

Language & Cognition

2004-2005

University Seminar #68I

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. Generalization gradients in perceptual memory STEPHEN D. GOLDINGER	7
2. Spatial and temporal differentiation in Nicaraguan Sign Language: The Emergence of structure ANN SENGHAS.....	17
3. Alternative minimalist visions of language RAY JACKENDOFF	27
4. Tuning the language organ: A New perspective on the role of Broca's area in language processing SHARON L. THOMPSON-SCHILL	41
5. Baby Bayesians? Exploring the bases of generalization in human language LOUANN GERKEN.....	53
6. Listener sensitivity to fine phonetic detail in speech perception JOANNE L. MILLER.....	65
7. How do bilinguals choose one language to speak? JUDITH F. KROLL	77

23 SEPTEMBER 2004

Generalization Gradients in Perceptual Memory

Stephen D. Goldinger

*Department of Psychology
Arizona State University*

For over 30 years, researchers in perception and memory have conducted experiments on long-term repetition priming effects. Although seemingly simple, repetition priming data have informed theories of lexical access, categorization, attention, and long-term memory. Among these studies, a continuing focus has been the specificity of newly created episodic memory traces, and their potential involvement in perception of later repetitions. From a theoretical perspective, it appears that traces are encoded with impressive fidelity, and that later word perception is at least partly mediated by collected perceptual episodes. However, the repetition priming literature reveals an unsettling lack of precision, relative to comparable studies in categorization, which rely upon measures of similarity. In recent research, we have been evaluating similarity-priming relations more closely, with a goal of specifying the function relating similarity to priming. In this presentation, I will review recent experiments, all measuring token-specific priming effects, plotted as functions of psychological distance between study and test tokens. The experiments cover a range of materials (printed and spoken words, pictures, and environmental sounds) and a range of behavioral measures. In addition to manipulations of stimulus similarity, we conducted a large-scale study of response similarities, and their effects on cross-task repetition priming. The results suggest that, when cognitive processes are reasonably controlled across study and test, quite similar generalization gradients occur across experiments. The results provide a benchmark for comparison against models, and several theoretical candidates will be discussed, including MINERVA 2, Hebbian learning networks, and Adaptive Resonance Theory.

This talk is like a wedding: I am going to begin by revisiting some old data from my dissertation; then, move on to something new; a lot of this is borrowed from Roger Shepard; and, the screen could turn blue at any second. (Here, Prof. Goldinger is referring to a distressing moment before his lecture, during which the projector momentarily colored his slides blue.)

The classic view of representation in memory, broadly speaking, is that there are fixed abstractions that are taken from experience and then stand for experience in memory. For example, prototype theories claim that we recognize a new instance of a dog by computing its similarity to the best representation of dogs. With respect to word perception, this is analogous to the notion of a canonical lexicon, in which numerous experiences of a word are stored as a unit of some kind, something better than any of the individual exemplars. Strong versions of this theory include a normalizing function where variable signals from individual tokens undergo some kind of computational fix to make them match up with the template or prototype.

This kind of theory has a certain appeal. It has great stability, and we know that conceptual knowledge and lexical knowledge are stable across a lifetime. It is a powerful system, in that it uses a relatively small number of computational principles to solve a problem of great variation. Although, solve is a strong word, since we have not solved it. From the perspective of someone like Chomsky, it has great economy of representation. In his minimalist manifesto, Chomsky argued that we do not want to clutter mind with lots of redundant information, we want to have a sparse representational system. All of which, I agree with.

Nevertheless, there are other ways to solve the problem that might provide some extra benefits. To begin with the polar opposite to canonical units, I will start with an exemplar model, or, a model with episodic traces. This model says that, in fact, we do accumulate multiple traces of words, sentences, or faces in memory. Then, during later recognition, the system computes its similarity across a set of traces, rather than using only one unified representation. The advantage here is that it is more robust in accommodating variation than a purely canonical system. It does not have to solve the computational problem of getting back to an abstract representation. It also relates perception and memory in a way that is natural, that seems right—It is obvious that perception is the achieving of something in memory and that memory is the recreation of something perceptual. They are simply different expressions of a common system. Exemplar representations elegantly provide a natural basis for context-sensitivity. Exemplar models can exhibit the classic prototype phenomena. Most importantly, they explain specificity effects in memory, which prototype models simply cannot explain.

This is an old issue, Richard Semon's theory of memory from the late 1800s, which was rediscovered by Dan Schacter some years ago, was the first attempt at a memory system that was both stable and could handle idiosyncratic events. Memory for specific tokens is quite good in some domains, for instance, pictures, music, event frequency, physical dynamics, and faces. These effects are difficult to explain without some kind of exemplar mechanism. Although it is the case that specificity effects arise in recognition memory, they show up most robustly when people are not actually trying to remember, as in perceptual priming. A study by Carolyn Cave in 1997 asked people simply to name objects in pictures, and up to 48 weeks later, there was still robust priming for naming the old pictures—people were faster at naming pictures of objects that they had seen just once before.

Specificity effects do emerge in memory for linguistic episodes, but they are more difficult to find. Kolers did some classic work on transformed text in which people had to learn to read as fluently as possible passages that had undergone different geometric transformations in the text. Even after one year, there was still some savings in relearning the task, either for the original or for a different form of transformation. Note that for these two findings, it is not clear whether there is a perceptual or a behavioral benefit, and I will return to this question later on in the talk.

If detailed memory traces are created from encounters with words, then people should be sensitive to later repetitions of those details. This sensitivity should be expressed as better recognition or priming when test words match study words. These kinds of findings have become common enough that people believe they are real. However, they are not always easy to find and they can even go away. One question

about the specificity required for self-priming that has escaped examination is: How different is “different?” In the vast majority of studies, only two variants of fonts or voices are tested in an experiment. This kind of design does not provide enough information to determine whether different magnitudes of change between study and test have any effect. By now, it is well known that specificity effects are greater when perceptual changes are larger, but I would argue that there are no systematic studies of the function that relates the magnitude of psychological difference to the likelihood of priming or memory.

What I really want to talk about today is Shepard’s Law. Over a 40 year span, Roger Shepard noted tremendous regularity in generalization data across different tasks and species: Generalization of a response from some stimulus was related to the likelihood of making that same response to a similar stimulus by a decaying exponential function. This figure from Shepard’s (1987) famous *Science* paper shows data from humans, pigeons, and monkeys performing many different kinds of tasks. Similarity among stimuli is measured as psychological distance, which is determined using scaling experiments. Shepard invented a method called nonmetric scaling, which, like principle components analysis, discovers the best-fitting representation for a set of points, fitting to dispersion in a way that minimizes the error variance. This kind of scaling of the relative similarity of a set of items can be used to predict the likelihood that an organism will generalize a response from one item in the set to another. For small differences, there is a large likelihood that the organism will generalize a response from the old item to the new; for large scaled differences, there is a very small likelihood; and there is an exponential function relating all the degrees of differences from large to small.

This next figure comes from the Ig Nobel prize website [<http://www.improb.com/ig/ig-pastwinners.html>]. It describes the work of a German physicist, Leike, who published a paper showing that the foam head on a glass of beer decays according to an exponential function. In a glass of beer, there are 100% of available bubbles at the start. In the first unit of time, some proportion of those bubbles will pop, e.g., 10%. Over time, there will be fewer and fewer bubbles, and the rate of popping will go down.

Shepard talked about generalization in psychological space as following the same principle. An organism has some notion of *close enough*, and the organism tries to find out if there is any way to make two points fall within the same region. If there are many ways to plot the two points in the same region, then the organism will respond with a very high likelihood. As the points are farther apart, it becomes less and less likely that the organism would randomly decide that any way of plotting the psychological space will encompass those two points.

This next example comes from bees. Ken Cheng trained honeybees to find sugar water near a landmark. The bees enter a room in which there is a landmark they are trained to fly to, and then they could fly from the landmark to a location containing sugar water, which they like. Later, the sugar water would be moved to locations at varying distances or orientations from the original location. Foraging was expected to decrease exponentially as distance increased, and using psychological distance derived

as a function of bee vision, the likelihood that a bee foraged over longer distances dropped off at an exponential rate.

Given that specificity effects in priming experiments vary, is there a consistent relationship, and does it follow something like Shepard's Law? To address this question, I revived the performance measures from my dissertation, in which I had tested implicit and explicit memory for words and voices. There were two big experiments using 300 words recorded by 10 people (5 males, 5 females). Because I was at Indiana University at the time, and Rob Nosofsky was one of my committee members, I had scaled the voices, but I did not do the data analysis for memory performance appropriately. I discovered this error recently while reviewing a paper that contained the same mistake I had made.

In the study, I asked people listen to pairs of words and to decide as quickly as possible whether the pair contained two instances of the same word; the words were always spoken in two different voices. The response time to report "same" from trials in which the word was the same was used to index how similar the two voices were to each other. These similarity measures were scaled to find the best-fitting 2-dimensional solution to the data. In the plot of similarity space, the dimensions were easy to figure out—one dimension was sex (man or woman), and the second dimension was gender-relative vocal pitch.

In one experiment, I tested recognition over three delays: 5 minutes, a day, and a week. People heard the words in a study session, and then heard the words again in a recognition memory test session, and they were asked whether each word was an old or new word. They were not asked about voices, but half the time, each old word was presented in the same or a different voice. The voice effect [the decrease in recognition rate or increase in response time when the voice is different between study and test] was large at a short delay, then diminished after a day, and disappeared after a week.

With respect to the different voice, there is variation around these points—Is there a meaningful pattern to this variation around the different voice responses? Does the variation in the difference between the new voice and the old voice have an effect? In this figure from my dissertation, you can see that there is a linear relationship between voice similarity and word recognition memory. This analysis showed that some aspect of specificity was preserved and effective, and was gone after a week.

However, this is what I did wrong—This analysis used raw hit rates without any consideration of the baseline hit rate for any of the talkers for any of the words. I recently reanalyzed these data, rescaling each priming effect as a percentage of its own logical maximum: How well each word was recognized when it was presented in the same voice, and relative to that, how well was it recognized in a different voice. Here are the analyses for the best-fitting exponential function to the data plots of the relationship between psychological distance and the newly scored recognition memory. After a week, there is nothing there, which does not change that part of the story. After a day there was a reliable fit, but not very remarkable, but after 5 minutes and with robust priming, there is a pretty strong exponential relationship between talker difference and priming.

In another experiment where priming did not fade over time, I asked people to perform a perceptual identification in noise over the same three delay conditions. Again, there were robust voice effects, but here, the effect was quite stable over the course of a week. After reanalyzing these data, they also fit an exponential function fairly well. In another study manipulating levels of processing, the same function fits the correctly analyzed data. Whenever the voice effect was bigger, the exponential relationship between similarity and priming was stronger.

In order to generate some new data, I recently ran some new conditions. This time, 12 talkers recorded 300 bisyllabic words twice each, and I collected paired similarity ratings among the voices. The scaling solution for these data yielded the same two dimensions as before, male versus female and gender-relative pitch. In one experiment, I collected explicit recognition data again, but I asked the subjects to respond to the word or to the conjoint of the word and the voice (the same word in the same voice). In another experiment, I used an auditory stem completion task in which I gave people the first syllable and asked them to complete it with the first two-syllable word that comes to mind. This is a Schacter implicit memory task. The data reveal reliable priming effects for the recognition memory tests, especially in the conjoint task, and for the stem completion task. Plotting these data against the perceptual differences reveals an exponential decay function for all three tasks.

Are there similar effects with other kinds of sounds, or is this specific to speech? Others have shown that form-specific priming occurs with environmental sounds. We recently tested listener sensitivity to the magnitudes of acoustic changes within different types of sounds. After a search of various sources of acoustic samples, we collected sets of around six stimuli each in four classes, dog barks, bird calls, sneezes (one horse, the rest human), and clock chimes. There were also filler sounds, two each from six classes. For the critical sets, we obtained similarity ratings among members of each class and scaling solutions with reasonable, but not great fits.

In a memory experiment with these sounds, people heard 10 sounds of around 3 s each and typed what they thought each one was into the computer. During the surprise recognition memory test, they heard another 10 sounds, five exact matches and five changed. The task was to say whether it was the same or a different sound from the one they heard before. The performance levels for recognition of same items was good, but identification of different sounds was poor and variable. Plotting error as a function of perceptual distance reveals a very good exponential fit of .96.

The next experiment examined memory for fonts. However, fonts and voices are different because there are reasons for having memory for voices, but not much reason for font memory. It is interesting that font effects show up anyway, although they come and go in the literature: There are stronger effects when more fonts are used and when more different fonts are used, but the relation is inconsistent. Font effects show up in many experiments, they seem real, so why are they so weak in the literature? If you do not want to find the effect, it is easy to choose two fonts that will make it go away. We did three experiments where we scaled similarity for 12 fonts, people named printed words displayed for 1500 ms in a study task, and then the same people named the words again in a memory test with the same or different fonts. We found the typical weak, but significant font effect that is usually found with these kinds of

materials. We plotted memory performance against similarity and once again, an exponential function fits the data quite well.

Taken together, there are similar results for auditory word perception and for verbal and nonverbal materials. This suggests that highly detailed, multidimensional memory traces are created in perception and affect later perception. There are many kinds of theories that could explain these data to varying degrees of satisfaction. The least satisfying class that I have used are the strong exemplar models. This kind of explanation is not satisfying because it is not perceptual—the modeler creates the memory traces and tests activation later. It would be nice to move beyond the sort of existence proof that exemplars have some effect and move toward something that is more meaningful relating memory and perception.

There are better accounts that might be able to do the same thing. Distributed networks, particularly the kinds of Hebbian learning networks that McClelland and Stark and McClelland and McClelland have been working with recently, are capable of creating prototypes while keeping enough variation from individual traces so that one can find exemplar-type effects with them. I am planning to run a simulation to see if they would follow an exponential decay function. Another alternative model has both the abstraction and the episodes. This is a different kind of connectionist system based on an analogy to a fast-learning hippocampal network and a slower-learning neocortical system; and over time, as the systems try to integrate recent memories into long-term memory the exemplar details are lost.

A third alternative involves self-organizing maps, including Grossberg's ARTMAP. In this conceptualization, a connectionist network exhibits an attractor space for the more stable aspects, but also allows things to move around in that attractor to yield form-specific priming effects later.

All the experiments that I have described so far focused on stimulus generalization. What about actions or responses? In 1958, Shepard predicted that if an organism has to make a new response to an old stimulus, the difference between those responses would yield the exact same function as before. We have just completed a large-scale study of cross-task priming holding all words constant—no font or voice differences here. We compared within-task and cross-task priming across 4000 participants. In a recent report by Franks et al., they did a similar thing, but we did it better.

This is a new experiment that my student Camie has been running for three years involving memory for mental operations. As a pilot test, we reanalyzed some of their data. First, we found the psychological distance between different kinds of tasks, i.e., comparing different kinds of judgments about the same word (object attributes, size, liking, animacy, etc.). Fifteen subjects performed all the tasks with a small set of words, and then reported how similar the tasks were in using the same mental resources or the same sorts of dimensions. Comparing the distances among the tasks to the priming data, there appears to be a noisy exponential with a fit of .55.

Next, we performed a more extensive experiment. We collected similarity judgments for 10 different tasks using the same set of 10 words. The tasks were binary judgments covering a large range, such as animate/inanimate, natural/man-made, masculine/feminine, etc. The subjects performed all the tasks with all 10 words and then provided similarity judgments for the tasks. We scaled the similarity space for the

tasks, and then we conducted more experiments where subjects performed one task, then another task with half old and half new words, and then performed the initial task again with half old and half new words. The dependent measure is relative priming [response time difference] within the same task compared to priming across different tasks for the same words. The data were scored taking the cross-priming performance as a percentage of the self-priming (same task) condition as the logical maximum. The self-priming conditions always yielded better performance than the cross-task priming conditions. When we plotted priming by perceptual distance, we found an exponential function with a nearly perfect fit.

Across different sorts of stimuli, different modalities, and now different tasks, there are similar generalization gradients that are consistent with Shepard's Law. This law has an ecological grounding and it connects these sorts of data to a larger enterprise involving learning and memory in both humans and animals. To me, the most interesting thing is that the same function occurs in response generalization that occurs for stimulus generalization, as Shepard predicted many years ago. Taken together, these findings provide a fairly dynamic sense of what memory is doing. What we see is a system of self-organizing memory traces organized not only on the basis of perceptual dimensions; they are organized in ways that reflect how those perceptual dimensions coalesce into a response.

To explain data like this requires a perception-memory system that can naturally create slightly different memory traces that reflect both what people take in and how they respond to the stimuli—a perception-action link. I am suggesting that something like ART or ARTMAP is a good candidate for that, but it is certainly not the only one. What is important is that we are starting to understand how stability and plasticity can coexist, and how these findings that come and go in the literature can do so for trivial reasons or for more important reasons.

Questions

Doctor Bruce McCandliss: When you investigate psychological distance, you could have relative distance or you could have another kind of distance, which is distance from your experience. When you think about reading experience, people are mostly reading print from a certain cluster of fonts. You could imagine there is an abstraction process that can only interpolate, it does not extrapolate very well. You end up with an abstract notion that is based on a very circumspect population of fonts. If you look at the work of some researchers, they are looking at a very different aspect of psychological distance, like, how far off from your experience certain things are. When you go very far off your experience, you are dealing with something like a Kolars's experiment because you have so few representations there. But, if you are in the center of your experience, everything looks really abstract because there are so many representations that you needed an abstraction process that allows you to come up with a particular set of representations.

Professor Goldinger: This is like a prototype enhancement effect. In studies of categorization, like Posner's work and Kuhl's perceptual magnet effect, when you get closer to a prototype, it has gravity and it shrinks the space. This is a sort of perceptual function where differences near the prototype are virtually ignored, and then they

grow. Looking at one side of that function, it has that characteristic growth. Also, looking at self-organization in dynamic systems, as deeper attractors form in a state-space, they will pull in things that are more and more different. If they run out of control, they get pathological and suck in everything, and that is when you get to Chomskian linguistics—everything is reduced to one single atom.

Doctor McCandliss: How do you reconcile these two different notions of psychological distance?

Professor Goldinger: I do not, really, I am forcing people through experimental manipulation to give me judgments that I can use to derive the space. When scaling was new, and the world was good, many social psychologists loved the method and would ask people to scale many different things. The norm always showed up in the middle of the space, it anchored the space and people were clearly using it.

Professor Robert Krauss: One of the things that you are doing that has traditionally been done in this area is to use categories or dimensions that are fairly well established in experience. A frequent experience that people have is to find themselves in situations where they have to figure out what the dimensions are, what are the differences that make a difference? That throws it off onto some functional, rather than a perceptual, focus.

Professor Goldinger: This could be tested as a difference between experts and novices, like in wine-tasting. An expert might show a steeper function than a novice, where small differences make a larger difference. The real question is to let them say what things matter, and you might find completely different ways of organizing the space.

Professor Robert Remez: We do a lot of work that presupposes similarity—we fix similarity as an *a priori*, and then we test to find out which things perceptually are correlated with apparent similarity. The dread that I have in working this way is that similarity judgments might only stable because test subjects readily resolve variation along ad hoc dimensions. In contrast, an assumption of Shepard's model and of several semantic models is that the perceptible dimensions of objects, and consequently the dimensions available for contrasts are fixed. I have not been able to relieve this problem, and I am still working this way, but I wonder if you could provide some reassurance that similarity is similarity whether the attributes are fixed or ad hoc.

Professor Goldinger: When, for example, Nosofsky would use scaling for category learning, he had a very psychophysical approach. Every subject would come into the lab for a few months to produce their own unique scaling solution. Because it is nonmetric, whatever dimensions they are using, the distances are distances, and he did not worry about it. One of the things that Shepard proved, although I have to just trust him because I could not work through the math, is that it does not matter if the dimensions are idiosyncratic or if organisms have different-sized functional regions they find important. He was concerned that this general law was an artifact of the math, and that is was not really telling us anything about psychology—If you combine any two functions and you get an exponential. That is not true, because if you combine random functions, you actually get a power function.

Professor Remez: Yet, judging the similarity of two objects depends on finding the dimension that is pertinent to the judgment of similarity, because two objects share an inexhaustibly vast number of dimensions on which they could be compared.

Professor Goldinger: Tversky wrote about these problems in scaling, one of them being the triangle inequality: Tanya Harding is similar to Nancy Kerrigan, and Tanya Harding is similar to Amy Fisher, but Amy Fisher is not similar to Nancy Kerrigan. When you have lots of triangle inequalities, scaling is actually not appropriate. If people switch dimensions from trial to trial, you will not find a meaningful solution.

Professor James Magnuson: Given that you have some fits that are more linear and some that are better fits with the exponential function, what does that tell you about these datasets. What is the distinction between a linear and an exponential a hallmark of?

Professor Goldinger: That is a good question. I asked Peter Killeen this question: The exponential fits the data better than a linear function by a half a percentage of the variance, but a linear function has fewer degrees of freedom and is a better fit than the exponential in a theoretical sense, so what does it mean? Peter's response was that the exponential was still superior because it has a reason to be there, it derives from something greater than itself. But, the question still bothered me for the exact reason that you are talking about. The less priming there is, when it is insignificant and you should not be picking it up, it is linear and nothing really happened. When there are stronger effects, there is a steeper function. It is flattening out as you are getting less and less priming. With less priming, you are still able to fit an exponential, but you are losing the power to pick up differences.

Doctor McCandliss: I think this really forces the position that it is nothing about the stimulus per se, but about the cognitive operations that are being applied to that stimulus. Kolers's original study was titled, *Remembering Operations*. So, when you try to look for structure among cognitive tasks, which has been the bane of cognitive psychology, I am wondering if there is a way to make sense of the structure among tasks by something like shared operations.

Professor Goldinger: This is exactly where we are, in fact, Logan's instance theory, which is a classic exemplar theory, claimed that what you create is an event memory.

Professor Remez: Let us thank Professor Goldinger and adjourn.

APPLAUSE

Place: Faculty House
400 West 117th Street

Time: 4:00 PM

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

Attendees: Joseph Cesario, Per Hedberg, Jim Magnuson, Urs Maurer, Bruce McCandliss, Lisa Son, Alexandra Suppes, and Jason Zevin.

