Journal of Phonetics 53 (2015) 141–152

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Unifying speech and language in a developmentally sensitive model of production

Melissa A. Redford *

Department of Linguistics, University of Oregon, Eugene, OR 97405-1290, USA

ARTICLE INFO

Article history:
Received 27 September 2014
Received in revised form
17 June 2015
Accepted 24 June 2015
Available online 21 July 2015

Keywords:
Speaking
Schema theory
Speech acquisition
Speech planning
Relative timing
Temporal patterns

ABSTRACT

Speaking is an intentional activity. It is also a complex motor skill; one that exhibits protracted development and the fully automatic character of an overlearned behavior. Together these observations suggest an analogy with skilled behavior in the non-language domain. This analogy is used here to argue for a model of production that is grounded in the activity of speaking and structured during language acquisition. The focus is on the plan that controls the execution of fluent speech; specifically, on the units that are activated during the production of an intonational phrase. These units are schemas: temporally structured sequences of remembered actions and their sensory outcomes. Schemas are activated and inhibited via associated goals, which are linked to specific meanings. Schemas may fuse together over developmental time with repeated use to form larger units, thereby affecting the relative timing of sequential action in participating schemas. In this way, the hierarchical structure of the speech plan and ensuing rhythm patterns of speech are a product of development. Individual schemas may also become differentiated during development, but only if the subsequences are associated with meaning. The necessary association of action and meaning gives rise to assumptions about the primacy of certain linguistic forms in the production process. Overall, schema representations connect usage-based theories of language to the action of speaking.

Published by Elsevier Ltd.

1. Introduction

This special issue on the cognitive nature of speech sound systems provides an opportunity to focus on the indivisibility of cognition and action in language production that is so salient in development. For example, the child's first attempt at referential communication demonstrates a *naming insight* and thus the influence of conceptual development on language acquisition. At the same time, the child's early attempts at words clearly demonstrate the influence of speech motor skills on the linguistic forms produced. Although the conceptual and speech motor domains are often studied separately, here I embrace their indivisibility in speaking to argue for a schema-based theory of language production. The argument takes shape as the outline of a model. The goal is to provide a framework for understanding the details of how we deliver language through speech at any age in a manner that (a) is consistent with usage-based approaches to grammar and language acquisition (e.g., Bybee, 2001; Goldberg, 2006; Tomasello, 2009; Vihman, 2014) and (b) offers continuity in representations across levels of analysis.

The informal model outlined here is motivated by several assumptions: (1) complex behaviors are best understood in terms of their development; (2) fluent speech production requires a plan; and (3) the structure of the plan emerges with production during language acquisition. The last assumption presupposes that production and perception are separate processes. Production is tuned to the individual's anatomy and goals, perception to accommodating variable input in relation to meaning. Phonology emerges at the intersection of production and perception from the abstraction and categorization of sound patterns across less abstract motor and perceptual forms that are themselves connected to quasi-bounded conceptual (semantic, pragmatic) information in the lexicon. While the phonology likely facilitates both word form acquisition and novel word creation across the lifespan, adults typically speak using words and phrases that we have uttered hundreds if not thousands of times before. This observation suggests that speech production need not rely on our most abstract knowledge of sound systems, but instead can be modeled as the activation of remembered forms. This paper focuses on the nature and organization of these forms. The central thesis is that the plan that guides fluent speech production (i.e., production at the level of the intonational phrase) emerges with language acquisition and with the

* Tel.: +1 541 346 3818.
E-mail address: redford@uoregon.edu

0095-4470/$ - see front matter Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.wocn.2015.06.006

extensive speech practice that acquisition entails. The units of execution in the plan are temporally structured chunks of remembered action and their sensory outcomes, called *schemas*. These are acquired and deployed in service of communicative goals. Individual schemas may combine with others to form larger units over developmental time, depending both on the frequency with which they have been deployed together and on the emergence of new or extended meanings. Schemas may also be differentiated to create smaller units of execution, provided these are associated with meaning and so can serve some specific communicative goal.

Although our focus is on the structure of the speech plan and how it evolves to reflect the complexity of adult language over developmental time, the point is to provide a framework that unifies higher- and lower-level processes in production. Let me therefore provide a higher- and lower-level context at the outset so that it is clear what types of representations are addressed in this paper.

