organization that I am going to be talking about. Larissa Chen, another student, and my colleagues and collaborators, Dani Byrd and Elliott Saltzman, have also been working on that project with me. Michael Studdert-Kennedy has engaged me in a number of useful conversations about phonological development, which have resulted in a recent paper that I will talk about a bit today. Catherine Browman and I have developed some of the basics of this approach to phonology, if there is such a thing. And last, I must mention the contribution of Vicki Fromkin who said, when I saw her about a year before she died, "Louis, I really like this articulatory phonology stuff, but nobody is going to believe in it until you deal with speech errors."

I would like to start with a general description of the problem. Historically, there have been two very different ways of describing human speech: 1) A phonological description in which a sequence of discrete symbols from a small inventory or alphabet combine to form the diversity of words in some language, a combinatoric system; and, 2) A physical description, in which continuous, context-dependent variation occurs in many parameters, whichever parameters one chooses to measure. The question to begin with today is: What is the relationship between those two descriptions?

The responses to this question historically are a kind of theoretical Rorschach test. One kind of response is that there is no problem—one description is cognitive, one is physical. That is the take on the problem that

was first given by Hockett, in the mid-1950s, in his Easter egg analogy of speech production. In the analogy, he likens the phonological structure of speech to a row of brightly colored, but unboiled Easter eggs. The act of producing the message on the part of the sender involves smashing the eggs into bits, which then puts the receiver in the position of being a kind of forensic expert, who attempts to reconstruct the distribution of eggs from the distribution of bits of eggshell and yolk. While Hockett would not have used the terms *cognitive* and *physical*, it is clear that his take on the problem is that there is no problem—there is no interesting relationship between the cognitive structure and the physical description of speech.

There are a couple of things wrong with this. One is that in the last twenty years, there have been a number of sources of evidence that the cognitive and physical properties of systems, particularly for action in the motor system, cannot be satisfactorily described independently of one another. Specific evidence from speech systems involves mutual reciprocity, which means that the phonological structure and the physical structure of speech seem to constrain each other. Everyone sort of agrees on the facts, but they do not always take into account what it means that they constrain each other. For example, phonologists and phoneticians agree that the physical properties of speech constrain the kinds of phonological inventories that languages may have. Conversely, the particular phonological system you happen to have in your language constrains your behavior differentially. Thus, the two descriptions are mutually constraining in a way that suggests their relationship may be more interesting than the Easter egg analogy implies.

The other major approach to the question is implicit in classical feature theory, which was the first attempt to unify these two descriptions systematically. It did so by applying the same theoretical objects, the features, in both domains. Rather than launch a full attack on classical feature theory on this occasion, I will say simply that feature theory has problems with generalizations based on articulatory timing. Here, I want to focus on promoting an alternative account, which stems from the work of Carol Fowler (1980) and colleagues, who point out that although the *products* of speech production are indeed continuous and context-dependent, as the physical description suggests, the *act* of speech production itself can be decomposed into discrete context-independent actions of the vocal tract.

In the approach that Cathe Browman and I developed, called articulatory phonology, we make the gestural hypothesis: The act of speaking can be decomposed into atomic units of actions that we call gestures, and gestures are dynamically controlled constriction actions of distinct vocal tract organs. As we have defined them, gestures are meant simultaneously to be units of action, involving the formation of constrictions, and units of information, that is, the contrast between different utterances and their combination into larger units. The time-varying and context-dependent properties of speech that we observe with an x-ray motion picture of articulation are hypothesized to result from the unfolding of these action units dynamically. What I want to do today is pose two questions from this framework: 1) What is the basis for discreteness in

gestural action units?, and 2) What empirical evidence can we find for the decomposition of speech into these action units?

One source for discreteness is that gestures control independent constricting devices, or vocal organs. If we examine the structure of the speaking device, we can see these organs: the lips, tongue tip, tongue body, the root of the tongue, the glottis, and the velum. These are all anatomically distinct body parts. The basic premise of this point of view is that gestures of distinct body parts count as meaningful differences. One side point to make is that not all contrasts that languages use differ in the organ employed. For words like TICK and SICK and THICK, the contrast occurs within the same tongue tip organ. I will address those kinds of differences later in the talk. However, the contrasts in languages that really are between-organ contrasts can be argued to be more common and near universal, whereas not all within-organ contrasts do appear in all languages. I am going to argue that there is something more basic about contrasts of distinctive organs than contrasts within the same organ units. Likewise, by extending this account to other accounts of phonology, such as feature theory, the features that describe between-organ differences always lie at the top of feature hierarchies that describe the functioning of features.

As far as empirical evidence for the primacy of oro-facial organ distinctions, Meltzoff and Moore have shown that even neonates are capable of facial mimicry. This is remarkable because the infant cannot see its own face and has no proprioception from the model's face. These two facts led Piaget to speculate that facial mimicry is impossible until an infant discovers mirrors. Some interesting properties of infant facial imitation are that it is informationally guided, that is, it is not reflexive. It is goal-directed in the sense that infants make successive approximations to a better imitation, it is flexible and generative, not stereotyped, and, most relevant to our purposes, it is specific to the facial organ involved. The imitation of the action is not always accurate, so an infant may protrude its tongue in the wrong direction, but the organ chosen is always correct. In their account of this phenomenon, perceived and produced acts are coded in a common supra-modal representational framework via a process of active inter-modal mapping that allows the infant to detect equivalence between its own proprioception and visual perception of the model.

The key aspect of this model is that the infant can identify the organ involved and specify the organ's relation to the rest of the body. If we extend this model to apply to the vocal organs, we should predict a primacy for between-organ contrasts in phonological development. In the early descriptions of phonological development provided by Ferguson and Farwell (1975), a child's production of its early vocabulary (at a stage of fewer than 50 words) contains a lot of variability in which consonants are not always correct. Michael Studdert-Kennedy and I reanalyzed the survey and found that all of the variants produced involved the same organ as the adult form. This is early evidence that infants match the organ first in order to differentiate words.

I recently completed a more quantitative analysis by examining the CHILDES database contributed by Bernstein-Ratner (1984), selecting 6 children in their first year of speaking (1;1-1;9). We then selected recordings of words that had a clear target referent in the context of the interaction and played them for a panel of judges, who were asked to transcribe the first consonant. When we compare oral constricting organ choice, lips versus tongue tip versus tongue body, we find that all 6 children matched the adult target with greater than chance frequency. For the glottis and velum, there was greater variability, with only some children matching the adult target. For constriction degree, the difference within an organ between a stop, fricative, or glide, no children showed any matching. Thus, this pattern supports the notion that organ identification is basic to phonological development. Again, if we apply Meltzoff and Moore's general model to vocal organs, this implies that auditory information can be used in the service of organ identification. We do not have direct evidence for that assertion, but members of our lab are working on the question of whether pre-productive infants exposed to purely acoustic information about vocal organ movement, either linguistic or nonlinguistic, show productive evidence for organ matching.

To turn to within-organ contrasts, we can begin by asserting that these differ in the exact degree and location of the constriction formed by a particular organ. For example, the words TICK and SICK and THICK can be differentiated by the degree and location of tongue tip placement. But, that description is metric and continuous, and the question is, how are those continua in location and degree of constriction partitioned into discrete regions? I hypothesize that it is possible to model the emergence of these kinds of discrete units through self-organization in a system of units that critically have the property of attuning their actions to one another. Ultimately, speech must be shared across members of a community, and in order for that to happen, we must attune our actions to one another. In that process of attunement, the continua become partitioned into discrete modes of action.

In order to demonstrate the plausibility of this proposal, I set up a model with a population of computational agents that play the "Attunement Game." These agents interact within sets of constraints or boundary conditions. The basic idea is to determine which constraints are necessary and sufficient in order for the system to evolve the kinds of properties we are interested in explaining. Using this model, we can ask whether the kind of partitioning we observe along the dimensions of constriction location and constriction degree can emerge from the random behavior of agents attuning to each other under two key conditions. The conditions are, first, that the agents are attempting to attune their constriction actions to each other, and second, that the agents must recover the constriction parameters produced by their partner from the acoustic signal.

Inevitably, recovery is noisy, so a range of values is recovered rather than a set of sharp cuts. The particular mapping between constriction and acoustic parameters is nonlinear in a systematic way. On each trial of the game, an agent picks a value from a continuum of constriction degree, at random. At the

beginning of the simulation, there is an equal probability of choosing any value of constriction degree. They recover the constriction degree from the acoustics that their partner produced, and then they compare what they recovered with what they produced. If they match, they increment the probability of producing that degree because they want to behave like their partner. That is all that happens in this game.

Following the work of Ken Stevens on the quantal theory of speech production, I use a nonlinear step function to map acoustics to constriction degree. Accordingly, as constriction degree varies in relatively small increments along a continuum, the acoustic consequences vary in a stepwise fashion. Thus, one set of values results in complete closure, another set involves generation of turbulence, and another set involves high amplitude output with no turbulence. I have modeled that in an arbitrary acoustic scale with a step function that looks like the one from all of Stevens's papers. The mapping procedure involves taking the acoustic value and looking up the corresponding constriction degree, and returning a range of values encompassing that value plus those within three acoustic units on either side.

After a number of iterations, both agents' behavior is nonrandom and production is clustered into three distinct modes, corresponding to *stops*, *fricatives*, and *glides*. Thus, when the mapping from acoustics to constriction degree has this nonlinear shape, it is possible for production to cluster into these three distinct modes. With respect to within-organ differences, in this case, it is possible for a continuous production parameter to partition into discrete units via a process of mutual attunement. Indeed, language differences are possible on some within-organ parameters, which may involve less nonlinearity than we observe here. Other kinds of within-organ contrasts, such as location, can vary a lot across languages—tongue tip stops in English and Spanish are quite different in their constriction locations. From this, we can see that self-organization in different situations can result in different partitionings. DeBoer's (2000) simulation of vowels shows how different languages could develop different vowel systems. On the other hand, the organs themselves cannot differ meaningfully across different languages, so the kinds of cross-language differences we observe involve within-organ contrasts.

One other aspect of constriction action that we need to consider is discreteness in time. To resolve this issue, it is useful to characterize motion using dynamical systems. The basic model is one of an oscillator, like a mass bouncing on the end of a spring. That oscillation can be described by a differential equation that characterizes the position and velocity of the mass at the end of the spring. The equation is fixed with respect to time, even though the object is moving. Thus, the units here are the things that characterize change over time.

The question is, what kinds of dynamical characterizations might be appropriate for constriction actions? Indeed, a springlike control might be useful because a spring always returns to its equilibrium position following perturbation. In a relevant sense, human muscles are springs, but they are magic springs because, unlike a model spring that only has one equilibrium

position, a muscle's equilibrium position can be changed through neural input. The motion of a single arm joint can be modeled by a simple change in the equilibrium positions of the muscles attached to the joint. Of course, real actions in a body involve multiple joints, but the entire system as a whole functions as a single virtual spring. Therefore, the model of constrictions is as virtual tunable springs. The actions for a given organ are associated with a distinct set of parameter values for this spring that is fixed over time.

Although this characterization is a plausible story, what evidence is there for units of this kind in the production of speech? Speech production theories typically distinguish two stages with very distinct representations and properties: 1) a planning stage which is couched in purely symbolic units; and, 2) an execution stage which involves a smooth trajectory of muscle controls that satisfies the goals of the plan. I am suggesting that gestural units are common to both planning and execution.

One kind of evidence that is consistent with this view is that of compensatory responses to laboratory perturbations of speech. When producing a constriction, numerous articulators and muscles coordinate to achieve closure. If you apply a perturbation to a part of the system, for example, by tugging on the jaw, the other articulators/muscles involved in that synergy compensate within about 10 ms to achieve closure appropriate to the target utterance. This rapid compensation is almost on the order of a reflex, but it is critically different from a reflex in that you only see compensation when the talker is in the act of producing the relevant utterance, and not a control utterance.

More recently, we have found new evidence for units of this kind by using errors in speech production. The systematic properties of speech errors have long been used to make inferences about the underlying process of speech production. In fact, typically they have been used to argue for abstract segments as units in planning. This is because the most frequent sublexical units in corpora of transcribed errors involve single phonological segments. According to the classical interpretation, these errors are evidence of abstract units because the transposed segments are accommodated to their new contexts during execution of the abstract plan. For example, normal production of the /p/ in SLUMBER PARTY includes aspiration of the /p/. In contrast, during production of the erroneous, LUMBER SPARTY, the /p/ is not aspirated, in accordance with its new context. The basic idea is that errors arise during planning when segments are inserted into the wrong positions within some phonological frame. The plans tend to be well-formed phonotactically, substituting consonants for consonants, and likewise, the execution of such utterances is not anomalous with respect to articulation.

There are some empirical problems with this standard hypothesis. For one thing, the evidence is based entirely on transcriptions of errors, and some potentially anomalous executions may fail to be represented in the transcriptional record. Indeed, there have been some hints over the last 10 years that partial or gradient errors might also occur, as observed in electromyography and more detailed acoustic analyses. Partial or gradient

errors may not be represented in transcription errors because transcription is inherently segmental, and such errors also may not be as perceptually salient.

In order to collect kinematic data during speech errors, we used the electromagnetic (EMMA) at Haskins Laboratories to measure the motions of receivers on a subject's articulators while performing a speech production task designed to induce errors. The task involves repeating phrases with alternating consonants, such as COP TOP and KIP TIP, as well as their non-alternating companion controls, COP COP and KIP KIP. From data collected in this way, we find what we have called gestural intrusion errors. That is, we see an extra copy of a constriction gesture produced at an inappropriate time: During production of the phrase, COP TOP, we observe tongue dorsum raising appropriate for a /k/ during production of the /t/ in TOP. These errors vary continuously in magnitude—when they are small, they sound completely normal, even though you can observe them in the movement data, but when they are large, they sound as if the phoneme has switched from a /t/ to a /k/. To illustrate this, I have some acoustic data that demonstrate the perceptual changes in phoneme identity, and the corresponding production data showing variation in the degree of tongue tip and dorsum raising during production of the /t/ in COP TOP.

[At this juncture, Professor Goldstein played speech samples illustrating the performance of a subject on this production task.]

Although we see intrusion of an erroneous gesture, these do not always involve reduction in the target gesture. In the example I just showed you, there is an increase in tongue dorsum activity, but she does not also lower the tongue tip, so it is not as if she is producing the /k/ instead of the /t/. We do see errors of reduction, but they are much less frequent. Given that intrusions are common and reductions less common, the system appears to be producing two gestures concurrently. Across all 6 subjects, the relative frequency of pure intrusions to pure reductions is consistent, with intrusions much more common than reductions.

These intrusion errors are systematic, but they cannot arise from misplacing abstract units within a well-formed plan because more than one unit is being produced concurrently, and that is not part of a well-formed plan. Likewise, errors produced at partial magnitudes cannot be part of a well-formed plan. Therefore, either planning or execution or both must have some different character than is typically assumed: Either plans do not have to be phonotactically well-formed—they could be gradient—or, execution consists of units whose activation can interact during errors.

