Linguistically Predictable:
When, How, and Why Do We Predict in Language?

This event is funded in part by the Linguistics Department and by the
Graduate Student Activities Fee and is
therefore open to the entire Graduate Student Community
Expectations in Continuous Speech: Evidence for a “Data Explanation” Approach
Michael K. Tanenhaus - University of Rochester, Department of Brain and Cognitive Sciences

I will briefly review three lines of work that focus on how listeners make real-time inferences mapping fine-grained acoustic/phonetic information onto potential referents. The first line of work establishes 200 ms as the earliest lag with which information in the speech signal can affect fixations. We then show that listeners use information from the first pitch period in a preceding vowel to generate lexical hypotheses (e.g., the shwa in “the” in “the lamp…”). The second line of work shows that the information structure of an utterance affects whether or not increased syllable duration is interpreted as evidence for a prosodic boundary in an onset embedded word (“hamster”). This result is unexpected given traditional approaches to spoken word recognition. The third line of work shows that changes in speech rate affect whether people hear the same acoustic input as “…a yellow duck swimming” or “…yellow ducks swimming”. These results demonstrate that speech rate rapidly modulates expectations, even within the first few syllables of an utterance. Ongoing work uses appearing and disappearing indefinites as a domain for exploring real-time adaptation and cue integration within a belief-updating framework.
Predicting language and connecting brains
Suzanne Dikker - New York University, Department of Psychology

It is widely assumed that prediction plays a substantial role in language comprehension, but only very few studies provide direct evidence pertaining to the neural correlates of the prediction process itself. This presentation discusses a series of experiments that investigate (a) how syntactic and lexical-semantic predictions may affect early sensory processing (MEG) (b) where and how predictions are generated (MEG), and (c) whether prediction facilitates neural alignment between a speaker and his/her listeners (fMRI), i.e., if the ability to predict language affects whether interlocutors are ‘on the same wavelength,’ so to speak.
There is strong evidence that human sentence processing is incremental, i.e., that structures are built word by word. Recent experimental results show that the processor is also predictive, i.e., it can anticipate upcoming linguistic material on the basis of previous input. However, the granularity of this prediction process is currently unclear. We present evidence from two visual world experiments that show that speakers can make fine-grained predictions based on the frame and the frequency of a verb. We use this data to motivate a computational model of human parsing that includes an explicit mechanism for generating and verifying predictions. Based on this mechanism, our model can capture both locality effects and surprisal effects, and thus unify empirical results that have so far been accounted for separately. We evaluate the model on the Dundee eye-tracking corpus and on relative clause data that existing models fail to capture adequately.

Joint work with Manabu Arai, Vera Demberg, and Roger Levy
An integrated theory of language production and comprehension
Martin Pickering - Psychology, University of Edinburgh

Currently, production and comprehension are regarded as quite distinct in accounts of language processing. In rejecting this dichotomy, we instead assert that producing and understanding are interwoven, and that this interweaving is what enables people to predict themselves and each other. We start by noting that production and comprehension are forms of action and action perception. We then consider the evidence for interweaving in action, action perception, and joint action, and explain such evidence in terms of prediction. Specifically, we assume that actors construct forward models of their actions before they execute those actions, and that perceivers of others’ actions covertly imitate those actions, then construct forward models of those actions. We use these accounts of action, action perception, and joint action to develop accounts of production, comprehension, and interactive language. Importantly, they incorporate well-defined levels of linguistic representation (such as semantics, syntax, and phonology). We show (a) how speakers and comprehenders use covert imitation and forward modeling to make predictions at these levels of representation, (b) how they interweave production and comprehension processes, and (c) how they use these predictions to monitor the upcoming utterances. We show how these accounts explain a range of behavioral and neuroscientific data on language processing and discuss some of the implications of our proposal.
How is the motor system involved in linguistic prediction processes? What are the possible computational advantages of motor-based predictions? In the context of speech processing, I model the prediction mechanism as an internal simulation-estimation sequence: articulatory commands are internally simulated to estimate the corresponding auditory and linguistic consequences. Three MEG studies – using mental imagery of speech as a model – are presented to illustrate this simulation-estimation computational architecture. Experiment 1 characterizes articulation imagery: a fixed activation pattern sequence is observed over somatosensory (parietal) areas followed by auditory (temporal) areas. This activation pattern supports the hypothesis that the speech target is estimated over the same regions that mediate auditory perception, and that such auditory estimation is carried out in a possible sequential prediction structure (Tian & Poeppel 2010). Two follow-up studies investigate the functional specificity of auditory estimation by investigating how internal estimation interacts with external feedback. Experiment 2 shows that auditory estimation derived from the internal simulation of articulation increases the neural responses to subsequent perception at the phonological level. These results demonstrate that auditory predictions actively modulate the sensitivity of auditory perception and reduce the variance in perceptual processes (Tian & Poeppel 2013). Experiment 3 shows that auditory estimation extends to acoustic level. The acoustic comparison is constrained by the differences between estimation and feedback in spatial and temporal features of speech. These results indicate that the prediction can provide an auditory target for detecting deviance during speech production (Tian & Poeppel under review). In summary, motor-based auditory predictions are available and interact with processes of external stimulation at multiple representational levels. Such internal simulation-estimation process and internal-external interactions provide the neurocomputational foundations for facilitation in speech perception as well as the temporally precise monitoring and finely graded online control of speech production (Tian & Poeppel 2012).
What does cloze probability mean?
Adrian Staub - Umass Amherst