Rapporteur: Jennifer Pardo

23 SEPTEMBER 2004

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

jsp2003@columbia.edu



28 OCTOBER 2004

**Spatial and Temporal Differentiation in Nicaraguan
Sign Language: The Emergence of Structure**

Ann Senghas

*Department of Psychology
Barnard College*

The recent emergence of a new sign language among deaf children and adolescents in Nicaragua provides an opportunity to study how linguistic features of a language arise and spread. New features that arise must be successfully transmitted from one generation to the next to survive as part of the language. During this transmission, language form is shaped by both the characteristics of ontogenetic development within individual users and by historical changes in patterns of interaction between users. To capture this process, changes over the past 25 years will be examined within two domains: expressions of the manner and path of movement in motion events, and expressions of spatial location. These data reveal that, as the new language is learned, holistic and analog expressions are being replaced by discrete, combinatorial expressions. It appears that these new form-function mappings arise among child learners who functionally differentiate previously equivalent forms. The new mappings are then acquired by their age peers and by subsequent generations of children who learn the language, but not by adult contemporaries. As a result, language emergence is characterized by a convergence on form within each age cohort, and a systematic mismatch in form from one age cohort to the cohort that follows. In this way, each age cohort, in sequence, systematically transforms the language environment for the next, enabling each new cohort of learners to develop further than its predecessors.

My talk today is about the origin of the structure in language. One of the hallmarks of language is the use of a discrete digital signaling function, as opposed to analogue systems. These elements are joined compositionally to give language its infinite productive power. Another attribute is that the symbols are arbitrary in their relation to their referents. It may be that these are part of an innately endowed language ability, or we may simply have very good pattern learning abilities that enable us to detect recurrence. Language involves both general cognitive abilities coupled with cumulative cultural complexity. Each generation builds upon the set of language tools provided by a particular linguistic culture. This learning is not exactly faithful to the environment, because it takes the current structure and adapts it, building in greater complexity. In a typical language learning situation, it is very difficult to differentiate the evolutionary and environmental contributions to structure in language. The presence of a fully formed language in the environment masks the contribution of the learner.

In order to examine the contribution of the learner, it is important to study language acquisition in environments that are not optimally rich. One example is research on *homesign*, which are systems that develop in the homes of deaf children who grow up without exposure to a sign language. Often, the child is the only deaf member of the household, and together with the family, the child will develop a gestural communication system. Susan Goldin-Meadow and colleagues have shown that these homesign systems can develop some of the kinds of structure that you see in mature languages, but not all.

Homesign systems are not mature languages, but they have 1) basic word order, 2) contrasts between different kinds of subjects and objects, 3) conventions for communicating who did what to whom in a very basic way, and 4) discrete referential symbols. Thus, it is possible to find some core aspects of language in homesign. Similarly, a deaf child of deaf parents who learned sign language late in life and who were not fluent signers was able to transition into a full grammatical language that was richer than the model. In both kinds of cases, the output language was richer than the input. However, without the rich bundles of structure provided by the deaf signers, the homesign children did not get richer than simple single-level constructions. This rules against the argument that children come into the world equipped with a full grammar that is ready to go. What abilities are applied for the child to arrive at a full language?

To address this question, I want to focus on a new sign language that has arisen in Nicaragua. This situation has an advantage over the homesign cases in that we can examine multiple transitions across a couple of generations. Moreover, the materials that the children here were exposed to were much more raw than those of the child with the late-learning deaf parents. In this case, they created a language from the gestures that were used in everyday life by hearing people speaking Spanish.

Before the 1970s, there were no social structures that enabled deaf people to come into contact with each other. In particular, there were no opportunities for inter-generational contact among deaf children and older deaf people. They were isolated in their homes, there were a few clinics and schools set up that never had more than 15 children together, and deaf children were never together for more than a couple of years. Before 1970, there was no sign language available in Nicaragua. Deaf people were not allowed to marry and in fact, they are still not allowed to own property.

In the late 1970s, a new school for special education was set up and 50 deaf children were admitted to the entering class. The school also served children with other disabilities, and it was somewhat segregated by disability. In particular, the deaf children were in dedicated deaf classrooms. The educational instruction was all in Spanish, but they had no success at learning Spanish—there was none of the kind of support that is needed to teach a spoken language to these children. That first cohort of children started to communicate with each other gesturally. It was easier for them to make up their own language than it was for them to have access to this language that they could not hear. In the early 1980s, a vocational school for special education was started, which many of the graduates of the elementary school attended. As adolescents, that first cohort became a key group because they were the ones who first started to sign what is now Nicaraguan Sign Language. Each year, new children have entered the school and learned the sign language from the older children around them,

and their language has continued to grow and expand. By the mid-1980s, there were 200 children in these schools, and now there are around 800-1000 signers of Nicaraguan Sign Language in Managua.

For research purposes, I divided this group into 3 cohorts, however, this is really a continuum because there are new children entering every year. The first cohort comprises those who entered from the late 1970s to the early 1980s. The second cohort are those from the early 1980s to the late 1980s. The third cohort are those from the 1990s to today. A cross-section of these groups today reveals a record, like rings of a tree, of the stages of the language. After the first cohort developed a language and stabilized to a degree, the second cohort quickly learned what the first cohort developed and expanded it, and the third cohort learned from the second cohort. To the degree that language stabilizes at adolescence, we will see differences between these cohorts. Any gap between younger and older cohorts reflects a contribution of the newer cohort.

I am going to show you three examples of sign language conversations among pairs of signers. The first example is of two first-cohort signers who have been friends since they were about four years old. One aspect of first-cohort conversation, even if you can not follow what is happening, has to do with the form. The size of the signing is quite large, the signs are relatively deliberate and easy to segment, and there is a lot of feedback from the listener. There is not always an assumption of understanding.

[Prof. Senghas shows a video of a conversation among two first-cohort signers of Nicaraguan Sign Language.]

The next video shows two second-cohort signers, and you should notice that the signing is more in the wrist and elbow, so the space is smaller, it is faster, and crucially, there is less checking for feedback.

[Prof. Senghas shows a video of a conversation among two second-cohort signers of Nicaraguan Sign Language.]

The next video shows a conversation of two third-cohort signers. The signing space is even smaller, faster, and it has a wonderful smoothness to it.

[Prof. Senghas shows a video of a conversation among two third-cohort signers of Nicaraguan Sign Language.]

They sign at the same time, something you can not do if you need a lot of feedback. Also, the language includes a lot of mouth movements.

These differences are interesting, but it is important to investigate the differences that contribute to linguistic structure. One is differentiation across time in describing motion events, and another is about space. Spoken languages also sequence across time, but they do not segment across locations. There are a couple of ways that languages could combine things. One is analogical, in which two elements such as red and white pigments are combined together to make pink, but you can no longer differentiate the elements. Another way is to have both red and white in a sequence.

I wanted to see how this language handled events that would lend themselves to analogue representation of motion events, like rolling down a hill. This is a complex

event in which there is a manner of movement, rolling, and a path, down. These attributes are necessarily bound together in the original event. Iconic representation of such events in the language would preserve the unitary nature of these events in the referential terms. Because co-speech gesture is the basis for these signs, I compared co-speech gestures of hearing Spanish speakers with the signs denoting these events.

The subjects watched a short cartoon that showed a character either rolling down a hill or climbing up a pipe, and then they would tell the story of the cartoon to a peer, while recorded on videotape. We analyzed how manner and path were coded for each of the three cohorts—whether manner and path were simultaneous or segmented in the gestures and signs.

[Prof. Senghas shows video clips of four different people gesturing and signing.]

Notice that the Spanish speaker's co-speech gesturing expresses the manner and path of movement simultaneously. The first-cohort signer did not combine manner and path, she left out the path and only designated the manner. The second-cohort signer described a lot about the manner, then, at the end, he showed a separate sign for the path. The third-cohort signer segmented manner and path.

For each of the participants, we looked at how often manner and path were segmented and assembled in a sequence. The Spanish-speaking gesturers never segmented manner and path. First-cohort signers did it almost a third of the time, but by the time the language got to the second and third cohort signers, this was the preferred way to describe manner and path. We can see a progression from the input in which all of these events are blended and holistic, to early sign which mostly takes structure from the event, to later sign in which there are mostly segmented combinatorial sequences. The later signers are not faithfully reproducing a motion pattern from the environment, they have learned to do something else.

Next, I will move to a different domain to observe segmentation in describing simple events. In events like a man giving something to a woman, there are many things happening at the same time—a man is giving, a woman is receiving, and something is being transferred. Different kinds of events have different numbers of participants involved in them, and so different verbs will require different numbers of arguments. For example, sleeping takes one argument, *I sleep*; pushing takes two arguments, *I push you*; giving takes three arguments, *I give you a cup*. Once you designate an event and participants, the grammar has to link those participants to the event to represent the roles the participants play in the event. Different languages code these in different ways—some use word order, and some sign languages use spatial differentiation.

In this study, first and second cohort signers watched very short events involving one or two participants. They were asked to produce a sentence-length sign describing each event.

[Prof. Senghas shows a video clip of a woman giving a cup to a man.]

The events varied in the number of potential arguments involved, and I was interested in how many arguments they linked to their verbs, and what devices they were using to link them. There was a difference between the first and second cohort. The first cohort used a very strict segmented word order, with one verb for each

argument. If there were two people involved, they each got their own verb—I did not observe things like *woman give man*, but, more typically, observed something like *woman give man receive*. If there was more than one argument with a verb, it would be an inanimate object, like *woman cup give*.

[Prof. Senghas shows a video clip of a person signing *woman give man receive*.]

Second-cohort signers used sentences like that, but they also tended to use more variable word orders, like *woman man give receive*. Interestingly, the order of the items was allowed to vary so that *man woman give receive* could refer to the same event as the latter example.

[Prof. Senghas shows a video clip of a person signing examples.]

In these new constructions, word order does not unambiguously designate the relations among these arguments. Sign languages typically use signing space for grammatical agreement—you can move signs to and from specified locations in order to link them with arguments that you have associated with those locations. In order to do this, there must be differentiation in the signing space, in which the left sign does not mean the same thing as the right side.

I looked at whether these signers showed any systematic differentiation in signing direction in their sentences. They could use an unrotated representation that follows the direction the signer sees, or they could use a rotated representation that goes in the opposite direction—the event has been rotated into the signer’s perspective.

[Prof. Senghas shows video clips of a person signing examples of this.]

For each of these utterances, I looked at how consistently signers used an unrotated representation, a rotated representation, or a mixed representation. The first cohort signers as a group do not seem to be using space in a systematic way. This is unexceptional for this group, though, because they rely on word order. The second cohort was systematically rotating. The next question is to ask what these different forms mean to these signers.

The next study looked at how these signers interpret these signed sentences. They were given a set of pictures to choose from and asked which ones could have been part of the event that the signers were describing. The set of pictures always included the original event and its mirror image, plus a couple of foils, and sometimes there was more than one correct picture. The first cohort signers would choose both the correct event and the mirror image event. The second cohort signers were more selective in choosing only one of the mirror-image events, the one that would yield a rotated sign. In some cases, these were not the same as what the signer intended because the signers were not perfectly consistent in using rotated signing space.

If the second-cohort signers were not getting consistent use of rotation in their input, then where did it come from? Next, I used a task developed at the Max Planck Institute for studying spatial language. The subjects in this procedure work in pairs, each in front of an array of 12 pictures. One signer describes a picture from the array, and the partner has to pick the matching picture from their array. The pictures were designed to use the same objects and the only differences were the relative locations

and orientations of the objects. To succeed in this task, the signers must share common symbols for the objects, but also must distinguish left from right in representing spatial relations. The first-cohort signers really struggled with this task; I was actually surprised by this. The second-cohort signers breezed through the task. In fact, half of the second-cohort signers used rotated and half did not, but they were able to negotiate the terms and work with through the task. The first-cohort signers did not differentiate at all and they had many errors. The use of spatial differentiation emerged during the time that the second-cohort signers were young, and they are not using a single system across both argument structure and contrasting orientation.

[Prof. Senghas shows video clips of pairs of signers performing this task.]

What happens when a first-cohort signer and a second-cohort signer do the task together? These two people are related; they are an aunt and nephew, and they live in the same household. During the task, he notices that there is something different about the way each of them signs and she keeps missing the items. He tries to explain to her how to do the task. Remember, his cohort learned the language from her cohort, and now, he is explaining the language in terms that she can not understand.

[Prof. Senghas shows a video clip of a pair of signers, each from a different cohort, performing the task.]

This difference between cohorts can not result from a hard-wired program for grammar because they should each have the same program, hence, the same grammar. It also can not come from faithful pattern learning because he is not reproducing the same kind of language she is producing. It also is not coming from generations of cultural evolution because this is only a single generation yielding so many changes. There is something right about the transition idea, but it may apply at a shorter time-scale. At the moment of transition, when a language is being passed down, that is the moment where any existing learning biases can have their impact. With a fragile language like this one—Nicaraguan Sign Language is very young and minimally structured—the learning biases have had a stronger impact than usual. The shared biases in this case will have a greater chance of sticking when the model can have less impact.

I want to step back for a moment to discuss iconicity and the arbitrariness of the symbol. In this case, it appears that segmentation is preferred by children and iconicity is disfavored. Languages end up being arbitrary because these other kinds of pressures are competing with the relationship between the form and its meaning. Iconicity is readily sacrificed to satisfy these other functions. If language were to preserve iconicity for meanings that are very similar, this would create difficulty when a linguistic expression aims to create a subtle distinction.

Where did the structure of Nicaraguan Sign Language come from? It comes from the nature of the development of the individual. As individuals are learning the language, the biases they have will give the language some of its structure—discreteness, combinatoriality, etc. At the individual level, there is great change, despite a relatively stable environment. Historical change is the consequence of repeated individual learning pressures. Evolutionary change is not the cumulative

effect of historical change; it is really more like the pressures on a vast time-scale favored individuals who could learn those kinds of things easily. Those who learned things in a segmented way held an advantage learning languages that are segmented. That advantage could accumulate over time. In modern humans, it would ensure that any new languages that emerge would have that kind of structure.

APPLAUSE

Questions

Professor Robert Krauss: In spoken English, when the perspective of the speaker and the hearer are different, there seems to be a preference for speaker to use the perspective of the hearer. But, that is not linguistic, that is really a convention of usage, and if you make the task hard enough, the speaker will revert to his/her own perspective. I wonder whether the arbitrariness that you point to in scene rotation is not a convention of language as much as a convention of usage.

Professor Senghas: So what you are saying is the language is arbitrary, but only in the concrete domain and not the abstract domain. You could still argue that the argument structure is not parasitic on talking about location, because in argument structure it is not arbitrary, despite arbitrariness in space. I am not really wedded to saying the distinction in the location of objects is grammatical, whereas the one in argument structure feels much more grammatical. One thing that makes it feel a little grammatical is how hard it is for them to do it the other way. It is more than just designating the directional terms, in fact, all of those pairs used opposing rotations, and three of them were best friend pairs. People are consistent in which one they use, but they do not notice that people are rotations from their own representation.

Professor Robert Remez: When I take the perspective of the young people, I see each cohort as improving the structural properties that they have to use. But, when I take the perspective of the old people, I see this more as a problem of language change, rather than language improvement. We know that all modern languages change, although it is hard to see the changes as improvements, especially when they produce things like assimilations. Labov says that if you want to know the way the language is going to change, think about the thing that young people do that you most deplore. Taking the perspective of the aunt talking to her nephew, my guess is that she does not see his more highly differentiated form of production as an improvement at all.

Professor Senghas: Actually, there are better examples of this than that one. One thing about all of these utterances I have shown you today that make a left-right contrast is that the later cohort produces a subset of the earlier cohort. In cases where she would sign either to the left or to the right, he will only sign to the left. So, it will never look wrong to her. It is always one of the things she could have done. What that means is he is rejecting an option in his input that is grammatical. When I ask the younger signers if they could do it either way, they said that you could not—one way was right and the other way was wrong. In this case, the younger people see the difference, and the older people can not see it. However, Shira Katseff and I have been doing work on the number system, and the number signs have changed dramatically.

One of the changes involves the number 15, which was originally something like this [*Prof. Senghas makes a gesture with her right hand corresponding to a raised index finger followed by all five fingers*], and then became something like this [*Prof. Senghas makes a gesture with her right hand corresponding to a flicking of the index finger from the thumb.*] I asked a second-cohort teaching assistant how the sign should be, and she made the first one, but when I asked her if she should make it the second way, she replied, “Uhk! I know the kids are doing that, and what is that? It is a flick, it is not even a number!” There is definitely this sense that some changes that do look like they are involving a loss of information are wrong or sloppy.

A member of the audience (through ASL interpreter): It is important for a receiver to know which perspective a signer is signing from. Once that is established, then you can go on and you understand. Each group has ways of working out their perspective—the younger signers seem to be establishing it in a more parallel manner. You need to focus on establishing perspective, and I think that that is incredibly important in any kind of developmental language. For signers, if we are seated next to each other, we have completely opposite perspectives. If I am signing, and someone is facing me, they tend to take my perspective. However, if they move next to me, they will take on the new perspective.

Professor Senghas: This is why I do the task side-by-side. Even though it is slightly less natural, I wanted to remove any possible justification for rotating to take on the receiver’s perspective. So you have the shared perspective, and I wanted to see if there would be rotation that is not motivated by differences in perspective. In this case, the children are taking on the adult’s perspective more readily than the reverse, despite the typical finding that children are notoriously bad at perspective-taking.

A member of the audience (through ASL interpreter): Do the deaf adults have regular social interactions with each other? If not, that could be a reason that he adults have trouble understanding the children. They are not able to learn as well as the children, and they do not have interaction with other adults.

Professor Senghas: That is an important reason why I segregated by cohort, and not age. The adults over 40 do not have a lot of contact with other deaf people and their signing is not very strong. Adults between 20 and 40 have a social life entirely among deaf people, and their primary language is this sign language. It seems like under 40, they all have equivalent deaf community contact. All of the people whom I discussed are part of this community, so those first-cohort people are failing for a different reason.

Professor Remez: Let us thank Professor Senghas and adjourn.

APPLAUSE

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

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

Rapporteur: Jennifer Pardo

Attendees: Allison Brooks, Debra Cole, Aili Flint, Peter Gordon, Shira Katseff, Robert Krauss, Michele Miozzo, Alexandra Suppes, unidentified others.

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

Jennifer Pardo
jsp2003@columbia.edu



2 DECEMBER 2004

Alternative Minimalist Visions of Language

Ray Jackendoff

*Department of Linguistics
Brandeis University*

The Minimalist Program proposes to rebuild a theory of the language capacity from absolutely minimal assumptions. While this goal is absolutely legitimate, I will show that the implementation adopted by the Minimalist Program is in many respects empirically and methodologically inadequate. An alternative minimalist approach, based on more robust basic principles and a constructionist view of the relation between lexicon and grammar, offers a more satisfactory starting point on grounds of empirical coverage, learnability, and possibly evolution

There are views of linguistics afoot outside of the field of linguistics that stem from a vision from the late 1960s or early 1970s. Then, there are some ideas that are floating around that are more recent. Most of this is associated specifically with the work of Chomsky. Although I was a student of Chomsky and I think he is 150% right about some things, over the past 10-15 years I have come to think that he was wrong about some very important things as well. There are alternatives on the market that meet his primary goals better than his own way of doing things. Some of this is in my book, *Foundations of Language*, and some of it is in a book that is now in press that I have written with Peter Culicover called, *Simpler Syntax*.

The primary goal of contemporary linguistic theory is an explanation of how a child attains adult competence in language, and I think this is Chomsky's most lasting contribution to intellectual life in general. Here, we construe competence as an ability to use this very rich combinatorial system in a creative way. A theory of the child's ability to learn language is under the constraint that it must show how the child arrives at the full complexity of human language, something I call the Descriptive Constraint. The theory must account for the full complexity of adult language. The more complex competence proves to be, the more there is to explain for a theory of acquisition.

One theoretical strategy to explain acquisition is to minimize what the child actually learns. There are two ways to do that that have been tried and true for a long time. One is sort of standard science, to say that underneath all this complexity, there are really much more general principles that interact in such a way as to produce complexity; and everyone is trying to pursue this strategy. A second possibility is to acknowledge that there is a lot of complexity, even when we pare it down to the most general principles, and it is hard to explain how the child learns this, so we can claim that a great deal of this complexity is part of the innate endowment of the child. The child comes knowing that language is going to be a certain way, and the learning process involves picking a way through a rather limited range of options that the heard

language presents. This is the strategy of Chomsky's theories starting around the late 1970s and 1980s, the so-called theory of Principles & Parameters.

This latter strategy of innate capacity is in tension with another constraint on the theory. If humans have an innate ability to learn language, and chimps do not, then there had to be some changes in the potential for language over evolutionary time. These changes might have been chance mutations, or they might have been driven by adaptation—we do not know, but we would like to have a theory that also minimizes the number of changes that occurred between chimps and humans. In particular, we would like to minimize the part that is special for language. There are other things we do that chimps do not—there is some debate, but the consensus seems to be that chimps do not have a theory of mind, they do not imitate and they do not point. Those are things that you need to acquire language, but they are not strictly linguistic.

Anything that serves the broad faculty of language is free and does not have to be explained as part of the theory of language. Whatever is left over is the part that is really special for language, which we might call the narrow faculty of language or Pinker's language instinct or what linguists sometimes call Universal Grammar. This is what we would like to minimize by saying that there were minimal changes in the course of evolution. This evolutionary constraint gives minimalist inquiry its empirical bite: Linguistic theory should be more than just elegant, there is an empirical question of how much change in the genome is necessary from chimps to humans to get the ability to learn language.

The hypothesis of the Minimalist Program, which is the latest incarnation of Chomsky's (1995) line of thinking, is that the narrow language faculty is in some sense perfect. It satisfies the Descriptive Constraint, it manages to map between sound and meaning, with an absolute minimum of machinery. The complexity of language arises only by virtue of interactions with independent properties of sound and meaning (that is, Broad Language Faculty). The idea is to get rid of as much of the richness of the Universal Grammar posited by Principles & Parameters, and still derive the same results. I think this ongoing program is not very successful, but some people think they are making progress.

For many people outside of linguistics, and certainly outside of Chomsky's narrow circle, this latest version has seemed like a recantation of his basic principles because he has spoken for years about a rich innate language capacity, and all of a sudden he is presenting a very stripped down theory. Is this the right move? It is one possibility, but it is discrepant with the actual complexity of language.