To take a gestural alternative to accounting for these data, which stems from the hypothesis that both planning and execution involve dynamical units of action, such errors could arise from a competition between lexical constraints on gestural coordination and intrinsic dynamical constraints on coordination. What I mean by lexical constraints is that we can view words as gestural structures, which contain a number of component gestures arranged in time and critically coordinated to one another. A particular pattern of coupling of

gestures reflects the lexical constraints on gestural coordination. Intrinsic dynamical constraints reflect the notion that some modes of coupling are inherently more stable than others.

In a repetitive task at least, and possibly more generally, the dynamics of constriction organs are oscillatory, and they function as a set of coupled oscillators. Coupled oscillators in general exhibit interesting phenomena such as *entrainment*. Entrainment was originally discovered by Huygens, a 17th Century Dutch physicist who noticed that even at sea, pendulum clocks attached to the same wall tended to synchronize, exhibiting 1:1 frequency locking. This kind of entrainment has also been observed in human bimanual coordination, in which case an initial difference in between-hand frequency and phasing eventually resolves through entrainment to unison phase and frequency. A transition from one mode of frequency locking to a more stable mode is typically abrupt and general to human and chemical oscillation. One-to-one frequency locking among oscillators is the most stable mode of coordination.

I want to suggest that these kinds of transitions in frequency mode locking can account for intrusion errors. Before the transition to error production in COP TOP, both the tongue tip and tongue dorsum gestures are in a 1:2 relation with respect to the lip gestures at the end of the words. After the transition, all of the gestures are in 1:1 frequency locking. Given that higher frequency oscillators dominate in transitions, 1:1 frequency locking will be achieved by increasing the frequency of the slower oscillators, in this case the tongue tip and tongue dorsum. Therefore, we can see a tongue dorsum intrusion error here as a spontaneous transition to a more stable mode of frequency locking. Furthermore, we should and do see such errors increasing over time, because the effects of coupling take time to build up; and more errors at higher frequencies, because coupling is weaker at higher frequencies. However, the system does not stay at the more stable state, as in bimanual coordination. The idea here is that the subject hears an error and resets, so this pattern is the result of a competition with lexical coordination. Likewise, we can ask, how do reduction errors arise, although infrequently? Again, these arise as a result of competition between the intrinsic and lexical modes of production.

Does this account generalize to non-repetitive task? Marianne Pouplier is investigating this question at the moment with the SLIP technique. In this method a subject produces the phrase only once, after silently reading a phrase that biases the subject in favor of making an error. Using this technique, we see similar kinds of gestural intrusion errors, as opposed to substitutions. Finally, we can ask whether this account generalizes to spontaneous speech. There are at least some asymmetries that have been reported in corpora of transcribed spontaneous speech. Those asymmetries also show up on our experiments when we take perceptual biases into account. When we use our erroneous productions as stimuli in listening tests, we find biases in perceptual identification. For example, a /t/ goes to a /k/ more so than the reverse, and we get these same patterns in our repetition data. Errors on initial consonants

are more common in transcribed corpora when they are in words that share final consonants.

What is the relationship between segments and gestures as they have been described here? Within this view, phonological segments can be modeled as gestures that are tightly coordinated with one another. If this tight coupling carries over into errors, then that could account for the fact that there is a relatively high frequency of segment errors in transcribed corpora, but we should still see some cases of individual gestures moving as well. In conclusion, I think there is some chance that we can account for the kinematics of speech errors by means of independent principles if we assume that speech is decomposed into dynamical action units which control the formation of constrictions in a system of discrete constrictors.

APPLAUSE

Questions

Doctor Jennifer Venditti: I am interested in a condition in which the two word sequences in the speech error experiments do not share either an onset consonant or a coda consonant. In all the EMMA data, they did share one of those consonants, but in the SLIP data, it looks like some of the examples did not, and from the frequency data, you show that generally they do share one consonant. I wonder, in the EMMA data, did you have any such conditions where the words did not share any consonants?

Professor Goldstein: We did have some in the SLIP, and the errors there are less frequent. But yes, we do want to look at that more closely. And, we do have some pairs that are designed to elicit more sub-segmental gestural errors; we just have not finished the analyses at this point.

Professor Michele Miozzo: The thing that I have always found intriguing about errors is that the vast majority involve word onsets. I do not know if you provide an explanation of why we find this distribution.

Professor Goldstein: That is a very good question, and I know that understanding that is probably important, but I do not have a story about that that I believe is worth anything. But I agree that it is something that needs to be explored.

Professor Miozzo: Is it because you did not work out the lexical component of your model?

Professor Goldstein: Not exactly. There is nothing in anything that I have thought about so far that explains it. Therefore, I have worked out the lexical component, and it is wrong. In a more developed model that I have not talked about here, the way that the gestures in the onset are coordinated to the rest of the word is different from the coda gestures, but whether that kind of difference can explain this behavior is not convincing to me. And, it is definitely a weakness of the model that it does not account for that in a more principled way.

Professor Robert Remez: There are some cases in which immediate compensation to perturbation fails. For example, Pascal Perrier asked the subject to speak while holding a tube between the lips; Shari Baum asked subjects to wear a pseudo-palate, and she found that they took a long time to compensate for the change in the topography of the hard palate. What does it mean that in some cases there is immediate compensation and in others there is a more gradual process of adaptation or learning?

Professor Goldstein: In the Perrier cases, they could not actually compensate because at the level of the constriction there was a big tube in their mouth. The Perrier cases are more relevant if you assume that the goals are acoustic, and since I do not, it is not a problem for my account. It would be more interesting and problematic if there were cases of his subjects that did show compensation, and he did have one such subject. With respect to the other palate study, it is hard to know exactly what the system is doing. There is no reason the system must succeed if you prevent it in really bizarre ways.

Professor Remez: There are some cases of compensation to bizarre perturbation that I know about. One was a subject filmed by Dennis Frye who had a saber wound of his cheek. This subject was incapable of making complete closures because he no longer had closed cheeks, yet in some respects, his speech remained intelligible. Likewise, glossectomy patients who have had radiation and resection of the tongue as part of their cancer treatment sometimes remain intelligible despite the formation of fibrous tissue that changes the spring-like properties of the tongue and prevents the tongue from hitting all of the targets where regions of closure exist for you and me, and speech remains intelligible.

Professor Goldstein: In one sense the fact that they are still intelligible suggests the primacy of the organ composition—if they are doing something with their larynx, it is good enough, whether they actually hit the target or not, they have got the right organ moving in generally the right way.

Professor John Saxman: Could you speculate a little bit about what the developmental sequence would be in motor control in infants? Would you start with organ differentiation initially, and then, attunement, and then go back to organ differentiation?

Professor Goldstein: I think that organ differentiation is always there. Perceptually, the infants know something about the organs from early on, and then at some point around the time that they begin to babble, they can move something like the right organs in roughly the right way. Then, I imagine the critical developmental sequence being two things: 1) The differentiation of organs using the same system, and 2) Learning how to coordinate gestures to them.

Professor Jeffrey Loewenstein: To what extent might this account extend to investigations of the development of sensitivity to word boundaries?

Professor Goldstein: I have not thought about that much beyond the finding that infants at least identify the correct organ at the beginnings of words.

Professor Remez: Let us thank Professor Goldstein and adjourn.

APPLAUSE

Place: Kellogg Center, Room 1512
School of International and Public Affairs
Columbia University
420 West 118th Street

Time: 4:00 PM

Chair: Prof. Robert E. Remez, Barnard College, Columbia University.

Rapporteur: Jennifer Pardo

Attendees: Josh Davis, Boris Gasparov, Robert Krauss, Serge Levchin, Jackson Liscombe, Jeffrey Loewenstein, Sameer Maskey, Michele Miozzo, Rebecca Passonneau, John Saxman, Michael Studdert-Kennedy, Jennifer Venditti.

Questions pertaining to this transcript should be sent to the rapporteur via email:

jsp2003@columbia.edu



30 JANUARY 2003

The Nature of social meaning in sociolinguistic variation

Penelope Eckert
Department of Linguistics
Stanford University

The field of sociolinguistic variation has not so far developed a coherent theory of the social meaning of variables. Variables are treated as having meaning, but the mechanism by which this meaning is constructed has not been a central issue. Analytical practice has implied that the meaning of a given variable is based in its social correlations, that it is enduring across space and speaker categories, and that it is independent of the larger style in which it occurs. To a great extent, this is a result of the dialectological roots of the field, and the resulting selective attention to regional variables. This talk proposes a theory of the construction of social meaning in sociolinguistic variation, shifting the focus from individual variables to the stylistic practice in which variables are deployed. A speaker's linguistic style is an ongoing project that involves the deployment of a large range of variables with heterogeneous meanings that index various aspects of the personae being constructed. The indexical potential of each variable is located in a conventional and relatively abstract meaning that is constructed at a level that may range from the local to the transnational; and vivified in a given style as it is combined with other variables.

Over the past 40 years, the study of sociolinguistic variation has come full circle—from a focus on local social dynamics through a focus on broad social categories, then back again to the local. I tend to talk about this phenomenon in terms of three waves of variation studies. That does not imply that anything has been left behind in the process as these waves move along. Rather, our progress through large-scale survey research, along with new ideas in social theory, allows us to take a fresh look back at the local. These three waves of variation study are really three modes of thought that are not mutually exclusive: They are different perspectives that are chronologically related, but that need to be resolved.

The earliest quantitative studies of variation indicated that sociolinguistic variables carry textured local social meaning. The first table on the handout shows a small sample of data from a large-scale study of children's conversational speech that was done by Fisher in the 1950's. This study examined the reduction of "ing" to "in" as in "walkin' and talkin' and fishin'." Fisher found that boys tend to exhibit more reduction than girls, and that

children tended to use more reduction when talking among themselves than when talking to adults. He also divided the boys into groups according to what the teachers referred to as the "model" boys versus the "typical" boys, and the latter category showed more reduction than the former. Later, Bill Labov's master's thesis (Labov, 1972), which was an ethnographic study of Martha's Vineyard, portrayed an even more textured view of the kinds of social meaning these variables can contain.

In Martha's Vineyard, there is a very well-known variable, the centralization of the nucleus of "ay" and "ow", so that they say, "fuyt (fight)" and "uot and about (out and about)." Under the influence of the mainland presumably, the nucleus has been in the process of lowering to "ay" and "ow" over the years. What Labov found was that resistance to this change was closely related to people's relation to the island, and particularly to the social and economic changes taking place on the island. In those days, the traditional local fishing economy was under threat from an impinging tourist economy and from a larger corporate fishing economy as well. The people who were invested in the local fishing economy were retaining the local centralized nucleus while those who were in favor of the changes on the island were lowering the nucleus. Labov also compared boys from 2 ends of the island, those from down-island who were planning to leave upon graduation, and those from the fishing end who were planning to stay. Those who were planning to stay had more centralized variants of these variables. Labov interpreted this as a reflection of local identity. That is, the local variants of these diphthongs were associated with the local culture on the island.

Labov also emphasized that local identity is not necessarily a homogeneous concept. In particular, he contrasted the sense of entitlement that the local Yankee Fishermen had to the island as opposed to the Native American population known as the Gay Head Indians. In brief, the local identity is not simply, "Martha's Vineyard," but it is people's construction of what Martha's Vineyard is, and what they are in relation to Martha's Vineyard and the outside. The orientation to Martha's Vineyard was very different for these 2 groups, and it appears to show up in their differential use of these variables: The Gay Head Indians led in the retention of the variants of "ow," and the Yankee Fisherman led in the retention of the variants of "ay."

How do these kinds of findings play out on a larger scale? Can we trace larger changes across larger populations? The study of variation, particularly in Labov's research at the time, turned to large-scale sociolinguistic studies. His dissertation on the socio-economic stratification of English in New York City was the first of what became series of urban studies by numerous people (Labov, 1966). In these studies, socio-economic status was the dominant social category of focus. Labov found that a range of variables in speech were stratified by socio-economic class. As an example, Figure 3 on the handout shows the stratification of the variable, fortition of "th" to "t," as in "dis ting, dat ting, and dee udder ting." This variable is not exclusive to New York, but it is a fairly simple illustration of the phenomenon. As you move downward in the social hierarchy, there is a greater percentage of usage of the variant. In

addition, as the style moves from casual to more formal interview and finally, to the most formal reading style, there is less use of the variant. Other variables that he studied, such as absence of post-vocalic “r” and tensing and raising of “ae” and of “ah,” all showed similar stratification, with some differences in detail that are not relevant here. The interpretation of these data was that the social meaning of the use of these variants was to distinguish between prestige and stigma, that is, you have prestige at the more standard end and stigma at the non-standard end of the continuum.

The notions of standard and vernacular and prestige and stigma tend to become loaded with many interesting ideologies and theories. First, the vernacular is seen as the locus of the initiation of change—changes originate in the working class and spread outwards through the socio-economic hierarchy. Also, the vernacular is not simply the speech of the lower class, but is part of everyone’s most natural speech. This concept connects with the critical period in language development as the vernacular is considered the first system that people learn, which is their most automatic manner of speaking. Labov interprets this variation as a function of the amount of attention paid to speech: As a person becomes more interested in and pays more attention to what they are saying, they pay less attention to the way they are saying it, and then they lapse into their most natural patterns.

The data also show that there is maximal local social differentiation lower in the socio-economic hierarchy. One interpretation of this finding is in terms of linguistic markets—those higher in the hierarchy function in a more standard linguistic market, in which they need to use more standard forms in order to be attended to. Certain institutions require the display of the use of cultural capital, which includes a standard educated language. The standard is designed to downplay local differences, that is, the standard lives in more cosmopolitan networks. Downplay of local differences does not simply create similarity among people in geographically vast networks. It displays a rejection of certain kinds of local loyalties because as we move up in the social hierarchy, we have to be able to convince people that our interests do not purely derive from our local origins. In some sense, we are taking into consideration a wider constituency. Meanwhile, within the local market, the use of the dialect shows membership and authenticity.

Labov distinguishes changes from above and changes from below, in the sense of above and below consciousness. The changes from above consciousness come from higher in the social hierarchy. On the other hand, changes from below are called natural changes, from below the level of consciousness—we can not control them. This fits into a view of linguistic change as something that is compelled by the linguistic system—we can not stop linguistic change, we can slow it down or resist it as they did on Martha’s Vineyard, but it is beyond our control. My goal is to resolve this view with the view that I am about to present, both of which I believe in. Among other things, this view represents a kind of consensual model of society and of the meaning of linguistic norms. In fact, Labov’s definition of a speech community is a community of speakers who share norms.

There are four main features of this first wave of variation studies. The first is in the locus of analysis in the geographically and dialectally defined speech community, so that what people are studying is regional and local dialect. The social focus has been the socioeconomic hierarchy as a map of social space, so that change spreads both across the geographic region and up the socioeconomic hierarchy. Other kinds of social categories are subordinated to class, so the treatment of gender, for instance, is considered in terms of gender relations to class. The meaning of the variables is seen as markers of these pre-determined social categories, based in a consensual stigma-prestige axis. Finally, speaker agency is limited to resistance to stigma on the one hand, and on the other hand, people argue that change is actually being accelerated at the lower end of the socioeconomic hierarchy as a way of laying claim to local membership.

