

port Levelt et al.'s assumption that words have metrical frames that specify a word's number of syllables and main stress position, but which do not include segmental content. Before elaborating further on this point and going on to the second, it is necessary to consider a study conducted by Meyer (1994), which is apparently inconsistent with this conclusion. Meyer's experiments considered whether the conclusions I drew in my 1993 paper would hold up in Dutch. Meyer varied certain characteristics of the critical words, such as whether vowels were short or long, and she found only partial support for my conclusions: pause compensation occurred but was only partial, so that the words with longer syllables had longer total durations. Two possible counterarguments can be made: one is that there simply are cross-linguistic differences in these phenomena (this is the conclusion Meyer draws). The other is that when duration intervals are created they ignore segmental content but not CV (consonant-vowel) structure. In other words, the difference between a word with a short vowel such as *tak* (branch) and a long vowel such as *taak* (task) might be represented as CVC versus CVVC, and that information might be available independently of segmental content. A critical experiment, then, would be to vary Dutch syllables differing only in whether a vowel is tense or lax; the model I laid out predicts that pause compensation will be total, while Meyer's arguments would lead one to expect any compensation to be only partial.

Second, regardless of when segmental content is accessed during the production of words in sentences, it is clear that word production depends critically on prosodic context. Words are produced differently when they occur at the edge of a major prosodic constituent compared with any other position, and their characteristics differ when they are semantically prominent rather than more discourse-neutral. The latter point suggests that the conceptual information that makes its way through the word production system includes not just stored knowledge but dynamically specified information as well.

Naming versus referring in the selection of words

Peter C. Gordon

Department of Psychology, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3270. pcg@email.unc.edu
www.unc.edu/depts/cogpsych/gordon

Abstract: The theory of lexical selection presented by Levelt, Roelofs & Meyer addresses the mechanisms of semantic activation that lead to the selection of isolated words. The theory does not appear to extend naturally to the referential use of words (particularly pronouns) in coherent discourse. A more complete theory of lexical selection has to consider the semantics of discourse as well as lexical semantics.

A successful theory of lexical selection in speech production must explain a huge number of phenomena, ranging from the semantic makeup of words to how the articulatory form of a phonetic segment accommodates neighboring segments. The theory presented by Levelt, Roelofs & Meyer tackles an impressive number of these phenomena with substantial conceptual analysis and empirical evidence, but it does not tackle all of them. In this commentary I will analyze issues associated with one particular lexical phenomenon, the use of pronouns, that Levelt et al. do not address and that provide some challenges to the framework that they present.

Pronouns are very common; "he" is the tenth most frequent word in English and other pronouns ("it," "his," and "I") are not far behind (Kuçera & Francis 1967). It seems likely that this class of words is frequent because it is very useful. Researchers from a number of disciplines (e.g., Brennan 1995; Fletcher 1984; Grosz et al. 1995; Marslen-Wilson et al. 1982) have focused on how pronouns (and other reduced expressions) contribute to the coher-

ence of discourse by implicitly marking the semantic entities that are central to a discourse. Do the mechanisms described by Levelt et al. provide a way of accounting for this use of pronouns?

Levelt et al.'s theory has been developed primarily to provide a detailed account of speakers' performance in the picture-naming task. A speaker who repeatedly exclaimed "it" in response to the stimulus pictures would justifiably be regarded as uncooperative, so it is not surprising that the task provides little evidence on pronoun selection. Yet during an ordinary conversation a speaker would be expected to use "it" frequently to refer to the kinds of objects represented in the stimulus pictures. Consider the following sentence, which Levelt et al. offered as a possible way that a speaker might describe a picture.

I see a chair and a ball to the right of it.

Here the expressions "a chair" and "it" are coreferential; they refer to the same thing. Presumably Levelt et al. would say that the word "chair" is selected because the conceptual and semantic preconditions stimulated by the picture lead to its being the most highly activated lemma. But then why is the same entity subsequently referred to with "it"? The semantic system used by Levelt et al. was selected to avoid the hyperonym problem, that is, in a compositional representation the activation of the semantic features of a concept will also activate any superordinates of that concept. From this perspective on semantics, pronouns might be seen as the ultimate hyperonyms, those specifying only minimal semantic features such as gender, number, and animacy. Levelt et al.'s use of a semantic system designed to avoid inadvertent intrusion of hyperonyms does not immediately suggest a reason for the widespread, systematic use of a particular class of hyperonyms.

