MONDAY June 2nd 09:25 – 10:40

Tuning the blueprint: How studies of implicit learning during speaking reveal the information-processing components of the production system

Gary Dell
University of Illinois at Urbana-Champaign

The ”blueprint for the speaker” is a diagram that occupies a place of honor in *Speaking*; it is Figure 1.1. Let us consider what a blueprint does, such as a blueprint for a house. It tells us how many rooms there are, and which rooms are next to which. It tells us which rooms are closed off by doors (e.g. the bathroom) and which rooms have been blended together in a trendy open-concept design (e.g. kitchen/dining-room/living-room). It may tell us about nearby structures that are important for the house, but not technically part of it (e.g. driveway). The blueprint for the speaker does, analogously, all of these things. It identifies processing components and their arrangement, with components existing with more or less isolation from others. It makes additional claims about how things outside of production (e.g. comprehension) coexist with production.

Much psycholinguistic research has tested the blueprint and later theoretical production models inspired by it. This talk will consider how we can investigate the modularity of processing components in production by studying implicit learning. We know that each act of language production (and comprehension) tunes the system. In production, the implicit learning of syntactic structures, lexical retrieval, and phonotactic constraints have all been identified and studied experimentally. The talk will give examples from phonotactic and structural learning paradigms in which the key result is the success or failure of learning that some production decision X depends on some feature Y. For example, can one implicitly learn that a new phonotactic constraint such as /f/ must be an onset (X), depends on an adjacent vowel (Y)? If so, that tells us /f/-onset decisions easily access vowel information. It’s as if the consonant-onset biases and vowel information are in the same room, or at least that one can see the vowel when standing in the consonant room. By considering such research we can test out particular blueprints and, if required, do some remodeling to our understanding of production.

MONDAY June 2nd 10:55 – 12:10

On the logic of speaking *in situ*

Herbert H. Clark
Stanford University

The reason people communicate with each other is to coordinate on activities they are jointly engaged in—from gossiping and getting acquainted to doing the tango and playing cards. The means by which they communicate should therefore reflect what they need to do in order to coordinate. Here I focus on a process I will call joint closure. The idea is that when Allan, say, is saying something in conversation to his friend Barbara, they are each *simultaneously* signaling the other, Allan by speaking and gesturing, and Barbara largely by gesturing. The two of them recognize that what they each produce is open, or provisional, until they reach joint closure on what they take Allan to have meant. Evidence shows that speakers repeatedly revise and annotate the content and timing of what they produce in tight correspondence with the actions of their addressees, and vice versa. The conclusion: speaking and being addressed *in situ* are inherently bilateral processes.

.

MONDAY June 2nd 13:55 – 14:40

Speaking **as inspiration**

Stefanie Shattuck-Hufnagel
Massachusetts Institute of Technology

With the publication of his landmark volume *Speaking* 1989, Pim Levelt accomplished something that few investigators have been able to do for their field: He summarized the current state of a complex and rapidly developing discipline, the cognitive science of language production, and at the same time laid the groundwork for a flood of new work, both by his own research team and by many others, to model the processes and representations that undergird communication via language. I share with many of his colleagues the experience of having my own thinking about these issues both broadened and sharpened under the prodding of such a comprehensive yet explicit and detailed model, as well as by the highlighting of what we don't yet know. An example of this effect is the motivation his work has provided to close the gap between modeling the development of the plan for an utterance and modeling the execution of that plan, inspiring a model of the process by which speakers select context-appropriate acoustic cues for the features of the planned utterance and generate utterance-specific values for those cues (Turk & Shattuck-Hufnagel, to appear). The breadth of his vision has encouraged me to expand my research focus to include phrase-level prosody as a contextual factor that influences the segmental feature cues that are signaled, and co-speech gestures of the hands, head, face and torso as co-planned aspects of the communicative act. Levelt's work has allowed us to envision for the future an even more comprehensive and explicit model that produces acoustic wave forms, with appropriate patterns of context-governed acoustic phonetic variation, co-speech gestures, disfluencies and errors.

MONDAY June 2nd 14:40 – 15:55

Reference production: efficiency or communication?