This first wave provided a picture of the context for variation and led people to examine smaller communities. The traditional grand surveys were then complemented by more local ethnographic studies. As an example, Rickford's (1986) study of Cane Walk, a sugar plantation in Guyana, illustrates how class is defined differently in different communities. He examined the use of standard English (acrolectal) variants in singular pronoun subcategories among estate workers and office workers, and showed that there is a clear split between the two groups in their use of the forms. Holmquist's (1985) study of communities in the Spanish Pyrenees examined a broader range from traditional mountain agriculture through dairy farming through factory workers. He found that where the local dialect has /u/, standard Spanish has /o/. The use of the more standard Spanish variant reflected the move to the more urban lifestyle. Interestingly, the women were ahead of the men in this particular change. This can be interpreted as resulting from the traditional mountain agricultural system not being particularly attractive to women.

I became engaged in ethnographic work on variation and the spread of change, and the relation between identity construction and the spread of change to address the question, why would someone pick up another way of saying something? Survey evidence has shown that it is adolescents who lead in use of the vernacular and in linguistic change. Note that this finding goes against a claim that change is solely a product of the acquisition process. Adolescents are farther ahead in linguistic change than their younger siblings. It also shows that change can take place after the larger process of language acquisition. Thus, it is not the first language system that endures.

Given these observations, I wanted to know what socioeconomic status means to adolescents. After all, all of the socioeconomic categories that researchers like Labov use are based on things that adolescents do not have, like occupation, income, and final education level. While adults' and children's use of variables are stratified according to socioeconomic status, adolescents are somewhat messier. In order to examine this, I followed a single graduating class through their last two years of high school in the Detroit suburbs, and I spent a month at each of four other high schools to confirm that these observations were not idiosyncratic to one school.

What I found was that there was a split over adolescents' orientation to their school, and in these schools, the social categories were called "jocks" and "burnouts." The jocks were those students who were involved with school as a corporate institution, not school as an academic institution. They were most interested in school activities like athletics and student council, and in order to be involved in these activities, they had to be at least respectable students. These students based their social lives in their school, and they had collegial relations with their teachers. They had institutional power such that they could affect the institutional life. These students were all college-bound, and mainly came from the higher end of the local socioeconomic hierarchy, but enough of them did not that socioeconomic class was not predictive of social class status. At the other extreme were those who were called burnouts. A decade earlier at this school, they had been called "greasers." At other schools they might be called "hoods" or "stompers" or "grits." It is important that local salient differences are embedded in this system, so that if you have schools that are at the edge of rural areas, the rurality fits into the picture as well. In some cases, the name derives from where the groups hang out, like the "loadies" from the loading docks. The use of the term burnouts came from the drug culture.

These groups represented two poles of relations to the school, and each of them saw the other as valueless and immoral. They did many different things to emphasize their opposition, from distinctions in social networks to more symbolic distinctions in clothing. Those who fit into these polar opposites referred to themselves as a member of one of these groups and constituted about half of the student population. The rest of the students referred to themselves as the in-betweens. The variables of interest in the local dialect involved a rotation of the mid and low vowels about the vowel space, which is typical for the Northern Cities region of America.

The Northern Cities vowel shift can be divided into older and newer changes. Some of the changes have been around for centuries and occur across all age groups in the area. The backing of "eh" and "uh," on the other hand, is fairly new and is not found in older speakers. Another variable is the raising of the nucleus of the diphthong, "ay," so that it sounds more like "oi." This variable is particularly noteworthy because it is not part of the Northern Cities shift, but it is quite dramatic in this school. The most dramatic versions of these variables occurred in the speech of those who lived closest to Detroit.

The first variable to examine is the raising of the nucleus of "ay." As shown on the handout, there is a clear split in the use of the variant between the burnouts and the jocks—the burnouts are far more likely to raise the nucleus of "ay" than the jocks. It also shows that gender interacts with group membership—the burnout girls are more extreme than the burnout boys, but the jock girls are more conservative than the jock boys. This appears to be an instance of what gender theorists term *multiple masculinities* and *femininities*.

This set of studies constitutes the second wave of variation studies. The locus of analysis moves from the community at large to the culturally defined speech community, which is not a city and tends to be a face-to-face

community. The social focus is in locally meaningful categories as locations in social space. These categories may be class-based, but not necessarily. The meaning of the variables is not very different from the first wave, indexing local groups and categories. Speaker agency is likewise not very different.

This second wave raised new questions, which are introduced by another set of data from the Detroit study. First, I compared the usage of the more extreme urban variables with the non-urban variables. The non-urban variables formed two classes, the old and new Northern Cities change shift vowels. The older vowels seem to differentiate the boys and the girls. Across both social groups the girls used more advanced variants of these vowels than the boys. The newer vowels in the non-urban variables group did appear to differentiate the jocks from the burnouts, with the jocks using more advanced variants than the burnouts. In the use of the more urban variables, the burnouts use more advanced vowel forms overall than the jocks.

These data tell us something about the life of a change. When a change is new, it has certain social utilities, and as it gets older, its social utility changes. However, we can also view this phenomenon as a function of style. Each one of these variables seems to contribute something different to linguistic style. These are not mutually exclusive views of the phenomenon. The newer variables may be useful for expressing certain aspects that are associated with being urban, while the non-urban variables may have some more generally expressive use.

The next set of data attempts to illustrate this point more clearly. Examining the urban variables more closely among all groups, we see that the use of certain urban forms correlates with participation in a particular activity, known as "cruising." This is an urban-oriented activity, which the burnouts are more likely to engage in, but it does not mean that the burnouts have more contact with people from Detroit than the jocks, many of whom have relatives or other family contacts from Detroit. Although amount of contact correlates somewhat with use of the urban variables, participation in cruising correlates strongly and distinctively with use of urban variables. Cruising refers to driving along certain routes in Detroit, for instance, Telegraph Road, and the teenagers go there not to see Detroit natives, but to see their urban-oriented suburban peers. Therefore, they do not appear to be engaging in this activity in order to bring Detroit variables home, but their use of these variables is related to their ideology and their sense of themselves. They are making a claim not on some category, but on their sense of that category.

In the next set of data, we can see a differentiation in the use of negative concord among four groups of girls in the school. The burnouts in general lead the jocks, but the more extreme burnout girls lead the less extreme burnout girls. In general, these burned-out burnout girls lead the rest of the school in the use of every variable, including non-speech stylistic variables like clothing and makeup. The point here is that we're no longer talking about categories as correlating with variables as much as practices and personae. What these girls constitute is a community of practice, the members of which mutually construct a community identity.

The interesting aspect of a community of practice is that it is both inward and outward looking at the same time. It is basically an aggregate of people who come together around some enterprise, whether it is a drug house or a garage band or a research institute. This brings us to a notion of variables no longer as solely markers but, as Ochs (1991) argues, the variables index stances and activities, which in turn index personae and categories. This stylistic enterprise brings us to a view of variables as not just category markers or parts of a dialect or linguistic change, but to a view of variables as stylistic resources. This orientation changes the focus from dialects to styles.

The next set of data takes this approach by examining more particular aspects of style. These data are from the work of Zhang (2001) on a new social category in China, called “yuppies.” The members of this group are the managers in the new foreign-owned financial sector. These yuppies have a consumption-oriented lifestyle, and live in stark contrast to styles from the pre-market era. Zhang contrasted managers in these new foreign-owned businesses with managers in the more traditional state-owned businesses. They all had the same educational background and training, but they worked in very different environments. She found that the yuppies have developed a style of speaking, not a dialect, that is suitable to their place in society; most importantly, they are a cosmopolitan group. They are outward-looking, but Beijing-based. They are not trying to be non-mainland, they are constructing a new way of being mainland Beijingers.

The variables Zhang studied were those that were already associated with classic Beijing characters, who live in the literature of the past century. The relation between the variables and these characters is fairly aligned in the attitudes of the speakers she worked with. First there are the “smooth operator” variables. This refers to a male character who transcends changes in regime and so forth, a Beijinger who can always manage regardless of the circumstances. This is an oily character, whose speech contains what linguists refer to as an oily tone due to the use of rhotacization of final syllables and the retroflex articulation of initial obstruents. Another character is the “alley saunterer,” who hangs around in alleyways and “sits in doorways playing with ants.” The variable associated with this character is the interdental realization of /ts/. A third kind of variable is the cosmopolitan variable, the realization of full tone in unstressed syllables. In mainland as opposed to non-mainland Chinese, unstressed syllables typically assimilate to the tone of stressed syllables.

Zhang found that the yuppies tended to use the full tone variable, whereas none of the state workers used full tone. Also, the female yuppies use full tone more than the male yuppies. In contrast, the state workers use the alley saunterer variable and the yuppies do not. For the smooth operator variables, the state workers use them more often than the yuppies in general, but the female yuppies use them much less. Zhang sees this difference as due to the perception that the smooth operator character is something that a man in business can get away with better than a woman. What these people have

done is construct a linguistic style that goes with their new lifestyle, but is also based in a whole history of social meaning.

The final part of the talk deals with variables like the reduction of "ing" to "in." These kinds of variables can signal different degrees of casualness or informality depending on the context in which they are embedded. Their use among certain groups can also signal that the members of the group value informality. This embedding of these variables enables speakers to create new meanings by appropriating these bits of meaning into new contexts. The last sets of data examine an under-examined variable that does not participate in any U. S. regional dialect, the release of final /t/. Standard American English has unreleased /t/ in final position, but there are some who use the released version as a style, not merely for emphasis.

The use of released /t/ exists in the geek subculture, certain Jewish communities, and in gay speech. Benor (2002) studied use of released /t/ in an Orthodox Jewish community in California. She found that boys used the variant more than girls, and particularly when they were making a point in a debate. When they went away to yeshiva, the boys' use of the variant increased considerably. She interprets this finding in terms of the construction of Orthodox masculinity that includes this use of a released /t/ that is related to articulateness and intelligence.

On the other hand, Podesva (2002) studies the use of the released /t/ among gay speakers. In particular, he studies the same medical student in two different situations: with a patient and at a barbecue with his friends. He found that this student used the release /t/ more with a patient than when he was with friends. Furthermore, when using the released /t/, the duration of the burst was longer when at a barbecue than when in a medical setting. In the medical setting, the use of the variant was related to articulateness, but at the barbecue, the extreme form of the variant takes on an analogue character and is related not to articulateness, but to prissiness—an exaggeration of articulateness becomes prissy, and is part of what he calls this student's "bitchy diva" persona. The next set of data shows this student's use of falsetto in three situations, at a barbecue, on the phone with his father, and with a patient. The main distinguishing factor is not how often, but the duration of the stretches of speech in falsetto. Again, the use of protracted periods of falsetto is related to performances of the bitchy diva persona.

Finally, this third wave of variation studies places the locus of analysis in communities of practice, with layered face-to-face communities of imagination and alignment. The social focus is on group and individual personae as the locus of social meaning. The meaning of variables is to index stances and social characteristics. Speaker agency comes from constructing social meaning through the construction of personae.

APPLAUSE

Questions

Professor Michele Miozzo: How easy is it and how long does it take for these changes to occur within a group? In a sense, I am asking about the plasticity of the system.

Professor Eckert: This is exactly where my problem lies, to reconcile the two views. First of all, all the studies that find a critical period for dialect acquisition (by 8 years old) are phonological studies—they are based on fixing word classes. They are not studies of phonetic variability. Once a range of phonetic variability is set, it is possible to produce variants carefully. For a jock to become a burnout or vice versa, is essentially more or less in the same thing. They do not have to learn new word classes. I think that there are parts of the system that are more plastic than others. Intonation is probably something that is more malleable and becomes automatic. For example, using rising intonation in declarative sentences can be adopted when young and dropped with age relatively easily. So, that is one of the questions that this research raises that is not yet answered.

Professor Robert Krauss: Not entirely facetiously, I want to suggest a fourth wave of variation studies. There is a small but growing literature that suggests that when two people randomly selected interact, some features of their speech become more similar over time. Often, there will be asymmetries in the rate at which these features approach each other. For example, people tend to converge on fundamental frequency, but the higher-status (or, more dominant) person will converge less. That approach seems to be taking interaction down to the smallest possible unit because in this kind of situation there are just two people, and one can think about what they come out of the situation with as a dialect that is shared simply by the two of them.

Professor Eckert: I do not know that I would call that a dialect, but what other features are you talking about, you mentioned fundamental frequency?

Professor Krauss: Amplitude and speech rate are the main focus of psychologists, not knowing very much about phonetics, but there are other aspects like lexical choice and at least the way certain words are pronounced.

Professor Eckert: Rather than being a fourth wave, I think that that kind of phenomenon is one of the fundamental dynamics in the second and third waves. I view these styles as being constructed within communities of practice, and the question of who are the movers and shakers in the construction of these styles and in the construction of the world view that gives rise to these styles is important. The interesting aspect is that there are a lot of different ways of defining dominance, and in an experimental situation you are limited in what you can do.

Professor Krauss: Well, less limited than you might think. There are also a lot of different ways of defining identity, and one thing we have learned how to do is to activate various identities so that you can create an experimental situation in which a person is keenly aware that she is a woman or that she is

a student. One would expect that that would be reflected in differences in their speech.

Professor Eckert: Yes, I agree.

Professor Boris Gasparov: I have two questions. One, all your examples look to the realm of phonology. Have you or others done studies about intonation patterns in their social patterns?

Professor Eckert: As you know, intonation is a problem. The problem is that segmental phonology comes with pre-existing categories for us, and intonation research is really in its infancy. The analysis of intonation contour is still a big problem. As far as I know, the only studies of intonation (aside from this falsetto study) are those on rising intonation in declaratives. There have been some interesting studies on that phenomenon, but the problem is that you can not locate that in an intonational system. I think intonation is one of the most important aspects of style, and we are really hobbled by our inability to do the work.

Professor Gasparov: My second question is about possible secondary social semiotic values associated with pronunciation. Can these values change so that for example a low prestige pronunciation becomes high prestige? Have you observed this kind of change, perhaps due to changes in society? Changes not in the pronunciation itself, but in its secondary interpretation? For example, in German there is a high prestige form, rough local dialectal pronunciation, that was suppressed at one time as being lowly and uneducated. But now, there is a preservation of local dialects as an act of defiance of an oppressive regime. In Berlin, it is most prestigious to use these forms.

Professor Eckert: I have a relevant example that I did not mention. That is, studies of white middle class adolescents using features of African-American vernacular. The important point is that they are not trying to be black, they are trying to borrow some of the things that they admire about those kids.

Doctor Ezequiel Morsella: In several cases the speaking style and the meaning attached to it was not arbitrary, so that if it is a more relaxed way of saying something, it is regarded as lazy. In the Chinese study, if a foreigner heard the smooth operator style in a yuppie, could they judge the speech to be more yuppie-like, or oily, even though they do not understand the language?

Professor Robert Remez: Are you proposing stylistic universals?