There is another strategy for satisfying Descriptive Constraint besides positing that the child actually has less to learn. Find a way to formulate complexity in adult grammar so that more of it can be learned. Set it up so that acquisition is more graceful than it is in the standard versions of generative grammar. One way to do that is to minimize the elements of syntactic structure for which the child has no evidence. For years we have taken for granted that syntax is full of pronominal elements that have no pronunciation. For example in a sentence like, *John tried to leave*, how do you know who is the subject of the verb, *to leave*? You know it is John, but the way that is incorporated in the theory is to say that *leave* actually does have a subject, an invisible pronoun PRO, which then refers to John. Now, the child has to know that there is a

PRO there, even though the child does not hear it. That ability requires more principles somewhere in Universal Grammar.

Another even grosser example is the standard stock and trade of Universal Grammar, the distinction between deep structure and surface structure. Surface structure is the order of words as you hear them, and deep structure is some other syntactic structure that is more canonical, more regular, closer to the meaning in various ways. Then, there is a sequence of operations on deep structure that moves things around and deletes and copies things and so forth, yielding the surface structure. To say that the child learns this implies that the child infers from the signal all of the hidden covert structure. If we can have a theory of linguistic structure that does away with that, then we can simplify the problem of acquisition by having things that are more learnable.

From the point of view of mainstream generative grammar of the 1960s, How is it possible to do without null elements & movement? That was the great advance of Chomsky's linguistic theory in its technical details. Over history, the main motivation for having these covert levels of structure is because we know that the surface structure of language does not conform to the meaning. The active and passive constructions have basically the same meaning with very different surface forms. The main advance in generative grammar was to account for that by saying that they are both derived from the same underlying form, and things moved around. So, there are all these mismatches between form and meaning, and those are now going to be encoded in the relation between this covert level of syntax and the surface structure. The covert syntax will be more-or-less homomorphic to the meaning, and that has been a presumption from the beginning. The strategy whenever you find some semantic distinction that is not expressed directly in surface form is to assume that the distinction is made in underlying syntax and things are moved around to get the surface form. This is a heuristic dating back to Katz and Postal in 1964.

An alternative is to encode these mismatches directly in the relation of meaning to surface form. Back in the 1960s, we did not have a theory of meaning that was stable enough to do that, so it was mostly a matter of speculation. It made more sense to do it in terms of syntax. Now, we have somewhat more robust theories of semantics and perhaps can bring it off. For example, instead of mapping the active and the passive from a single underlying syntactic structure, which corresponds to the meaning, you start with the meaning and say that there are two different ways of mapping it into syntax. The default way is an active, but then there is this other way that you can use if for some reason you want to put the patient in subject position, perhaps to *topicalize* it.

Is that the same thing as a movement transformation? It is, because you have to encode the same mismatches one way or the other, but you save a component in the grammar. A transformational grammar has a simple mapping between meaning and covert syntactic structure, and a complex mapping between covert syntax and surface form. In a direct mapping theory, you go right from meaning to a complex mapping to surface form. From the point of view of minimality, you should prefer the direct mapping theory, if you can bring it off. Can you bring it off?

There are theories that do it that way, the most prominent is Head-driven Phrase Structure Grammar. It does this technically, formally, and very precisely as a direct mapping theory. The problem with it is that the theoretical framework in which it is

usually couched does not say much about the primary goal of explaining learnability. There is not much language acquisition research and no discussion of evolution. Nevertheless, if one could adapt those techniques of description to a theory that takes the primary goal seriously, then we would be in business.

We could say that direct mapping is a priori superior, but is it empirically superior? As one example, we can look at the passive form. If you think of the passive as a syntactic movement that exchanges the subject and object positions, that commits you to say that you should always find a subject. For example, *The dog chased the cat* becomes *The cat was chased by the dog*. However, there are many passive verb phrase constructions where there is no evident subject:

1. Dick had John followed by the FBI.
2. The man followed by the FBI is my brother.
3. My brother heard insults shouted at him by the cops.
4. Followed day after day by the FBI, John went slowly nuts.

These are all cases where there is no evident source for the passive in a corresponding active sentence.

The movement theory can be salvaged by positing a null (or deleted NP) that has undergone movement, but that makes the theory less than minimal—adult is obliged to know more about syntax. The theory of the passive in terms of syntactic movement commits you to its being semantically blind. The whole point of syntactic transformations was that they are indifferent to semantics. Unfortunately, there are cases of passives where the lexical semantics plays a role, the less canonical prepositional passives:

5. The bed was slept in/on/*under/*beside by John.
6. The telescope was looked through/*inside by the technician.

The good constructions are about the proper function of the surface subject. There is some semantic dependence between the surface subject and the verb and preposition, which is very hard to characterize in any terms other than semantics. In the classic approach, the transformation can not know about that at all. So, a theory of transformational movement has no way to handle this problem. In a theory of direct mapping, you still have to say that there is a semantic dependence for this kind of passive, but at least you can account for it because you are mapping directly from the semantics, where that information is available.

Another problem with transformational movement, which was recognized very early in the learnability literature, is that learning structural descriptions of ordered transformations is one of the most severe obstacles to language acquisition (Wexler & Culicover, 1980). Learning sequences of operations that can be potentially unbounded presents severe difficulties. Wexler and Culicover (1980) showed that grammar learning could only occur by supplementing the theory of movement with very heavy restrictions on where things could move, which had to be innate. In order to have transformations, you have to have a repertoire of rich constraints on movement that somehow emerge from the genome.

Next, I want to look at the implementation: If there are these tree structures, how does the theory describe their construction? The heuristic from the evolutionary constraint is to have the most minimal mechanisms. Let us see how minimal the

Minimalist Program actually is at ending up with complicated recursive structures. The simplest possible way is to take two constituents and stick them together into a tree and give it the name of one of the elements. Then, keep adding more trees, an operation called Merge: Take A and B and create either [_A A B] or [_B A B] or [_A B A] or [_B B A].

You can see how applying this over and over again results in big trees, but how do you get this process started? You start with a numeration: Take a set of elements from the lexicon, in a bag like *Scrabble* letters, then pull something out and pull something else out and build those into a tree, and so on. The basic process is pulling items out of the lexicon. What is in the lexicon? Minimally, the lexicon comprises words and/or morphemes, coded nonredundantly. All redundancy in the lexicon is squeezed out into rules. Chomsky quotes Bloomfield here: “The lexicon is really an appendix of the grammar, a list of basic irregularities.” Bloomfield says exactly that, but he does not say it is nonredundant, I think Chomsky reads that in.

That sounds very plausible and minimal, but let us see what that approach presumes.

- i. Organization of syntactic structure is to be characterized in terms of putting pieces together one after another sequentially. (Derivational, a.k.a., proof-theoretic)
- ii. Binary branching is the optimal and minimal kind to have in syntactic structure.
- iii. The lexicon is nonredundant.
- iv. There is a strict division between the lexicon and the grammar: They are entirely different beasts, the grammar is all the regularities and the lexicon is all the irregularities.
- v. Semantics is strictly locally compositional (Fregean): The meaning of a sentence is constructed word by word, combined according to structural branching.

There are some alternatives to these assumptions that have become prevalent in other frameworks. The rules do not construct trees, they license trees. You can check whether each piece of structure and each relation among pieces of structure online is licensed by a relevant principle. This allows the possibility that pieces of structure could be licensed by multiple constraints at the same time. The constraints could conflict with ways to resolve conflicts. These are called constraint-based, or representational, or model-theoretic grammars. Is this different from derivation? Instead of building the structure, you just check the structure at every point. Chomsky usually says it is a notational variant, and that derivational is right. To a first approximation, they look like notational variants, but I want to show you that they are really different.

Here is a simple case of the difference between the two from Paul Postal’s (2004) book, *Skeptical Linguistic Essays*. He points out that there are sentences that contain pieces of non-English:

7. The space alien said ‘klatu barrada nikto’ to Gort.
8. [teenspeak:] And then, I am all like, [gesture of exasperation].
9. The sign @ was invented in 1451.
10. *Sklerf* does not rhyme with *nikto*.
11. *Jean est mangé le pain* is ungrammatical in French.

How do you turn these into sentences? These cannot be described using a derivation starting from enumeration. One possibility is that these are not English sentences, which is unlikely. Another possibility is that they are all in the lexicon, too, but the lexicon is supposed to be knowledge of English. The idea of building structure from enumeration of the lexicon gets into trouble. In a constraint-based model, there can be particular contexts that do not constrain constituents to items of English. It is a weaker theory, but it allows you to describe sentences such as these in a way that you can not in a derivational theory.

Instead of the notion of Merge, the fundamental combinatorial device for these kinds of grammars is *unification* (sort of like Boolean union on feature structures). If you take two things that share some features, when you unify them, you get something that coincides where they are the same and still has the differences sticking out: Unification of [V, +past] and [V, 3 sing] = [V, +past, 3 sing]; Unification of [_{VP} V NP] and [V, +past] = [_{VP} [V, +past] NP]. Unification is different from Merge, but you can state Merge as a special case of unify: Unification of A and [x, y] = [A, y]; Unification of B and [A, y] = [A, B]. Unify cannot be reduced to a special case of Merge. So Merge is not the conceptually simplest combinatorial operation, as claimed.

What about binary branching? If you have a ten-word sentence, and you are restricted to binary branching, then you will have a pretty high tree with many nodes in it. On the other hand, if you allow ten-way branching, you can have a tree with just one node and ten branches. Multiple branching trees require fewer nodes than binary branching trees. Which is minimal, fewer nodes, or fewer branches per node? You do not know in advance. Multiple branching recursion is present elsewhere in cognition, so it arguably comes for free. Here is an example from visual grouping:

```

xxxxx  ooooo  xxxxxx      xxxxx  ooooo  xxxxx      xxxxx  ooooo  xxxxx
ooooo  xxxxxx  oooooo     oooooo  xxxxxx  oooooo     oooooo  xxxxxx  oooooo
xxxxxx  oooooo  xxxxxx     xxxxxx  oooooo  xxxxxx     xxxxxx  oooooo  xxxxxx

xxxxxx  oooooo  xxxxxx     xxxxxx  oooooo  xxxxxx     xxxxxx  oooooo  xxxxxx
oooooo  xxxxxx  oooooo     oooooo  xxxxxx  oooooo     oooooo  xxxxxx  oooooo
xxxxxx  oooooo  xxxxxx     xxxxxx  oooooo  xxxxxx     xxxxxx  oooooo  xxxxxx

xxxxxx  oooooo  xxxxxx     xxxxxx  oooooo  xxxxxx     xxxxxx  oooooo  xxxxxx
oooooo  xxxxxx  oooooo     oooooo  xxxxxx  oooooo     oooooo  xxxxxx  oooooo
xxxxxx  oooooo  xxxxxx     xxxxxx  oooooo  xxxxxx     xxxxxx  oooooo  xxxxxx

```

This is recursive in the standard sense (and could be further embedded in an array of arrays). But there is no justification for binary branching here. The pattern is not composed of two xs and then another x and another x and another x to make a group of five xs. There is no empirical reason to suppose binary branching here, although, there is an empirical reason to suppose that there is three-way and five-way branching. What this means is that cognition gives you multiple branching. Saying that binary branching is more economical is denying yourself something you ought to be able to get for free. The alleged simplicity of binary branching is spurious.

There are some claims in the literature that binary branching simplifies binding theory, the way pronouns find their antecedents. It turns out that it simplifies binding theory by not using linear order as a principle for determining the referents of pronouns. However, linear order comes for free, so why throw that out?

There is another claim that binary branching simplifies acquisition. If you hear a string of three things, and you know that the language has to be left-ward binary branching, you get a unique structure for it. That is fine, but if you have a string of three things, and the assumption is that unless there is evidence to the contrary, it is just a three-way branch, that is just as easy for acquisition. As long as you choose one as a default, there is no reason binary branching is any cheaper than any other kind. There is no fair argument for uniform binary branching as the minimal operation in terms of either the Descriptive or the Evolutionary Constraints.

Although Chomsky often asserts that the lexicon is nonredundant, no one (to my knowledge) has ever formulated how all redundancy is squeezed out of the lexicon into rules. The best guess is that redundancy is characteristic of brain processes, it helps stabilize them and make them more reliable. In terms of simplicity, the general environment in which language sits favors redundancy.

To return to the main question: What does the child have to learn in addition to the grammar? The child has to acquire lots and lots of words. In addition, you need a learning procedure that acquires them, whether general or special or some combination. It is a very difficult task, and it is certainly not done by setting a bunch of parameters.

What else do you have to learn? Thousands of clichés, titles, etc. A few years ago, my daughter was devoted to *Wheel of Fortune*, and just for fun, I asked her to write down all the answers. After about six months, she had collected about 600 of them. They are all clichés and idioms, and titles of songs and names. These are all part of an American English speaker's knowledge. Otherwise, they would be useless as puzzles on this show. *Wheel of Fortune* shows no sign of running out of them after 25 years. What is interesting about idioms is that their meanings are not compositional, some of them are discontinuous, and they are redundant in the sense that they use known words to create a nonredundant meaning. This is a problem for using syntax to arrive at these idiosyncratic meanings, if words get inserted only at the bottom of the tree (e.g., *kick the bucket* means dying). But, idioms fall out very nicely in a constraint-based theory.

There are also noncanonical utterance types to be learned, which are not predicted by X-bar theory. These could be prepositional phrase with noun phrase constructions, like *Off with his head!* or *Into the trunk with you!* Here are some more examples:

12. *How about X?* (How about a cup of coffee? How about we have a little talk?)
13. *NP+acc Pred?* (What, me worry? Him in an accident? John drunk?)
14. *NP and S* (One more beer and I am leaving. One more step and I shoot.)
15. *Scores* (The Red Sox 4, the Yankees 3)
16. *The more ... the more* (The more I read, the less I understand).

These have to be stored as exceptional pieces of syntax, complete with some sort of special interpretation. It was these kinds of things that led to the theory of Construction Grammar. Syntax alone can have meaning attached to it, without

particular words being involved. The point is that you can not derive these pieces of syntax by deletion and movement in any account of standard phrase structure.

Finally there are other noncanonical pieces of syntax. Here are some examples for names of geographical features: the Atlantic *Ocean*, the Hudson *River*, the Mediterranean *Sea*; the *Bay* of Biscay, the *Gulf* of Aqaba, the *Sea* of Azov; Arrowhead *Lake*, Wissahickon *Creek*, Laurel *Hill*, Loon *Mountain*; *Lake* Michigan, *Mount* Washington. These are idiosyncratic productive systems, which follow rules for feature and name positions. The grammar of numbers is another idiosyncratic element: three hundred fifty-five billion, fourteen million, one hundred twenty-five thousand, six hundred thirteen. Another system is *focus reduplication*:

17. You make the tuna salad, and I'll make the SALAD-salad.
18. Would you like some wine? Would you like a DRINK-drink?
19. Do you LIKE-her like-her?
20. Are you guys, um, LIVING-together living-together?

These indicate a generic member of a category or a special marked member. Unlike most standard phonological reduplication rules, this one can copy phrases as well as words. Another one is the N-P-N construction: dollar for dollar, face to face, house by house, month after month, volume (up)on volume of phonology texts. These are productive in some respects, but they are riddled with special features. How is meaning related to the meaning of the preposition, if at all? All these little patterns have to be learned. They are rules. But they do not follow from any standard notion of Universal Grammar. Presumably every language has lots of these sorts of things.

There is a possible objection from the minimalists to consider, that these are merely peripheral aspects of grammar. The problem of language acquisition and the goal of perfection apply only to core grammar (that is, argument structure, passive, raising, long-distance dependencies, basic cases of binding). So the sorts of phenomena I have been describing are irrelevant to these accounts. Such an approach explicitly abandons the Descriptive Constraint: The theory is no longer responsible for the structure of the adult grammar (or, postpones it indefinitely). Moreover, if you have a learning procedure that can acquire words and all these peripheral grammatical phenomena, can the same learning procedure not acquire core grammar as well? Without an account of the learning of the periphery, you can not tell.

The research strategy of Minimalist Program is to idealize away from periphery, not to mention its acquisition, so it will never investigate this question. However, similar peculiarities are found in indisputably core areas of grammar. For some verb phrase constructions, where the verb normally determines all the arguments, in English these are often infested with parasites. If you look at something like, *He sang/drank/slept/laughed HIS HEAD OFF*, you can insert many verbs, but they can not be transitive. You can not say, *He drank scotch his head off*. Why? Because *his head* has taken over the object position. The idiom, *his head off*, takes over the head and particle positions and you can put a verb there, but there is no room for anything else, and it means something like excessively. There are many of these examples in which the verb phrase is parasitized by peculiar constructions to demote the verb to a kind of modifier.

One of the big victories of core grammar is to unify long-distance dependencies like relative clauses and topicalization, etc., by saying there is only one principle, *move*

wh/alpha, that moves to the front and satisfies the same constraints every time. Early on, we were originally concerned about getting the front of these long-distance dependencies right. So that the relative clause had the right stuff at the front, and it was really hard to move something and then make sure it came out the right way. The ones that were particularly vexing were the infinitival relative clauses, like *the man to whom to speak* and not *the man who to speak to*, or *the man for you to hire* and not *the man with whom for you to talk*. The interesting thing is that nobody looks at these any more. Once the movement was unified, everyone stopped worrying about them.

These particulars are not predictable from a general rule that says to move things to the front. They have to be learned. It is not easy to write rules that come up with these configurations after fronting. The attempt was abandoned with the onset of Principles and Parameters (i.e. in disregard of Descriptive Constraint). In a constraint-based (non-movement) theory, these can be learned as idiosyncratic configurations associated with surface forms – i.e. syntactic idioms with particular constructional meanings. Generalizations about long-distance dependencies are not a consequence of movement, but a consequence of relating the signature to a gap within the clause.

There is an unbreakable continuity between core and peripheral phenomena, and between core generalizations and complete lexical idiosyncrasy. There is probably a multi-dimensional space from the most general rules of verb phrases all the way to individual verbs. A theory that posits a principled difference between them is missing a deep and important fact about language (not to mention abandoning the Descriptive Constraint). A derivational movement-based theory does not lend itself to expressing this insight. A constraint-based theory does. Therefore, derivational and constraint-based theories are not notational variants, and constraint-based theories are more adequate for expressing insights about the texture of linguistic structure. Virtually all the basic properties of implementation of phrase structure in the Minimalist Program are either formally non-minimal, empirically inadequate, or methodologically unsound.

With respect to learning and innateness, the Minimalists could object that a constraint-based theory requires a proliferation of rules in order to meet the Descriptive Constraint. How does this approach address acquisition, so as to comply with Evolutionary Constraint (that is, to reduce the volume of innate components of a Narrow Language Faculty)? Let us look at the difference between a word and a rule: Both are pieces of structure stored in memory, but a rule has variables as part of its structure, which must be satisfied by unification with something else. To examine this smooth transition from idiosyncrasy to maximal generality, we start out with a fully specified piece of structure, like *kick the bucket*, we also have idioms with variables, like *take__ to task*, and we can go on up the line to more general frames with more variables, until we see things that start to look like the head parameter or X-bar theory. The core principles of phrase structure are general schemata; idiosyncratic rules and fully specified items are specializations. There can also be idiosyncratic rules that are not specializations of more general schemata (e.g. N-P-N).

What does this formulation of rules say about learning? This presents a possibility of a learning procedure that people as different as Tomasello (2003), Culicover, and Nowak (2003) and also Martin Braine in the 1970s have proposed. A child learns particular constructions holistically at first. When multiple items share a part, create a new item (that is, a rule) that consists of the constant part plus a variable

corresponding to parts that differ from item to item. More and more general schemata arise by recursive application of this process. This is much easier in a theory without movement.

What is the role of Universal Grammar in this: How is this different from plain analogical learning? It is different from analogical learning in that there is an extraction of variables. That is really important, it is the thing that has yet to be added to connectionist learning. The way I see Universal Grammar playing a role is as sort of attractors for generalizations—you are aiming in a particular direction with your generalizations, if you can get there.

What kinds of generalizations would the child be looking for?

Some aspects of Universal Grammar:

- I. Basic organization of conceptual structure, growing directly out of primate cognition (hence part of Broad Faculty of Language).
- II. The notion of words being used symbolically to communicate intentionally about perceived world – the evolutionary breakthrough (Deacon 1997). The rest is refinement.
- III. Use of Unification plus variables in stored structures to permit productivity and recursion.
- IV. Basic principles of phrase structure:
 - A. X-bar theory.
 - B. Other common alternatives such as conjunction schema.
- V. Basic default principles of syntax-semantics interface:
 - A. Semantic heads map to syntactic heads, semantic arguments to syntactic arguments, semantic modifiers to syntactic adjuncts.
 - B. Agent First order preference.
 - C. Topic First, Focus Last.
- VI. Basic principles of morphological agreement and case-marking.
- VII. Basic principles of long-distance dependencies.

Not to mention Universal Grammar aspects of phonology and morphology.

This is not a perfect system by any means, but it appears relatively minimal, given the need to satisfy the Descriptive Constraint. Unlike the Minimalist Program, this conception of grammar allows for proliferation of learned rules, under a potentially realistic learning regimen. Learning rules is mostly an extension of learning words. The narrow language faculty is sort of a toolkit that results in tendencies toward language universals.

If any approach to language is eventually going to satisfy the primary goal of linguistic theory, to satisfy the Description and Evolutionary Constraints, and to make meaningful contact with cognitive neuroscience and evolutionary biology and psychology, it will be an approach growing out of constraint- and construction-based minimalism, not out of the Minimalist Program.

APPLAUSE

Questions

Professor Gary Marcus: It seems like the point you were making with the Martian language example is equally a problem for any theory; what you need is a *use-mention* distinction. I do not understand why this distinction is more of a problem for the Minimalist Program than it is for any other account of grammar. These are mentioned instances of language, rather than using the language as such.

Professor Jackendoff: Unless you have something in the lexicon that says *mention*, and there is an invisible item in the syntax that allows you to put anything in it, this suggestion will not work. In some cases, for direct quotes in your own language, those get semantic interpretations, and the truth conditions matter.

Professor Boris Gasparov: Your discussion of the proliferation of clichés and the impossibility of separating grammar from the lexicon strikes me as being rather close to the ideas of Charles Fillmore.

Professor Jackendoff: Absolutely, he is one of the originators of construction grammar, and that is one of the tenets, that words are one kind of construction and there are other chunks that come in all sizes.

Steven Frisson: When you are attempting to build more abstracted rules, and you want to distinguish things that are idioms, which can not accept variables, from things that can, this is going to be difficult. Could it come out of the semantic-syntax interface?