In almost all research on predictability effects in comprehension, a word’s predictability is defined as the probability of the word being provided as a sentence continuation in the cloze task. We suggest, however, that the meaning of the cloze probability variable is not transparent, and that understanding its meaning requires an explicit cognitive model of the process of generating a cloze response. We present the Cloze Race Model (CRM), which is a simple evidence accumulation model in which potential sentence continuations independently race toward a response threshold. A version of the CRM in which sentence context modulates the distribution of potential responses’ starting points captures detailed data patterns from a large-scale cloze experiment in which response time (RT) was also recorded. Most notably, RT is faster when a response has higher probability, and when the item is more constraining. The CRM naturally simulates these independent effects. The data and model have a range of implications. Among other things, they suggest that variable cloze responses may arise simply from the stochastic nature of the response process, rather than from inter-subject variability, and that cloze probability may not be a linear, or even monotonic, function of a word’s activation by context.
Better than expected: The dynamics of prediction-based processing in younger and older adults' language comprehension
Kara D. Federmeier
Emanuel Donchin Professorial Scholar
Department of Psychology,
Program in Neuroscience, and
Beckman Institute for Advanced Science and Technology
University of Illinois

Compelling electrophysiological evidence that language processing can be facilitated by expectations for semantic, lexical, and sensory features of likely upcoming words has dramatically changed views of language comprehension over the past decade. This evidence for facilitative effects of prediction is complemented by findings of processing consequences when predictions are disconfirmed. The field's increasing appreciation of the role of expectation-driven processes in language learning and comprehension raises new questions: what are the neural and cognitive mechanisms underlying predictive processing? what are the benefits and consequences of prediction? what types of information are predicted and under what circumstances? In my talk, I will briefly overview the history of empirical investigations into predictive processing in language and will review current evidence about the neural and cognitive bases for such expectation-based processing. I will then describe emerging evidence showing that multiple language comprehension mechanisms are implemented in parallel and that the brain dynamically adapts its use of these mechanisms, both over the long-term, in response to changing neural and cognitive abilities with age, and over the short-term, in response to situational and task demands.
Predictability and probability in language comprehension
Nathaniel Smith - University of Edinburgh, School of Informatics

It's well known that comprehenders process words differently depending on how predictable they are in context -- for example, more predictable words are read more quickly. But it's not clear how such predictions are represented, why they affect processing time, or how they are arrived at in the first place. I'll discuss two studies which may shed some light on these questions.