Doctor Morsella: It seems as if some associations are not purely arbitrary.

Professor Eckert: Yes, Zhang has played samples for us, and people agree that the speech sounds oily, but she is already proposed oily as a way of hearing it. Also, the full tone speech sounds staccato, but the actually association of those features with characteristics like laziness or oiliness are more culturally embedded.

Doctor Rebecca Passonneau: I have a question about the data collection. What procedures do you use to elicit and collect the speech samples?

Professor Eckert: The general practice is to do a sociolinguistic interview, which is designed to take people's attention away from the way they are

speaking. In the jocks and burnouts study, my data are based on an interview that lasts from 1–3 hours. I sample the tokens evenly throughout the tape, after dividing the tape into an equal number of sections, taking 50 tokens per variable per speaker.

Doctor Passonneau: It seems like most of these data are phonological. Has there been any work on lexicons?

Professor Eckert: Well, I can tell you that the burnouts use “fuck” a lot more than the jocks, and the jocks use the discourse marker, “like,” a lot more than the burnouts. But, lexical studies are going to require the use of much larger corpora. As far as grammatical variables, I use negative concord because it is such an overtly evaluated variable. Generally, the studies on syntactic variables have focused more on internal than external constraints.

Doctor Passonneau: This is just a comment. I find it interesting that jocks have always been jocks, but that there are so many more terms for the burnouts category.

Professor Eckert: Actually, I sort of slid over that issue. There are other terms for jocks that I have not mentioned. However, the jock term does seem to be a little more constant because, of course, sports are the quintessential service to the school.

APPLAUSE

Place: Kellogg Center, Room 1512
 School of International and Public Affairs
 Columbia University
 420 West 118th Street
Time: 4:00 PM

Chair: Prof. Robert E. Remez, Barnard College, Columbia University.

Rapporteur: Jennifer Pardo

Attendees: Antonina Berezovenko, Miranda Cleary, Erin Curren, Josh Davis, Bridgid Finn, Aili Flint, Ofelia Garcia, Boris Gasparov, Anna Hahn, JoAnne Kleifgen, Robert Krauss, Jackson Liscombe, Jeffrey Loewenstein, Michele Miozzo, Ezequiel Morsella, Rebecca Passonneau, Anna Popovitch, Owen Rambow, Ann Senghas, Diana Vanlancker-Sidtis.

Questions pertaining to this transcript should be sent to the rapporteur via email:

jsp2003@columbia.edu



27 FEBRUARY 2003

Fearful symmetry: *Similar* and similar concepts

Lila Gleitman

Department of Psychology &
Institute for Research in Cognitive Science
University of Pennsylvania

This work analyzes English symmetrical predicates such as *collide* and *match*. Its point of departure is an analysis of the concept 'similar' from Tversky (1977) that appears to show that similarity is psychologically asymmetrical. One basis for Tversky's claim was that the sentences *North Korea is similar to Red China* and *Red China is similar to North Korea* are assessed as differing in meaning by experimental subjects. This seems to imply that the symmetrical entailment ($R_{x,y} \square R_{y,x}$) fails for this concept. I will present several experiments showing that (1) the apparent asymmetry of *similar* is reproduced for all of the hundreds of English predicates that, on intuitive grounds, appear to be symmetrical, including *cousin*, *kiss*, and *equal*; (2) unique inference rules hold for symmetricals, including *similar*, e.g., *John and Bill meet* is interpreted reciprocally while *John and Bill eat* is not; (3) general principles of linguistic interpretation of subject-complement constructions explain the apparent predicate asymmetries observed by Tversky; (4) the structural positioning of nominals in these constructions—rather than their inherent semantic properties—sets their interpretation, as shown by the semantic attributions that experimental subjects place even on nonsense words in constructions like *The zup is identical to the rif*. Most generally, I claim that a deeper understanding of symmetrical terms comes from analyzing the semantics of the syntactic structures in which they appear.

I am talking about a topic today although this work was completed many years ago, around the mid-1990s. I am always happy to discuss these issues, actually it goes all the way back to my dissertation, around 1890 or something like that. I am not sure that it will hold together too well because I have forgotten half of it, but I put something together that we can talk about.

This work is titled in honor of two enemies of symmetry. The first is Blake, in his wonderful tale about tigers, "Tyger tyger burning bright...." This gives me an excuse to decorate this talk with tigers; there will be no other relationship to tigers, but there will be tigers here and there. The other is Amos Tversky, whose wrote about a concept called *similarity* and argued that it was not symmetrical. When I talk about symmetry, I am referring to it in a perfectly orthodox sense. That is, the relation between x and y that, if x is to y, then y is to x. What Tversky famously denied is that "similar" is a symmetrical concept.

Why did Tversky think that similarity is an asymmetrical concept? Because he did some experiments in which he asked people to assess the degree to which two concepts are similar to each other. For example, assess the degree to which North Korea is similar to Red China. This is a fortunate example, as North Korea is actually in the news again. Then, he asked, significantly, *other* subjects to assess the degree to which Red China is similar to North Korea. The subjects gave different ratings to the two presentations, and you can see why, as users of the language, that might be the case. From this example, it appears that the symmetrical entailment fails. If x is more similar to y than y is to x , then symmetrical entailment fails, therefore, similarity is not a symmetrical concept. This example has the flavor of the irrationality of human nature that permeates the work of Tversky and his colleagues. Logically, we might argue that similarity is a symmetrical concept, but psychologically, it is not.

According to this finding and Tversky's interpretation of it, you might think that similarity is just like any other asymmetrical concept. If similarity is like other asymmetrical concepts, such as *hostility*, then why are we surprised to learn that it behaves asymmetrically? After all, Tversky was awarded the Nobel prize for his work on similarity. No such surprise is elicited by studying the asymmetry of concepts such as hostility. There is a difference between similar and other asymmetrical concepts, like *drown* and *father*. Tversky pointed out that the asymmetry of similarity is a small effect, but that it is significant and reliable.

If similarity is allowed to be an asymmetrical concept, you have to ask whether there are any truly symmetrical concepts at all. For example, you might want *equal* to be a symmetrical concept. However, the phrase, "The least of the citizens is equal to the president," is not the same as, "The president is equal to the least of the citizens." Either there are no symmetrical concepts, or the problem has little to do with similarity in particular, as I will argue today. Notice that you are more likely to say, "The bicycle is near the garage" than "The garage is near the bicycle," even though the garage is just as near to the bicycle as the bicycle is to the garage. Why should there be this preference? Even for phrases like, "Sam met the Pope" versus "The Pope met Sam," there is a tendency to think of one as more natural than the other.

I think that Tversky was trying to think about concepts, like the concept of similarity, but he thought that you could do so by asking people questions about words without making a distinction between the way words behave and the way concepts are, as though the concept of similarity shines out through the language uninfluenced. I am going to try to describe what really makes these words symmetrical words of English, and it is their behavior in sentences. I hope those properties will make clear why Tversky was finding these asymmetrical results. It is because he is talking about the way English or any language works, as opposed to what this concept is.

Here are the real diagnostics of symmetry in predicates of English. You have a symmetrical predicate if it appears in a sentence that is intransitive and plural, and if the plural is interpreted reciprocally. But, because similar behaves this way, it is symmetrical in English. To demonstrate, here are some

asymmetrical predicates and some symmetrical predicates. We will use DROWN and MATCH as examples here. As you can see in these examples, plurality has no effect for asymmetrical predicates like *drown*. You can say “The woman drowns,” and that is a perfectly acceptable sentence in English, but you can not say “The button matches,” and have as natural a sentence. Likewise for “The button is similar.” The intransitive is odd in singular form for a symmetrical predicate—it is a two-place predicate. The other diagnostic is that conjunction entails its conjuncts for asymmetrical predicates. For example, if you say “The man and the woman drown,” this phrase necessarily entails the conjuncts, “The man drowns” and “The woman drowns.” However, if you say “The shirt and the button match,” this phrase entails that “The button matches the shirt” or “The shirt matches the button” or “The shirt and the button match each other.”

In order to demonstrate that these phenomena are true of people, we asked undergraduate students to rate a set of predicates on their symmetry. The ratings verified that half of the predicates were symmetrical and half were asymmetrical, as we expected. However, there was some variation in the ratings within the symmetrical and asymmetrical predicate classes. It could be that some symmetrical predicates are more symmetrical than others, but I hope to convince you that what is really happening is that every word has several properties that enter into these ratings. The problem with this particular design was that we mixed all the word classes together, and we should have done each class separately. For example, we found that the words COLLIDE and HIT fell into different categories. Collide was rated symmetrical, while hit was rated asymmetrical. This occurred despite the fact that a dictionary listing for collide lists hit as a synonym. Therefore, the explanation in terms of concept structures is not simple.

After establishing that the predicates are indeed symmetrical or asymmetrical, we set out to examine their properties in more detail. In the first experiment, we tested the dimensions of singularity and plurality. We asked subjects to choose which of two constructions sounds better; for example, “Sam met” versus “Sam and the Pope met.” The students chose the plural form for symmetrical predicates more than for asymmetrical predicates like “The frog eats” versus “The frog and the flea eat.” Therefore, we established that there is a preference for plural forms for symmetrical predicates. Next, we asked students to rate how close in meaning are two constructions of predicates. Symmetrical predicate pairs like “North Korea and Red China are similar” and “North Korea and Red China are similar to each other” were rated as closer in meaning than asymmetrical predicate pairs like “The principal and the pupil hurried” and “The principal and the pupil hurried each other.” These experiments established the seemingly obvious properties of symmetrical and asymmetrical predicates in ordinary language users.

Nevertheless, we must ask why Tversky found a preference for nominal ordering in symmetrical predicates. We do find that there is a bigger change in meaning when switching the order for asymmetrical than for symmetrical predicates. [“The man drowns the woman” versus “The woman drowns the

man" is a much more extreme difference in meaning than that found in "Sam met the Pope" versus "The Pope met Sam."¹

Finally, I want to examine the distinction between active and stative predicate types. Some predicates describe ongoing events or activities, while others describe some state of mind or being. You could order someone to do some activity, but you cannot order someone to be in a particular state of mind, that would be thought control. It follows that you can order a person to accuse someone of treason, but you cannot compel a person to suspect someone of treason. Likewise, it is more natural to say "He is giving the answer" than "He is knowing the answer." These are the diagnostics of the active-stative distinction. I will show that symmetrical predicates are more like statives than actives because with actives you are changing the causal agent-patient relation, and that matters more to people than subtle changes in state.

What is driving these changes in construal for symmetrical predicates? There are two main possibilities. Superficially, it appears that the properties of the nouns are central to the phenomenon, and that is the most common explanation in the literature. According to Tversky, in comparisons like "North Korea is similar to Red China," the more prototypical noun, Red China, takes second position. The linguist Len Talmy makes a similar distinction concerning spatial predicates. In his examples, he compares phrases like "The bicycle is near the garage" versus "The garage is near the bicycle." Talmy attributes the preference for the former phrase to our concept of space not in metric terms, but as a figure moving on a stable ground. Both of these claims state that it is the properties of the nouns that legislate where they appear in these constructions.

Examining this proposal more closely, we see that for almost any two nouns, they will invite some relationship. As Lewis Carroll points out in "Cabbages and Kings," it is odd to talk of these two things because they do not invite any sort of relationship. However, when thinking about drunks and lampposts, especially when thinking about anything having to do with movement, there is a preference because of our natural conception of the relationship between these things. In the most extreme cases, reversing the order is even worse. For example, there is a huge difference between saying "Your eyes are like limpid pools" and "Limpid pools are like your eyes."

What kind of role are the nouns playing here? They play a big role in inviting a particular comparison. Kelly and Bock (1986) showed that without the predicate, the order preference is the opposite, with the more powerful/prototypical item coming first. Also, we scrambled the predicates so that we had constructions like "The bicycle married the garage," we still found a preference for a particular ordering of the nouns. However, as you will see, the nouns invite an ordering, but they are not decisive.

Following Tversky, we asked students to tell us which of two constructions they would prefer to say. Of course, they provide the typical preference and degree of agreement. However, we also asked students to tell us if they could think of a situation in which they would want to say it the other way. The students enjoyed coming up with interesting situations in which they would

say it the other way, and this reversal usually involved changing the relative valence of the nouns along the dimension of comparison, for instance, size, importance, mobility, etc. Why do the subjects change the invited order in the sentence context? Perhaps there is something in the syntax that cannot be ignored that reverses the order so that the ground must come after the figure.

It could be that there is something about syntactic position that gives some inherent semantics to nouns. If so, then we should find these effects with nonsense words. We presented subjects with a set of comparisons using nonsense words like, "The zup/rif is similar to the rif/zup." Taking the dimensions that people provided as their reasons for ordering the nouns, size/mobility, age, fame/importance, we asked the subjects to tell us which item was larger/older/more important. The subjects agreed with each other on which item is more extreme based on the ordering of the nouns in the predicate comparison for all symmetrical predicates, and the order effect was the same across all the predicates.

Therefore, it is the relationship between the subject and the complement that is driving the ordering relationship. The semantics of these syntactic classes supercedes the semantics of the predicate itself. The predicate SIMILAR is symmetrical, but the language takes the vacuity out the comparison. The language provides information about the representation under which the comparison is to be made, by placing the items in the subject versus complement position. In order to preserve the symmetric relation, English requires that one use the intransitive, "North Korea and Red China are similar" or "Red China and North Korea are similar." The importance of this distinction is revealed by the fact that every language has a way of expressing symmetrical predicates without directionality.

I want to discuss one final point about how one might describe symmetry in a grammar because I wrote my dissertation on this topic. I wrote about coordinate conjunctions, and symmetrical predicates pose a problem for the kinds of grammars that were popular in the mid 1960s. For example, if you say "John and Bill drowned," the sources of this construction derives from the sentences like, "John drowned" and "Bill drowned." However, if you say "John and Bill met," you cannot attribute the sources to "John met" and "Bill met." The simple grammar involved in forming coordinate conjunctions involved taking two sentences that differ only in the noun phrase and combining the noun phrases with a single predicate. For symmetrical predicates, both noun phrases appear in both the simple and coordinate constructions, with a substitution of the reciprocal pronoun for the repeated noun phrases: "John met Bill" and "Bill met John" becomes "John and Bill met [John and Bill] each other." With symmetrical predicates, you can further prune the reciprocal pronoun, and be left with "John and Bill met," because you can recover it uniquely. On the surface, then, the asymmetrical and symmetrical predicates appear similar to each other, but underlyingly, they are different because there is a mental trace of the deletion in the case of the symmetrical predicate.

At the time, no one believed my story until I stopped believing in it. Why is that the wrong story? First of all, the semantics of the words invites a particular interpretation (symmetrical or asymmetrical). Second, the two constructions ought to be interchangeable across a range of applications, and they are not: "John and Mary kissed" is not the same as "John and Mary kissed each other," if you mean that John kissed Mary's hand. Here is a secret I did not tell you all along: Although it is true that no asymmetrical predicate behaves in this way, not all symmetrical predicates behave themselves. You can not say "John and Mary resembled" as a way of saying "John and Mary resembled each other." I found another anomalous example, "encounter." There are several reasons to believe that one should never have proposed the underlying derivation explanation in the first place, that one is derived from the other, because the findings indicate that symmetrical predicates are lexically stipulated—not all follow the pruning rule. Since you have to consult the mental lexicon to use the rule, why bother having the rule, when a lexical reference will do just fine? The fact that the subject-complement position affects the meaning of the symmetrical relationship is simply a property of deriving semantics from phrase structure trees.