Professor Jackendoff: It seems like a lot of these sort of grow out of one canonical case that is of much higher frequency than any other instance of the construction. There has to be some sort of frequency and variety sensitivity. In general, I do not know the answer to that.

Professor Gary Marcus: I am trying to figure out why I am not quite satisfied with the argument, even though I agree with so much of what you are saying. Learners could easily be seduced by certain kinds of distributional generalizations. The field of acquisition has been down the distributional path before. I do not think they should search the entire search-space, I think there are critical parts of the space they did not search. Like, what if you did not just learn the strings of syntax, but you learned them with the semantics. There is a way in which what you are proposing is at least reminiscent of these old distributional approaches. The difference between you and Tomasello, and I think you are on the right side, is to say that Universal Grammar is integrated in this; you are not just making any generalization. What is somehow funny about the structure of your talk is that it makes arguments for the construction side, but it does not really give the classic examples of why that is not enough by itself. One needs more of an integration of Universal Grammar to figure out why it is that this approach not going to fall into the same trap.

Professor Jackendoff: I think that is exactly right, which is why I am saying that if this works out, it is at least believable. One of the early points of Wexler and Culicover is that you can not get to first base without having a meaning to correlate with it. That presumes that the system of meaning is in place for the child, and that Universal Grammar is at least telling you a default way to say something. At this point, I do not think we have any hope of a learning theory for derivational grammar, but it remains to be seen whether we can work it out in this case.

Professor Robert Remez: I think this is possibly the same question that has been proposed a few times in slightly different form, but it is a psychologist's question again about learning. There are two steps in your proposal that have elements that are completely incommensurate. In one, the infant is described as learning something holistically, and in the next, the infant is said to make various parsings over sets of

elements, and the elements can not be given in a holistic description. So, where do the elements derive from?

Professor Jackendoff: I think the set of primitives have to be given, you have to know that there are syntactic categories. I think the notion of a syntactic category is not discoverable, the notion of a phonological decomposition of a syllable is not discoverable. Maybe the notion of a syllable is discoverable. I think the basic bedrock has to be given.

Professor Robert Remez: So, if you tried to describe the initial state, it would be a capacity to elaborate a sample according to a set of elements that are given *a priori*.

Professor Jackendoff: Yes, and then the question is how to do the segmentation and classification.

Professor Robert Remez: But, you can do it every which way—it is inherently multi-stable.

Professor Jackendoff: Yes, it is going include statistical properties of the environment plus the attraction of Universal Grammar plus whatever you can bring to bear from the meaning. So, you are coming at it from a number of different angles. The hope is that maybe this might work. I think there are serious questions about which things sort of go regular because you are just hearing samples of both. How does the learning extract that regularity from a few examples? I think that is a fundamental problem.

Mr. Martin Jansche: Everything that you have sketched out is in HPSG and some other formulations, and there is some acquisition literature within HPSG, by Georgia Green and some of her students. I am hoping that your approach will garner some respect for these endeavors.

Professor Jackendoff: One of the things that has bemused me for some time is that there are all these little schools of grammar who define themselves in opposition to Chomsky, and do not spend much time talking among themselves about what they have in common. I actually tried to get a number of these people together, and they did not know how to talk to each other. What I am trying to do through my work in part is to say there are elements that all these things have in common; let us sort through the additional presumptions in each framework and see what is right and not right. I have some problems with HPSG's emphasis on the sign, in that I think it is more heterogeneous—not everything is a matching of syntax, meaning, and phonology. Also, there are problems with the emphasis on heads. Some people in HPSG have been moving in the direction of construction grammar. This dialogue is a move in the right direction.

Professor Robert Remez: My corner of the field is bedeviled by this notion of *parsimony*, or what really, in fact, is *prescriptive parsimony*. The way we actually learned about this in the olden days is that a pre-requisite for application of Occam's Razor is functional equivalence. Unless you have two functionally equivalent accounts, an assessment of parsimony is premature. You can not aim to be more parsimonious, you can only make a more parsimonious choice. I would like all claims about parsimony to be postponed until functional equivalence is established. This is a cranky comment.

Professor Jackendoff: I see what you mean. I think in the book with Peter, we take the position that a lot of times people invoke Occam's Razor and they are not looking

at the true price. They are invoking it locally. We can take this condition out of binding, but what you do not see is that you need that thing anyway for seven other things, or you have to add more things elsewhere.

Professor Gary Marcus: One of the puzzles that you raise is how you can have a lot of things that seem law-like and seem to follow Universal Grammar, and then you have all these idiosyncrasies, and there is almost a continuum between them. I wonder whether this notion of analogy constrained by Universal Grammar escapes from that problem. Do you suffer the same problem if you bring Universal Grammar back into the analogical system. It is nice that you can now represent the constructions given the notation, but there is still this continuum, and you still have to wrestle over things like the gradation from law-like to idiosyncratic.

Professor Jackendoff: Suppose the input does not let you make a generalization in that direction, so you get stuck with something like, *how about some lunch*, which does not follow from anything particular in Universal Grammar. So the generalization goes as far as it can and it gets stuck in an ecological *cul-de-sac*. If there is some way that it can join the main stream, and get up to X-bar theory, it will do so, but if it can not, well, Universal Grammar does not have to tell you what the class of possible grammars is. It is more like the theory of markedness: grammars are going to be easier to learn or more robust if they have certain properties.

Professor Bill Benzon: I want to get back to semantics—how do you incorporate the notion that we inherit so many cognitive abilities with meaning in the real world? I notice that you have a forthcoming paper with Steve Pinker, and I wonder how you reconcile the evolutionary framework. Some propose that languages evolved to be different so that group members could identify each other.

Professor Jackendoff: I have been pushing the view for 20-some years that the semantic system is not restricted to language. If language is to be used for anything useful, it has to connect with the way you understand the world. At what point does language stop and this other stuff begin? There have been these huge arguments back and forth in semantics as to whether there is a specifically linguistic semantics and then world knowledge, or whether to go straight to world knowledge. I tried to argue that it goes straight to world knowledge, perhaps with a particular slant put on it by the fact that it is expressed in language. Not every attribute of language can be cast as an adaptation. Evolution can not code a whole language on the genome. It got as far as it needed to so that we could learn the rest. The fact that we use language for identification is just a side-effect of the fact that we use anything we can for identification.

Pinker and I wrote that article in response to an article in *Science* by Hauser, Chomsky, and Fitch. They claimed that all you need to add to the chimp cognitive repertoire might be just recursion. What happened to words, phonology, morphology, grammatical functions and stuff like that? They posed the right question—what do you need to add to the chimp repertoire—but I think it is the wrong answer. In fact, I think you get recursion for free. It is the words and the phonology and so on that you need.

Professor Remez: Let us thank Professor Jackendoff and adjourn.

APPLAUSE

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

Time: 4:00 PM

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

Rapporteur: Jennifer Pardo

Attendees: Bill Benzon, Dawn Chan, Frank Enos, Simon Fischer-Baum, Molly Flaherty, Aili Flint, Steven Frisson, Boris Gasparov, Peter Gordon, Nizar Habash, Julia Hirschberg, Martin Jansche, Robert Krauss, Fred Lerdahl, Chaille Maddox, Gary Marcus, Michele Miozzo, Ezequiel Morsella, Smaranda Mureson, Katherine Nelson, Rebecca Passonneau, Jeffrey Postman, Carolyn Ristau, Andrew Rosenberg, Richard Schwartz, Ann Senghas, Valerie Shafer, Michael Studdert-Kennedy, and Athena Vouloumanos.

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

Jennifer Pardo
jsp2003@columbia.edu



27 JANUARY 2005

**Tuning the Language Organ:
A New Perspective on the Role of Broca's Area
in Language Processing**

Sharon L. Thompson-Schill
*Center for Cognitive Neuroscience
University of Pennsylvania*

For more than a century, lesions to the left frontal operculum have been implicated in a constellation of linguistic deficits affecting the production of words and sentences and the comprehension of certain syntactic structures. However, the preponderance of the evidence fails to support the link between this structure, Broca's area, and this syndrome, Broca's aphasia. Rather, numerous neuroimaging and neuropsychological studies have converged on the hypothesis that Broca's area is involved in selecting information among competing alternatives. Here, I explore the possible link between this putative selection mechanism and some deficits that are commonly observed in nonfluent aphasia. The ability to explain certain linguistic deficits as a failure of a more general selection mechanism may have far-reaching implications for the study of language.

I am going to start with a very brief video clip from an experiment that I will return to at the end of this talk. You will see a patient trying to follow the instruction, "Put the cow in the bowl onto the plate." First, you will hear the experimenter say the instruction, then you will hear the patient, N. J., repeat the instruction correctly, but you will see that he does not perform the task correctly.

[Professor Thompson-Schill plays a video clip of a man picking up a cow doll and putting it in a bowl, then he says, "Uh oh."]

As you can see, he makes an error putting the cow on the plate, which he also realized at the end. Why does he make this error? It is possible that he did not understand some part of the instructions, like what "the plate" or "the bowl" means. It is possible he is having trouble parsing the syntactic structure of the sentence—understanding that "in the bowl" modifies "the cow," instead of being a destination. Today, I am going to argue that something else is going on here. Specifically, I want to suggest that the error this patient is making is better described as nonlinguistic. Rather than considering this a semantic or syntactic problem, this patient has a problem with a more general cognitive control ability, an inhibitory control process. Furthermore, this deficit is linked to the part of the brain known as Broca's area.

Paul Broca is credited not only with describing this area, but also with the birth of neuropsychology. In 1861, he wrote, "A faculty that can perish alone without those

that are nearest to it being altered is obviously a faculty independent of all the others.” That is to say, a special faculty. He outlined the idea of looking at disorders and carving cognition at its joints. The specific disorder that he was describing is captured in this quotation, “Somewhere in these frontal lobes, one or several convolutions holds under their dependence one of the elements essential to the complex phenomenon of speech, which must not be confused with the general faculty of language.” Broca was actually very clear in this paper that he was not talking about a language disorder; he considered it to be a speech disorder, and he cautioned people about thinking about this as a language area. Nonetheless, a confusion with the general faculty of language happened, as demonstrated by the name of the disorder normally associated with this area, Broca’s aphasia. Broca did not use that term, he described an *aphemia* speech disorder, but today people talk of Broca’s aphasia.

Some of the symptoms of Broca’s aphasia include problems with speech articulation (apraxia), reduced utterances (nonfluent/telegraphic speech), problems understanding syntactically complex sentences (receptive agrammatism), problems retrieving single words, and so on. The main thing to note about this classic description of Broca’s aphasia is that it has become synonymous with Broca’s area—especially as it is described in just about any textbook you can find. As it turns out in the research literature, patients with Broca’s aphasia do not all have lesions in Broca’s area. Perhaps more critically, patients with damage to Broca’s area do not all have Broca’s aphasia; according to one estimate in a 1985 paper, approximately 35% of patients with lesions to Broca’s area do have Broca’s aphasia. What exactly is this area doing? Why is the association so weak?

One possibility is due to the fact that Broca’s aphasia is really an umbrella term for a variety of disparate symptoms. They may hang together only by virtue of the fact that these patients tend to have huge lesions, affecting lots of different structures. A more sensible approach to a structure-function mapping is to pick out a single symptom and see if you can find the necessary neural substrates, rather than to try to localize the syndrome. One example of this approach examines the articulatory deficit referred to as apraxia of speech by looking for correspondences in the overlays of their lesion profiles. In a group of about 20 patients who had apraxia of speech, there was a very small area that was damaged in 100% of these patients. In a comparable sized group of patients without apraxia, the same area was implicated in 0% of the patients’ lesion profiles. Those kind of numbers are a whole lot better than 35%. Trying to find a relationship between a brain structure and a very specific function is a much more promising approach to the neural bases of language.

The other thing I want to point out about this area is that area critical for apraxia of speech is not Broca’s area, it is part of insular cortex, which is hidden between the frontal and temporal lobes. Converging evidence from a number of neuroimaging studies has supported the hypothesis that a region of insular cortex is necessary for the successful articulation of speech. Broca originally hypothesized that some portion of the frontal lobe was important for speech production. We now know that his classic patient, like the patients in this study, had a lesion that included insular cortex—it was a very extensive, deep lesion. What about the part of the brain that Broca could see from the outside, the cortical lesion? Does damage to that area cause a language deficit? What would the deficit be if the area were damaged all by itself?

There are a number of hypotheses currently circulating about the function of Broca's area. Many of these are driven by neuroimaging studies conducted over the past 10-15 years. Early on, a number of studies concluded that Broca's area plays a critical role in either representing or retrieving semantic information. Others have argued that Broca's area plays a critical role in phonology—representing or retrieving speech sounds. A somewhat related idea is that Broca's area is part of an articulatory loop for verbal working memory. From the patient literature on aphasia, there is a very prominent theory which states that Broca's area is the seat of syntax. As you can see, all of the classic sub-divisions in language have been attributed to Broca's area.

I am going to argue that all of the evidence points, instead, to a nonlinguistic function, and that the better hypothesis is that Broca's area is important for *selection*. What I mean by that is that Broca's area is important for guiding the selection among competing alternatives, in at least two circumstances: 1) in a weakly constrained situation, when indeterminacy warrants selection among possible responses; or, 2) in a situation in which one act might be the most likely but another is possible. These are each situations that would produce a conflict that needs to be resolved. My suggestion is that Broca's area is functioning to facilitate that conflict resolution process.

This idea is not entirely new with respect to the part of the brain surrounding Broca's area: the prefrontal cortex. Luria wrote, "The frontal lobes are in fact a superstructure above all other parts of the cerebral cortex, so that they perform a far more universal function of general regulation of behavior than that performed by the posterior associative centers." A more recent statement along these lines comes from Miller and Cohen (2001), "To deal with this multitude of possibilities and to curtail a confusion, it is commonly held that the prefrontal cortex is particularly important." When you have lots of competing things activated, you have to resolve the resulting conflict.

When people have talked about what Broca's area does, it has always been considered an island in prefrontal cortex. The fact that this bit of tissue is sitting in prefrontal cortex, that people have described in these terms, tends to be ignored. Instead, because of an historical association with something like Broca's aphasia, people are looking for a language-specific explanation. An alternative is to examine the evidence in the literature using these more general notions.

With respect to semantic abilities, this analysis begins with a classic paradigm dating back to the earliest neuroimaging studies, using verb or action generation tasks. These tasks involve prompting a subject with a concrete noun and asking the subject to produce an action word associated with that noun. In 1988, Steve Pederson and his colleagues reported that compared to just reading the words, the verb generation task led to increased activation in the vicinity of Broca's area. In response to this finding, we reasoned that one of the things that happens in the verb generation task is that you have to ignore all the other things you know about the noun that is used as the eliciting prompt. For some items, those other things would be very strongly activated, and for other items, they would be weakly activated. For example, when I say SCISSORS, you say CUT. When I say PIANO, you say PLAY. When I say CAT, you say ...DOG. I think I noticed an error there. The point is that many of these items have a strong association with an action response, if I asked you to say the first word that

comes to mind, you would probably say CUT for SCISSORS, and PLAY for PIANO, but there is no strong action response associated with CAT.

Another experiment manipulated the amount of competition involved in the task, and you can see that over three different experiments with the same group of subjects, we found an effect of the amount of competition (a *Selection* effect) on brain activation in Broca's area. When we take this task to patients with brain damage, comparing those with damage including Broca's area to controls with frontal damage sparing Broca's area and to undamaged elderly controls, we observed normal performance for the low selection items and higher error rates for the high selection items. An error was defined as an omission after 20 s with a reminder after 10 s, or saying something that was not an action response, like seeing CAT and saying DOG. This finding suggests that the ability to do this task is a consequence of cognitive function under competition. And, it appears that there is a tight linking of that ability to Broca's area, specifically. Furthermore, there is a correlation between the number of errors in the high selection task and the amount of damage to the posterior portion of Broca's area, Brodmann's area 44. We can explain over 90% of the variance in behavior by knowing the amount of damage to Broca's area. In contrast, overall lesion volume does not explain any of the variance in these data, nor does amount of damage to adjacent regions.

One concern pertinent to these kinds of experiments is the task difficulty. It could be the case that it is just more difficult to make a response to CAT than to SCISSORS—increasing selection demands is something that would make a task more difficult. To rule out task difficulty, we constructed a task that is not more difficult with respect to response time and accuracy, but would still be subject to selection demands. This is a priming study where we asked people to generate actions or colors, and over the course of the experiment, some items repeated in one of two conditions. For example, if I say TAR, you might say BLACK for color or SPREAD for action. In one repetition condition, subjects reported the same attribute as the first time they saw an item, and they showed a priming effect relative to novel items (approximately 200 ms reduction in response time). In another repetition condition, the attribute switched and there was still a priming effect, albeit smaller, relative to the novel items. Despite the fact that the different attribute task is easier (faster and more accurate) with a repeated than a novel item, we predicted that there would be an increase in activity in Broca's area, and only in Broca's area, because we would be making this other information about the repeated items more available. Now, when you are asked the color of TAR, the fact that you can spread tar, for example, is just a little more available, and there is a little more competition in contrast to the same attribute condition when you just reported the color of tar. Relative to novel items, we found a decrease in Broca's area activation for the same attribute condition, but an increase for the different attribute condition. So, this activation pattern is not just due to task difficulty.

Another approach to assessing the role of Broca's area in accessing semantic information uses picture naming tasks. Most models of picture naming include not only a semantic activation/retrieval process, but also a process that can be described as selecting between all of the various representations activated by the picture. Perhaps it is this latter process that is driving the activity in Broca's area, the selection process,

and not the retrieval of semantic information. In order to manipulate selection, we used normative name agreement data as measure of competition. Some pictures have high naming agreement, in which case everyone agrees to call this picture an APPLE and that picture a HAMMER, while other pictures have lower agreement, in which case this picture could be a BLOUSE, a SHIRT, a TOP, etc. This study used normative name agreement as a proxy measure of competition in a picture naming task, and there was increased activation for the low agreement pictures relative to the high agreement pictures, and the effect was specific to Broca's area.

In other semantic studies, there is evidence of increased activation in the naming of pictures of manipulable objects after naming pictures of tools. Because Broca's area is sitting in pre-frontal cortex right next to the pre-motor cortex—right next to the hand and mouth area of the motor strip—it was proposed that maybe Broca's area is storing motor representations, among them the semantic representations about what to do with tools. Following up on our previous experiment, we examined whether controlling for name agreement between pictures of manipulable objects and pictures of animals would eliminate the effect, and the answer is provocative. In prefrontal cortex, there is a big name agreement effect, but no effect of picture category. In contrast, just posterior in pre-motor cortex, there is an effect of category with a lower effect of agreement.

This outcome is appealing, because many of reports in the literature on tool-naming tend to find a swath of activation that includes both pre-frontal and pre-motor areas, and discuss the finding as if it were a single function. The current result is useful because it shows that these areas might be contributing in different ways. Moreover, half of the subjects were left-handed and varied the degree to which they used their right hand for tools, and there was a very high correlation (.88) between estimates of their use of the right hand and how much activation was found in left pre-motor cortex. This effect was specific to pre-motor cortex and also to naming tools. This is consistent with the notion that our conceptual representations are tied to sensorimotor systems, so that when we think about these manipulable objects, we are activating the part of the brain that manipulates them; and, for people who use their right hand, that is left pre-motor cortex, but we do not see that relationship just anterior in Broca's area.

We also used the name agreement paradigm (not just tools and animals) with the patient that I opened the talk with, N.J. This patient is a 63 year old male who had a single infarction to the pre-central branch of the middle cerebral artery. His lesion includes all of Broca's area, and it is relatively specific to Broca's area. He is a highly educated and very intelligent subject. Broca's aphasics tend to show slight impairments in picture naming. If we take into account picture name agreement, we can predict when those impairments will show up. Plotting percent correct picture naming (a correct response would be anything that normal controls produce) as a function of high versus low name agreement, we see that N.J.'s performance is in the normal range for the high name agreement items (low selection) but several standard deviations below the normal controls for the low name agreement items (high selection). The role of Broca's area in picture naming tasks appears to be modulated by variation in selection demands (name agreement) across different sets of items.

Next, I will turn to another way of manipulating competition in picture naming within items, using semantic context effects. In this task, subjects name items one at a time as quickly as possible. Over time, the items are repeated. In one condition, the repeating blocks are arranged in groups of semantically related items (semantically blocked). In another condition, the repeating blocks contain items that are semantically unrelated to each other. For normal subjects, the response times to name the picture are slower in the related than in the unrelated case—there is interference to name a particular item when a subject has been naming other related items. For a group of control patients with damage to other frontal areas (not Broca's area), there are some unreliable differences in the error rates for picture naming in the semantically related and unrelated conditions. For a group of patients that we believe have damage to Broca's area (we are currently verifying the anatomical speculation), there is a much bigger difference in error rates between the related and unrelated items. In other words, they have greater errors in the semantically related condition than brain damaged controls, suggesting that they are experiencing particular problems with interference. The effect grows over the course of the experiment; they start off the same and the error rate grows over each repetition.

The last experiment in this section uses a verbal/category fluency task, which is a standard test to assess frontal lobe function and semantic knowledge. In this task, subjects are asked to name as many members of a category that they can think of within 20 seconds. Previous research compared fluency for big categories, like ANIMALS and FURNITURE, versus small categories, like FARM ANIMALS and BEDROOM FURNITURE. Intuitively, it seems easier to name big categories rather than small categories, because there are more candidates. However, for big categories like ANIMALS, there are so many options that there is much more competition, and you could get stuck in semantic space. We reasoned that prefrontal cortex might be important for getting you unstuck. Normal subjects can provide many more members of ANIMALS than FARM ANIMALS. Patients with Broca's area damage might get stuck in a small semantic region and would show no increase in the number of items for the larger category. For patient N.J., the difference in the number of items retrieved for large versus small categories is much smaller than that for normal age-matched controls or for non-Broca's area frontal damaged controls. This is consistent with the notion that N.J. is getting stuck in a subcategory when asked to name items from the larger category.

Across all of these experiments, we find that in situations where there is little competition or conflict and the demands for selecting among different representations are minimized, Broca's area seems relatively unimportant. This is not consistent with the notion of this area as the Semantic organ. We should see activation there whenever there is semantic processing. Even when the semantic task is easier, and subjects can perform it more quickly, if you increase the amount of competition, there is an increased reliance on Broca's area.