Jennifer Arnold
University of North Carolina, Chapel Hill

In Levelt's (1989) book *Speaking*, he highlights the dual nature of the speaker as both information processor (chapter 1) and a social participant in communication (chapter 2). In this talk I examine how these themes affect current theories about one domain of language production: reference. Successful reference requires selecting a form that is both informative enough to communicate the intended referent, and contextually appropriate. Thus, speakers have a wide range of choices, including pronouns (*he*), names (*Pim)*, descriptions (*the professor*), or modified expressions (*the production researcher).* Yet researchers disagree about what mechanism drives these choices. Most scholars agree that explicit forms are needed when the context does not help identify the referent, but it is less clear why speakers would ever use a reduced form like a pronoun. I focus on two theoretical explanations: 1) reduced forms are **pragmatically required**, and 2) reduced forms are **easy**. I demonstrate that both views make predictions that are not borne out by data, and present a new model of reference production.

Levelt (1989) describes the classic view of reference production, where pronouns, definiteness, and acoustic variation are driven by the conceptual accessibility of the referent. This is presented as a grammar-driven mechanism, following rules like "IF the referent is in the addressee's current focus, THEN Assign it the value + in focus" (p. 149; see also Schmitt, Meyer, & Levelt 1999). This view overlaps with Grice's communicative view of language production (chapter 2), as well as models by Ariel, 1990; Chafe, 1976; Gundel et al., 1993). However, actual productions are more variable than would be predicted by this account (Zerkle & Arnold, 2016).

An alternative view of language variation comes from information-theoretic models (e.g., Aylett & Turk, 2004; Mahowald et al., 2013), which propose that referential explicitness reflects the pull between two forces: the need to provide enough information to be understood, and the speaker's desire to be **efficient**. These models are typically used to account for acoustic variation or word choice, but the models are broad and might be assumed to explain pronoun production as well. However, they make the incorrect prediction that pronouns are produced because they are efficient or easy. I review evidence against this view.

I propose instead that both information-theoretic and traditional models have lost sight of an old idea, namely that some reference forms serve the function to link the current utterance with the context. Levelt points this out: "…definiteness, pronominalization, and (de-)accentuation… are anaphoric grammatical devices for linking the interpretation of the current utterance to what was said earlier or what is about to be said." (1989, p. 283). Thus, reduced forms like pronouns carry a heavy discourse requirement, which may incur a memory load. By taking this function seriously, production models must account for the memory load of maintaining the discourse context, which may explain both discourse context effects and production processing effects on reference production.

MONDAY June 2nd 16:10 – 17:25

From conceptual representation to linguistic meaning

Anna Papafragou
University of Delaware

How do we talk about what we see? Current models of language production assume that the act of speaking begins with conceptualization (e.g., the preverbal apprehension of the broad details of an event), followed by further processes of information selection and linguistic formulation that culminate in speech execution. This view on how conceptual representations make contact with language faces two challenges. First, the underlying preverbal perceptual/conceptual representations themselves can be elusive. Second, there have been few explicit attempts to connect the process of building conceptual representations to the mechanisms of language production. Here I address these challenges, focusing on the domain of events. I show that the cognitive representation of events is finely articulated in a way that aligns with the demands of the language production system. Furthermore, I provide evidence suggesting that, in both adults and young children, preparing to talk about a dynamically unfolding event changes the way speakers inspect the visual world, and does so in a way specific to the speaker's language community.

.

TUESDAY June 3rd 09:00 – 10:15

What can the bilingual speaker teach us about the generation of surface structures?

Sarah Bernolet
University of Antwerp

When Levelt's book *Speaking* was published, syntactic priming research was only starting to emerge. Now, 30 years later, hundreds of studies have used this method to investigate the representations and the mechanisms involved in the production and comprehension of syntax. Many of these studies have shed light on central assumptions of the Grammatical Encoder that is part of Levelt's model of the speaker (e.g. about the mapping between semantic and syntactic representations or the possibility of feedback between the phonological and the syntactic level of production). The use of cross-linguistic syntactic priming studies has broadened the range of questions that could be answered, but it also raised additional questions and concerns. Between-language priming appears to be modulated by second-language proficiency, and response tendencies may be more heavily influenced by participants' strategies and memory limitations during L2 syntactic production than during production in the L1. In order to be able to tell how much between-language priming data can tell us about the syntactic representations that are accessed during L2 production, we need an explicit theory on the development of syntactic representations during L2 learning.

