

MAYFEST 2006

Language Learning Fest: Counts, Cues, Constraints and Computation

May 12 & 13, 2006



UNIVERSITY OF
MARYLAND

About Mayfest

This two-day workshop is an annual event organized by the graduate students in the Department of Linguistics at the University of Maryland. This workshop presents the current state of research in a particular domain of interest to linguists from a variety of perspectives. This is the sixth year that Mayfest is being held.

Mayfest 2006 Organizational Committee

Christopher Dyer, Jobin Mathew, and Akira Omaki

Acknowledgements

Mayfest 2006 is the result of a lot of hard work by many dedicated individuals. The organizational committee would especially like to thank Paul Pietroski, Jeff Lidz, Bill Idsardi, Colin Philips, Amy Weinberg, Philip Resnik, Rob Magee, Kathi Faulkingham, Stacey Conroy, Lydia Grebenyova, Yuval Marton, Eri Takahashi, and Heather Taylor. We also gratefully acknowledge the supplementary funding provided by the University of Maryland Graduate Student Government.

Language Learning Fest: Counts, Cues, Constraints and Computation

Whatever underlying cognitive mechanisms for language usage are presupposed by any particular theory of language acquisition, the fact remains that the limited and noisy data children are exposed to is sufficient for them to learn their native language. This workshop brings together researchers from a variety of areas within the field of language acquisition who are working on the problem of how this primary linguistic data is processed by children learning language. We focus on two active and often independent subfields in language acquisition research, phonological development and syntactic development. In both fields, a series of talks representing contrasting theoretical and methodological perspectives (e.g., statistical, cue-based, computational modeling) are presented side-by-side. During the workshop, Bill Idsardi and Jeff Lidz will offer a summary of the viewpoints presented and facilitate a discussion between the participants and the theories they represent. By juxtaposing these divergent theoretical positions, the underlying problems facing each of these domains will be brought into sharper focus. Additionally, by presenting theories from subfields of language acquisition that are often researched independently, participants will have the opportunity to seek cross-domain commonalities by exploring general questions about the role of, for example, probability distributions, Bayesian statistics, and cues for parameter setting in a variety of linguistic domains.

Friday, May 12

		page
8:30 – 9:30	Registration & breakfast Location: 0108 Marie Mount Hall (next to Maryland room)	
9:30 – 9:35	Welcome	
9:35 – 10:35	<i>The Acquisition of Phonological Representations</i> Elan Dresher (University of Toronto)	4
10:40 – 11:40	<i>The Subset Principle: Consequences and Conspiracies</i> William Sakas (CUNY/Hunter College)	5
11:40 – 1:30	Lunch At local restaurants (refer to enclosed map)	
1:30 – 2:30	<i>Distributional learning and phonetic development</i> Daniel Swingley (University of Pennsylvania)	6
2:35 – 3:35	<i>Bootstrapping Grammatical Categories from Structure-Independent Distributional Cues</i> Toben Mintz (University of Southern California)	7
3:35 – 4:05	Coffee & cookies	
4:05 – 5:05	<i>Bayesian learning of grammars</i> Mark Johnson (Brown University)	8
7:00pm	MayFest Reception Location: Paul Pietroski's house (refer to enclosed directions)	

Saturday, May 13

		page
8:00 – 9:00	Breakfast Location: 0108 Marie Mount Hall (next to Maryland room)	
9:00 – 10:00	<i>Language and Learning</i> Charles Yang (Yale University)	9
10:05 – 11:05	<i>Learning to Hear a Language: A statistical account of developmental speech perception</i> Jessica Maye (Northwestern University)	10
11:10 – 11:40	Summarizing Remarks Bill Idsardi (University of Maryland)	
11:40 – 12:10	Discussion led by Bill Idsardi	
12:10 – 2:00	Lunch At local restaurants (refer to enclosed map)	
2:00 – 3:00	<i>Rapid Language Change as a Window on Language Learning Mechanisms</i> Carla Hudson Kam (University of California at Berkeley)	11
3:05 – 3:35	Summarizing Remarks Jeff Lidz (University of Maryland)	
3:35 – 4:05	Discussion led by Jeff Lidz	

The Acquisition of Phonological Representations

B. Elan Dresher

University of Toronto

How do children acquire mental representations of the sounds of their language? That is, how do they acquire phonological representations? One might think that they can be acquired directly from the acoustic signal, but I will argue that this is not the case. Phonological representations make up one part of a complex interacting system. Acquisition of representations must proceed together with, not prior to, the acquisition of the rest of the phonology. The notion that phonological representations — at least, surface or phonetic representations — can be acquired directly from the signal might seem appealing on a number of grounds. For one thing, it would make the task of acquisition simpler. There is a respectable-looking line of argument that leads to this conclusion, and it goes this way:

- Speech sounds are bundles of distinctive features.
- Children have specialized feature detectors designed specifically for language.
- If these detectors are working properly, children ought to be able to extract the correct phonetic features from an acoustic signal consisting of utterances in their language.
- The result of this extraction will be a representation of words in terms of their phonetic features, i.e., a phonetic representation.