Let me end with a final example, and then we can argue. Suppose I say, "The man is tied to the tree." What are you imagining? Something like this, right? [Shows drawing of man tied to tree.] Suppose I say, "The tree is tied to the man." You get a different picture, right? [Muttered agreement from audience. Shows picture of small tree tied to large man's back.] The tree probably comes up out of the ground, but at minimum, the tree gets smaller. The syntax has the last word about how these relations are going to behave.

APPLAUSE

Questions

Professor Peter Gordon: If you think of the complex including the complement as the predicate, rather than thinking of just SIMILAR as being a predicate, then the truth conditions under which you would confirm that are different. I wonder if thinking about it in this way would do more of the work, rather than having to rely upon syntactic position.

Professor Gleitman: You would have an awful lot of predicates, you would lose all generality. It would be infinite, right? Maybe that is ok, but you would have to have some way of recovering this generally for novel combinations.

Professor Gordon: You bring up the Talmy description, in which you have to use the figure/ground distinction. You could ask, where does that all come from conceptually?

Professor Gleitman: The problem is that you have to talk about both of the noun phrases in the end no matter what.

Professor Gordon: I am just wondering if there is something about a complex predicate that makes the more stable ground item a better fit.

Professor Gleitman: However, it turns out that that is not the case because you can say “The house is near the bicycle.” So, you would end up with every possible predicate combination anyway. There is something to the packaging of semantics, so that the subject is more separable from the complement-predicate combination.

Professor Robert Remez: My question is about RESEMBLE and ENCOUNTER. Are they just astonishing facts about the verbs of English, or is there an analytical account that will explain why they are exceptions?

Professor Gleitman: I just do not know because I have asked people who know other languages, and they mainly say that RESEMBLE acts weird in their language, too.

Professor Robert Krauss: My recollection of Tversky’s series of studies was that he was really addressing something called the Limen Theory—that you calculate similarity by extracting a set of properties individually. I think what you are trying to say is that this is a really a half-assed way of doing the experiment, if that is what you are interested in. Supposing you did the experiment correctly, by varying the order of presentation so that you see the items without a sentential context, do you think you would find the same kind of asymmetry that you find when you put it in the context of a sentence?

Professor Gleitman: Tversky did find it in other situations, but he took the opposite interpretation from it. So if you compared Hungary, Sweden, and Norway to Bulgaria, people thought Bulgaria was more similar to Hungary. Depending on the countries that you chose, you could change the similarity relation. Again, he seemed to think that you were comparing these countries like pieces of earth, and then it is incoherent that you would change the relationship. But, if you are comparing them as representations, it all makes perfectly good sense. In my opinion, what the syntax is doing is influencing the representation under which you make the comparison. And, do not forget that he did this between subjects—when you do it correctly, people seem more reasonable and coherent in their regard. The syntax is indirectly influencing the regard because it is organizing the representation.

APPLAUSE

Place: Kellogg Center, Room 1512
School of International and Public Affairs
Columbia University
420 West 118th Street

Time: 4:00 PM

Chair: Prof. Robert E. Remez, Barnard College, Columbia University.

Rapporteur: Jennifer Pardo

Attendees: Rejine Barzilay, John Chen, Josh Davis, Pablo Duboue, Carol Dweck, Inge-Marie Eigsti, Noemie Elhadad, Bridgid Finn, Peter Gordon, Julia

27 FEBRUARY 2003

Hirshberg, Shira Katseff, Robert Krauss, Jackson Liscombe, Kathy McKeown, Michele Miozzo, Smaranda Muresan, Ani Nenkova, Rebecca Passonneau, Lois Putnam, Barry Schiffman, Andrew Schlaikjer, Ann Senghas, Adam Shavit.

Questions pertaining to this transcript should be sent to the rapporteur via email:

jsp2003@columbia.edu



27 MARCH 2003

The Biological basis of speech: Talking to the animals and listening to the evidence

J. D. Trout

*Philosophy Department
and the Parmly Hearing Institute
Loyola University Chicago*

Language research has established the powerful theoretical hypothesis that speech perception and production are uniquely human adaptations. The mechanisms they invoke and the psychological generalizations they follow are tuned to linguistic rather than general auditory phenomena, leading to the view that speech is special (SiS). Despite the progress made under the theoretical guidance of SiS, there is now a growing “auditorist” literature that attempts to refute the idea that the mechanisms subserving speech are unique to human biology and the psychology it grounds. Auditorist critics of SiS conscript nonhuman animals such as quail and chinchilla, and, using discrimination tasks on speech stimuli, infer a common mechanism from similar cross-species performance. I argue that attempts to undermine SiS must demonstrate not just cross-species isomorphisms of behavior, but also either common biological mechanism or common functional organization of the organisms compared. The refutation project does neither. I further argue that the critics’ appeal to the principle of parsimony is insufficient to shift the evidential balance against SiS, and that respect for the total available evidence, including evidence of homology, leads to the rejection of auditorism. Auditorist inferences are frail and yet growing in popularity, a troubling combination that often indicates motivations and interests beyond the search for truth, and I will venture an explanation for this trend.

Roughly half of this talk comes from a review paper and the other half from new work on specialization of speech processes. In the interest of full disclosure, if I seem negative toward primitive attempts at comparative research on mammals, it could be because my son did not compare favorably to a 3-year-old Congo Gray on some object permanence tasks at my 11-year-old niece’s science fair. The speech is special (SiS) thesis has a long history, and we’ll see throughout that this thesis is sometimes confounded with a broader thesis about language being special, but I want to defend the thesis that speech is special. We can construe that doctrine as the doctrine that production and perception of speech are uniquely human adaptations, rooted in human biology. The history of this notion goes a long way back to Eric Lenneberg and Al Liberman, early and mid-career Noam Chomsky, and Steve Pinker.

Auditorism is offered as an alternative hypothesis to the claim that there are specialized mechanisms for processing sequences in language. Auditorism is an implicit hypothesis that only general auditory mechanisms are required to explain the distinctive achievements of speech perception. I say implicit in part because there is no programmatic or systematic defense of auditorism so named, but its propagators are identifiable. The basis for the hypothesis trades on the idea that we share auditory capacities, such as frequency analysis, with organisms beyond our evolutionary lineage. That hypothesis fuels another, that if you can make non-human organisms perform similarly to humans on selected speech tasks, then there is no reason to postulate specialized speech mechanisms.

There is diverse biological evidence for the SiS hypothesis. There is PET and fMRI evidence that shows that there are different brain areas used when people perform auditory tasks, like pitch judgment tasks, versus speech tasks, like phoneme identification. The area that activates most distinctively for the speech task is Wernicke's area, which regulates preparation for articulation. By making a simple phonetic judgment, one is in effect preparing for articulation. Next, there is a well-established critical period for language and speech acquisition, but not for audition. This is a general phenomenon, not just for speakers, but also for signers. If signers do not begin prior to the end of this period, they end up with correlative forms of agrammatism and other sorts of disorders. They can learn as speakers can learn after that, but they deploy general problem-solving strategies that are much more onerous. In addition, there is an autonomy of aphasia from auditory function, showing that there is no necessary auditory basis for a speech-relevant disorder, as such patients typically show normal audiograms. Moreover, dichotic listening tasks demonstrate a right-ear advantage for speech tasks. The effect is small and complicated, but nevertheless real. A recent article by Hickok makes it clear that there is a certain amount of speech processing done in both hemispheres, but the connections are better to the left side. That places some restrictions on the kind of animals that you can use in a comparative paradigm because some animals' brains are not even lateralized.

In terms of speech perception research, the most prominent paradigm is categorical perception, and this task is the one that was used by researchers at Wisconsin and Texas on quail to form the basis for auditorist theses. In this paradigm, a series of sounds varying along a single acoustic dimension, such as voice onset time, is created so that there are items of intermediary values, and people did not identify these middle versions as different from the extreme versions. They would only identify the items as "ba" or "pa." The idea was that if you could train quail to respond similarly to humans, then they are displaying categorical perception as well.

There is also auditory-visual speech perception evidence that is absent from animal research except for a rough orientation to the face. In humans, the ability to comprehend running speech is enhanced by access to movement of the lips. In the standard experiment, researchers add noise to the sound or visual image (using sanded plexiglass) and shift modal reliance on audition or

vision. Others use cognitive load as well—one set of researchers presented a passage of Kant's *First Critique* spoken by a talker whom listeners either could or could not see, and the listeners had to shadow the passage. There were fewer shadowing errors in the audio-visual condition. The integration appears to occur early on in perception, not as a result of independent auditory and visual perception. Finally, there is the characteristic symmetry between speech perception and speech production. There is a relationship between being able to perceive certain categories of speech and being able to produce them. This finding is confirmed by some of the PET findings connecting phoneme identification with the area associated with preparation for articulation.

The main effort to debunk SiS and the motor theory of speech perception began with the categorical perception task. The method was the operant training of quail. Another version was done earlier with chinchilla by Miller and Kuhl, but their conclusions were much more hedged, and they interpreted the similarity in the performance functions as an evolutionary predisposition to mark category boundaries for perceptual reasons. In the quail studies, the training proceeds by presenting the animals with CV combinations in which the F₁ onset frequency is manipulated to create a continuum of variable CVs between the two relevant speech sounds. The dependent variable is pecking behavior, and the animals are rewarded when they identify the item correctly. They are trained on the clear cases, then given a generalization task where they are offered a few novel intermediary versions. The quail do show similar identification performance functions to humans in this kind of task.

There are a couple of points to keep in mind about the task. There is a lot of individual variability, believe it or not, among the quail in their ability to perform the task. Some birds are given up on, which is always sad. Some quail require three times as much training as others. Each quail in any experiment requires a minimum of 4,000 trials over relatively brief periods of explicit reinforcement. In the 1987 study, in which there were only 3 or 4 quail, one quail required over 12,000 trials. In the 1991 paper, only one required as few as 4,000 trials.

The SiS rejoinder to this debunking design is as follows. Humans learn to speak and to comprehend speech, and an aspect of this knowledge is phonetic. Quail are trained to identify sound patterns outside of a linguistic domain. Human psychophysical performance reflects only a small part of human competence in language. Quail psychophysical performance in these studies is their *only* contact with language. Therefore, the suggested analogy between human and quail psychophysical performance is inaccurate, and the comparison is unilluminating.

I refer to the auditorist projects as a refutation project. The structure of the argument and the motivation for the experiments as stated in these articles does not have anything to do with understanding the quail auditory system any better, nor with understanding quail perceptual organization. It is an argument for reorganizing and reorienting the research program on the contemporary scene in speech perception. That is, we should shift attention to the mechanisms of human hearing because we know a lot more about hearing

than about speech mechanisms, if there be such. That promises to allow us to solve problems in the relatively near future.

The auditorist refutation projects states something like the following: Nonhuman animals display the same response patterns as humans when presented with CV syllables. Parsimony demands that we infer that similar mechanisms are responsible for similar response patterns. Therefore, the same mechanisms are employed by humans, chinchilla, and quail. This argument is invalid without the following additional premise: And the only evidence bearing on this hypothesis is the behavioral evidence in these experiments. However, if you include any of the developmental evidence or any of the biological evidence and much of the behavioral evidence in humans, you are not driven to the conclusion that auditory mechanisms are the only ones at play in speech processing. Even if that additional premise is true, the argument is unsound because other premises are false.

The structure of the refutation argument looks like the standard behaviorist argument, that similarity of behavior indicates sameness of mechanism, but similar behavior does not imply similar function or biological mechanism, yet that is the conclusion that is drawn—the title of the 1991 paper uses the word “mechanism”. In the famous Breland & Breland paper, mere behavioral isomorphisms come cheap, too cheap to earn their explanatory keep. There are elephants that are trained to walk bipedally, horses that can count in base 10, and fleas that pull carriages, and nobody suggests that that shows that humans do not have specialized anatomy for bipedal gait. But, that is the logical structure for the auditorist refutation argument.

Compounded with these problems is what you could refer to as misleading attributions. When the quail’s behavior is described by the individuals who do this research, they often say things like, “for birds trained to peck to CVs with low F₃ onset frequencies” and “peck rates increased for CV syllables following [all,” papering over the difference in perspective one might say exists between the quail and the human. Everybody would agree that it is misleading at least in a certain sense to say that Oedipus wanted to marry his mother. He wanted to marry Jocasta, and it happened that Jocasta was equivalent to his mother, but that fact was opaque to him. So, transparent attributions are overly generous, they make the subject of the study appear wiser than it in fact is—CVs do not appear in the vocabulary of quail, so this attribution makes it appear as if they are conceptualizing the pattern in some way.

The fact is, we do not really know what it is about the stimuli that they are responding to—they may be responding to some acoustic feature that they are really good at picking up on. I think that is the key to a really interesting experiment that could be run: To train quail to perform on speech materials in a way that is better than humans can do. Then there could not possibly be an explanation for why they would have that adaptation. It would be obvious that some other feature in their auditory system is available for them to draw on that allows them to perform the categorical perception task. It should also be noted that the experimenters start collecting data as soon as the quail hit

90% performance in training, which is considerably worse than what a child could do under similar circumstances.

Professor Jont Allen: Are you arguing against perceptual features in speech? Because that is what it sounds like. If you take a particular feature and the fact that a quail can recognize it, it is an argument for the primitiveness of that feature, but there are many other ways of getting at this question, you do not have to use quail. I would say that there is a lot of evidence that there are some basic features that are pre-phonemic that are recognized, and this goes back to Miller and Nicely and some others. You are arguing against them.

Professor Trout: No, my point is that if there are these basic features, they are features that may be useful to the quail in their environment, but they are certainly not features in the sense of phonological features, for example. It might be that animals statistically track changes in distributional information, and some of them can do that relatively well, but what they can not do very well is learn the homorganic constraint—Anything that requires the recognition of a phonological feature or a hierarchical relationship.

Professor Allen: They can use pitch to communicate, they have primitive pitch codes. It is obvious to me, I do not know if you agree, that that is a form of a feature. Also, durations and things like that, in some kind of a binary representation.

Professor Trout: Yes, I should just make clear that I do not deny that there is some ability that these quail are using, but when humans make the discriminations and identifications, it is likely, given the specialized processing that I have been talking about, that they are making it on a different basis, a different mechanism is involved.

Professor Allen: I think that there is strong evidence to indicate that that is not the case, like Miller and Nicely.