Recently, Rob Hartsuiker and I put forward a theoretical account of late second language syntactic acquisition. We assume that structural priming effects in L2 (and between L1 and L2) depend on the structure of this developing network but also on explicit memory processes, and we speculate that these memory processes might aid the formation of new representations. In this talk, I will present the account and its predictions and focus on the way in which explicit memory processes might influence priming results in L1 and L2. I will talk about ongoing studies testing the model using a miniature language paradigm. Finally, I will argue why it is important to document the different phases in L2 syntactic development with ecologically valid learning data.

TUESDAY June 3rd 10:25 – 11:40

Semantic guidance of attention during language production: Evidence from scene meaning- and saliency maps

Fernanda Ferreira
University of California, Davis

Speakers often use language to describe the world around them. The current project attempts to specify how the linguistic and visual information processing systems interact during language production. A new, innovative approach to scene processing which precisely quantifies visual salience and meaning has demonstrated that attention in scenes is more strongly controlled by semantics than by visual salience, as shown by eye movements made to objects in scenes during a memorization and aesthetic judgment task (Henderson & Hayes, 2017). In the present work investigating language production, subjects were shown the same 30 complex, full color scenes (e.g., a kitchen or street) and asked to describe them or to suggest a set of actions a person might perform in the scene. Using detailed salience and meaning maps, we observed that fixations to scene regions are driven more by semantics than by visual salience, a pattern that held for almost every scene and across the entire viewing period. In addition, and consistent with results reported during scene viewing in non-linguistic tasks, we found that meaning had a particularly strong effect during the first couple of fixations, possibly reflecting the initial stage of extracting scene gist (Castelhano & Henderson, 2007). In vision-language tasks, this stage has been viewed as one that reflects holistic apprehension (Griffin & Bock, 2000), a suggestion that is reinforced by our finding that the advantage of meaning over salience was stronger in the action-suggestion task than in the simple description task, as was the time to begin speaking. These results provide new evidence in support of language production models which assume that the representation driving formulation is meaning-based, and which assume an initial holistic apprehension stage during which the speaker formulates a global plan for the entire linguistic sequence. This apprehension stage reflects processes related to utterance formulation as well as extraction of scene gist.

TUESDAY June 3rd 11:50 – 13:05

From blueprints to brain maps: The status of the Lemma Model in cognitive neuroscience

David Kemmerer
Purdue University

When Levelt's pioneering book about the Lemma Model of speech production appeared in 1989, cognitive neuroscience was just starting to take off. During the 30 years since then, remarkable progress has been made in refining this influential framework and relating it to the brain. My main goal is to provide a broad (and, due to time constraints, very selective) overview of these advances, drawing primarily on data from single word production experiments with both healthy and brain-damaged participants. I will first provide a synopsis of the model itself, starting with the original "blueprint of the speaker" proposed by Levelt (1989), and then turning to the more detailed framework elaborated by Levelt et al. (1999) and Levelt (2001). Next, in keeping with previous meta-analyses and reviews (Indefrey & Levelt, 2004; Indefrey, 2011; Price, 2012; Roelofs, 2014; Roelofs & Ferreira, in press), I will show that, according to numerous studies employing diverse brain mapping methods, spoken word generation depends on a predominantly (but not exclusively) left-hemisphere circuit in which the different levels of representation and computation postulated by the theory are hosted by different cortical areas, and the flow of information between them is largely sequential. Finally, I will point out, albeit very briefly, that even though this view of the neural substrates of speech production has many virtues, it also faces several challenges involving issues like the ontological status of lemmas, the nature of cortical dynamics, the mechanisms that transform sound-based phonological representations into motor-based ones, the initiation of utterances, and the precise ways in which auditory and somatosensory feedback modulate articulation. By covering all of this material, I hope to set the stage for further discussion of specific topics.

TUESDAY June 3rd 14:50 – 16:05

Meaning what we say: Semantic influences on lexical selection

Rasha Abdel Rahman
Humboldt University Berlin

The ultimate goal of speaking is to convey meaning. However, while semantic-categorical relations and their influences on lexical access are well-investigated, much less known about the role of other dimensions of meaning. I will present evidence on how different meaning aspects and contexts shape lexical selection. These include non-categorical relations as associations and thematic links, influences of the semantic distance of contexts, as well as emotional and social aspects of meaning in communicative situations. They also include message inherent variations such as the richness of their semantic representations and the density they inhabit in lexical-semantic space. This multifaceted nature of meaning processing goes together with dynamic adjustments of lexical selection to situational requirements and meaningful contexts, ranging from ad-hoc relations to associations and social-communicative affordances. Together, these findings highlight a high degree of plasticity and a complex interplay between lexical selection and semantic, emotional and social-communicative aspects of meaning processing.