I will argue that the above general approach, which I will call the Direct Access theory of phonological acquisition, is mistaken in a number of ways. First, phonetic representations are not simply bundles of features. Much work in modern phonology leads to the conclusion that surface representations contain in addition structures at various levels of what has been called the prosodic hierarchy (Selkirk 1986). I will briefly show why these structures (prosodic phrases and prosodic words, for example) cannot be identified directly from the acoustic signal.

A second dimension of phonological representation includes syllable structure and stress. I will recap the argument of Dresher 1999 that the acquisition of stress patterns requires a complex acquisition procedure, and cannot be read off of or abstracted from the acoustic signal. I will then turn to segmental representations, the heart of the Direct Access theory I will argue that the evidence from infant and child perception of speech sounds (e.g., Kuhl et al. 1992) bears not on the acquisitions of phonemes, as is sometimes claimed (Pinker 1994), but rather of surface phones. While early ‘tuning’ of a child’s perception of speech sounds could well be an important first step in the acquisition of phonology, it does not amount to the acquisition of phonological representations. There is thus no evidence that children can acquire phonemes by Direct Access.

I will argue further that acquisition of the phonemes of a language involves acquiring a set of contrasts (Jakobson, Fant and Halle 1952): in every phoneme, the contrastive distinctive features must be distinguished from redundant features. Therefore, even if learners have built-in feature detectors to process speech sounds, these detectors do not yield representations, as assumed by the Direct Access theory. Rather, early phonetic perception of speech serves only as an initial input to a much more complex acquisition process that results in phonological representations.

Selected References

- Dresher, B. E. 1999. Charting the learning path: Cues to parameter setting. *Linguistic Inquiry* 30: 27-67.
- Jakobson, R., C. G. M. Fant and M. Halle 1952. *Preliminaries to Speech Analysis*. MIT Acoustics Laboratory, Technical Report, No. 13. Reissued by MIT Press, Cambridge, Mass., Eleventh Printing, 1976.
- Kuhl, P. A., K. A. Williams, F. Lacerda, K. N. Stevens, and B. Lindblom. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255: 606-608.
- Pinker, S. 1994. *The Language Instinct*. New York: William Morrow.
- Selkirk, E. O. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3: 371-405.

The Subset Principle: Consequences and Conspiracies¹

William Gregory Sakas

Hunter College/CUNY

Gold (1967) showed that the Subset Principle (*SP*), though not by that name, is necessary (but not sufficient) for learning from positive data. *SP* is most familiar to linguists through the work of Baker (1979), Pinker (1979), Dell (1981), Berwick (1985), Manzini & Wexler (1987), Clark (1992), and others. The penalty for not obeying *SP* is very well understood. In this talk, we examine problems that emerge as we begin to think about how to apply *SP* in various contexts.

We begin by adopting the psychologically attractive assumption that the learning mechanism (*LM*) is memoryless; during the course of learning *LM* has no ability to recall past hypotheses that were entertained² or prior sentences that were encountered. Given this memoryless assumption, we present a safe definition of *SP*:

SP: When *LM*'s current language is incompatible with a new input sentence *s*, *LM* should hypothesize a UG-compatible language which is a smallest superset of {*s*}.

By "smallest superset", we mean a language that contains *s* and has no proper subset that also contains *s*. Although *SP* as defined above is safe (i.e., will not lead to chronic overgeneration errors), it is problematic since previous facts that were correctly learned may have to be abandoned if the *next* input does not exhibit them. Intuitively, in order for a learner with no memory for past learning events to abide by *SP*, each newly encountered sentence is essentially the first sentence the learner has heard.

In the worst-case, we prove that even a finite (e.g. parameterized) domain is not learnable unless every potential target language in the domain contains a *subset-free-trigger*: a sentence *s* such that the target language includes no proper subset languages that also contain *s*. Note that a subset-free-trigger is not necessarily an unambiguous trigger, so a single encounter may not correctly pinpoint the target language. But a learner with a fair text and no bias against the target will eventually converge. We give examples of how and to what extent languages and/or parameter settings conspire to create a scenario in which learning fails. It is surprising that *SP* and memoryless learning are incompatible (except in a domain with unnatural properties, e.g. where every language contains subset-free-triggers and there are no conspiracies).

Finally, we consider how memory in the form of a partially-ordered enumeration of grammars that places subsets before supersets, together with a psychologically plausible means of exploiting it, holds promise for *SP* compliancy. Preliminary data drawn from simulation experiments run on the CUNY CoLAG word-order domain will be presented.