Professor Trout: As long as it turns out that the chinchilla are exploiting a mammalian feature that other mammals have as well, an innate predisposition of some sort, we may share a common basis, but there are many linguistic contrasts that we mark that are not of that sort—that are not just a pitch change or something like that. And there is a reason for that as well. When we listen to language, it is easy for us humans to track formants in the sense that we are sensitive to formant changes at particular points, it is very hard for us to track F₂ alone, for example. That is because the linguistic nature of the speech suppresses your efforts to hear it as a mere acoustic property. Yes, there are acoustic features that we may be able to track, like pitch contours for pragmatic implicatures, but much of the information that we do track is higher-level linguistic information that is above the fray of quail.

Here are some examples of the attempt to paper over the differences by making some transparent attributions. The more grotesque versions would be to say something like, when I come home, my dog is happy to greet the lover of earthy poetry. He is responding to something in me, but he is not conceptualizing me under that mode of presentation. In the same sense, if you could train quail or starling to discriminate frequencies in a musical score, it

would not necessarily show that they are conceiving of it as anything other than a sequence of acoustic properties, they are not appreciating any subtleties of Puccini. It is important to try to represent the mode of presentation under which the organism is capable of understanding the stimuli presented.

There are cognitive approaches that I regard as largely above the fray of this dispute. Many of the studies on numerosity in apes and object naming by Alex, for example, the Pepperberg Congo Gray, and studies of the sort that Hauser is doing on tracking hierarchical information in humans compared to tamarins do not suffer from the same problems. They may suffer from methodological problems, but if they do, they are of a different sort. First of all, there is no effort in those studies explicitly to state behaviorist inferences of the sort that I am discussing about similarity of performance and sameness of mechanism. They have wisely chosen species for which an argument from homology could be made. Finally, they involve largely cognitive tasks—again there is no effort to eliminate appeal to cognitive mechanisms as in traditional behaviorism.

I offer a defense of auditorism that memorializes the distinction between speech and language. This happens early on in any dispute—that all of the difficult problems are with language. Among these are the hierarchical relations in syntax that other animals clearly are not able to appreciate, but speech could be understood in terms of hearing. Speech is a waveform that the auditory system is suited to process, but language concerns higher-level cognitive or central processes. One reply is just to point out that it is not clear what the distinction is, or where the distinction is drawn between speech and language. When you have phenomena like effects of coarticulation or hierarchical relationships among even short sequences of speech. The homorganic constraint does not range over very wide sequences of speech—it is relatively local when you consider the rule that an obstruent at the termination of a syllable sequence that is preceded by a nasalized consonant must share the same place of articulation: ankle, amble, antler, angle. If one considers a phonological rule of this sort determining the articulation of such short sequences, it becomes difficult to distinguish a proprietary speech stimulus from a proprietary language stimulus.

There are other reasons to distinguish between speech and language, and one of them just has to do with good housekeeping. We need a taxonomy to work from, and working with one set of stimuli may be more appropriate for some purposes than for others, but be clear that it is as conventional taxonomy, and that you need not expect the natural world to cooperate with the taxonomy that you are imposing. You have to be open to discovering that the world does not respect the contours of your theory.

In a very recent article in *Science*, Hauser, Chomsky, and Fitch construct a new theory to approach the problems of comparative method and offer a new explanation for the distribution of these psychological features in language. They propose that there is a faculty of language in the broad sense (FLB), which says that the basis for speech perception is shared with non-human animals. Faculty of language in the narrow sense (FLN) would be a subset of that. FLB

includes 1) a sensorimotor system, 2) a conceptual-intentional system, and 3) computational mechanisms for recursion. They think that #3 is the only uniquely human feature of language, and that #1 is so well-established as shared among humans and non-human animals that it deserves presumptive favor, or status as the null hypothesis. It is worth asking in a situation like this, what are the alternative hypotheses to the one that deserves presumptive favor here?

An interesting aspect of the paper is that they are very clear that elaborate training of the non-human subjects is a threat to validity—even in the close of the article, they mention that, importantly, they are engaged in the process of running non-human animals (tamarins in one case) with a minimum of training. It turns out that there are some linguistic tasks that the tamarins just cannot perform, even after a lot of training. If elaborate training is a threat to validity, why is #1 asserted as the null hypothesis? All of the evidence that we have from non-human animals is the result of elaborate training. Another point is that if one considers the fMRI or PET evidence, if areas devoted to articulation are activated as part of a sensorimotor system, that must be a disanalogy with quail because they do not have the motor system for producing speech.

Professor Allen: You seem to be ignoring all the neurophysiological evidence of what goes on in the auditory cortex. It is obvious that if you stick an electrode in a cat, you can get a lot of information that is clearly similar to what is going on in humans. There is a lot of overlap based on neurophysiological evidence and that is also being support in the PET and other sorts of imaging studies.

Professor Trout: That is right, and even in the research by Hickok and Poeppel that I mentioned earlier, the auditory system is very richly involved in speech processing, along with other areas of the brain. You do not get the same kind of activation if you use sub-lexical items that you get if you use entire words. I am going to argue that there are important contributions that audition makes, and you can see where those contributions are important when you look at the reading disorders literature as well, which I will discuss later. I am not ignoring the contributions that audition makes, I am just saying that if your argument is that humans do not have specialized speech mechanisms because quail have a similar performance function on a categorical perception task, or if you are going to argue that the basis for speech perception is shared with non-human animals' sensorimotor systems, then it is peculiar that the motor area gets activated in humans, but there could not be any correlate in quail for the motor system that is associated with speech production.

The faculty of language in the narrow sense (FLN) includes only the computational mechanisms for recursion, and it is the only uniquely human component of the language faculty. These mechanisms are responsible for some of the more syntactic features of language, namely, discrete infinity, embeddedness, and hierarchical structure. Hauser et al. argue that FLN could have evolved independently to serve the adaptive advantage that might be

conferred on an organism that can play Machiavellian social schemes. This ability might allow such organisms to survive to reproduce more of their kind. The problem is that they offer little more detail, so you have to worry about the evolutionary explanations being "just so stories" without more constraining information.

Attempting to constrain such evolutionary explanations is a healthy project, in part because it allows one to be clearer about whether this performance across different species can really be explained by biological homologies, or whether there are homoplasies, or analogies, in organisms. In the classic comparative example, the human eye and the octopus eye are analogies rather than biological homologies that evolved because the constraints of the environment on an ability to focus a clear image on a sheet of cells converges on similar solutions. This is one way that you could think about different species having similar performance, but not homologous mechanisms for the solution of the problem.

Because the auditorist position has not been defended in much positive detail, though there are allusions to hearing being relatively better understood than speech, there are auditorist promissory notes about how reorientation of a research program from speech perception to the auditorist view would be healthy. The reason most often given is that speech perception is a tractable problem on the auditorist view. If we understand so much about hearing mechanisms, and speech perception is reducible to hearing, then it is at least possible to solve some of the problems, like that of acoustic variability. If you are searching for the content of the claim that speech perception is a tractable problem, one thing they might mean is that it is a natural process and if speech perception is a natural process, then of course it is tractable in a certain respect. It is part of the world to be understood with experimental methods and theoretical enquiry, and it is only on the condition that the real solutions to problems of speech perception outstrip our computational capacities permanently that it would not be a tractable problem. Perhaps more is meant than that.

One thing that might be meant is that, understood in terms of audition, the mechanisms of which are well-documented, speech perception is a tractable problem. This trades on the understanding that we have of hearing mechanisms. Another possibility, and I think this is getting closer to the bone, is that clinical problems could be solvable in 10 years or so. The suggestion is more like we are at the horizon of solving some stubborn clinical problems that require the development of distinctive technologies. It is easier to imagine how those problems could be ameliorated with clinical solutions if speech perception is reducible to hearing because there is a lot of technology for hearing. Part of the thrust behind the speech perception as a tractable problem view may also have to do with the hopes for the construction and patenting of prosthetics.

One issue that has oftentimes given solace to the auditorist's understanding of speech perception is that there is evidence that there is an auditory basis for selected reading disorders. The finding that is most frequently mentioned in

the reading disorder literature is a difficulty associated with processing of duration information. In these cases, researchers identify people with specific reading disorders and examine performance on audition tasks. For these subjects with selective reading disorders, they also perform poorly on triad tone tests, in which the tones vary in duration in two increments (250 ms versus 500 ms), and the task is to report the sequence of short and long tones. Such subjects have difficulty processing short-term auditory information. This finding has led some auditorists to recruit this as evidence that speech perception is an auditory process. In particular, the phonological encoding that is crucial to reading is hampered by duration processing disorders. On that view, however, the basis for reading is not identical with the basis for hearing because if x (hearing disorder) causes y (reading disorder), then x cannot be identical with y .

Any experimental manipulation is done in the context of a particular theoretical and hypothetical construct, so one cannot simply resort to striking the pose of pure empiricism. In a well-ordered science, there are priorities set for what the big questions are and how they should be funded, among other things. In many domains, including cellular development, there are payoffs expected in some unknown future, and the decision to fund basic research cannot hinge on what those might be because they are unknown. Yet, decisions about the scientific significance of particular theoretical approaches are typically made in terms of how the resources would be used and how much success could be expected based on those allocations. There is a certain value that could be applied to the goal of truth in theoretical understanding, but it may be that funding priorities under limited resources really drive people's theoretical rhetoric about which positions are most plausible.

If you have a view that you can hold out that might contribute to patents and might contribute to certain kinds of clinical solutions, it is not only good press, but it is also good for the closing paragraph of a grant application. Some of it may be true, but the point here is that there is no reason why a theory of biological specialization for speech should perform much worse at that task than an auditorist theory. In addition, some care and responsibility must be exercised when you promote views based on the hope for patents and clinical promise because you are talking about a vulnerable population that is desperate for solutions. We should be concerned that the standards of peer review are different from the standards of accreditation.

Those are some of the reasons that the SiS dispute might matter. It may help define what an ordered science in the area would look like—what a unified theory would look like that is general, that draws on diverse fields, rather than simply audition, and that takes a proper perspective on the relation between basic research and technological application.

In summary, cross-species behavioral similarities recruited by auditorism are unintelligible until we know more about what the quail is actually experiencing. The structure of the refutation project, with its premature appeal to parsimony, is invalid. Most support for SiS is untested on non-human animals. The theoretical perspective of SiS is by contrast with auditorism

extremely complicated, which may make it less attractive for politically interesting reasons. The auditorist perspective is comparatively simple, in casting these long-standing problems in speech as practical. The impact on funding priorities, like big questions versus local applications, may drive the discussion, but SiS can promise to organize the big questions.

APPLAUSE

Questions

Professor Robert Krauss: This is a question that I would only ask a philosopher. That is, not whether SiS raises a question that is important, but whether it is a question that is answerable. Unless you are a creationist, you have to assume that at some remove at least, there is continuity between what we do and what other species do, and if you believe that the species is identified by the fact that it is in some respects unique, it is unlikely that you are going to find another creature that is going to do exactly what we do. So, to say that speech is special in some sense is a no-brainer because we do not see very much speech in other species, in the same way that virtually everything about any species is special. Then, the really interesting possible question is what is there about humans that enables this facility that is not present in other species? I want to know how special does it have to be in order to be interesting?

Professor Trout: I do not know whether this is largely a rhetorical effect, but the thought that it is would probably be cynical. Evidence from brain localization that some find persuasive does not simply suggest that these mechanisms are innate. Rather, over the process of learning we may develop an ability to do something that is species-specific. It is not just that we come endowed with certain mechanisms, but that those mechanisms are time-locked in an interesting way. One quick answer to your question would be if you could find mechanisms associated with developmental features that seem to be significant only for humans, but the kind of evidence there is thus far for speech cases relies on behavioral isomorphisms.

Professor Krauss: I just want to make two points, and then let someone else get in. One point is that you are able to make that argument without referring to any other species, which I think is fine. The other point is that saying that there is a structure in humans that seems to be associated with speech and that does not occur in other species is not, I think, convincing evidence that that structure is responsible, or even necessary for speech.

Professor Trout: That is right. The second stage of what Hauser et al. refer to as the conceptual-intentional process is something that can be interestingly inquired about across species. You can ask whether an animal is really capable of referring, which is some of the research that Hauser has done. If they can, then they have achieved some intentional state. Of course, there is a question about what constitutes referring, but the idea is that you can look for the same features in human language, but notice that I have changed from speech to

language. Yet speech has indexical characteristics—you are using words to pick out objects.

Professor Krauss: Well, within the domain, vervets do the same thing. That is why putting the issue as to the specialness of speech seems to be an uninteresting question, or maybe an unanswerable question. What is answerable is what we do with the capacities that we have, and what other species do with what they have.

Professor Trout: You might use what kinds of predictions the theory makes as a measure of its fecundity. It would be a serious objection to SiS if you were to find that it vacuous, and it has never been subjected to that critique, rather than that it is unduly substantial. There have been predictions that motor theorists have tested, and the particular motor theorist response to the quail findings was that there was no reason ever to predict that mammalian hearing could not mark boundaries that might be auditorily important, but it is surprising that the ones that were auditorily important to them happened to mark categorical boundaries in human speech.

Professor Allen: I have a great confusion with what you are saying about SiS. At some point, you have language, and that is special—there is a big difference between English and Chinese and French, those are special. At one end of the spectrum, clearly there is something that is special because there is a difference between French and English, and that is special. But at the other end, you have the input device, which is speech goes into the ear and is processed by the auditory system, that is one input to language. You also could sit down and read something, that is another input for language. So you can get your information by reading it off of a page or by hearing it in the auditory system. It seems to me most of the time you are talking about speech, so that means it is the oral form of communication, and then sometimes you mix in the visual cues, and then you end up talking about language, which is special. It seems to me the real question here is at what point is it specialized? Clearly the auditory system is not specialized, and actually, speech is a very very simple signal compared with listening out a violin in an entire orchestra. Speech is a very primitive form of auditory signal, you could do a much more complicated task than listening to speech. So, it starts off as an acoustic waveform processing problem by the auditory system, and eventually it is presumably decoded into some basic units, which may or may not be special. I would argue that they are not special, those are very basic feature detectors, just like vision has very basic feature detectors. Once it gets up into the modality of language, it becomes special, but it does not have anything to do with speech.

Professor Trout: There are a couple of ways of responding. I do not think that everything that you are saying is unfriendly to the proposal that I am making, but I have to do some assimilation. In the sense of special that I am using here, the differences between Chinese and French and American English are not special. There are differences across those languages that have to do with tonality and lots of other things, but the traditional arguments for the specialization hypothesis always had to do with things like the argument of the

poverty of the stimulus as well as motor theory kinds of arguments. There is a distinct pace and sequence with which children learn syntax, and it can happen in a relatively impoverished linguistic environment, whether you are Chinese, French, or English-speaking. I had a grandma in Buffalo who was always saying, "No, no, no, it is 'I ain't going to the store,'" so this is really bad feedback. So the differences between languages are not special, but other aspects are special. I do not recall the second part of your question.

Professor Allen: I am really looking for where is the definition of what is special? It is not special at the output of the cochlea, and maybe it is starting to become special when you get to the auditory cortex, but we do not even know that. By the time you get into the realm of language, and understanding, and syntax, you know, high-level context, then it can be special, but I do not think that is surprising. Nobody is arguing about that. When you say speech is special, what do you mean is special?