The example sentence points to the powerful ways in which sentential and discourse context can influence lexical selection, a topic I have worked on with my colleague Randall Hendrick. Building on the work of others, we (Gordon & Hendrick 1997; in press) have argued that a discourse model consists of a set of semantic entities and a series of predications, of which the entities are arguments. With respect to lexical selection, we have argued that the primary purpose of names (and other unreduced referring expressions) is to introduce semantic entities into a model of discourse, whereas the primary purpose of pronouns (and other reduced expressions) is to refer directly to entities that are prominent in the discourse model. On this view, results from a task like picture naming may provide evidence about the mechanisms that a speaker uses to initially introduce an entity into the discourse. Competitive selection based on activation levels is an attractive mechanism for accomplishing this, one that builds on the substantial explanatory use of such mechanisms within cognitive psychology to account for a variety of semantic and lexical processes. However, this appealingly straightforward mechanism may have to be substantially augmented if it is to account for the selection of words that refer to entities in an evolving discourse model as well as words generated from long-term memory. For the example sentence, we would argue that the selection of the pronoun reflects the prominence of the semantic entity CHAIR in the discourse model. To the extent that we are right, a theory of lexical selection must not only address the semantic organization of lexical concepts in long-term memory (as Levelt et al.'s theory does); it must also address the semantics of models created to represent the interrelations of entities and concepts in a discourse. The principles governing lexical selection from these two domains may be quite different.

categories (e.g., animates vs. artifacts) are clustered in distinct (but contiguous) cortical regions.

In this view, lemmas would also have a syntactic role. It is clear that syntactic properties cannot be directly attached to concepts, because semantic features do not directly map onto syntactic features. The syntactic properties could be attached to the phonological or orthographic word forms; however, this is computationally inefficient because syntactic information is modality-independent (but see Caramazza 1997). Therefore, the intermediate lemma level is the most adequate for accessing syntactic information.

Lexical concepts acquire syntactic properties relatively late in development (between the ages of 2.6 and 4 years; see Levelt et al., sect. 1). This process is termed *syntactization* by Levelt et al. and refers to the development of a system of lemmas. However, the explosive growth of the lexicon takes place between the ages of 1.6 and 2.6. This means that an efficient mapping between concepts and phonological word forms is already established at that onset of the syntactization process. Within the architecture of Levelt et al.'s model, such mapping would presumably involve conceptual nodes and word forms, thus bypassing the yet-to-be-developed lemmas. Therefore, the later development of the lemma level would mean a massive rewiring of the lexical system. We believe that such a process is truly unlikely (both from the neural and computational standpoints). By contrast, if lemmas develop as a necessary component of the mapping between meaning and phonology, syntactization is simply the process of linking syntactic features to the existing lemma representation.

ACKNOWLEDGMENT

Preparation of this paper was supported by NSF grant SBR 9729118 to Gabriella Vigliocco and by NATO grant CRG 971002 to Gabriella Vigliocco and Marco Zorzi.

NOTE

1. This very restricted notion of lemma is what led Caramazza (1997) to argue that lemma nodes are contentless representations (the "empty lemma"), and as such they are dispensed with in his model of lexical access.

Authors' Response

Multiple perspectives on word production

Willem J. M. Levelt,^a Ardi Roelofs,^b and Antje S. Meyer^a

^aMax Planck Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands; ^bSchool of Psychology, University of Exeter, Washington Singer Laboratories, Exeter EX4 4QG, United Kingdom. pim@mpi.nl
www.mpi.nl a.roelofs@exeter.ac.uk asmeyer@mpi.nl

Abstract: The commentaries provide a multitude of perspectives on the theory of lexical access presented in our target article. We respond, on the one hand, to criticisms that concern the embeddings of our model in the larger theoretical frameworks of human performance and of a speaker's multiword sentence and discourse generation. These embeddings, we argue, are either already there or naturally forgeable. On the other hand, we reply to a host of theory-internal issues concerning the abstract properties of our feedforward spreading activation model, which functions without the usual cascading, feedback, and inhibitory connections. These issues also concern the concrete stratification in terms of lexical concepts, syntactic lemmas, and morphophonology. Our response stresses the parsimony of our modeling in the light of its substantial empirical coverage. We elaborate its usefulness for neuroimaging and aphasiology and suggest further cross-linguistic extensions of the model.