Selected References

- Baker, C. L. (1979). Syntactic theory and the projection problem. *Linguistic Inquiry* 10, 233-280.S
- Berwick, R.C. (1985) *The Acquisition of Syntactic Knowledge*. Cambridge, MA: MIT Press.
- Clark, Robin (1992) The selection of syntactic knowledge. *Language Acquisition* 2.2, 83-149.
- Dell, François (1981) On the learnability of optional phonological rules. *Linguistic Inquiry* 12.1, 31-37.
- Fodor, J.D., and Sakas, W.G. (2005) The Subset Principle in Syntax: Costs of Compliance. *Journal of Linguistics*, 41, 513-569.
- Gold, E. M. (1967) Language identification in the limit. *Information and Control*, 10, 447-474.
- Manzini, M. Rita and Wexler, Kenneth (1987) Parameters, binding theory, and learnability. *Linguistic Inquiry* 18.3, 413-444.
- Pinker, S. (1979). Formal models of language learning. *Cognition*, 7, 217-283.

¹ This is joint work with Janet Dean Fodor, City University of New York much of which is drawn from Fodor & Sakas (2005).

² It is important to note that Gold-style learnability studies standardly assume that *LM* has access to an enumeration, or ordering, of hypotheses where subsets appear before supersets. Under our memoryless assumption, we disallow such an enumeration in much of the discussion here. Thus, some of the conclusions we draw may seem to contradict well-established formal results when this is not the case.

Distributional learning and phonetic development

Dan Swingley

University of Pennsylvania

Everyone agrees that infants learn phonetic categories from speech sound distributions. This learning yields one mechanism for guiding lexical development, namely discrimination failure—a surefire way to exclude some spurious phonological contrasts. But discriminable phonetic variation still needs to be interpreted. This problem will be discussed with reference to experiments testing 1.5-year-olds' interpretation of novel words, and analyses of acoustic distributions in child-directed speech. The assumption that an unbiased distributional learner would succeed in identifying phonological categories will be questioned based on speech data and experimental results using vowel duration as a test case.

Bootstrapping Grammatical Categories from Structure-Independent Distributional Cues

Toben Mintz

University of Southern California

Grammatical categories such as noun, verb, and adjective, are the building blocks of linguistic structure. Identifying the categories of words allows infants and young children to learn about syntactic properties of their language. Thus, understanding how infants and young children learn the categories of words in their language is crucial for any theory of language acquisition. Even if the knowledge of the existence of lexical categories is innate—e.g., that there are nouns, verbs, etc.—learners still must learn the assignments of words to categories. While most theories assume that learners rely on distributional information to categorize at least some words, in many theories, distributional analyses are thought to occur only in a highly constrained fashion, using knowledge of phrase-structure and previously categorized words (e.g., Pinker, 1984). Such *structure-dependent* distributional learning requires that the learner has already acquired language-specific syntactic knowledge, and thus such a mechanism could not serve as a means of initially categorizing words.

In this talk I will discuss a proposal that young language learners use distributional information contained in lexical co-occurrence patterns—specifically, *frequent frames*—as an initial means of categorizing words, and that they do so without recourse to prior syntactic knowledge (Mintz, 2003). A frequent frame is defined as any two words that co-occur frequently in a corpus with exactly one intervening word position. Studies by Gómez (2002) show that frame-like patterns are detectable by infants. I will discuss an analysis of six corpora of speech to English learning infants and toddlers that demonstrated that the words that are “framed” by a frequent frame are almost exclusively members of the same grammatical category. For example, the frequent frame *you__it* contains verbs, almost exclusively, and *put__in* contains nouns. Thus, frequent frames could provide an initial source of information, or a bootstrap, into the category assignment of words. I will also briefly discuss extensions of this work to French (Chemla, Mintz, Bernal, & Christophe, under review), and Mandarin (Xiao, Cai, & Lee, 2006), which suggest that frequent frames might be an informative distributional environment in any natural language. Finally, I will present behavioral evidence that 12-month-old infants use frequent frames to categorize novel words (Mintz, 2006).

Taken together, these studies suggest that while structural knowledge *per se* does not guide a distributional analysis using frequent frames, this particular environment captures enough structural regularity to be informative to a learner. Thus, infants innate processing biases (Gómez, 2002), while not specifically linguistic, seem to select distributional environments that are structurally constrained, and may serve as a proxy for structure-dependent learning.

Chemla, E., Mintz, T. H., Bernal, S., & Christophe, A. (2006). Categorizing Words Using Frequent Frames: What Cross-Linguistic Analyses Reveal About Core Principles. Ms. under review.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, 13(5), 431-436.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90(1), 91-117.