More briefly, now I will turn to the notion that Broca's area is the seat of phonology, or phonological processing. In previous fMRI studies where subjects are asked to make rhyming judgments, numerous researchers have observed activation in Broca's area, and argued that it is engaged by phonological processing. Our study used two

different working memory tasks. In the semantic working memory task, subjects had to remember five words in the order in which they saw them, and were later probed whether a target item occurred in a particular position (for instance, “Was KNIFE the third item?”). Moreover, the subjects were prompted to make the judgment on the words either in forward order, or they had to reorder the items according to increasing size. In contrast, the other working memory task (which I consider a phonological task) the subjects had to remember nonwords in order or they had to alphabetize the items (more or less phonological, but clearly non semantic). There was no difference between the semantic and phonological tasks in activation in Broca’s area. There was an equivalent increase in activation for performing either re-ordering task over the forward ordered task. Why did we find no difference in the phonological task?

In our recent review of studies that directly compared phonological and semantic tasks, some found a difference in activation in Broca’s area and some did not. Two of the studies that did not find a difference, like ours, used nonwords, and there were no studies reported with nonwords where they did find a difference. The differences reported for phonological and semantic tasks could be due to the fact that the phonological tasks used words, which have irrelevant semantic information, and the subjects have to suppress this information. In our study, we compared two phonological tasks—one used words and the other used nonwords. The task was to chose which word or nonword contained the same the vowel sound as a target item. If we are correct in our interpretation of the findings in the literature, there should be more activation in Broca’s area for the task that used words versus the one using nonwords. In addition, we had an analogous set of conditions that manipulated the degree of competition for two semantic tasks, one high and one low. Broca’s area showed greater activation for the high versus low competition semantic task, and more activation for the phonological task using words versus nonwords. If this area is responsible for retrieving phonological representations, there is no reason to find more activation for words than for nonwords. There is increased activation when there is a greater task demand to inhibit irrelevant semantic information.

A related hypothesis claims that Broca’s area is an important part of the phonological loop for verbal rehearsal. This claim has the advantage that it seems to be related to other things that people have said about the prefrontal cortex, unlike any of the other hypotheses. Unfortunately, it does not have the data on its side. In a meta-analysis of spatial and nonspatial span tasks and also delayed response tasks, where there was an unfilled interval, only one study with nonspatial/verbal tasks found impairment in a simple working memory task. In contrast, when the interval was filled with some sort of distraction, there are more studies reporting impairments. The authors of this study argued that working memory is a multi-component set of processes that contribute to performance on these tasks. Some of those might best be described as mnemonic, the memory part, but some are non-mnemonic, the working part. Those could involve processes like selection and inhibition, which are very important when there is distracting information.

Like most good ideas, Luria’s 1973 book on neuropsychology said it first: “Destruction of the frontal lobes leads not so much to a disturbance of memory as to a disturbance of the ability to inhibit orienting reflexes to distracting stimuli.” This statement was based on findings with monkeys that had lesions to prefrontal cortex. It

is well-known that they are impaired in working memory tasks, but if you turn the lights off during the delay period, their performance is normal. That can not be explained by a simple mnemonic account.

This next paradigm I would like to discuss tests specifically whether Broca's area is important for a non-mnemonic component, prevention of interference. This is a simple delayed response task with a twist. The subject first sees a display with four letters; then, there is a brief delay; and, last, a probe is displayed. The subject simply has to say whether that probe was in the memory set. On some trials where the correct answer is NO, the probe item had been lurking in the previous set (Recent No). On other trials where the correct answer is NO, the item had not been present in the previous 2 sets. The logic is that in the Recent No trials, the familiarity of the item acts as a competing source of information about the status of the probe in the current trial. In both a recent PET study and an fMRI study, activation in Recent No trials was associated with increased activity in Broca's area.

We examined this effect in two ways. One way investigated inhibitory control and performance on this task in an individual difference study with college students. We gave them a self report measure called the Disexecutive Questionnaire (DEX). We were interested in whether subjects' scores on this test would be correlated with their performance on a working memory task. For overall performance, using all the items, we found no correlation. Instead, if we correlated performance on the DEX with the magnitude of the interference effect in the working memory task, we found a striking correlation. Subjects who show really big interference effects are the ones endorsing items like, "I act without thinking."

Stronger evidence linking Broca's area specifically comes from a case study with another patient, R.C., who suffered a ruptured aneurism resulting in a lesion centered in Brodmann's area 45. This is a little more anterior than I have been showing, but still in Broca's area. Comparing this patient's performance to a number of control subjects, his performance on a working memory task was similar to a number of other frontal patients. In terms of baseline working memory performance, items with no interference, R.C. performs normally. For the interference effect, his performance is 6 standard deviations outside of his age control group, both in response time and in error rate. A simple manipulation that has a probe item being shown in a previous trial was enough to reduce his performance from somewhere around 90% to around 70%. This was a huge effect. Also, across all subjects there is no correlation between baseline performance on this task and the interference effect, which shows that these are really two dissociable systems. And, the non-mnemonic component is linked to the inferior frontal gyrus. It is not the memory component per se, but something to do with competition.

Finally, I want to return to the paradigm that I began with today, and to examine syntax. The syntax argument has been one of the strongest arguments in the literature about the function of Broca's area. This is based on the fact that Broca's aphasics exhibit not only production deficits, but also comprehension deficits limited to syntactically complex sentences. Can these deficits be explained as a result of a problem with competition? We decided to study temporary syntactic ambiguity as a means of seeing the effects of competition. The sentences were structured like this

one: *Put the frog on the napkin in the box*. This is not a globally ambiguous sentence, but at the point that you hear, *on the napkin*, there is an ambiguity. When looking at this scene, *on the napkin* could be a destination for the frog or it could be a modifier of *the frog*. There is a strong prepotent bias to interpret *on the napkin* as a destination because whenever you hear *put*, you will always hear a destination. You do not always hear a modifier for *frog*, and in this display, you do not need a modifier for *frog*. However, as the sentence continues, and you hear *in the box*, you have to override that interpretation.

There is another way to override that bias, by presenting a scene with more than one frog in which one of the frogs is posed on a napkin and the other is not. In this situation, there is a bias to interpret *on the napkin* as a modifier, overriding the lexical bias for *put*. In eye movement tracking experiments, with two frogs in the display, subjects never look at the napkin as a destination. Interestingly, research with 5-year-old children shows that they are not influenced by context in this way—they express a strong preference that they do not override. This sounds just like what we think the frontal lobes are doing during language processing, and furthermore, we know that the frontal lobes are one of the last areas to develop in childhood.

We examined the possibility of competition in a couple of ways. One study with college students asked them to look at these scenes and to listen to these sentences, and we examined whether they made errors of the sort the patient and children made. Under time pressure, we found that some college students made this kind of error (13 out of 40), failing to revise their initial commitment to *on the napkin* as a destination. Looking at their performance on the proactive working memory task, the error-prone students show larger interference effects (4% more errors). Finally, we looked at performance on this task with patient N.J. He was able to complete the task with a syntactically complex but unambiguous sentence. This shows that his problem is not due to syntactic complexity. Note that he was also able to repeat the sentence, which indicates that his problem is not a sentence comprehension deficit resulting from a failure of verbal working memory. The error that N.J. makes with the syntactically ambiguous sentence is the same one that the children and the college students make. Compared to frontal damaged controls, N.J. shows a large effect of ambiguity on his sentence comprehension ability.

In all of these different areas, semantics, phonology, verbal working memory, and syntax, activation of Broca's area or deficits following damage to Broca's area are affected by increasing demands for selection. Damage to Broca's area is associated with a pattern of deficits best described as an impairment in selecting between competing sources of information. The function of Broca's area is, like other regions of prefrontal cortex, one of cognitive control, rather than a unique linguistic specialization or operation.

APPLAUSE

Questions

Professor Kevin Ochsner: Every single one of the examples you gave is verbal, yet you are arguing that the functional specialization is not devoted to language. Have you

done any experiments where you manipulate selection demands using completely nonverbal materials?

Professor Thompson-Schill: I am so glad you asked because I only had an hour, but Melissa Brandon, who is sitting right here, did exactly this experiment with pictures of faces. We used the working memory paradigm I just showed you, presenting four faces, a delay, and then a face prompt comes up. Some subjects apply a verbal label to these items, but the ones who label more are not the ones showing greater interference effects. We found increased activation in Broca's area for the Recent No (interference) trials with this nonverbal task. However, the activation was not specific to the left inferior frontal gyrus, it was also significant in the right inferior frontal gyrus, which I have never seen reported before. We do not know yet whether the left activation is necessary, so we have N.J. scheduled to perform this face task with us next month. The main point I want to make is that the function can be described as nonlinguistic, but there could be still, within prefrontal cortex, specializations, based on connectivity for the kind of input it is modulating. It could be doing this nonlinguistic function, but because of its connectivity, it tends to be acting on representations that we think of as linguistic. This is an open question.

Doctor Ezequiel Morsella: As you know, substantial research shows that the anterior cingulate is important for interference resolution, like in the Stroop task; what role does it play in your framework?

Professor Thompson-Schill: Some people have argued that the anterior cingulate is detecting the conflict, and then sending the signal to prefrontal cortex to do something about it. I think that is not entirely consistent with some of the data that are out there because you can get dissociations between them. If that is the story, then you should never get prefrontal activation when you do not also have anterior cingulate activation. However, that is not what is found. I prefer another interpretation based on some research with the Stroop task and a proactive interference task. The anterior cingulate is specialized for resolving conflict at the output level—it is really a response conflict. Prefrontal cortex is responsible for handling conflict that you could think of as conflict among representations, but not at that last stage of output. This would be consistent with the observation that you can get prefrontal conflict responses when you do not get anterior cingulate responses. I should also add that we tried to look at some of these effects in patients with anterior cingulate damage, and we did not find any of these sorts of effects.

Mister Christopher Summerfield: In the beginning of your lecture, you showed a slide in which it seemed that the overlap between Broca's aphasia and brain damage area was in the anterior insula, rather than in Broca's area. A number of recent fMRI papers talking about decision uncertainty have isolated activity which correlates with decision uncertainty to precisely that region. Are you saying that Broca's area is mediating the underdetermined responding in these types of tasks, or is Broca's area and this sort of adjacent bit of anterior insulate cortex doing it? Are they doing the same thing or are they doing different things?

Professor Thompson-Schill: I do not know the studies you are talking about, so I can not comment on them specifically. The function that others have described is not one that I would call a problem with competition, generally. It might be a problem

with competition at the phonemic planning level. Again, another sort of output level. These patients get the phonemes, but they put them in the wrong order—that is what apraxia of speech refers to. That would definitely not be described as a general conflict function. We tend not to see activation in insular cortex. It is kind of surprising because it is the kind of thing that some people have said responds as things become more automatic, which sounds like the opposite of being important for decision making.

Professor Michele Miozzo: You have presented a very impressive array of data, but it all seems to have in common the processing of word phonology. At the conclusion of your talk, I would come out with the impression that you demonstrated that Broca's area is an organ of phonological processing, or at least, involved in the interface between semantics and phonology. It deals with selecting the right word form at the right time. So, it is really a language-specific organ.

Professor Thompson-Schill: I am not arguing whether or not these findings are language-specific. I am questioning whether the *process* is specific to language. Whether all of the data can be explained by phonological processing is more problematic. I do not see how you can explain a difference in saying that tar is black when you have just said that tar is something you spread versus when you do not have that previous information by virtue of phonological processing.

Professor Miozzo: In all of these tasks you can imagine that you have activation of multiple phonological forms as demonstrated, for example, in your data about name agreement. So, it is just a question of how many other names or words you have activated in the background, picking among a set of activated phonological forms.

Professor Thompson-Schill: So you are saying that because all of these representations that I have described are verbal, when I say there is multiple representations active, it could be specifically the phonological representations. You are saying that any time there is more information, that information could be characterized as more phonological information. So, there is no study that I could do with language that could rule out that hypothesis. I think that is why the studies moving to the nonverbal domain are important.

Professor Miozzo: I am not saying that, I am saying that the evidence that you provided indicates that phonology is really a factor in all the data.

Professor Thompson-Schill: I would say that I can not rule that out right now with the exception of nonverbal tasks. Right now, I can not distinguish between any sort of representation becoming active versus only phonological representations becoming active. Any time I manipulate how much competition there is, you could describe that as purely phonological competition. I think the test for that is with these nonverbal stimuli, and the fact we are seeing activation in this area with a nonverbal working memory task is not consistent with that story.

Professor Miozzo: You may also have selection at the level of Wernicke's area, where the hypothesis is that there is representation of lexical information. So, you also have to select at that level.

Professor Thompson-Schill: Let me clarify—I am not arguing that selection is happening in Broca's area. That is critically important—there was a slide where I said that it is guiding the selection process. For people who think of selection as a stage,

any type of information processing step is going to have a selection component; information is passed through some series of stages, and you have to select something to go on. That is happening everywhere. The question is, when that process fails because you do not have a winner—either because there is too many weakly activated things (indeterminate representations), or because you are using the wrong cue so you have to shift the pattern of activation—Broca's area becomes involved with this process that is happening somewhere else. I am critically not saying that the selection is happening in Broca's area.

Professor Miozzo: This begs the question: What is the grain or the kind information on which Broca's area resolves the competition? If Broca's area does not store information that is semantic in nature, how can it decide the correct response?

Professor Thompson-Schill: I think I should refer you to Matt Botvinic's work on this, because it is the best work I know. It is about computing energy in a Hopfield net as a measure of conflict among representations, and you do not have to have knowledge of the representations themselves to make that computation. It is a beautiful work that may help clear that up.

Professor Remez: Let us thank Professor Thompson-Schill and adjourn.

APPLAUSE

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

Time: 4:00 PM

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

Rapporteur: Jennifer Pardo

Attendees: Hannah Bayer, Melissa Brandon, Yi-Chun Chen, Chaio-Wen Deng, Amy Endo, Simon Fischer-Baum, Molly Flaherty, Boris Gasparov, Sarah Gilman, Peter Gordon, Robert Krauss, German Kyrychenko, Jackson Liscombe, Michele Miozzo, Ezequiel Morsella, Kevin Ochsner, Ann Senghas, Anja Soldan, Christopher Summerfield, Alexandra Suppes.

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

Jennifer Pardo
jsp2003@columbia.edu



24 FEBRUARY 2005

**Baby Bayesians?
Exploring the Bases of Generalization
in Human Language**

LouAnn Gerken

*Departments of Psychology and Linguistics
University of Arizona*

Recent research on infant language development suggests that learners have remarkable abilities to extract a variety of statistical information from their input. Most of these studies have focused on conditional probabilities among elements in strings. In my talk, I will focus on another way in which infants use statistics — to decide which generalization is most appropriate, given a particular input set. I will describe recent experiments from my lab examining infants' generalization in both phonological and syntactic domains. These experiments suggest that infants are able to use statistics as a tool for evidence evaluation.

I am about to show you a sets of five slides each containing three colored bars in a rectangle. The first four slides exhibit a property that I want you to learn, and for the fifth slide, I want you to decide whether it has that property.

[Professor Gerken shows five slides, and an audience member correctly answers, "No."]

For these slides, the property in question was whether the area covered by the bars was greater than 50% of the rectangle. Do not feel bad if you did not realize that, and I hope that you were just guessing, because it turns out that that kind of generalization is more natural for pigeons to learn than for humans.

[Professor Gerken shows another set of five slides, and an audience member correctly answers, "Yes."]

In this series of slides, the property in question was whether the three bars were unequal in height, and that is something that pigeons are not so good at, but humans are.

[Professor Gerken shows another five slides, and an audience member incorrectly answers, "Yes."]

In the final series, you only saw examples where the bars were decreasing in height from left to right, but the correct generalization, because I say so, was simply that the bars had to be unequal in height again.

Your frustration with my answer there illustrates the generalization problem that I am going to talk about today. Any set of input data potentially allows for an infinite number of generalizations: Or, at least two. A successful learning mechanism needs to converge on an adaptive generalization with a reasonable set of randomly selected data. There are two loaded terms in that statement—one is *adaptive*, and the other one is *reasonable*—and I am just going to focus on reasonable. What is reasonable must be defined with regard to some knowledge of the computational prowess of the learner. Recent studies suggest that, in fact, infants possess considerable computational abilities. I think there is a misunderstanding in the field about what those computational abilities might buy, in the sense that I think that they do not solve the generalization problem. They do allow the theorist to consider a wider range of solutions to that problem than otherwise might have done.

There are essentially two classes of solutions to the generalization problem. One is that learners are so tightly constrained that there is something about the properties of the input that simply triggers the correct generalization. This kind of approach does not need to assume very much computational prowess on the part of the learner because there is a sort of reflex or triggering mechanism in place. The second approach is to provide learners with a mechanism to compare the goodness of multiple generalizations, assuming randomly selected input.

In all of the studies I will describe, we are interested in infant learning in the laboratory. In order to study infant learning in my lab, we use a setup in which a baby is held inside a sound-proof booth facing an amber light, and there are two loudspeakers on either side of the baby with red lights above each. There is an observer outside the booth who watches the baby on a monitor, via a camera focused on the baby's face. Once the baby is settled inside the booth, the center light starts to flash, and as soon as the infant looks toward the center, one of the two side lights will start to flash. Regardless of whether the infant looks at that light, the observer pushes a button that starts a familiarization phase that lasts about 2 minutes. During that time, we present the materials from both loudspeakers, despite the fact that only one light is flashing. Then, the test phase begins with the flashing center light, followed by a flashing red light. As soon as the infant orients to that light, the sound comes out of that loudspeaker, and the sound is played for as long as the infant orients in the direction of that loudspeaker. On alternating trials, the test items are either consistent or inconsistent with the familiarization materials. Our dependent measure is the infant's listening time before looking away, and any significant difference between consistent and inconsistent looking times counts.

I am going to start by discussing a groundbreaking study that got a lot of people interested in infants' ability to use statistics to organize their input. Here is a demo of the materials.

[Professor Gerken plays an audio sample of rapidly presented nonsense syllables].

There were three 3-syllable words embedded in that synthetic speech string, BIDA KU, GOLABU, and PADOTI. If you heard that series long enough, you might have been able to tell that DA follows BI 100% of the time, whereas PA follows BU 33% of the time. It appears that infants have the ability to calculate these conditional

probabilities and use that information to segment the continuous stream of speech. They listen longer to statistical part-words like BUPADO (BU plus PADO), than things that were together 100% of the time, like PADOTI. This study tells us that infants have some ability to calculate statistics over the surface strings that they are exposed to.

More recently, Rebecca Gomez and I (and, also Gary Marcus) have demonstrated that infants are not only able to attend to the surface properties, but they are also able to generalize beyond the input received in the laboratory. In this case, they are noticing the form of repeating or alternating syllables. In the study by Marcus, there were four A-words and four B-words, combined into 16 three-syllable strings. Here are some examples of AAB stimuli: LELEDI, LELEJE, WIWIJE, etc. Other infants were exposed to ABA strings, so that half the infants were familiarized with AAB and half with ABA items. Over 12 test trials, infants heard new AAB and ABA strings instantiated in new syllables (like POPOGA and POPOBA, or POGAPO and POBAPPO). Marcus found that infants listened significantly longer to the inconsistent trials than the consistent trials. Considering our two accounts of generalization, we could ask whether infants are simply constrained to consider algebraic generalizations.

In a direct follow-up of Marcus's study, we looked at infants' ability to select among possible generalizations. This study used the same stimuli, but I noticed that these stimuli have an interesting property. If you examine the items along the diagonal—LELEDI, WIWIJE, JIJIBI, and DEDEWE—you discern that they have an AAB pattern, and no other generalization jumps out. In contrast, if you look at the first column—LELEDI, WIWIDI, JIJIDI, and DEDEDI—what do you notice, DI. So, these are truly ambiguous in the sense that they have a more abstract and a more surface property. What do infants notice? Do they notice both, or only one?

To examine this question, we familiarized two groups of 9-month olds with synthesized tokens from either the diagonal or column. Within the diagonal or column groups, half of the infants heard AAB and half heard ABA combinations, for a total of four groups of babies. At test, all of the infants heard two AAB and two ABA strings instantiated in new syllables. The infants in the diagonal condition, where the only obvious generalization is the AAB or ABA one, listened longer to consistent than to inconsistent test items. Infants exposed to the column, where there are two possible generalizations, did not seem to make the AAB or ABA generalization. What I guessed about the infants in this condition is that they were, in fact, noticing the more surface property. To make sure that that was the case, I tested another group of infants, familiarizing them with the column stimuli again; but in the test, I gave them the surface property that they were probably looking for, using items like POPODI and KOKODI or PODIPO and KODIKO. When the DI is still there, the effect comes back—they listen longer to consistent than to inconsistent items.

This kind of result is what got me thinking about the Bayesian issue. If a grammar is an AAB grammar, and a subset of the input is randomly selected, the likelihood that all four strings contain DI as the B element is very low. Infants' behavior in this study is consistent with computing how well different potential generalizations are supported by the input. That is, they have an ambiguous situation, just like you had with the bars example, where the likelihood that there was another interpretation seemed really low.

To illustrate a similar point, let us move on to another study, but first we have to get linguistic. Here are some multisyllabic words of English with the primary stress

marked on them, so we have FLUID and ELEPHANT with stress on the first syllable. Then we have MUSEUM, BALLOON, BUCCANEER, HEXAGONAL, and POLYTUNAL. What is the generalization? One way to talk about the stress system of English is to look at the right-most three syllables, and stress the last syllable if it has a long vowel, and that explains BALLOON and BUCCANEER. You stress the second to the last syllable if it has a long vowel or if it ends in a consonant (a heavy syllable). Otherwise, you stress the second to the second to the last, and that is how you get ELEPHANT and HEXAGONAL. These rules have an alternate expression in something called Optimality Theory, which I will return to shortly.

We already know that 9-month old infants can discriminate the predominant stress pattern of English, which is primary a strong-weak language. These infants prefer strong-weak words over weak-strong words. Do infants show any evidence of learning something more abstract than that? Can they exhibit evidence of learning these stress principles or rankings? We tested this with two groups of infants, half were familiarized with Language 1 and the other half with Language 2. They were then tested on the withheld cases of their respective language. I will say at the outset that I actually stole these stimuli from somebody else. There was a study with adults by Guest, Dell, and Cole that examined adults' ability to learn these kinds of principles. I have actually got some modeling data with Tom Schultz at McGill suggesting that this linguistic description is not necessary in order to get generalization. You do not have to buy the linguistic story, but there is something interesting going on here nevertheless.