Professor Trout: One point that I was trying to make was that it is not clear where you draw the line between speech, where a lot of the sequences are relatively tractable to issues of hearing, and language. The boundary is a natural boundary, so it may be unclear. In the cases of short sequences that are governed by phonological rules, you have something that is a relatively simple sequence that is still constrained by higher-order rules that are specialized for linguistic processing. There is not going to be a simple answer to your question, like, it has to happen before 200 ms if it is going to be special. But, there are answers in terms of what kinds of predictions you would make about ease of processing that implicates the use of the lexicon. For example, I have just completed a set of studies that any auditorist should love. They basically create 5-channel noise-band speech, Shannon speech, with easy (high frequency, low neighborhood density) and hard (low frequency, high neighborhood density) words. The performance is quite different for these words—much higher recognition rates for easy words. There is an independent finding about cochlear implant patients, that the most successful patients are the ones who have a more developed lexicon to begin with. There is a straightforward proposal to make—if you are going to create training materials, do so from easy words. Now, the output of the electrical leads in the cochlear implant are noisy, but because of the tonotopic organization of the basilar membrane, it is producing a signal which subjects can remap by drawing upon their lexicon. They are performing the remapping by drawing upon the words that they already know. Are they implicating general learning strategies to do that? We are talking about noisy outputs of electrical leads and the use of the lexicon, we are not really sure at what stage that interface occurs.

Professor Allen: You are really out on a limb if you are trying to understand normal human speech recognition by looking at cochlear implants. The articulation index, from 40 years of study at Bell Labs, proved that starting with just 20 bands of signal-to-noise ratio, you could predict nonsense phones, and from the nonsense phones you could predict words, and from the words you could get a certain probability correct on sentences. So the whole thing is a hierarchical description from a probabilistic point of view that describes how

you can go from some very very primitive measurements to predicting performance of human speech, a complex human speech in noise. This is from something like 50 year's worth of research.

Professor Trout: Well, I am not out on a limb making the argument without pediatric cochlear implant patients, I would add them to the fold, because I used normal hearing subjects hearing 5-channel speech. The expectation is that normal individuals who are challenged by 5-channel speech will display the same advantage for easy words, and they do.

Professor Allen: *Entropy* in language is important, and what you are studying is about the entropy in language. If you use a high-entropy task and a low-entropy task, then you are going to see a huge change in performance.

Professor Trout: So, there are a couple of different ways you can go, and some of them may seem like they are trivializations of a more substantial view that you are offering, but one way that people might traditionally have made this point in the 1980s and early 1990s is that specialization consists in how precompiled the system is—if there are very complicated access codes already in place that transformed stimuli.

Professor Allen: You mean hard-wiring, do you not? If we were hard-wired, we could only learn one language, right?

Professor Trout: It depends on at what stage hard-wiring occurs. People agree that at the sensory stage, humans are relatively similar input, and normal humans are going to react in similar ways at the sensory periphery, but even though hard-wired, their systems do not lack plasticity as well. So, that they can learn different languages.

Professor Allen: They can learn different phonemic features, even, just like the quail.

Professor Trout: If what the quail are learning is phonemic features, as opposed to correlated markers.

Doctor Ezequiel Morsella: One quick question. Your listing of the fallacies is very well-done because you see those fallacies everywhere in research on function. Another fallacy that may be related is that, as you know, early in behaviorist research, they used to believe that behavior is due to very simple mechanisms. Then, they found several cases in which behaviors could not be explained by simple operant principles, so then they said rats have cognitive maps. What has happened is not that now people doubt whether such complicated mechanisms are working in simple things, but that now people accept that once they find a complicated mechanisms, they apply it to the simpler things as well. The point with the auditory account is the idea that because quail are more simple, the process is more simple.

Professor Trout: If you have really simple animals, then one could argue that S-R psychology would not do such a bad job. Two points that I underplayed were the discussion of parsimony, and what I did not talk about at all, the principle of total evidence. The principle of parsimony has to be very carefully applied because it is a methodological principle that could be misapplied if you have a really good theory. People are oftentimes opportunistic in the way

that they apply it. For example, I teach at Loyola, so I often hear the argument that believing in God is actually more parsimonious than not believing in God, because not believing in God requires a relatively complex account of the world, whereas monotheism commits you to the existence of just one entity. It is parsimonious in ontology, but methodologically, it is pretty compromised. So, there are different ways that one could invoke parsimony arguments, so you have to be careful when you invoke parsimony. In the auditorist case, it is invoked by saying, "all I know is that these quail respond in the same way that humans do, I am just a poor farm boy, it looks like the same performance function, it is the same mechanism." Certainly people believe that what is simple also has to be true, and so they may be drawn to simple explanations, but a principle that is widely accepted about considering total available evidence says that you should look at the greatest diversity of evidence that there is bearing on the hypothesis.

Doctor Michael Studdert-Kennedy: Maybe one way of arguing that there is more to the SiS issue that you throughout seem to be raising is the idea that humans get more or different information out of the speech signal than other animals. In the same way, I am quite sure that you could train quail to distinguish a phrase of Mozart from a phrase of Bach; in fact, I am quite sure that if you worked at it, you could get them to make innumerable distinctions about music, but there would be no question that they are not getting the same information out of it as a human listener, and I think that exactly the same is true for speech. The amount of information that is sitting in the speech signal, non-semantic, non-syntactic information, is vastly greater in human comprehension.

Professor Trout: But, to take the auditorist side, I think the interesting question there is, why would the information that the quail get out of speech signal lead them to mark a boundary in the same way that humans do? If they are getting different information, or at least a human is getting more information, why would the quail mark the same boundary?

Professor Allen: You are asking the quail to do a one-bit task, and they find out that they can do a one-bit task, but let them do a five-bit task. Can they do a five-bit task? If you use nonsense speech, and you play it to L1 Chinese and L1 English they will both perform the task quite well if you pick the phone, feature set from their language, and they will not perform it well if you do not. So, how much information you get out of speech depends on which language is being spoken.

Doctor Studdert-Kennedy: Absolutely, as long as you are human, you can always get information out of some language, and the fact that you get different information is not as relevant as the fact that you get the same class of information.

Professor Allen: Yes, it is language, you get knowledge about what the person means.

Doctor Studdert-Kennedy: No, there is different information because I get different information about how it is pronounced in English, and I do not get that information for Chinese.

Professor Allen: I am certainly not saying that people are not getting information from vocal or auditory inputs, it is just a matter of how much context they can take advantage of, it is how deeply they can process it. But, this question of SiS, I am asking, at what point are you trying to say it is special? Certainly at some point when you get down to certain languages, it is clearly special, but my question is still, at what point is it supposed to be special, and I still do not think I have an answer.

Professor Trout: Well, there are properties that other animals have that humans do not, and you could ask the same question about those. I probably can not discriminate above 16 kHz, but crickets apparently can discriminate categorically beyond that.

Professor Allen: Are you saying that it is special when it is at the phonemic level because your example of the quail is with simple CV stimuli, so that must be sensitive to the point of your argument because speech is special because people can categorize and discriminate CV sounds?

Professor Trout: You could also talk about lexical items, longer sequences of sounds, or you could talk about parts of the spectrum, shorter than CV sequences. You could run them on noise-band speech or on sinewave analogues, you could do any number of things and see how they react. Just because a sequence is suprasegmental, for example, does not mean that it is not a good candidate for being in a specialization experiment.

Professor Remez: In the Introduction to *Acoustic Phonetics*, by Martin Joos, published in 1948, he says something like this: We have been stuck with articulatory descriptions of the phoneme inventory at least since Jespersen, and now with the advent of acoustic analysis technology, it will be possible, finally, to give a rendition of the phoneme inventory in physical acoustical terms. That was the opening fanfare of a campaign of research that essentially failed. That is, the phoneme inventory of no language has been well rationalized by the distributional states of acoustic properties. When I think about the claim that Kluender has made on the basis of the quail, it seems to contradict directly 50 years of research showing how phoneme inventories fail to coincide with auditory categories. In fact, the little model case that they considered with place variation in quail is a rather selective presentation of an argument that attempts to refute this tradition—even leaving out all of the localization and developmental evidence for phonemes. I always took the rude version of their claim to refute SiS as: the quail, having access only to auditory experience, shows us that phoneme categories can indeed be auditory categories. If you doubt their thin evidence you are left with two sets of incommensurable categories: one auditory experience and another of the phonemes that languages use to mark distinctive contrasts. What is wrong with that way to frame the argument?

Professor Trout: Your question raises another question—Are you supposing that the similarity of performance is an artifact of the training regimen?

Professor Remez: It is the same as Tinbergen's stickleback in a tank on the window sill adopting the threat pose when the postal van drove down the lane. No ethologist would claim that the stickleback was threatening a postman, even though that was the performance. It is essentially a kind of bait and switch, exactly as you say.

Professor Trout: My question is that if the similarity in performance is somewhat more complicated than a threat or non-threat posture, how do you explain the specific character of the similarity in performance?

Professor Remez: Why do human psychophysics and quail psychophysics look the same? Well, if you gave me a bunch of quail and a function to match, I would figure out a way to do it. And, if you gave me 30,000 trials to do it in, I would be able to do it there, too. You know, I tried to find out what the test items actually were, but I learned that they had been lost to the mists of history, so it is now impossible to conduct additional tests with the items that the quail actually heard. My guess is that the quail were listening to apical release bursts, an extremely prominent spectral feature, and that they were not hearing place variation at all. So, it was an accident. Not only that, of course, but the quail put labial and palatal place items in the same category.

Professor Trout: I guess there is another question why they stopped when they did as well. With humans in the McGurk effect, you can play that as long as you want and still find the effect. It would be interesting to see what would happen to the quail if you continued training them beyond the 4000 trials required to get them to 90%. You wonder why the organism is exposed for that particular period, because the identification point at which they stopped does not have human significance.

Professor Allen: Do you know the work examining the auditory nerve of a gerbil using whispered voiced and unvoiced speech? Even in the auditory nerve, the voicing was very clearly represented. This is just as important as an F2 transition, if not a lot more important, and it is represented early on in very simple creatures. Is that special?

Professor Trout: It depends on how specialized humans can become on those tasks as well. For example, in an ethology experiment where you are looking at non-human animals responding to some speech contrast, it may be that there is a significant amount of brain plasticity that is required for just learning, even though the learning itself displays highly specialized features. Rather than simply saying that some non-human organism can do this, so is it special when a human can do it? You have to be able to see what the differences are. Part of the question is whether the organisms are responding to the same features, that is, the one is responding to acoustic correlates, and the other is responding to a linguistic feature.

APPLAUSE

A Joint Meeting of the University Seminars on Cognitive and Behavioral
Neuroscience & Language and Cognition

Place: Kellogg Center, Room 1512
School of International and Public Affairs
Columbia University
420 West 118th Street

Time: 4:00 PM

Chair: Prof. Robert E. Remez, Barnard College, Columbia University.

Rapporteur: Jennifer Pardo

Attendees: Jont Allen, Peter Balsam, Colin Beer, Bill Benzon, Gina Cardillo, Joseph Cesario, Martha Chaiken, Josh Davis, Mike Drew, Bridgid Finn, Jessica Goldberg, Nate Kornell, Robert Krauss, Jeffrey Loewenstein, Dustin Merritt, Ezequiel Morsella, Tammy Moscrip, Rebecca Passonneau, Lois Putnam, Kelley Remole, John Saxman, Eric Schoenberg, Ann Senghas, Yaakov Stern, Michael Studdert-Kennedy, Andrew Sunshine, Robert Thompson, Athena Vouloumanos, Cynthia Yang.

Questions pertaining to this transcript should be sent to the rapporteur via email:

jsp2003@columbia.edu



24 APRIL 2003

The Mapping of Sound Structure to the Lexicon: Evidence from Normal Subjects and Aphasic Patients

Sheila Blumstein

*Department of Cognitive Science
Brown University*

This research explores how listeners map the properties of sound to the lexicon (the mental dictionary) and investigates the neural basis of such processing. Our experiments with both normal subjects and aphasic patients examined the effects of phonological and acoustic-phonetic structure on lexical processing. Specifically, we investigated the extent to which phonological and acoustic-phonetic modifications of an auditorily presented prime stimulus affect the magnitude of semantic priming to a real word target in a lexical decision task. Results from normal subjects suggest that:

- activation of the lexicon is graded;
- both phonological and acoustic-phonetic structure influence lexical activation;
- the prototypicality of an exemplar member of a phonetic category influences the degree of lexical activation; and
- acoustic-phonetic structure activates not only its lexical representation and lexical network but also the lexical representation and lexical-semantic network of its competitors.

Results from aphasic patients suggest that they have deficits in the dynamics of lexical activation. Broca's aphasics appear to have an overall reduction in lexical activation, whereas Wernicke's aphasics appear to have an increase in lexical activation or a failure to inhibit lexical candidates. The neural systems underlying lexical activation will be considered.

I am going to talk today about the mapping of sound structure to the lexicon, largely focusing on data from both normal subjects and aphasic patients. I want to start with a mini-presentation on language in the brain because until the advent of technologies like neural imaging, most of what we knew about the neural basis of language came from research in aphasia, the so-called "lesion-based model" of language processing. Accordingly, the left hemisphere is dominant for language, and within the left hemisphere, language functions are located primarily in the perisylvian area, the supramarginal gyrus, and Broca's area. Damage to those areas will produce a language deficit, but the nature of the deficit varies as a function of the location of the lesion.

The classical model (Geschwind, 1965) makes the claim that there are structural/functional dichotomies where damage to anterior language areas, Broca's area, will result in expressive language deficits, including difficulty with articulation and production agrammatism, in which grammatical function words are deleted or substituted. Damage to the posterior areas, Wernicke's area, produces deficits in language perception and comprehension. Having said that, my talk today will be about the extent to which current research supports such a anterior-posterior distinction for language functioning. I will be talking about how auditory language input maps onto the lexicon, and I hope to convince you that we can demonstrate lexical processing deficits for patients with damage in both Broca's and Wernicke's areas. In other words, the lexical processing system is distributed across anterior and posterior areas. Then, the question becomes, what is the nature of the deficit—is there a difference in the functional architecture of these anterior and posterior areas?

Here is a sort of cartoon characterization of auditory lexical processing:

1. Sound input at the ear.
2. Spectrotemporal mapping.
3. Phonetic/linguistic mapping into a featural/phonemic domain.
4. Lexical mapping for word candidates.
5. Word selection from among word candidates.

The stages of lexical processing can be studied using the semantic priming paradigm. The general idea is that you have a word like CAT, which is converted into a phonetic representation that is mapped to the lexicon to activate a lexical entry for CAT. This lexical activation includes the meaning of CAT and the lexical-semantic network associated with CAT, which might include other lexical entries that are similar in meaning. Therefore, this mapping onto the lexicon could lead to partial activation for the lexical semantics of DOG. That is the essence of the semantic priming paradigm.