In the first, we combine a state-of-the-art computational language model, two large behavioral data-sets, and non-parametric statistical techniques, to measure for the first time the quantitative form of the relationship between reading time and (estimated) word probability. We find that -- as hypothesized by surprisal theory -- it is logarithmic over six orders of magnitude. The result lends partial support to an existing optimal perceptual discrimination account of word recognition (Norris, 2006); I'll also describe a novel model which combines a domain-general account of predictability effects on processing time with a language comprehension system that is highly incremental well below the level of individual words, and show that it predicts both the shape and time-course of this effect. More generally, the existence of such a dramatic systematicity suggests that probability is the correct method for quantifying linguistic predictions, and its logarithmic form means that even for words with very small absolute probability (1-in-10,000 and below), relative differences in probability still produce substantial reading time effects. This allows us to use a single underlying mechanism to explain predictability effects that reflect different levels of linguistic structure (lexical, syntactic, discourse coherent, etc.). It also suggests that predictability effects arise from an explicit, graded representation of probability, rather than implicitly from a discrete anticipatory guessing process, and thus provides challenges for both semantic integration and anticipatory processing accounts.

But if the comprehension system has access to explicit estimates of word probabilities, then where do they come from? In a second study, we used the cloze task to elicit participants' subjective probabilities, compared them to objective probabilities from corpora, and found large and systematic differences. Some of these differences appear to be artifacts induced by the task or by the non-representativity of the corpora; others, however, turn out to
match the principle biases intentionally introduced in computational linguistic models to better estimate word probabilities from sparse and limited data. Humans similarly must estimate probabilities from sparse and limited data, so these biases may reflect a statistical learning mechanism which uses savvy methods to avoid overfitting. Together, these results paint a picture of a highly incremental comprehension system which performs explicit and principled probabilistic computations across multiple representational levels.

Joint work with Roger Levy
Growing Expectations
Akira Omaki - Cognitive Science, John Hopkins University

Much psycholinguistic research has investigated the relation between distributional information and expectations in language comprehension. One way to study this relation is to examine how expectations arise during sentence processing as a function of relevant distributional information. An alternate approach is to explore how such predictive behaviors grow in children as a result of relevant distributional information that accumulates over a longer period of time. This presentation will discuss a series of studies that explore the second approach, with a focus on non-adjacent dependencies that allow us to examine how the beginning of such a dependency can serve as an effective cue for expecting the end of the dependency. The first part of the talk explores the role of transitional probability information (and its breakdown) in learning a dependency between is and ing (e.g., John is kick-ing the ball). It is shown that 15-month-old infants, who generally do not demonstrate sensitivity to the co-occurrence of is and ing, can learn to detect the co-occurrence relationship after a few minutes of exposure to input in which the local transitional probability between is and the verb becomes very small. The fact that children rely on distributional information to discover non-adjacent dependencies suggests that it is very likely for children to also use probabilistic information during parsing of non-adjacent dependencies during comprehension as well. The second part addresses this question by reviewing cross-linguistic studies on filler-gap dependency processing in children, and demonstrates that children's syntactic expectations in filler-gap dependency processing may be slightly different from the type of expectations that have been observed in the adult sentence processing literature, despite the fact that the distributional information available for children should fully support adult-like expectations. These findings highlight the fact that distributional information is probably not the only factor that contributes to predictive behaviors during language comprehension. I will discuss other linguistic and cognitive factors that may need to grow before adult-like predictive behaviors emerge.
Predicting When to Attend in Speech Perception
Lisa D. Sanders - NeuroCognition and Perception Lab, University of Massachusetts - Amherst

Temporally selective attention, the preferential processing of stimuli presented at expected times, is critical when information in presented too rapidly for all of it to be processed in detail. Under these conditions, sounds and images presented at attended times elicit better behavioral performance and larger sensory event-related potentials (ERPs) compared to identical stimuli presented at unattended times. When presented with continuous speech, listeners direct temporally selective attention to the initial portions of words resulting in better performance and larger amplitude ERPs for sounds presented at these times. However, recent data from both natural and artificial language studies suggest it is the decreased predictability of the sounds at the beginnings of most words, rather than word boundaries themselves, that drive the allocation of attention. That is, listeners attend to the sounds in continuous speech that they are least able to predict from the preceding context. Ongoing research is focused on whether predictability calculations directly cue temporally selective attention, or if listeners use other cues such as stress and rhythm that may be associated with the predictability of upcoming speech sounds. Further, to the extent that the predictability of the difficulty of processing upcoming speech sounds is related to the predictability of lexical items and syntactic structures, predictability at these levels is also hypothesized to play a role in the allocation of attention during speech processing.