In the approach to stress assignment that I am taking here, there are ranked principles in each language. One set of ranked principles says that syllables ending in a consonant should get stress; this principle is called *Heavy*. Here, it is more important to stress a heavy syllable than it is to stress the second to the last syllable. The second set of rankings say that stressing the second to the last syllable (the principle called *Penult*) is more important than stressing every other syllable from the left (the principle called *Alt Left*). What is important about this setup is that infants hear words that exemplify the principle that *Heavy outranks Penult*; and, they hear words that exemplify the principle that *Penult outranks Alt Left*; but, they never hear any words that would lead them to know that *Heavy outranks Alt Left*; yet, they should be able to infer that from what they have heard. The test item is the inferable, but unattested, ranking. Note that this stress pattern, with stress on second and fourth syllables, never appears in the familiarization set. They have to learn something beyond simply which syllables in a string are stressed. Language 2 works the same way—it is the mirror-image of Language 1. The test items can only be discriminated if the infants generalize across multiple words and they are abstracting away from stress patterns. This is actually the first study with infants that examines these kind of generalizations with stimuli that are related to language.

Infants listened longer to test stimuli that were inconsistent with their training grammar than the ones that were consistent. However, that just leads to the interesting part. In all of the stimuli that I showed you, the only Heavy syllable that I used was TON [pronounced like the word, TONE]. *Stress* TON is not a possible rule in a natural language, but *Stress Heavy* has been argued to be possible. Are infants able to generalize to other Heavy syllables, or, is their encoding of the situation that I gave

you in the previous experiment just about the syllable TON? This time, I familiarized the infants using the Heavy syllable, BOM [rhymes with the word, HOME], and then I tested them with TON. I thought these were so close, that they would surely be able to do it, but they do not show any evidence of generalizing at all.

Is it possible to get infants in a two-minute laboratory exposure to behave as if they know something about Heavy syllables? In the next study, which is just completed, I familiarized them with three different Heavy syllables, BOM, KIR, and SHUL, and then they were tested on TON. There are two things to note about the results. One is that I got successful generalization to a new Heavy syllable. However, the preference is in the opposite direction from the original finding. I had not thought about this initially, but I think that this is another example of the same kind of principle I have been talking about. If the grammar involves Heavy syllables, and the subset of input is randomly selected, the likelihood that the only Heavy syllable that the infant hears is BOM is extremely low. The likelihood is somewhat higher that they will hear three different Heavy syllables. An important question to pursue here is what likelihood is high enough for them to generalize? Right now, I have another set of infants being tested on the principle that there is stress on Heavy syllables that start with /t/. This is an unnatural principle that should yield the same sort of generalization, if all they are doing is considering the input.

This last study also fits into this Bayesian perspective, and in it, infants appear to be avoiding overgeneralization. This study uses an English noun determiner paradigm. A learner might hear A DOG, A TREE, A SHOE, A BOAT, and THE DOG, THE TREE, THE SHOE. If the learner never heard ~~THE~~ BOAT, given this paradigm, they should be able to infer that THE BOAT is a legal phrase in the language, but there is peril in that approach. Consider this example, THE SWING, CAN SWING, THE SLIDE, CAN SLIDE, etc. If you hear CAN EAT, should you be able to infer THE EAT? No. This is an illustration of what has been viewed as a serious problem with using distributional information to discover categories in language. If you test adult humans in these experiments, they do not seem to learn anything about which elements can go together.

We did a lot of experiments like this about five years ago and kept getting null results. We could have saved ourselves a lot of trouble if we had read the old literature because there was something called the MNPQ problem that we stumbled upon. If you give adults phrases like MN and PQ, adults appear to extract that M and P can precede N and Q, but they do not seem to learn that M can not occur with Q and vice versa. They do not seem to learn the exclusivity relationship. That is good because they do not overgeneralize, but how do people ever learn word classes? It turns out that you can avoid the overgeneralization problem if you require that a subset of the items have simultaneous double-marking for categories. Martin Braine looked at this with respect to a semantic marker and a morphological marker, but you can also have two morphological markers.

Can infants use distributional information to form categories, and are they also restricting themselves to correlating cues? Russian gender is a great domain for examining this question because first of all, it exhibits a category that did not exist in English, so gender is good, and Russian has a very rich marking system, but it also has a great deal of double-marking. A subset of feminine nouns in Russian end in K, and a

subset of the masculine nouns in Russian end in TIL. These items are double-marked. We exposed a set of 17-month old infants to these kinds of items with some of the possible combinations withheld, and they heard the withheld items and some novel ungrammatical items during test. Another group of infants got the same test, but they were familiarized with items in which there was no double-marking. The infants who were familiarized with double-marked items showed a preference for the ungrammatical items (novelty preference). The infants that only had a single marking cue (that is, they faced the MNPQ problem) do not seem to learn, just like adults.

In natural language, it turns out that the chance of two morphemes co-occurring with the same set of other morphemes is very high, for example, M₁ and M₂ occurring with N. There is rampant overuse of the same morphemes in language. The chance of A₁ and A₂ co-occurring with B₁ and then with X is very low in natural language. It seems as though both adults and infants only notice co-occurrence relations or distributional information in the latter case.

From these studies, it appears that human infants behave in accord with the likelihood of a particular generalization given random input. We saw that in the AAB versus AA-DI case, in Heavy syllables versus TON, and in avoiding category overgeneralization. Should we characterize infants' behavior in terms of choosing which generalization is more likely? Or, can we approximate the behavior that we are seeing in infants with some simpler computations plus some constraints on which dimensions they detect? One way of approaching that question is to find evidence that learners actually discern the generalizations they do not ultimately choose. Another question arises concerning constraint. Can we use cross-linguistic data to provide a source of evidence for constraints on what generalizations learners are likely to entertain, as in triggering approaches to generalization? The one that I am pursuing right now is: Can infants learn a generalization like stressed syllables that start with /t/?

How does frequency of input affect the generalizations that infants make? In a lot of neural network models, if you gave infants 90% column stimuli of the AAB sort and only 10% diagonal, the diagonal stimuli would be treated as some kind of noise. The generalization that all of these end in DI would be the one that the model would make. A Bayesian learner would treat these counter-examples as important information, and make the more abstract generalization. We can begin to assess whether infants are taking into account counter-evidence as well as the most frequent pattern in the language.

The last thought is an epiphany that I had on the airplane, which may damn the whole enterprise. There is an interesting relationship, I think, between the amount of input and the kinds of generalization you would predict from a Bayesian learner. An infant needs a certain amount of exposure to learn a pattern well enough for me to find evidence of it at test. The more input I give them, the greater the base against which they are comparing different subsets of the input. Here is the example that made me upset on the airplane: Consider the likelihood of getting only three different Heavy syllables when you have only 20 input words versus 200 input words. Even though the learner might learn more in some sense from the 200 input words, the relationship between the number of types they get in the 20 and the 200 case is really very different. It makes me think that I should just stop all this, if there is something

interesting to be explored here in the relationship between sheer number of examples and how they are distributed with respect to type and token frequency.

It is clear that infants have the ability to make a variety of complex generalizations with very little exposure. Moreover, the particular subset of the input that infants receive influences the generalizations that they make. There is not a clear triggering relationship between a type of input and the generalization, but the particular subset matters. Although infants may not, in fact, compare among possible generalizations, thinking about the input as containing multiple generalizations is turning out to be a profitable short-term approach.

APPLAUSE

Questions

Professor Peter Gordon: Given that there is a lot of ungrammatical language input that children hear, would a Bayesian learner make too much of the errors?

Professor Gerken: That is a good point, and I think that we do not want a learner to completely change its generalization based on a single input. It has to be something like a 90/10 sort of ratio where it can be infrequent but more than once. It will be really important to work out those kinds of details to make something like this work. You do not want them to go off on a tangent with an ungrammatical item.

Professor Michele Miozzo: You mentioned the importance of enlarging the number of stimuli that you present, but there is another dimension to pursue, and that is time. You test the infants right after two minutes of exposure, and this raises the question, what kind of representation do they form? Is it of the same kind that is the basis of language acquisition, where you have a more protracted exposure?

Professor Gerken: I agree that time is important. One question is whether we are even approximating language learning in the laboratory, and only time will tell. There is an interesting study that we have been working on in which we have been exposing infants to an artificial grammar and then sending them home; but, half the infants are nappers and half are non-nappers. When they return the non-nappers do not seem to have remembered anything, so napping is good. When the nappers return to the lab, they do not seem to remember the exact grammar that they learned before their nap, rather, they prefer the first test item we give them after the nap. That sets the pattern they choose for the rest of the experiment. What they seem to be learning, and what they seem to be consolidating over time is not the specifics of the language, but that they should pay attention to certain elements, without remembering the relationship. When they get into the lab, they just need one example, and they go with that. I think they are learning something like what they learn with real language in the lab, but what they end up consolidating from that is not exactly what you exposed them to, but some vague representation of that, which can either be reinforced if they get more exposure like that, or it can just go away.

Professor Miozzo: It is strikingly similar to what you find in memory research, it is a kind of primacy effect.

Professor Gerken: I think it is a reinstating kind of effect. If we had just given them that one example and no previous exposure, they would not have learned anything.

Clearly, the pre-nap exposure does something for them, but basically it just says where to look for the information next time they encounter an example like this.

Professor Robert Remez: I am convinced by the warm-up that if you were to present the items to me as a subject, my performance levels would not look that different from the kids. In a way, that is satisfying, and in a way, that is not satisfying. At least the kids are on the adult trajectory—you can see evidence of that really early on. When I think about the dimensionality of the analysis that the kids must be performing, that is why I am disappointed. In some ways, you have got a kid who has got the phoneme inventory of English nailed already, and it has the meter and melody also set. There ought to be something about the linguistic needs of this child that are satisfied by applying the method that of tracking incidence in this particular way. What is it about this particular moment in language development that makes this particular strategy the one that gives the kid what it needs?

Professor Gerken: I do not think there is anything about this particular moment. My sense is that what I am looking at is something that is common to the species and possibly common to a variety of species, but the fact that infants behave as though they are taking into account distribution tells us that the induction problem, or the generalization problem, does not require such a constrained solution as we have often considered within linguistic frameworks. Learners have some other computational machinery so that they can actually sort out cases where there are multiple generalizations and come to the same conclusion that you would.

Professor Remez: Suppose you had a type mismatch error—in the ABA cases, some of the Bs were of the phoneme inventory of English, but others were Serbo-Croatian. The distributional properties that the kid would supposedly be assimilating are identical, but I am guessing that there is something about the kid's performance that would say, "No sale!" to the Serbo-Croatian infiltrations. What is the context that establishes the legitimacy of the child's analysis, independent of the distributions?

Professor Gerken: I have not solved much of the problem in the sense that they still have to figure out which dimensions are staying the same, and which ones are changing. If we are looking at this as a dynamic problem, I think they are laying down a lot of the featural information, what is relevant, all the way from birth, and possibly before birth. By the time I see them, I can set up this question because it makes sense at this point, but it would not have made sense before. In the stress experiment, for example, you can not even get 6-month olds to generalize to the TON cases. The Bayesian part is not developmental, but everything else is.

Professor Remez: What is the Bayesian part resting upon?

Professor Gerken: Invite me back!

Professor Remez: I think we should.

Professor Gerken: You could still be applying it, and they could still be acquiring the other information via the same kind of mechanism. In the acoustic-phonetic domain, people are looking at those issues and doing some kind of hierarchical clustering analysis taking into account base rate data.

Professor Gordon: How do you think meaning interacts with any of this? Is it just like another feature?

Professor Gerken: No, I do not think it is just like another feature. The hypothesis that I am working on right now is that one of the reasons infants are such good language learners is that they spend a good period of time being pattern detectors before they actually associate those patterns with meaning. Although you could construe the semantic domain in the same kind of multidimensional way that you could construe the surface string they get, the match between those two is computationally horrendous. What makes infants so good is they are actually solving the surface sequential information before they ever try to associate that with meaning. One reason infants look so good in these experiments when 2 and 3-year olds are so lacking in their ability to deal with generalization is that once you overlay meaning, the task becomes much more difficult.

Professor David Elson: I am curious how much you think that these techniques are tied to language, and how much they are a carry-over from something like shapes or colors, like the example you made for us.

Professor Gerken: All of the AAB-type experiments can be replicated with shapes, lights, tones, etc. I do not think that what I am tapping is language-specific, although, getting back to Robert's question, as the domain builds up from the bottom, and what starts to get isolated are those linguistic dimensions that go together, language emerges as its own specialty. What we are looking at is fairly general abilities being applied to a specific domain.

Professor Gordon: I thought Scott Johnson and Gary Marcus were finding that you can not get the rule learning in the shape domain; you could only get it with linguistic input.

Professor Gerken: I do not buy it. We know that with syllables, cotton-top tamarins can do it, that is the Hauser and Marcus finding. It turns out that there is a lot of other animals that can do things that are very similar. I stumbled across a pigeon study from the 1960s where the display contained three lights in a row, and the pigeon's job was to peck the outermost light when there were two lights in sequence. If it was blue-blue-green, they had to peck the first blue, or if it was blue-red-red, they had to peck the right-most red. The pigeons could learn this task without a problem. That is an AAB/ABB task. It would surprise me very much if human infants require syllables when pigeons can do it with lights. It seems like it is not very linguistic at all. Also, bees can learn to turn right if there are two blue pieces of tape in a row.

Professor Gordon: But, do pigeons generalize to novel stimuli?

Professor Gerken: Yes, they do generalize. If you train them on blue, red, and green lights, and then you throw in yellow lights, they can generalize.

Professor Ann Senghas: How do adults do on the AAB type task?

Professor Gerken: We have never tested the AAB type tasks on adults. In the stress experiment that I stole from Guest, Dell, and Cole, they would show adults the word and ask them to produce it to see if they got the stress pattern correct. In production, adults are not very good at that task. They only showed evidence of generalization if, at test, you gave them 7-syllable words and they had never seen 7-syllable words before. If you gave them word with a different syllable number, they would generalize, but what they would try to do it just apply a stress pattern that they heard before. That

could be something about production versus perception. The adults can do the Russian experiment, and it takes them longer than it takes the infants, but we are using different measures, too.

Professor Senghas: But, in terms of whether they are attentive to these surface-specific attributes or some more abstract version, are they like infants?

Professor Gerken: It is not clear how you would go about testing adults. I think that if you put them in a situation where you give them more surface input (where the last syllable is always DI), they might figure out during test when you give them POPOGA, that even though it is not the generalization they made, it is consistent. You would have to come up with a tricky way to avoid letting adults think about it because I think they would realize that there are two possible generalizations.

Professor Remez: I want to return to the dread you described as the consequence of your epiphany. That was about something like the number of cases that were contributing to the characterization of the distribution. I must have looked at hundreds of 3-bar histogram presentations of data over the years, but when you showed your example with the bars, I reset the counter to zero, even though I was completely familiar with that kind of display. My guess is that the babies are also doing this when they come into the lab and hear your items. I am not sure why I did it, or why a baby should do it, but I have a feeling that the distribution, although related to a prior experience, still warrants setting the counter to zero. I think that whole circumstance presents an interesting puzzle to solve.

Professor Gerken: So, what is the time-window and what instances get put together as part of the same problem?

Unidentified member of the audience: Would it make a difference if the materials were spoken by a familiar voice?

Professor Gerken: Yes, but I think if we did a familiar voice with an unfamiliar stress pattern, for example, we might get more fuss-outs than usual. We know from other work by Peter Jusczyk and Derek Huston that changing the voice between familiarization and test in certain ages leads to no transfer. Even though it seems as though infants are able to assume a certain level of analysis and then proceed to the next level, they are not throwing away the calculations from the previous one because they are taking into account talker voice as well.

Unidentified member of the audience: It seems as if what you are saying supports Optimality Theory.

Professor Gerken: It does to a certain extent, but returning to the stress experiment, it is critical for the optimality story that infants are exposed to words that support this principle. Otherwise, there would be no inference they could make to the test items. Tom Schultz and I did a lot of work with a model of infants' behavior in this experiment. One of the things that we did was to remove from the model these stimuli. If Optimality Theory is correct, and these stimuli are what allow the inference, then it should not generalize any more. We compared that to another set of runs with the model where we just took out other items that should not make a difference because there is still plenty of evidence for that principle. Although the generalization we got when we took this latter one out was the same as when we left

everything in, and it was somewhat worse here, the model still generalized. What that tells me about optimality theory or other kinds of more formal models of generalization is that there is something in the input that we can analyze in terms of ranked constraints or stress rules, but there is some other statistical underpinning for that that the model is detecting, and I do not know if infants are detecting. We are pursuing that, and it is turning out to be very interesting. There are other ways of construing what learners are doing than that approach.

Professor Remez: Let us thank Professor Gerken and adjourn.

APPLAUSE

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

Time: 4:00 PM

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

Rapporteur: Jennifer Pardo

Attendees: Stefan Benus, David Elson, Molly Flaherty, Sarah Gilman, Peter Gordon, Norma Graham, Augustin Gravano, Julia Hirschberg, Robert Krauss, Sarah Malcolm, Michele Miozzo, Katherine Nelson, Ann Senghas, Michael Studdert-Kennedy, Lauren Wilcox.

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

Jennifer Pardo
jsp2003@columbia.edu



24 MARCH 2005

**Listener Sensitivity to Fine Phonetic Detail
in Speech Perception**

Joanne L. Miller
Psychology Department
Northeastern University

A widely held assumption in the speech perception literature for many years was that during the course of processing listeners derive an abstract phonetic representation and, in doing so, discard information about the fine-grained detail of the speech signal. However, more recent research has shown that the representations of speech are much richer than this emphasis on abstract categories would suggest, and that listeners retain in memory a substantial amount of fine-grained acoustic-phonetic information. One line of evidence for the richness of phonetic representations comes from research showing that phonetic categories are internally structured in a graded fashion, with some members of the category perceived as better exemplars (as more “prototypical”) than others. In this talk I will describe selected findings from our research program concerning the characteristics of these internally structured categories, with a focus on perceptual sensitivity to variation arising from acoustic-phonetic context and individual talker differences.

The work I will be talking about today is concerned with the way in which human listeners process spoken language. Of course, spoken language covers a very broad domain. It includes, for example, work on sentence processing: how listeners derive the syntactic and semantic structure of an utterance. It includes work on word recognition: how we recognize the individual words that make up sentences. And, it includes work on speech perception, that is, how we as listeners the individual speech sounds, the individual consonants and vowels, that make up the words of the language. That is what I will focus on today, the level of language processing that we traditionally call speech perception.

To delineate a theoretical framework for this talk, let us consider a particular example. When I say the word, PEACE, I generate acoustic energy. For you as the listener to recognize that the word I said was PEACE, you have to analyze that acoustic signal and recognize that I said three phonetic segments in a particular order: the /p/ at the beginning, the /i/ in the middle, and the /s/ at the end. Within this framework, in order to perceive speech, listeners have to map a continuously varying acoustic signal onto discrete phonological categories. The real issue of speech perception is how listeners do this. What kind of perceptual mechanism allows this mapping from the acoustic physical signal to the discrete linguistic signal?

For many years, a prevailing view of this problem was that when listeners perform this mapping from the speech signal, they derive an abstract representation of the individual phonetic segments. In the course of doing so, according to this view, they

discard all kinds of detailed information about the speech signal. What you have left are just the individual, discrete, abstract categories of language. This view held for many years, but there is now a lot of evidence to suggest that this view is wrong. Instead, what seems to be the case is that we as listeners have as part of our mental representation of language, not only information about discrete phonetic units, but also a lot of fine detailed information about the speech signal. We do not discard that detailed information in the course of performing the mapping from the acoustic signal to the phonetic level of language.

Phonetic categories are not simple abstract structures. Instead, they have a fine-grain, graded internal perceptual structure. This structure embodies a lot of detail about the speech signal. As I will describe in this talk, some findings from our research program show that these internally structured, graded categories are themselves highly sensitive, in a very systematic way, to acoustic-phonetic contextual factors. Finally, I will describe some very recent evidence that suggests that these graded, internally structured perceptual categories might even be sensitive to differences among individual talkers' voices.

The first point that I want to make is that phonetic categories are internally structured. In particular, these categories have a graded internal structure. It is not the case that any given exemplar of a category is perceived to be an equally good exemplar to all others. Rather, some exemplars of a category are perceived to be better than other exemplars. Just as every instance of red is not an equally good instance of red, so too, every instance of a particular consonant or vowel is not an equally good instance of that consonant or vowel.

There are a number of different techniques that one can use to demonstrate this graded internal perceptual structure. One of the techniques that we have used involves a very simple goodness rating paradigm. To use this paradigm, we have to start with a particular linguistic contrast. For this example, I am using the contrast between word-initial /b/ and word-initial /p/. This contrast is what linguists call a contrast in *voicing*—/b/ is considered a voiced consonant and /p/ is considered a voiceless consonant. Voiced consonants are articulated so that very shortly after the consonant is released, the vocal folds are vibrating. Voiceless consonants are produced so that there is a short, but measurable delay between the release of the consonant and vocal fold vibration. This distinction in production is reflected in a property that linguists call voice onset time, or VOT. This particular contrast will be the focus for my talk today, although, the conclusions are general to other phonetic contrasts.

VOT is defined precisely as that interval between the release of a consonant and the onset of vocal fold vibration. When the vocal folds start vibrating, it causes a change in the acoustic signal so that there is a sudden onset of much higher amplitude energy that is periodic in nature. For voiced consonants like /b/, the VOT interval is relatively short, and for voiceless consonants like /p/, the VOT interval is longer. Using any number of speech synthesis or editing techniques, it is possible to generate a series of items that ranges from /b/ to /p/, simply by keeping everything else constant and incrementing VOT values in very small steps. You start with /b/ and end up with /p/, and somewhere along the line, you cross a phonetic category boundary and there is a

qualitative shift in perception. For this particular example, the boundary is located at about 40 ms VOT.

To show evidence for a graded internal structure, we had to modify the classic VOT series. To do so, we did not stop at the long VOT value of around 70 ms for a good /p/, but we kept going. We made additional syllables that still sounded like /p/, but they had longer and longer VOT values. The series now ranges from /b/ through /p/, and then the /p/ items start sounding really weird. They sound something like a very breathy, exaggerated /p/. In these experiments, we generated many tokens of each exemplar from this long extended series, and presented them randomly to listeners. We asked the listeners to focus on just one phonetic category (in this case, the voiceless category /p/) and to judge each token one at a time for how good an instance of /p/ it is, using a rating scale from 1 (lousy) to 10 (good). Listeners are always judging goodness of a single phonetic category, in this case, /p/.