Here is how the semantic priming paradigm works: A subject hears 2 successive items, a prime and a target, and is asked to make a lexical decision for the second (target) item. The dependent measure is the response time to make the lexical decision for the target item. Half of the time, the target is a word, and half of the time, it is a non-word. When presented with prime-target pairs like CAT-DOG versus NOSE-DOG, response latencies to DOG are faster when the prime is a semantically related word like CAT. This semantic priming effect is relatively easy to show and very robust under many conditions. Underlying this is the notion that the lexical processing system is a network of connections exhibiting activation and inhibition. When CAT activates its node in the network, the activation spreads to semantically related items like DOG, and the response to DOG is faster when preceded by CAT because it is already partially activated.

We can use this paradigm to study phonological encoding in lexical processing by manipulating features of prime words. These experiments address how changes in phonological input affect lexical mapping. The primes could be of three types: a word related to the target (CAT-DOG), a word unrelated to the target (NOSE-DOG), or a non-word that is phonetically similar to

a prime (GAT–DOG or WAT–DOG; each non-word is one or more distinctive features different from a semantically related word). We already know that CAT–DOG shows semantic priming, and the question here is whether phonetically similar non-words show semantic priming? This test has the potential to tell us something about the nature of the mapping from the phonological code to the lexicon—are words the only entities that map onto the lexicon? The first set of results, with normal subjects, indicated that non-words can serve as primes, although the priming effect is smaller in magnitude than when real words serve as primes. This finding suggests that the activation of the lexicon is not all-or-none—the extent of the activation is graded as a function of the goodness of the input’s sound structure. There is graded activation of a lexical representation, which is influenced in this example by the phonological distance of an item from a lexical entry.

In the second series of experiments, we examined the performance of aphasic patients on semantic priming tasks. Now, Broca’s and Wernicke’s aphasics behave very differently on many tasks, particularly with auditory input, so we should expect to find some differences. If anything, perhaps Broca’s aphasics should show normal performance. The results show that both Broca’s (nonfluent) and Wernicke’s (fluent) aphasics show regular semantic priming with word primes. Their performance diverges when non-word primes are introduced. Broca’s aphasics do not show priming with any non-word primes (GAT, WAT, or NOSE). Wernicke’s aphasics do show priming effects with phonologically similar non-word primes (GAT and WAT, not NOSE), and, interestingly, these primes were as effective as word primes. The results suggest that both groups have deficits in the dynamics of lexical activation: Broca’s aphasics have an abnormally stringent criterion for lexical activation, and Wernicke’s aphasics have an abnormally loose criterion for activation.

Many researchers who study lexical processing appear to assume that the mapping from sound structure to phonological features is relatively straightforward and uninteresting, and much of their work focuses on higher levels of processing. I love speech, and the acoustic-phonetics of speech, and in my own view, it has been vastly underrated because it is so important. The fine acoustic details are not just extra stuff to be discarded as chaff in this world of wheat that we are searching for. I hope to show that the spectral mapping is important enough to influence not only the lexicon, but lexical-semantic activation. How can we begin to assess this proposal? It is usually assumed that words prime both semantically related (for example, CAT–DOG) and phonologically similar (for example, CAT–MAT) words. It has also been shown through pioneering work at Haskins Laboratories that the categorical sound structure of language is relatively stable. For instance, sounds labeled BA and PA differ by only a few milliseconds in the acoustic parameter voice onset time (VOT). The perceptual boundary between BA and PA is very sharp, in that only a small change in VOT can completely alter the perceptual experience. However, let me point out that when we speak, we are not like machines, we produce a range of VOTs for different productions of both BA and PA. Thus,

each talker produces variable input for a potential listener. Although perceptual processing is categorical, activation of the lexicon is graded and depends on the quality of the prime. Therefore, if perception preserves intra-category variability in some form, then some category members ought to be better exemplars than others, and show graded priming relative to poor exemplars.

These next experiments asked how variations in phoneme quality affect semantic priming. Quality was manipulated by varying VOT in voiceless consonants. For example, we removed about $\frac{2}{3}$ of the VOT from the middle of a good /k/ in CAT, leaving the burst and other fine acoustic details intact, to make a poor C*AT. We also confirmed that all of our manipulated stimuli were still labeled as voiceless despite these manipulations. The idea is that these poor exemplars will still activate the semantic network, but that activation will be somewhat dampened relative to the good exemplars. If you believe in spreading activation, and if you believe that acoustic-phonetic structure should show some influence, then a poor exemplar of CAT will activate the lexical entry and all of its competitors less than a good exemplar, but it will still show priming, just at a lower magnitude.

In the experimental design, we had three priming conditions, good CAT-DOG, poor C*AT-DOG, and neutral NOSE-DOG (and again, half of the trials had non-word targets). Note that the voiced alternative to the manipulated token in this condition would produce the non-word, GAT. We also introduced another condition, in which the alternative produces a voiced lexical competitor, as in TIME versus DIME—in this case, T*IME-CLOCK. Here, we are asking whether the presence of a voiced lexical competitor might have an additional influence on the magnitude and/or the latency of priming.

In order to examine the dynamics of priming in greater detail, in one experiment there was a 50 ms interval between prime and target, and in a second experiment there was a 250 ms interval. Overall, our acoustic manipulation did produce a significant reduction in the magnitude of semantic priming. Also, in the competitive versus the non-competitive comparison, latencies were slower when there was a lexical competitor. However, these effects were only found at 50 ms ISI; by 250 ms ISI, there was no effect of the acoustic manipulation. These results suggest that acoustic-phonetic structure and lexical competition affect lexical activation, but the effects are short-lived (less than 250 ms).

We also examined the performance of aphasic patients on this task. Broca's aphasics performed like normal listeners in the 50 ms condition: They showed significant priming and there was a significant reduction in priming for low-quality stimuli. This finding shows that these patients are sensitive to acoustic-phonetic structure and can map onto the lexicon just like normal listeners. However, in the presence of a lexical competitor, Broca's aphasics no longer show semantic priming by low-quality stimuli. This failure may be due to the increased lexical competition induced by the voiced competitor, which activates its semantic associates. Another experiment more explicitly showed that acoustically modified words prime their lexical competitors in normal

listeners (for example, T*IME primes DIME, which primes PENNY, but good TIME does not prime PENNY). Therefore, this acoustic modification is having a cascading effect all the way up the lexical identification system. Moreover, this effect persisted over the 250 ms interval. Broca's aphasics, however, did not show semantic priming of lexical competitors with the acoustically degraded stimuli. The experiments illustrate the following:

1. the lexical network is a highly interactive system, involving both the expressive and receptive elements of the classical system.
2. there is graded activation at multiple levels of the network.
3. low-level acoustic information affects lexical semantics.

In the next series of experiments, we started looking more directly at the neural substrates of lexical processing using fMRI. There has been a great deal of imaging work on lexical semantic processing, and this work has shown both anterior and posterior activation duration lexical decision tasks. Nevertheless, the work has focused primarily on frontal activation. Some researchers have gone so far as to claim that lexical semantics is a frontal-mediated process. Many of these studies used poorly designed behavioral tasks—in neuroimaging I cannot emphasize enough that tasks are so critical. The areas that are activated are critically tuned to the subject's task performance. Most studies to date have asked subjects to make overt decisions about words. For instance, one kind of task, verb generation, asks subjects to provide a related verb for each noun presented (for example, CAT–MEOW), and another asks subjects to decide whether a noun is abstract or concrete, or to indicate when a word is “a member of the category of animals that resides in domestic homes in the United States.”

The question is whether this frontal activation is due to active involvement of the lexical-semantic network, or whether it can be attributed to some executive functions required for holding this information and making some overt decision. We decided to look at semantic priming in a lexical decision task because the semantic relationship between the words exists, but a subject does not make an overt decision about the semantics of the prime word. We hypothesized that frontal activation would be minimal in the semantic priming task because the semantics are implied and no explicit semantic decision is required.

Professor Remez: Does anyone have a model of the processes that intervene between the experience of identification and the report that a word was heard?

Dr. Blumstein: I do not think so, not any that I know of.

In the specific semantic priming task used with these neuroimaging experiments, all primes were words, either semantically related or unrelated to the target. Half the targets were words and half were non-words. We tested 15 subjects in event-related fMRI, which allows us to present the materials in the same manner as we would in a behavioral study with random presentation of the stimuli. Subjects were required to indicate whether the targets were words or non-words. All subjects showed semantic priming, in that related targets

decreased choice latency. Unrelated, related, and non-word conditions all activated the same neural systems to different degrees (relative to a resting baseline): largely the left posterior and frontal language areas. It appears as though the unrelated primes produced more activation than did the related primes. Using more quantitative analyses, we found that word targets produced more activation than non-words in the middle temporal gyrus, the cuneus, the angular gyrus, and the left anterior cingulate. In no area was there significantly greater activation for non-words than for word targets. For unrelated versus related priming conditions, we found greater activation for unrelated than for related in the pre-central gyrus, the superior temporal gyrus, the middle frontal gyrus, and the cingulate.

We were also able to examine the hemodynamic time-course of activation across 8–10 s in the superior temporal gyrus. This analysis also indicated greater activation for unrelated than for related words.

Dr. Sidtis: It is common in event-related fMRI to impose a hemodynamic response function. Do you know if that was done with this dataset?

Professor Blumstein: We did that in the earlier analyses, but for the time-course we just show the time-series that does not impose a gamma-function.

Professor Remez: Another aspect of your data is not only that the unrelated peaks are higher, but that the decay lasts longer—how should we interpret that difference?

Professor Blumstein: We interpreted it as modulation of neural activation as a function of whether words were semantically related or not. We did not go any further because there was no way to establish the reliability of the function statistically. Most of the time in the literature, researchers simply show a time-series with no analyses, but that won't last.

Dr. John Sidtis: Just to follow up on this point, it is actually meaningless to do that because fMRI is absolutely not quantitative, and I think the confusion in your earlier slides where it looked like more activation across conditions is due to a difference in the baseline activations. This difference in the peaks here may likewise reflect a shift in baseline. I would say it is not possible to tell and that is a general problem with fMRI datasets. We can talk about this more later, though.

Professor Blumstein: The pattern of greater activation for unrelated than related words is consistent with repetition priming data indicating that the magnitude of activation decreases with successive presentations of the same word. This pattern seems counter-intuitive—why would there be less activation with identical or semantically related repetitions? It does make sense if the spreading activation enables the second item to be accessed with greater neural efficiency, hence less activation.

To conclude, the fMRI data suggest that posterior (temporal parietal) brain structures are the areas where lexical semantic representations are stored. There was no significant activation of the inferior frontal gyrus, which was

implicated in earlier studies of lexical processing. This means that the inferior frontal gyrus may only be activated in tasks where there is an overt semantic decision. The inferior frontal gyrus may mediate semantic working memory or other semantic executive functions. Our next step is to determine whether the inferior frontal gyrus would be activated if subjects were required explicitly to compare the relatedness of the prime and target. I would like to stop here and take some questions now.

APPLAUSE

Questions

Professor James Magnuson: Have you tried to think of tasks other than the lexical decision that would help you get away from the frontal demands? The lexical decision task still requires the subject to make a meta-linguistic decision.

Professor Blumstein: I am so new in functional neuro-imaging that I have not really tried. It is going to be hard to find a task that is not meta-linguistic. I think it is a good idea not to just do lexical decision tasks. Some people actually have looked at phoneme monitoring, but that is also meta-linguistic.

Professor Magnuson: How about eye-tracking?

Professor Blumstein: There is a graduate student in our department doing that with aphasics right now. Yes, that is a great idea; a much more natural task.

Professor Michele Miozzo: It is intriguing that you did not find areas that are more activated for non-words, whereas if you look at the reaction times, it takes longer to respond to non-words. Is it possible that it takes longer to respond to non-words because there is overall less activation?

Professor Blumstein: Sure. And a nice feature of the data is that although non-words take longer to respond to, there is no increase in neural activation. In many paradigms, the amount of neural activation tends to be correlated with the difficulty of the task.

Dr. Jont Allen: Has anyone looked at the confusion matrices of the nonsense CVs? That is basically where I thought you were going. Why not just look at phone errors as a function of your VOT distinction, and also add noise and damage the words that way. This would get around the whole reaction time paradigm, which is hard to model in a quantitative way.

Professor Blumstein: The problem is that the aphasics make many production errors, and so we do not want them to have to speak.

Dr. Allen: You could use a touch screen or a mouse.

Professor Blumstein: Well, we would have to make sure they can read. A lot of aphasics have trouble with reading. One thing we did do, which is not exactly the same as what you are suggesting, is establish that our results are not simply due to any manipulation that makes the tokens less good. Therefore, we looked at whether noise would affect the magnitude of priming, and it does not. We have also looked at changing the speaker between prime

and target, and that also did not affect the magnitude of priming. A confusion matrix might tell us something more about the perception of the sound structure, but not about the mapping onto higher levels of meaning beyond that initial perception.

Dr. Morsella: I have two questions. First, is semantic priming due to feature overlap as opposed to being simple associates? Two, what if you presented a bad /t/ but also presented visual information suggesting a /d/, as in the McGurk effect? Would this prime both /t/ and /d/ or would this only prime /d/ because the perceptual experience would be of /d/?

Professor Blumstein: As for the first question, I do not have a clue; we are not making any claims in that regard. As for the second question...

Professor Remez: You can actually get priming of PENNY with TIME even if it is the best example of TIME. By moving the phone to a slightly imperfect example, you are just increasing the ambiguity and pushing the priming of PENNY higher. This is Luce's work on neighborhood activation along phonetic similarity even with clear cases.

Professor Blumstein: We do not get priming with TIME-PENNY. In theory we should get it, but perhaps we did not have enough power to detect the effect.

Professor Magnuson: I think Luce's work has looked at just the phonetic priming, not semantic.

Professor Remez: Yes, that is right.

Ms. Gina Cardillo: I noticed that in your behavioral task, the primes could be either words or non-words, but in the fMRI task the primes were all words. Is there an asymmetry in activation with respect to non-words priming words versus words priming non-words? In other words, why were there no non-word primes in the fMRI study?

Professor Blumstein: We have done that. We did not see any inhibition by non-word primes. If you vary the proportion of related words, you see that Broca's aphasics are very much driven by strategic processes. So if they know that the outcome is predominantly one or the other, their behavior is very sensitive to that.

APPLAUSE

A Joint Meeting of the University Seminars on Cognitive and Behavioral Neuroscience & Language and Cognition

Place: Kellogg Center, Room 1512
School of International and Public Affairs
Columbia University
420 West 118th Street

Time: 4:00 PM

Chair: Prof. Robert E. Remez, Barnard College, Columbia University.

Rapporteur: Jennifer Pardo

Attendees: Jont Allen, Gina Cardillo, John Chen, Josh Davis, Simon Fischer-Blum, Kristin Geiger, Jessica Goldberg, Robert Krauss, Stephen Lowery, Jim Magnuson, Michele Miozzo, Ezequiel Morsella, Lois Putnam, Owen Rambow, John Sidtis, Diana Vanlancker-Sidtis, Jingtian Wang

Questions pertaining to this transcript should be sent to the rapporteur via email:

jsp2003@columbia.edu