R1. The larger context

R1.1. Lexical access in utterance generation

The stated aim of our target article was to present a theory of lexical access covering "the production of isolated prosodic words." Several commentators didn't like this straitjacket and preferred to consider lexical access in the larger pragmatic context of utterance generation. We share this preference, and a historical note may clarify our position. The larger framework was developed by Levelt (1989; henceforth *Speaking*), which outlines a theory of speaking, the skill that gets us from communicative intentions within ever-changing pragmatic settings to articulated utterances. A major observation in reviewing the relevant literatures was that a core aspect of this skill, lexical access, was theoretically deeply underdeveloped. Our team at the Max Planck Institute set out to fill this gap, and the target article reports on a decade of research dedicated to unraveling the process of normal lexical access. This required, among other things, the invention (by us and by others) of appropriate reaction time paradigms. Of course, the pragmatic context is limited in the laboratory (although none of our subjects ever said "this is not the real world"), but we always had the larger, explicit theoretical framework at hand. Several comments on our target article can be handled by referring to that framework.

Gordon addressed the issue of pronoun generation, indeed something we hardly touched on in the target article. He suggested that the primary purpose of unreduced referring expressions is "to introduce semantic entities into a model of discourse, whereas the primary purpose of pronouns (and other reduced expressions) is to refer directly to entities that are prominent in the discourse model" – almost a citation from *Speaking* (sect. 4.5.2). Meanwhile **Schmitt** (1997) in her Max Planck dissertation project, elaborated this notion in terms of a processing model and performed the relevant reaction time experiments.

Hirst's well-taken discussion of language-dependent conceptualization, as it may occur in bilinguals, is foreshadowed in section 3.6 of *Speaking*, referred to in section 3.1.2 of the target article. Unlike Hirst, though, we are not worried by the thought that many, or even most, lexical concepts are language-specific. The empirical evidence for such a notion is rapidly increasing (see, e.g., Slobin 1987; 1996; 1998). We agree that this is a special challenge for modeling the bilingual lexicon.

Ferreira correctly notes (and has demonstrated experimentally) that a word's prosody will vary from context to context. For instance, a word tends to be relatively long in phrase-final position or when it is contrastively stressed. Chapter 10 of *Speaking* outlines an architecture for the modulation of phonetics by such higher-level types of information. In that architecture there is indeed *parallel* planning of prosody and lexical access, as **Roberts et al.** would have it. These commentators could have read in section 10.3.1 ("Declination") of *Speaking* how declination and resetting are conceived within that framework. Indeed, these phenomena are not handled through conceptual input, and neither are the setting of amplitude and duration. It is curious to be confronted with a caricature of one's theoretical framework and subsequently to be accused of not justifying one's assumptions.

Another issue directly related to the larger sentential

in this way, errors of types (1) and (2) may also find their explanation in these terms. Several examples in the commentary and the relevant papers suggest a reorganization of the lexicon towards “fixed expressions.” These are stored “lexical” items with complex lemma structures. Very little is known about their production.

R6. Suggestions for future research

Several commentators have alerted us to issues that we have not treated yet. One of these is the embedding of our theory in a comprehensive theory of development (Roberts et al.; Zorzi & Vigliocco). Here we respond to those homework assignments that may be manageable within another decade of research. We do, for instance, whole-heartedly agree with Cutler & Norris that comprehension and production researchers should join forces to create a parsimonious model to account for both capacities. In section 3.2.4 of the target article we proposed (Assumption 3) that, from the lemma level up, the two networks are identical; however, we have argued for one-way connections at lower processing levels. Cutler and Norris similarly argue for one-way bottom-up connections in the perceptual network. Hence the two networks are arguably different at that level of processing. What is in urgent need of development is a theory of their interconnections. Carr (personal communication) proposes to focus on the word-frequency effect. It arises in both production (see our sect. 6.1.3) and perception, and in both cases it involves access to the phonological code. Can there be a unified theoretical account? Hirst wonders whether syntactic operations on the lemmas can feed back to the lexical concepts. Assumption 3 invites the answer “yes.” Chapter 7 of *Speaking* presents examples of syntactic operations checking conceptual argument structure (see also sect. R2.1.2). But Hirst’s point for future research is whether speakers will adapt their lexical choice if the developing syntactic structure turns out to have no solution for the current set of lemmas. We would expect this to happen, but an experimental demonstration will not be easy. In addition, Hirst invites a psychological account of how pleonasm is (usually) prevented in our lexical choices. Theoretical solutions to both issues may naturally emerge from Kempen’s (1997) recent work. It already provides a principled theoretical solution to the conceptual primacy effects in the generation of syntax, as referred to in Dell et al.’s commentary. **Gordon’s homework assignment is to further relate lexical choice to the constraints arising in discourse. This obviously involves pronominalization (extensively discussed in *Speaking*) and other reduced or alternating forms of reference. The experimental work is on, both in our own laboratory and other laboratories (see Schmitt 1997 and Jescheniak & Schriefers’s commentary).** Another important area of investigation is the use of what Clark (1998) calls “communal lexicons.”