To portray the finding, we plot the mean goodness rating for the group as a function of VOT value. Looking at the very short VOT values, the ratings are very low, which makes sense because those items are really /b/, so they should get really low ratings for /p/. When you cross the category boundary (as determined in a separate experiment), there is a nice increase in the rating function—very systematically those instances of /p/ are judged to be better examples of /p/. Then, critically, the ratings very systematically decline so that those items with very long VOT values are again heard as lousy examples of /p/. There is a gradual change in perceived category goodness throughout the category, and that is what I mean when I say that these categories have a graded structure—the perceived goodness of the tokens varies quite systematically as a function of the acoustic variable we are manipulating, in this case, VOT.

We have done the same kind of experiment on many different kinds of linguistic contrast, not just voicing, and we have manipulated a number of different kinds of acoustic variables. In every case we have tested so far, we have obtained these kinds of graded goodness functions. We have not been able to find a case in which the ratings stay very high. It is not the case that the function looks identical in each case—the functions can stretch out or be more narrow, depending on the distinction and the acoustic variable you are working with. There is always a gradual difference in ratings throughout the category. Also, these kinds of group data represent very accurately what happens when you look at an individual listener. This is not a result of averaging—individual listeners give you these kinds of graded categories. Moreover, this phenomenon is not specific to English—we get exactly the same kind of function for contrasts in other languages.

The second point I wish to raise for you draws evidence from our investigations of a number of different characteristics of these categories. Specifically, we have found that phonetic categories are context-sensitive as well as graded. One of the most striking characteristics of speech concerns the fact that the boundaries between categories do not remain fixed at a particular value of an acoustic dimension. For example, it is not the case that the boundary between /b/ and /p/ is always at 40 ms VOT. Rather, the location of the boundary varies systematically as a function of a host of different contextual factors. We sought to determine whether this kind of context dependency

is limited to the region of the category boundary where, by definition, ambiguity exists. Perhaps context dependency is such a fundamental aspect of speech processing that you also see evidence for it within the category. More specifically, is the location of the best exemplars a function of contextual variables?

One acoustic-phonetic contextual variable, speaking rate, was the focus of our investigation of this question. When people talk, they do not maintain a constant rate of speech—they speed up and slow down. If you actually measure conversational speech, you find that those changes in rate, even within an individual talker, can be quite dramatic. This is interesting for phonetic perception because when talkers change speaking rate, one of the things they do is stretch and compress the individual syllables of speech in complex nonlinear ways. In doing so, they alter many of those properties that themselves specify phonetic distinctions, for example, voice onset time. Recall that short VOTs specify voiced consonants and long VOTs specify voiceless consonants; but what counts as short and what counts as long? It depends on the rate at which you are talking—VOT values change very systematically with a change in speaking rate, particularly for voiceless consonants.

We knew from earlier work that listeners are indeed sensitive to this change in speech production, and they reflect this in their phonetic category boundaries. The listener's perceptual boundary between /b/ and /p/ actually changes as a function of speaking rate. The boundary occurs at shorter VOT values for short syllables and longer VOT values for long syllables. In terms of VOT length, what it takes to turn a /b/ into a /p/ depends on syllable duration. Are listeners also sensitive to variation in speaking rate insofar as the best exemplars of the category are concerned?

Lydia Volland and I conducted an experiment to investigate this question. We created two matched, extended series that varied in VOT. The two series differed in syllable duration as given by the length of the vowel. All of the items in the short series had syllables of about 125 ms in duration. All of the items in the long series had syllables of about 325 ms in duration. For any given VOT value across the two series, the syllables were identical except for whether there was a short or long vowel. We presented multiple tokens of each item randomly to listeners performing the category goodness task.

We expected that for both series, we would obtain an inverted U-shaped function; the question was whether those functions would be displaced with the peak of the function at longer VOT values for the longer syllables, and that is exactly what happened in the data. Both series led to very orderly goodness functions, but the best exemplars of the long syllables are systematically displaced to longer VOT values compared to those of the short syllables. It is not a huge displacement, but it is highly significant and very reliable across subjects. Listeners are sensitive to this contextual variable of speaking rate, and we would argue that they have retained information in memory about the particular VOT values that would be the best exemplars for a particular speaking rate. They have context-sensitive graded phonetic categories.

The third point concerns the sensitivity of listeners to the speech of individual talkers. It has been known for a very long time that talkers differ from one another in precisely how they pronounce particular consonants and vowels of their language. Even within the same dialect, there are idiosyncratic differences from one talker to another.

However, it was thought that this kind of variation plays no important role at all in speech perception. The traditional view has been that when listeners map the acoustic signal onto phonetic categories, they discard all the detailed information about individual talkers' voices. Indeed, that has always been considered noise. The issue was about how listeners process language despite all this noise.

However, there are two recent lines of evidence that suggest that there is a problem with this view. First, listeners have an easier time recognizing spoken words when those words are produced by a talker whose voice they are familiar with compared to an unfamiliar talker's voice. One interpretation is that listeners encode in memory precisely how individual talkers pronounce individual consonants and vowels that make up words—they have a fine-tuning to individual talkers. They use this detailed information when they are trying to recognize new words spoken by those talkers.

There is also some very interesting evidence, largely owing to Robert Remez and his colleagues, that listeners use this detailed information, not only to recognize which words the talkers produced, but also to recognize the talkers. There is information about the characteristic idiosyncratic ways that talkers produce the consonants and vowels that makes a talker sound just like that talker, and not like another talker. That is part of the information that we use when we recognize individual voices. Taking these two lines of research together, we find the evidence convincing that when listeners map from an acoustic signal onto a phonetic representation, they do not discard detailed information about talkers' voices. They retain this information somehow in memory.

We were very intrigued by these findings, and we started thinking about them in terms of our own work on internal category structure. Listeners are very sensitive to differences in the phonetic properties, such as VOT, that specify linguistic contrasts, such as consonant voicing. Clearly, listeners are very sensitive to small differences in VOT, and we saw that in the graded nature of the categories. Listeners are also very sensitive to the contextual factors operating on speech. We wanted to know whether a listener's sensitivity goes beyond a sensitivity to overall contextual factors, but also includes a sensitivity to individual talkers' voices.

Are a listener's phonetic representations fine-tuned or customized for different talkers? Say we have two talkers, Talker 1 and Talker 2. Let us focus on the talkers productions of words that begin with three voiceless consonants in English, /p/, /t/, and /k/. Let us assume that Talker 1 typically produces these voiceless consonants word-initially with somewhat short VOT values. That is part of what makes Talker 1 sound like Talker 1. Let us say that Talker 2 characteristically produces the same words with somewhat longer VOT values. This is part of what makes Talker 2 sound like Talker 2. Let us say that we have a listener who is familiar with both talkers' voices. Perhaps as the listener became familiar with the two talkers' voices, our listener learned/encoded information about how those talkers produce many aspects of the speech signal, including VOT. Our listener simultaneously learned, because she was familiar with those voices, that Talker 1 produces words with short VOTs and Talker 2 produces words with longer VOTs. Perhaps listeners have customized phonetic categories in memory so that those for Talker 1 have shorter VOT values than those for Talker 2. Can we find evidence for such a scenario, in which listeners use

customized categories that are displaced from one another, reflecting what happens when these talkers produce the words?

We are just starting to look at this question, so we do not know the answer yet. We have some initial evidence that I find very encouraging. This work was part of Sean Allen's Ph. D. thesis in my lab. The first study investigated whether individual talkers do in fact differ characteristically in VOT when they produce words that begin with these voiceless consonants. On the basis of the available literature, we thought that this might be the case, but it was not absolutely clear. For this study, we asked eight different talkers to produce 18 different words, and the words all began with /p/, /t/, or /k/; for example, PILL, PUSH, TIME, TOWN, CAVE, KISS. The talkers produced 30 instances of each of the 18 words, and we measured the VOT value and the overall duration for each token.

We found that talkers do indeed vary characteristically in their VOT productions for these kind of monosyllabic words. If one looks across all of the different words, some talkers characteristically use shorter VOT values and some characteristically use longer VOT values. Importantly, this difference in VOT values across talkers remains even if you control for speaking rate. When you control for speaking rate so the durations are equalized, you still get these characteristic differences in VOT across our talkers.

Given this evidence that talkers produce characteristically different VOT values, we turned our attention to the main question, which concerns listener sensitivity to these individual talker differences. If listeners have customized speech categories, then it must be the case that when the listeners are listening to a talker's speech, they must have a way of tracking the VOT values of individual talkers and a way of storing this information in memory. That is, they must somehow be able to learn that Talker 1 has short VOT values and Talker 2 has long VOT values. If they could not track VOT in production as they were listening to speech and somehow store this information, they would not be able to develop these customized categories. This is really a prerequisite.

Are listeners able to track the VOTs of individual talkers and code them so they know that one talker produces voiceless consonants with short VOTs and another talker produces them with longer VOTs? For this study, we used a speech synthesis technique that allowed us to alter productions of natural speech, by only tweaking the VOT value and leaving everything else constant. We used this technique for two talkers that we fictitiously called Annie and Laura. We ended up with a set of items for Annie and a set of items for Laura, where Annie's voice really sounds like Annie and Laura's voice really sounds like Laura; we just tweaked the VOT values within each talker.

We created two subsets of items for each talker. One subset consisted of the word DOWN and two different variants of the word TOWN: One variant had a short VOT and one had a long VOT. In pre-testing, we ensured that both the long and short variant of TOWN did indeed sound like TOWN, and we ensured that both variants sounded like reasonably good exemplars of TOWN. The other subset consisted of the word DIME and two different variants of the word TIME. Again, the variants had a short and a long VOT and were heard as reasonably good exemplars of TIME.

The study was run over two days. The Day 1 session focused on the DOWN and TOWN items and consisted of two kinds of trials, training trials and test trials. During

the training trials, all of the listeners heard both Annie and Laura saying DOWN and TOWN, but the version of TOWN that they heard depended on which subgroup they were in. One subgroup, the Annie long/Laura short Training Group, always heard Annie's long VOT version of TOWN and Laura's short VOT version of TOWN. The other group, Annie short/Laura long Training Group, always heard Annie's short VOT version of TOWN and Laura's long VOT version of TOWN.

At the beginning of the Day 1 session, the subjects simply listened to the speech of the two talkers saying DOWN and TOWN all mixed together without making any response. On each trial, a listener would hear a token and then the talker's name would be displayed on a computer screen. There were 24 of these initial trials with no response—they were just trying to learn which name went with each voice. This was easy to do because we deliberately chose voices that were very distinct. Next, there was a series of trials in which the listeners would hear a word and indicate both the word identity (which was trivially easy) and who the talker was, Annie or Laura. After they made their judgment, we gave them feedback indicating whether their response was correct or incorrect. It was very easy for the listeners to recognize the talkers, and they reached ceiling on the task very quickly.

After 48 of these training trials, we introduced our first set of test trials. The purpose of our test trials was to determine whether our listeners had picked up anything about how Annie and Laura produced TOWN. During each of the test trials, a subject heard a pair of TOWNS from one of the talkers—one was a short VOT variant and one was a long VOT variant. Their job was to say which variant sounds more like that talker, the first or the second, based on how they heard the talker during the first part of the experiment. There was no feedback given on these trials. We ran a few test trials for each talker's voice and then went back to another training block of 48 trials, exactly as before. After that training block, there were eight test trials for one of the talkers' voices. Then, there was another set of training trials followed by another set of eight test trials for one of the talkers. We kept alternating between these long training blocks of 48 trials and these short test blocks of just eight trials, until we had four blocks of test trials for each talker's voice.

We used this rather complicated procedure of alternating training and test blocks because we did not want testing to interfere with and contaminate the memory. Remember, on testing, one of the variants was the one that was not part of training. If there are too many test trials, then that variant will start sounding like one that they have been trained on. We wanted to minimize the possibility that the testing itself could interfere with what we were trying to measure.

At the end of the first session, we collapsed the data across all of the test blocks, and we looked to see if our listeners had coded something about how the two different talkers pronounced TOWN. When trained on long VOT variants for Annie and short VOT variants for Laura, listeners chose the long variant for Annie more often and the short variant for Laura. For the other group, who were trained on short VOT variants for Annie and long VOT variants for Laura, the reverse pattern of results occurred. Listeners chose the short variant for Annie and the long variant for Laura. In all cases, they were choosing the variant that they were trained on. This provides initial evidence that listeners can simultaneously track two different talkers producing that

initial consonant in the word TOWN, and they code that information in memory at least long enough to give us this kind of performance on the test trials.

What did the listeners really learn? Have they learned only how each talker produces the word TOWN, or have they learned something more general? Perhaps at least they have learned something about how the talkers produce that initial consonant /t/ at the beginning of words. Will the listeners generalize to another word spoken by the two talkers, TIME? To examine this, we tested them on Day 2 in a paradigm run exactly like the Day 1 session. We alternated long blocks of training, 48 trials, with short blocks of testing on a single voice for eight trials. The training was the same as on Day 1, they were always trained on TOWN, but the test trials were on TIME, with short and long VOT variants. Now, the listeners had to choose which version of TIME sounds more like that talker, based on what they learned about how the talker says TOWN. The data show the same pattern for these trials on TIME as those on TOWN—listeners were able to generalize their knowledge from TOWN to TIME. Listeners were able to track how talkers produce /t/, and they were able to code it in such a way to use it not only to make judgments about the same word, TOWN, but also a different WORD, time.

We find these data very encouraging. At the least, they suggest that listeners have the ability to track this variability in talkers' voices. Clearly, it is not variability that is automatically discarded, listeners can track it, and it generalizes across words. On of our next steps will be to use a version of the goodness rating paradigm to assess whether after this kind of training, listeners have coded customized categories with the best exemplars actually being displaced from one another along the VOT continuum for the two talkers depending on training.

In summary, we now have strong evidence that the phonetic categories of language have a fine-grain graded perceptual structure. Some members of a category are better exemplars than others. Our representations of the phonetic categories of our language are indeed highly detailed categories, they are not simple abstract entities. Furthermore, we have evidence that this structure is highly dependent on acoustic-phonetic contextual variables. Again, more evidence for the complexity and the sensitivity of these representations to the kind of context effects that exist in speech. Finally, we have at least initial evidence consistent with the idea that these internally structured categories might be customized for different talkers' voices. If that turns out to be the case, that would provide even more striking evidence for the detailed rich nature of phonetic categories. Taken together, these findings suggest that during the course of spoken language processing, there is a highly complex finely tuned mechanism working that serves to map the continuously varying acoustic signal onto the phonetic categories of the language. Critically, it does so in such a way that very fine-grain information about the speech signal is retained.

APPLAUSE

Questions

Professor Robert Krauss: I was expecting that the next study you would do would be to test them on something like FINE and VINE. If the question is what they are

learning, you really would want to know whether they are learning something about VOT, rather than VOT in this particular pair of phonemes.

Professor Miller: You are exactly right! One line of research moving ahead is to try to see if we have these internally structured categories, and a whole other line of research is looking to see exactly what they have learned. For example, we tested on the same initial consonant in the two words, but there are three voiceless stop consonants in English, /p/, /t/, and /k/. One question is whether this generalizes to p and k; another is whether it generalizes to something like fricatives, /f/ and /v/, which have slightly different properties. Another thing we want to know is whether this generalizes across different speaking rates. Do you learn something relative about a talker's VOT, or something that is relativized to the speaking rate? In addition, will something learned on monosyllabic words generalize to multisyllabic words? This is just the start, we have a whole slew of generalizations studies that we want to evaluate.

Another thing is to learn one talker from a particular dialect and to see if you transfer that to another talker of the same dialect, but not to a talker of a different dialect. Actually, we are now doing work related to this on perceiving speech with a foreign accent.

Professor Gordon: There used to be a sitcom on television called, *Dear John*, and one of the characters on it called, Kurt, used to have very short VOTs on his voiceless consonants. I think all of his voiceless consonants had short VOTs, and I wonder if it is true in general that you see it across the board when you have that in your dialect.

Professor Miller: That is interesting. I do not know the answer to that question, but my guess would be yes. What we found with our individual talkers in that first study I mentioned was that they had short VOTs for /p/, /t/, and /k/. Because you see it in individual talkers, my guess is it would go across the board for dialect, too.

Professor Joseph Jaffe: This should reproduce when animals listen to speech samples.

Professor Miller: I am not so sure. Humans express an attention guiding mechanism sensitive to particular properties in the speech signal that are relevant for phonetic contrasts. It is not so clear to me that animals would pick up on just those same properties. There are lots of studies these days with animals perceiving speech. If you train animals over thousands of trials in many cases you can get performance levels in animals to look sort of the way human performance data does, but often it takes a lot of training. Whereas, even if you look at human infant research, very young infants seem to home in on just those particular properties that are relevant to specify phonetic contrasts in the language.

Professor Jaffe: I have a vague memory from years ago about this continuum in which perceptually there was a sudden change from BA to PA.

Professor Miller: There is evidence contributed by Pat Kuhl and her colleagues from a long time ago with chinchillas showing phonetic category boundaries in some, but not all cases. A more recent finding of Kuhl used a different paradigm from the one I have talked about here to examine the internal structure of vowel categories. She found evidence that adults perceived some exemplars of vowels as better exemplars, and it was a little harder to discriminate variants around those best exemplars than it was for variants from a different part of the vowel category. She found the same effect

with infants, suggesting that infants have at least the precursors for these phonetic categories. Peter Eimas and I have also looked at consonant categories with very young infants, and we found evidence for internally structured categories. When Pat tested monkeys, in this crucial control comparison she did not find evidence for internal structure of categories. That suggests that human languages might have been built on some auditory discontinuities at certain places along these acoustic dimensions, but it might be that this internal structure of categories is something that is uniquely human and that you would not find in the animals.

One other piece of data comes from a study I did with Larry Brancazio which capitalized on the finding that if you are looking at a talker's face, it will influence what you hear. We found that if you measure the best exemplars of the category using the same kind of series I have talked about today, but you alter what the subject hears by changing what they are looking at in articulation, you can actually alter what the best exemplars of the category are. These best exemplars are not some hot spots in acoustic space, but they are probably hot spots in some phonetic space. Presumably that is something that would be uniquely human. We do not know yet whether these hot spots are there in the beginning so that infants have these structured categories which are fine-tuned according to the particular language environment, or they have the categories and the structure arises through language learning.

Professor Boris Gasparov: I think phoneticians tend to distinguish voiced and voiceless consonants by two distinctive features. One is the working of the vocal cords, and the other is the difference in the degree of intensity with which the articulatory organs work. The relative weight of each of these features is different in different languages. Also, within one language, in different positions the relative weight of these two features is different. Have you taken this into account?

Professor Miller: Yes, we have. That is a very good observation because I have been talking about voice onset time as if it is this unidimensional property, but even voice onset time is not. You have the duration of that aperiodic interval, but you also have a difference in the trajectory of the fundamental frequency—it starts low for voiced consonants and higher for voiceless consonants. There are lots of things varying. We are coding this as VOT, but really, there are lots of things varying across the board. What I think we really have is a multidimensional space for any of these contrasts. This was just very simplified.

We actually looked at that in another study that I did with Philip Hodgson that was part of his Ph. D. thesis. We looked at the contrast between having a stop consonant or not, as in the contrast between SAY versus STAY. In STAY, there is closure that you do not have with SAY, and there is a lower first formant. It had been known that you could trade off these two cues, the first formant onset and closure duration, in terms of where that SAY-STAY boundary is located. We wanted to know whether they would also trade off in the middle of the category. We were able to show that the best exemplars of STAY actually depend on both of those properties trading off against one another. It is not the case that a best exemplar is going to be at one place along a particular continuum, and that is it, but it is really going to be multidimensional.

Professor Robert Remez: My question is about how you correct your estimate of the eccentricity of voicing for the rate at which the syllable train is produced. How do you handle this in the laboratory to be able to come up with a normed value? Perceptually, I am doing something like this when I listen to you to tell whether you are Professor Miller long or Professor Miller short given your rate of speech.

Professor Miller: I do not know how we can actually use that kind of accommodation in online processing in discourse. I do not know how much experience we need with a particular talker's voice to be able to come up with that.

I do know that at least in this procedure, at the point that the listeners had the first test trials, they did not know that they were going to be tested. They had not been trying to pay attention to anything specific about how these words were pronounced. We analyzed those first practice test trials separately and they look just like the overall data from the combined test trials. At that point, they had had 24 familiarization trials and 48 training trials, half were DOWN and half were TOWN, and only half in Annie's voice and half in Laura's voice. They had very few trials of each talker saying TOWN before we collected that first set of data. Also, we did the experiment again with training on time and testing on TOWN, and it came out basically the same. When we analyze the practice data from that experiment, we find the same pattern again. They picked it up with just a little bit of exposure.

To get back to your question, I would really like to do this with training on a different speaking rate. I want to train them on TOWN at a slow rate and see if they can pick it up and apply it to TOWN at a fast rate. I can say that it is much more complicated than simply determining some kind of ratio of VOT to duration, that does not work. Our production studies show that even that would not work in theory because for the same duration words, some talkers have longer and some shorter VOTs. There is something else in the signal that is going to code it or not.

In terms of how we did this in the lab, in our production study, we tried to set it up so that people would speak at the same rate. We had each of the words appear on the computer screen with very systematic timing. We measured the duration each talker saying each of 30 instances of a particular word and looked at the distribution of durations across all instances of that word for that talker. We compared these distributions to find another talker that had basically the same distribution of word durations for that same word, and we found some cases like that. We looked at the VOT distributions for some of those cases, and even though the overall word duration distributions were overlapping and the means did not differ, the distributions of the VOT values were displaced from one another with differences in their means. There was only about 10% of the data that allowed us to do that comparison. We used hierarchical linear modeling to apportion variance—we found a significant effect for our talkers and we pulled out the variance that was due to variance in speaking rate, which accounted for most of the variance in VOT. Then we were able to look at the remaining variance and there was a significant effect of talker. Even after statistically taking out the variance due to speaking rate, there was still variance that was accounted for significantly by the different talkers.

Professor Remez: Let us thank Professor Miller and adjourn.