With regards to the immediately higher-level context, speech planning is viewed as the sequential activation of conceptually-linked schemas within a temporal window that is defined by a domain-general constraint on attention or working memory. The order in which schemas are activated is dependent on the abstract construction activated as a function of the propositional content the speaker intends to deliver (i.e., on language planning). Constructions are understood not only in the traditional manner as form-meaning pairings (Goldberg, 2006), but also as habitual or routine trajectories through a lexical-semantic space. As for the lower-level context, speech plan execution involves setting one or more control parameters that will affect overall time and force, determining both the global rate at which action will unfold and movement amplitude. The control parameter(s) can be reset with the activation of each schema in a sequence. This allows production to be modulated in response to the communicative context. Once parameters are set, schemas are executed. Coordinated articulatory movement is triggered at the relative timing intervals specified by the schema. Movement accuracy is controlled by efferent copy, generated from the sensory information stored within the schema. Between the activation of a particular construction and the details of motor control are the structured representations that guide speaking, intonational phrase by intonational phrase. These representations are the focus of the current paper.

2. Speaking as the implementation of language

A process-oriented model requires the identification of a starting point. The choice is important because it determines the type of description that will follow. Models of fluent speech production typically start with idealized adult language (e.g., Cooper & Paccia-Cooper, 1980; Dell, 1986; Garrett, 1980; Levelt, 1989; Shattuck-Hufnagel, 2015; Turk & Shattuck-Hufnagel, 2014; Wheeldon, 2000). Because adult language is so complex, theorists who start here have adopted linguistic representations derived by others from a 'pure' linguistic analysis; specifically, from the transcription-based analysis of linguistic structure isolated from context. Although this type of analysis makes the problem of (adult) language description tractable for the linguist, adopting abstract linguistic representations to model fluent speech production means positing units of action that make no reference to speech. To generate speech, the theorist is left to assume that these abstract representations are translated into ones that the motor system is able to reference. This translation process has come to be known as phonological and phonetic encoding in the psycholinguistics literature, and it can result in incongruent marriages between theories.

Consider, for example, the marriage between Metrical Theory and Articulatory Phonology in Levelt's (1989) influential model, which proceeds in stages from lexical access and syllabification to morphological and metrical spellout. After metrical spellout, resyllabification is required. The resulting syllable structures are used to ensure appropriate serial ordering during segmental spellout. Once phonological plans are specified in this way, they are "*enriched* by prosodic information" (p. 409, emphasis in the original) and then matched, syllable by syllable, to motor programs that take the shape of Articulatory Phonology representations (Levelt cites Browman and Goldstein, 1986). It is the reference to Articulatory Phonology in the context of Levelt's information processing model that is jarring since the aim of that particular research program is to unify speech and language representation:

> Gestures are characterizations of discrete, physically real events that unfold during the speech production process. Articulatory phonology attempts to describe lexical units in terms of these events and their interrelations, which means that gestures are basic units of contrast among lexical items as well as units of articulatory action. From our perspective, phonology is a set of relations among physically real events, a characterization of the systems and patterns that these events, the gestures, enter into (Browman & Goldstein, 1992:23).

Like Browman and Goldstein (1986, 1992), the model I propose grounds language in speech to avoid the problem of translation that occurs in current psycholinguistic models of speech-language production. Translation is a problem because it requires postulating additional mechanisms for mapping one type of representation onto another in the process of rendering meaning into action (i.e., phonological and phonetic encoding). This kind of labor is at odds with the sheer speed and automaticity of speaking, where automaticity in behavior is defined as a process that requires no conscious attention or working memory resources dedicated to the selection and manipulation of information[1,2].

[1] The definition of automatic versus controlled processes that I adhere to here is taken from Shiffrin and Schneider (1977) who define these processes as follows:

An automatic process can be defined… as a sequence of nodes that nearly always becomes active in response to a particular input configuration, where the inputs may be externally or internally generated and include the general situational context, and where the sequence is activated without *the necessity* of active control or attention by the subject (155-6; emphasis added).

A controlled process utilizes a temporary sequence of nodes under the control of, and through attention by, the subject; the sequence is temporary in the sense that each activation of the sequence of nodes requires anew the attention of the subject (156).

[2] Gathercole and Baddeley's (1993:Ch. 4) conclusions support the contention that phonological and phonetic encoding are at odds with the automaticity of speech production. In particular, they find that the evidence from working memory studies on speech production fails to support specific predictions that follow from the hypothesis of encoding. The predictions they consider follow from Garrett's (1980) model.