Mintz, T. H. (2006). Finding the verbs: distributional cues to categories available to young learners. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action Meets Word: How Children Learn Verbs* (pp. 31-63). New York: Oxford University Press.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Xiao, L., Cai, X., & Lee, T. (2006, March). The development of the verb category and verb argument structures in Mandarin-speaking children before two years of age. Paper presented at The Seventh Tokyo Conference on Psycholinguistics, Keio University.

Bayesian learning of grammars

Mark Johnson
Brown University

While the most famous applications of statistical learning are perhaps word associations and neural networks, in the past decade we (i.e., the computational linguistics community) discovered how to extend these learning algorithms to grammars that generate linguistically-realistic structures. These techniques currently learn phrase-structure and similar grammars, but there is no principled reason why they can't learn other kinds of grammars as well. Bayesian approaches are particularly attractive because they exploit "prior" (e.g., innate) knowledge as well as statistical generalizations from the input. Structured statistical learners have two major advantages over other approaches. First, because the generalizations they learn and the prior knowledge they utilize are both expressed in terms of explicit linguistic representations, it is clear what was learnt and what information was exploited during learning. Second, because of the "curse of dimensionality", learners that identify and exploit structural properties of their input seem to be the only ones that have a chance of "scaling up" to learn real languages.

Of course, developing explicit computational models that actually learn language is more difficult—and more interesting—than constructing the kind of "in principle" arguments given above. Time and audience permitting, I will describe the Markov Chain Monte Carlo techniques we have developed for sampling from Bayesian posterior distributions over syntactic analyses and grammars, and our use of Dirichlet Process models to address over-dispersion in lexical and morphological acquisition.

Language and Learning

Charles Yang

Yale University

For divergent methodological orientation to converge, a prerequisite is to take each other's results and goals seriously. In this talk, I will provide an assessment of some recent studies on language acquisition—phonology, morphology, and syntax—from developmental psychology, computational modeling, and theoretical linguistics, to demonstrate that there is still a great deal of disconnect. As a remedy, I suggest that the understanding of language learning ought to be partitioned in two parts: language and learning. I argue that what's remarkable about language learning is really language, not learning.

Learning to Hear a Language: A statistical account of developmental speech perception

Jessica Maye
Northwestern University

Different languages utilize different sets of phonemes and phonemic contrasts; for example, English distinguishes the phonemes /l/ and /r/, but Japanese does not. Further, in two languages that utilize the “same” phoneme category the particular phonetic realization of the phoneme is unique to each language; for example, /d/ is generally characterized as a dental prevoiced stop in Spanish but an alveolar short-lag stop in English. In other words, the set of actual sounds (‘phones’ or ‘phonetic categories’) that speakers produce in a given language is unique to the language. In addition to these segmental properties, each language follows a unique set of constraints (typically referred to as ‘phonotactic’ constraints) on what order the sounds may go in; for example, Russian allows the sequence of /rt/ to occur at the beginning of a syllable (e.g. *rta* ‘mouth’), whereas in English the sequence /rt/ may only occur at the end of a syllable.

In this talk I will discuss how both segmental and sequential properties of the native language sound system affect listeners’ perception of speech, such that speakers of different languages actually hear speech differently (an effect that may be thought of as foreign accented perception). I will present research investigating these perceptual effects in adult speakers of different languages, as well as the development of these effects in infants as they begin to learn the phonetic and phonological properties of their native language. The findings of this research indicate that infants begin to acquire the phonetic properties of their native language on the basis of statistical properties of their speech environment, and that the algorithm by which infants track these statistics must incorporate information about phonetic and/or phonological context.

Rapid Language Change as a Window on Language Learning Mechanisms

Carla Hudson Kam

University of California at Berkeley

One of the most notable features of first language acquisition is the almost universal success that children have; children world wide seem to easily acquire a language that looks very much like the language spoken by the people who provided their input. Occasionally, however, this is not the case. Instead, learners with completely normal cognitive capacities change the language as they learn it. These “failures to learn” are in apparent conflict with a great deal of work showing the incredible sensitivity of human learners to the statistics present in the input, and as such, are often pointed to as evidence supporting the operation of innate domain specific-learning mechanisms. In my work I focus on a particular kind of change that is often seen when children learn from non-native speakers—regularization—and ask about the nature of the mechanisms that might be responsible for such changes. I will present results from a series of artificial language experiments showing that children and adults appear to differ as to what they learn and what they regularize: Data from a production task shows that adults are relatively good at learning variation present in their input, whereas children are very likely to regularize probabilistic variation, suggesting that the children are doing something different from adults when learning from input containing inconsistent patterns. However, data from judgment tasks show that children (and adults who regularize) are sensitive to the underlying probabilities, despite the fact that their productions do not reflect the statistics. But why then are their productions regular? I propose that regularization actually emerges from processes involved in language production, rather than directly reflecting learning differences. Thus, regularization may reflect the operation of a more general learning mechanism that is sensitive to statistics, in interaction with a domain specific production system that is less so.