APPLAUSE

24 MARCH 2005

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

Time: 4:00 PM

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

Rapporteur: Jennifer Pardo

Attendees: Lauren Aguilar, Colin Beer, Molly Flaherty, Boris Gasparov, Peter Gordon, Joseph Jaffe, Robert Krauss, Ann Senghas, Michael Studdert-Kennedy, Alexandra Suppes.

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

Jennifer Pardo
jsp2003@columbia.edu



21 APRIL 2005

How do Bilinguals Choose One Language to Speak?

Judith F. Kroll
Psychology Department
Pennsylvania State University

Until recently, cognitive science virtually ignored the fact that most people of the world are bilingual. In the past decade this situation has changed markedly. There is now an appreciation that learning and using more than one language is a natural circumstance of cognition. Not only does research on second language (L2) learning and bilingualism provide crucial evidence regarding the universality of cognitive principles, but it also provides a sensitive tool for revealing constraints within the cognitive architecture. Recent studies investigating adult second language performance have shown that even among the most proficient bilinguals, there is parallel activity of both languages when only a single language is required. The observed activity of both languages and the interactions between them, even once bilinguals achieve a high level of skill in the L2, suggests that successful acquisition is not a matter of developing an encapsulated representation for the L2 that becomes functionally automatic and independent of the first language (L1). Instead, there appears to be a restructuring that renders the bilingual distinct in some respects from his or her monolingual counterparts. Under ordinary circumstances, bilinguals do not suffer from the consequences of cross-language competition, suggesting that they have in their possession an elegant mechanism of cognitive control that allows them to effectively select the language they intend to use. The focus of much of the current psycholinguistic research on adult bilinguals and L2 learners is to understand how bilinguals negotiate the parallel activity and interactions of their two languages, the cognitive consequences that result in response to the need to resolve potential competition across the grammar and lexicon of the two languages, and the constraints that remain as a function of the context in which the L2 was acquired, the context in which it is used, and the properties of the specific language pairings. In this talk I present a series of studies on bilingual language production that will serve to illustrate the models and methods that cognitive psychologists adopt to examine these issues.

In the last ten or fifteen years, cognitive research on bilingualism has increased dramatically. One reason for this expansion is that the field has begun to understand that experimental psycholinguistics has something to contribute to understanding the nature of bilingual experience. This is one approach among many others that contribute to that effort. Cognitive psychologists and cognitive scientists have a selfish reason for studying bilinguals. Bilinguals enable us to provide a universal account of how cognitive systems develop and interact with one another. On one hand, bilinguals in ordinary experience do not randomly mix up the languages they speak. On the other hand, they can fluently alternate between languages, or *code-switch*, with other bilinguals. Understanding how this cognitive control is effective promotes a universal account of how cognitive systems interact and compete with one another.

I want to start in Amsterdam, in Centraal Station. The scene you see as you exit the station is a confused jumble of bicycles. In the two years that I lived in the Netherlands, I never figured out how a Dutch person ever found his/her bicycle. My theory was that it was similar to having children in daycare—as long as they came home wearing clothes, it did not matter whose clothes they were. The question is, how do you find your bicycle? This seems like a silly problem, but it provides an interesting metaphor for thinking about the problem that bilinguals solve all the time when they have to decide which language they are going to speak. For example, a Dutch-English bilingual faced with a bicycle would have to decide what to call this object, a BIKE or a FIETS. This is a task that most 3-year-olds can handle readily, yet cognitively, it is quite a complex task.

My talk today will examine three issues in bilingualism. First, I am going to try to convince you that bilinguals are active in both languages all of the time. In fact, it is virtually impossible to turn off one of the two languages. Then, I want to describe how deeply into speech planning that activity extends. Then, I will propose a preliminary answer to the question of how that activity might be resolved. I will present two potential solutions to the problem. One solution involves the idea that bilinguals are sensitive to the cues that signal the use of one of their two languages. The other solution appeals to inhibitory mechanisms that might be activated as a means of shutting down one of the languages when a bilingual intends to speak the other language.

Researchers of bilingualism have borrowed the theoretical framework that has been developed in the monolingual literature. The problem is to decide whether the bilingual can function as if he/she were two monolinguals, effectively selecting one language and shutting the other one down, or whether candidates in both languages are active all of the time. The Selective Access view holds that the intention to speak in one language is sufficient in and of itself to create a situation so that the bilingual can function as a monolingual. The alternative Nonselective Access view holds that candidates in both languages become active in parallel, even when the person intends to speak in only one of his/her two languages, and those active candidates then compete for selection. One possible way to resolve that competition is to propose that distinct cues to language membership eventually bias access for candidates in the intended language. Another possibility is that those cues eventually allow the unintended language to be inhibited.

According to the Selective Access view, the Dutch-English bilingual looking at a bicycle has decided in advance to speak in Dutch and simply activates that part of the lexical representation exclusively. The alternative view includes the possibility that a bicycle might have some properties that make it Dutch-like (it exhibits features corresponding to Dutch bicycles), either enabling the speaker to select the Dutch word; or, the Dutch word wins following active competition with the English word.

Models of bilingual production are very similar to models of monolingual production. The commonplace models differ in only one respect. One assumes that when both alternatives are available to the bilingual, they are available all the way to the phonological level. The other assumes that the process resolves at an earlier planning stage, at a stage with distinct abstract lexical representations. Note that you

have alternatives active in both of the languages, but former asserts that the language conflict is resolved at the lemma, while the latter does not. I am going to argue that highly proficient bilinguals have this problem, it is not specific to second-language learners. If it is a problem for highly proficient bilinguals, then it is really a problem for second-language learners, for whom the first language is much more dominant.

To investigate this question, we use a variety laboratory tasks. We present simple materials on a computer screen and ask subjects to make responses. Note that in all these cases, the response is the same, the speaker must say FIETS, for example. The difference lies in what elicits the response, and what is hypothesized to occur in between. In order to understand the intervening processes, we use a modified Stroop interference paradigm. In the classic Stroop task, a word is printed in colored ink, and when the printed word conflicts with the name of the ink color, the response to say the color of the ink is slowed; or, there are errors in speech production.

In our variant of the task, interference is the delay in naming a picture when it is accompanied by a distracter word that is either heard or seen. The distracter word is presented at varying stimulus onset asynchronies (SOAs) from the presentation of the picture, and we simply measure the delay in the time to name the picture. If the ability to produce in one language were truly selective, a distracter presented in the other language should have no effect on picture naming. If it is nonselective, then we should find semantic interference from the distracter word, regardless of its language of presentation. Using this paradigm across a variety of language combinations, the results consistently show an interference effect, suggesting that both languages are active and available to some extent.

Another approach asks, what is the consequence of language mixture? If language production is fundamentally selective, forcing both languages to be active should be problematic. To investigate this issue, we developed a cued picture naming task. In this task, the speaker sees a picture and does not know in advance which language to respond in until a high or low tone cue occurs. Here, we are essentially forcing activation of both languages, at least until the cue, which occurs at varying SOAs. This condition is compared to a blocked control condition in which the speaker is told to respond to one tone with the name (in the specified language) and to the other tone with the word, NO. Here, there is still uncertainty about when to name, but there is no uncertainty about which language to speak.

For the speaker's second language, it makes no difference whether the speaker knows in advance, the first language is active anyway, even for highly proficient Dutch-English bilinguals. The data for the speaker's first language are quite different. There is a cost to first language production when the second language is forced to be active, and our subjects are slower to speak in their first language when the language is uncertain. These results suggest that both languages are active, and possibly that the speaker's second language inhibits the first language.

In another demonstration of language mixture, we used a somewhat different task. We used a simple picture naming task, but on a small percentage of trials, the speaker was interrupted with a word and had to say the word instead. We varied whether the speaker was interrupted with a word in his/her first or second language. When the speakers were interrupted while using their second language, it did not matter if the word was in their first or second language. When they were interrupted while using

their first language, the language of the distracter mattered—there was a cost for picture naming when they were interrupted in their second language as opposed to their first language. This suggests that maybe bilinguals are nonselective in their second language, but selective for their first language. I am going to argue that this is not the case, that the system is fundamentally nonselective, but the time-course of processing and automaticity associated with the first language makes it possible for it to escape the effects of the second language.

So far, the data suggest that there is activation of both alternatives at least to the abstract lemma level. The consequences seem to be more severe for the second language than for the first. Can we be sure that this competition ends at the abstract lexical level, or does it perhaps filter down to phonology? One way to examine this issue is to exploit the fact that many languages have cognates, words that are virtually identical in both languages. In Dutch the word for BED is something more like BET, but the two words are very similar phonologically. Does phonological overlap facilitate picture naming? If so, this would suggest activation all the way to phonology.

In these studies, it is important to remember that the subjects are naming pictures, and they never see the words. The speakers are naming a picture whose phonology overlaps between the two languages. Bilinguals show reliable facilitation effects for naming cognates as opposed to noncognates, and monolinguals show no effect at all. Phonological overlap facilitates picture naming, suggesting that the two languages are active all the way to the level of phonology. Other research in our lab shows that the same effect occurs for languages that share phonology, but not orthography, so the effect is not due to orthographic overlap.

Returning to our cued picture naming task, we can examine the time-course of these phonological effects. For the speaker's second language, there is cognate facilitation regardless of whether the speaker knew the language in advance. Over time, cognate facilitation goes away. Initially, there is momentary activation of the other language all the way down to phonology. This suggests that bilinguals have a problem because there is activation of both languages and the other language's phonology is potentially on the tip of their tongue.

How is this conflict resolved? Can bilinguals exploit cues that are available in the event that initiates planning in speech production, which might enable them to resolve this problem early on? I would argue that if you see a more realistic picture than the canonical line drawings used here, it is more likely to bias one or another language. In one study, we used a translation task, which shares many attributes with picture naming, like accessing semantics. However, in a translation task, the speaker knows that the language of the word is not the language to be spoken. Here, we presented the word to be translated followed by a distracter word in one or the other language (e.g., Spanish GATO/CAT followed by English DOG). On some trials the distracter word was a word that would interfere with the translation word (DOG interferes with CAT), and it was presented in the language of production (as above) or in the same language as the original translation word (e.g., Spanish GATO followed by Spanish PERO/DOG).

The prediction is that we might not see interference in the second case (Spanish-Spanish) because the first word switched the language of speech production planning. We found that the interference occurred only when the language of the distracter

word matched the language to be spoken, unlike the picture-naming interference, where it is symmetric. We want to argue that a cue in the language input allowed production to proceed selectively.

What about more natural contexts of language use? Perhaps the language of a sentence context provides a cue that would eliminate these nonselectivity effects. We used a set of words that had previously been shown to elicit parallel activation, and we examined the effect of a sentence context on cognates with similar or different phonology. As a baseline measure, we found that people are faster to name cognates with phonology convergence than to name cognates with distinct phonology. Next, we used a rapid serial visual presentation (RSVP) paradigm to present the words in sentences (each word of the sentence is presented one at a time rapidly), and the speaker has to say the word that is presented in red out loud. The words in the sentence always occurred in only one language, but the critical words printed in red were these cognates with similar or different phonology. In addition, some sentences were highly constrained with respect to the critical item and some were unconstrained: For example, 'The composer sat on the bench and began to play the PIANO as the light dimmed,' versus, 'We noticed there was a very large PIANO by the window'.

In the high constraint sentences, the effect of cognate phonological similarity disappears. There is no longer evidence for parallel activation of the wrong language. When sentence constraints are low, when anything could appear, the effect was the same as it was when the words were out of context. The language of the context alone is not sufficient to override parallel activation, but when it is coupled with semantic constraints, it is sufficient for these effects to go away.

The data that I have shown so far suggest that language-specific information during an event that initiates speech planning might work to reduce interlanguage competition. The data from these sentence-processing experiments in context suggest that there must be convergence between that language-specific information and meaning to reduce cross-language competition. Interestingly, there are a whole set of studies in the literature that seem to show that cues that you might think would work, do not. Sentence context itself is not sufficient. The intention to use one language alone in both word recognition and production is not sufficient. Instructions are not sufficient. Cross-language script differences do not make a difference.

If it is so hard to reduce activation through all these cues, there must be some other means for resolving the conflict. Although this work is still in its infancy, I want to argue that the mechanism is most likely to be some form of inhibition. Do bilinguals inhibit one language when they speak another? There is some recent research suggesting that bilinguals (Spanish-Catalan) do not require inhibition when switching between their two languages. In two studies in our lab, we tried to capture what happens when someone is immersed in a second language. We hypothesized that one of the many things that happens through immersion in a second language is reduced activation and potential inhibition of the first language.

In the immersion environment, there could be both suppression of the first language and potentially unique cues available to the second language. The first experiment tried to simulate immersion in the laboratory with a group of students performing a vocabulary acquisition study. We taught a group of American college

students who knew no Dutch or German a set of 40 Dutch words. The students either associated the Dutch words with their English translations, or associated the words with pictures of objects to which they potentially referred. Some of the time, the pictures were presented in a noncanonical view, for instance, upside-down. From many studies of visual cognition, it appears that naming objects in noncanonical views involves mental rotation of the object to its canonical view. This additional process might inhibit retrieval of the name in the first language. If the problem with second language acquisition is an inability to inhibit the dominant first language name, having the object in a noncanonical view might help. Moreover, the noncanonical view might provide a unique cue for the second language.

At test, the subjects had to translate words from English to Dutch or they had to name pictures in Dutch (presented in both normal and noncanonical views). There is typically a cost to naming objects in noncanonical views. Will the cost be observed for learners who learned the words by associating them to the noncanonical views? People who were trained on words showed the typical naming costs associated with noncanonical views at test. Remarkably, for people who were trained on the pictures, there was facilitation when tested on pictures, and people who were trained on pictures, but tested on translation, also showed facilitation. This suggests that these subjects had acquired an abstract representation of the meaning of the Dutch word that has been facilitated by having this odd view of the object. It might be possible to inhibit briefly the first language in order to learn the second language.

In the second study, we looked at real immersion. This is difficult to study because when you look at students who travel to another country, they do so for a variety of reasons, they spend time with a variety of people, and it is very easy in study-abroad programs to spend time with other students who are not really speaking the second language. There was a group of students in Spain and another control group of classroom learners, and we equated them on a number of variables. The control group actually had more semesters of Spanish than the immersed group, and they were equated on working memory span.

Both groups of subjects performed a translation recognition task in which two words are presented, each in a different language, and the task is simply to say whether they are translations of each other (e.g., HOMBRE-MAN). For the trials in which the words were not translations, there could be one of four different kinds of foils: 1) The Spanish foil could be lexically related to the English translation (MANO/HAND-MAN), 2) The Spanish foil could be lexically related to the actual Spanish translation (HAMBRE/HUNGER-MAN), 3) The Spanish foil could be semantically related to the English translation (MUJER/WOMAN-MAN), or 4) The Spanish foil could be completely unrelated to the English translation (CASA/HOUSE-MAN). We compared the time it took the subjects to reject the mismatched translations for the three related conditions to the unrelated control condition.

The classroom learners at Penn State showed interference across all three comparison conditions, but the immersed learners had a different pattern of results. The lexical form condition showed no effect for the immersed learners, they are no longer showing interference from the first language for lexically related items. Interestingly, they showed an increase in the amount of semantic interference. This

suggests that they are more readily able to derive the meaning of the word in the second language than their counterparts in the classroom.

The students also performed a verbal fluency task, in which they were given a category label and asked to generate as many exemplars as possible over 30 seconds. They performed the task for four different categories in each language. Both the classroom and immersed students provided more exemplars in English than in Spanish, but the immersed students provided fewer in English than the classroom students and more in Spanish. It is not just that their Spanish is getting better, but their English is affected as well. This result rebounds when the students return. This suggests that there is at least temporary inhibition of the first language.

So far, the results suggest that lexical access in language production is language nonselective. There appears to be parallel activation of both alternatives. There is significant activation in the first language when speaking the second language, even for highly proficient bilinguals. That activation may produce active competition that needs to be resolved. The ability to negotiate that competition may come in part from cues that reliably signal the second language and in part from an ability to inhibit that irrelevant information. This work holds promise for beginning to understand how bilinguals are so clearly activating both languages in this out of context environment, yet in actual language performance bilinguals very clearly have a great deal of control over the language they intend to speak.

What are the questions that need to be asked next for this research agenda? I have not said anything about language-specific factors. Languages differ in a variety of ways, and some language-specific factors may not have much effect on this process, but presumably some do. For example, morphology may turn out to be very critical when we move to that level. A second issue is to consider the implications of the findings from this research for cross-linguistic research. It is often the case that the English speakers in cross-linguistic studies are monolingual and non-English native speakers tend to be bilingual. In many cases, it is possible that cross-linguistic studies are really comparing monolinguals and bilinguals rather than two different native languages. Another question that I have touched on that is beginning to emerge at many different levels is that the first language is affected by the second language, and we do not currently have a good psycholinguistic model for how this works. I have said almost nothing about the consequences of age and context of acquisition or of language maintenance. We also need to know more about the cognitive consequences of this competition. Finally, I have not said anything about the neural basis of bilingual performance. For the most part, recent neuroimaging studies show that the same neural tissue is activated by both languages, but there are still many issues to be resolved.

APPLAUSE

Questions

Professor Michele Miozzo: I think your data are convincing in demonstrating that activation of both phonological forms occurs, and I also have data that converge with yours. However, I think that this might not be sufficient to demonstrate separability of the mechanisms. I think there is still room for showing that the trick in very

proficient bilingualism is explained by separate and distinct lexical mechanisms. We have shown in the picture-word interference paradigm that there is an identity effect. I think the only way to explain this effect is by assuming a language-specific selection mechanism. You can still have activation of both phonological forms, but the selection takes into consideration only the forms of one lexicon.

Professor Kroll: I think it is very clear that selection eventually works. As you know, the question that is theoretically difficult to answer is whether the activation of these alternative forms results in active competition. When that selection mechanism operates, it might be taking into account, however momentarily, alternatives that are not intended for production. But, the evidence cannot resolve the alternate conceptualization, in other words, whether these alternatives are active and must be inhibited, or whether there is an a priori mechanism that creates bias so that competition does not happen. I think it is possible that both alternatives are true, and that they occur under different circumstances. One of the things that bilingual work teaches us that we would not learn from studying monolinguals alone is that the complacency encouraged by the monolingual speech production literature is blocked: the experimental measures discourage a model of fixed loci at which specific effects occur. The bilingual work shows that there are no fixed loci. There are a variety of factors, in this case having to do with the proficiency of the language, the relative activation of the language, or the context. Instead of fixity, we see something more like a moving window. I think the results from the cued picture naming task that look so selective for the first language and so unselective for the second language show that it may in fact be that way.

One problem with using a picture naming task is that when you look at a picture, you activate a whole cohort of visual and semantic items. Unlike processing words, the visual cohort is correlated with the semantic cohort. There may be a front-end process in picture naming that delays object identification, and any delay in that process for a bilingual will increase the likelihood that both languages are active. Therefore, this kind of task may not be appropriate for investigating early selection.

Professor Boris Gasparov: Some people learn a second language late as opposed to those who spoke both languages from childhood. This developmental difference can produce differences in proficiency, which might influence this quality of nonselective treatment. But, is there a qualitative difference between the two kinds of second language learning?

Professor Kroll: I did not mention this at the beginning of the talk, but cognitive psychologists in general are quite broad in their definitions of what counts as a bilingual. Basically, it can include almost anyone who is attempting to use a second language. The small amount of research that has been done suggests that this nonselectivity occurs right away. There is more of an effect of the first language on the second language the more dominant you are in the first language. In these kinds of experiments, learners are more dependent on the presence of the translation in the first language during the early phases of late second language acquisition. With increasing proficiency, that reliance on the first language translation drops out.

Ms. Jane Piliti: When bilinguals switch between languages, they often mention a motivation to hide their conversation as a reason for switching.

Professor Kroll: There is a rich tradition in sociolinguistics looking at code-switching. The two main bodies of code-switching look at grammaticality effects or the effects of a more social context. We know very little about how those kinds of factors might interact with the kinds of language processing effects I have been talking about.

Professor Remez: Psychologists are used to thinking of the lexicon as something that begins with semantic categories, and that phonology is the unimportant part—it is just the part that you use to control articulators when you want to talk. However, phonologists think of the phonology as an addressing scheme in which markers are used to distinguish words; without the phonology, all words become homophones. English and Dutch can be thought of as languages that have very similar phonologies. If you simply used the phonologies as addresses for all your lexical nodes, you would probably get blending as a simple consequence of similarities in phonologies. Which languages would you have to use to see if very different phonological organizations might actually reduce competition intrinsically?

Professor Kroll: You might want to look at Chinese-English speakers, or speakers of other tone languages to ask that question. This has not been studied very much, but I know of very little evidence in the few studies that have been done that have looked at languages that differ in that way. There is no evidence so far that I know of that suggests that this could be used as a cue to sort out the two languages at that level.

Professor Suparna Rajaram: I am sympathetic to the nonselective activation view for which you have presented strong evidence, but I am thinking of a particular type of code-switching that occurs within a sentence. Just based on informal observation, it seems that people do this sort of code-switching for two reasons. One is for effect—it is just much more effective to use certain expressions and words in one language rather than the other. The second happens when they fail to access the words or expressions that they need in a particular language, so they insert from the other language. In either case it seems that it is not nonselective activation, they are really choosing the best or the only available language for that particular item in the sentence.

Professor Kroll: If you have a case where there is a retrieval failure, then it defaults to selectivity. It is not clear that it is defaulting to selectivity because nonselectivity is incorrect or whether there is simply nothing to compete with. What I know of the code-switching literature is that even in the within-sentence cases, it is very grammatically constrained. You only see code-switching within certain types of grammatical structures. The idea that the grammatical structure could be used specifically to constrain nonselectivity is a nice result. Ultimately, we want selection.

Professor Remez: Let us thank Professor Kroll and adjourn.

APPLAUSE

Place: Kellogg Center, Room 1510
School of International and Public Affairs
420 West 118th Street
Time: 4:00 PM

21 APRIL 2005

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

Attendees: Stefan Benus, Lila Braine, Kerry Fischer, Boris Gasparov, Peter Gordon, Jill Grose-Fifer, Robert Krauss, Jackson Liscombe, Janet Metcalfe, Michele Miozzo, Jane Politi, Lois Putnam, Suparna Rajaram, David Rosenbaum, Ernst Rothkopf, Adam Shavit, Alexandra Suppes.

Rapporteur: Jennifer Pardo

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

Jennifer Pardo
jsp2003@columbia.edu