TUESDAY June 3rd 16:25 – 17:40

Phonological encoding of single words

Niels Schiller
Leiden University

Speaking involves turning thoughts into controlled movements of the jaws, the lips and the tongue. That is, speakers start with an abstract, amorphous idea – not yet in the form of a lexical item – that is converted into something pretty concrete such as an acoustic signal. This process is fascinating but also complex. The current paper will focus on the lower end of the process. Once speakers have selected and retrieved the intended lexical items (lemmas) from their mental lexicon, they need to phonologically and phonetically encode those lexical items. Levelt (1989) described the corresponding processes in chapters 8 and 9 of his seminal book "Speaking". In those chapters, Levelt laid the foundation for further research that has been carried out in the last 30 years. The amount of research on phonological and phonetic encoding has been enormous. I will review this work selectively and try to sketch what the 1989 model looks like today.

.

WEDNESDAY June 4th 09:00 – 10:15

Phonetic encoding in utterance production

Marina Laganaro
University of Geneva

In the psycholinguistic literature on language production, phonetic encoding refers to the transformation of an abstract linguistic code into motor programs which guide speech articulation. The generation of 'phonetic plans' is described in Chapters 9 and 10 (Levelt, 1989) where it is embedded in the architecture of 'phonological encoding', whereas in later publications *phonological* and *phonetic* encoding are described as distinct processes (Levelt, Roelofs & Meyer, 1999). I will review the questions and results from the last 20 years focusing on *i.* the debates on the distinction between phonological and phonetic processes and *ii.* the nature and size of phonetic plans by integrating evidence from psycholinguistic, phonetic, neuropsychological and neuroimaging studies.

WEDNESDAY June 4th 10:25 – 11:40

Articulating: The neural mechanisms of speech production

Frank Guenther
Boston University

Speech production is a highly complex sensorimotor task involving tightly coordinated processing in the frontal, temporal, and parietal lobes of the cerebral cortex. Historically, the study of the neural underpinnings of speech has suffered from the lack of an animal model whose brain activity could be measured using invasive electrophysiological techniques. The development of non-invasive structural and functional neuroimaging techniques in the latter part of the 20th century has led to a dramatic improvement in our understanding of the speech network. Techniques for measuring regional cerebral blood flow, including positron emission tomography and functional magnetic resonance imaging, have illuminated the neural regions involved in various aspects of speech, including feedforward control mechanisms as well as auditory and somatosensory feedback control circuits. To better understand these processes, we have designed, experimentally tested, and iteratively refined a neural network model detailing the neural computations performed by the brain regions involved in speech. Because the model's components correspond to neural populations and are given precise anatomical locations, activity in the model's neurons can be compared directly to neuroimaging data. Computer simulations of the model account for a wide range of experimental findings, including data on acquisition of speaking skills, articulatory kinematics, and brain activity during normal and perturbed speech. Furthermore, "damaged" versions of the model are being used to investigate communication disorders, and the model has been used to guide development of a brain-computer interface aimed at restoring speech output to individuals suffering from locked-in syndrome.

WEDNESDAY June 4th 11:50 – 13:05

Monitoring and control in language production

N. Bonnie Nozari
Johns Hopkins University

Three major theories of monitoring currently exist for language production: the perceptual loop, the forward models, and the conflict-based monitor. I will briefly review findings from behavioral, neuropsychological, neuroimaging, and electrophysiological studies of error detection, and discuss how the conflict-based monitor accommodates these findings, while pointing out the important roles that the other two types of monitor play in the monitoring process. The conflict-based monitor is a framework combining principles of signal detection theory with the operations of a conflict detector, a mechanism that measures the amount of conflict between the activation levels of representations competing for selection. According to this framework, detecting an error is a probabilistic event, and the probability of that event is determined by the amount of competition in the production system. I will further show the utility of this framework for exploring broader concepts such as metacognition and its link to learning. Time permitting, I will preview a model of error repair, and discuss the relationship between conflict monitoring, probability of errors, recruitment of cognitive control, and probability of repairs.