Dell et al. close their commentary by recommending the study of attentional mechanisms that control the timing of the activation of conceptual and linguistic units. We have already taken up this challenge. Assuming that there is a close relationship between gaze and visual attention, we have started to register speakers’ eye movements during the description of pictures in utterances such as *the ball is next to the chair* (Meyer et al. 1998). In a first series of experiments, we found that speakers have a strong tendency to

fixate on each object they name and that the order of looking at the objects corresponded almost perfectly to the order of naming. Most importantly, we found that speakers fixated on each object until they had retrieved the phonological form of its name. This suggests that at least these simple descriptions are generated in a far more sequential way than one might have expected. Whether more complex utterances are generated in the same sequential manner must be further explored.

Some homework assignments failed to come forth. We were somewhat surprised to find that two core issues of word-form encoding, the generation of morphology and the generation of metrical structure, invited very little reaction. The experimental findings are nontrivial, and in fact the first of their kind in language-production research; they cry out for cross-linguistic comparisons. It is by no means obvious that the generation of morphology involves the same mechanisms in languages with limited morphological productivity (such as Dutch or English) and languages whose generativity largely resides in morphology (such as Turkish). The storage/computation issue that we addressed for the generation of a word’s metrical phonology will most certainly be resolved differently for stress-assigning languages (such as Dutch or English) than for languages with other rhythmic structures. The production mechanisms will probably vary as much as the corresponding comprehension mechanisms (see Cutler et al. 1997 for a review).

References

Letters “a” and “r” appearing before authors’ initials refer to target article and response, respectively.

- Abbs, J. H. & Gracco, V. L. (1984) Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech. *Journal of Neurophysiology* 51:705–23. [aWJML]
- Abdullaev, Y. G. & Posner, M. I. (1997) Time course of activating brain areas in generating verbal associations. *Psychological Science* 8:56–59. [rWJML]
- Abeles, M. (1991) *Corticomics - Neural circuits of the cerebral cortex*. Cambridge University Press. [FP]
- Arnell, K. & Jolicoeur, P. (1997) The attentional blink across stimulus modalities: Evidence for central processing limitations. *Journal of Experimental Psychology: Human Perception and Performance*. 23:999–1013. [THC]
- Baars, B. J., Motley, M. T. & MacKay, D. G. (1975) Output editing for lexical status from artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior* 14:382–91. [aWJML, JS]
- Badecker, W., Miozzo, M. & Zanuttini, R. (1995) The two-stage model of lexical retrieval: Evidence from a case of anomia with selective preservation of grammatical gender. *Cognition* 57:193–216. [TAH, aWJML, CS]
- Baddeley, A., Lewis, V. & Vallar, G. (1984) Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology* 36A:233–52. [aWJML]
- Baumann, M. (1995) The production of syllables in connected speech. Unpublished doctoral dissertation, Nijmegen University. [aWJML]
- Beaugregard, M., Chertkow, H., Bub, D., Murtha, S., Dixon, R. & Evans, A. (1997) The neural substrate for concrete, abstract, and emotional word lexica: A positron emission tomography study. *Journal of Cognitive Neuroscience* 9:441–61. [HMM]
- Béland, R., Caplan, D. & Nespoulous, J.-L. (1990) The role of abstract phonological representations in word production: Evidence from phonemic paraphasias. *Journal of Neurolinguistics* 5:125–64. [aWJML]
- Berg, T. (1988) *Die Abbildung des Sprachproduktionsprozesses in einem Aktivationsflußmodell: Untersuchungen an deutschen und englischen Versprechern*. Niemeyer. [aWJML]
- (1989) Intersegmental cohesiveness. *Folia Linguistica* 23:245–80. [aWJML]
- (1991) Redundant-feature coding in the mental lexicon. *Linguistics* 29:903–25. [aWJML]
- Berg, T. & Schade, U. (1992) The role of inhibition in a spreading-activation model of language production: I. The psycholinguistic perspective. *Journal of Psycholinguistic Research* 21:405–34. [LF]