Language Sciences 53 (2016) 58–74

Contents lists available at ScienceDirect

Language Sciences

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

How spoken languages work in the absence of an inventory of discrete units

Michael Ramscar, Robert F. Port*

[a] Eberhard Karls Universität, Tübingen, Germany
[b] Indiana University, Bloomington, IN, United States

ARTICLE INFO

Article history:
Available online 26 September 2015

Keywords:
Language
Speech perception
Language understanding
Linguistics
Learning theory

ABSTRACT

Historically, linguists and psychologists have generally assumed that language is a combinatoric process, thereby taking the idea that language users have access to inventories of discrete, combinable units (phonemes, morphemes, words, etc.) for granted, despite the fact that these units have tended to resist formal definitions. We propose a new approach to language understanding based on the psychological mechanisms that underpin context-sensitive processing. This new method is surprisingly simple, in large part because it embraces a view of learning that has been developed from studies of animal behavior and neuroscience. From this perspective, learning is seen as a systematic, discriminative process that seeks to reduce a learner's uncertainty in making moment-to-moment predictions. We suggest that language processing employs all the information available to the listener at any given moment to predict what will happen in the next moment, in the next couple of sentences, etc. This approach does not rely on any of the ambiguous traditional linguistic units because continuous-time processing simply acts to reduce a hearer's uncertainty about an actual message in relation to possible messages, rather than building up an interpretation out of elemental components. From this perspective, the conventional units of language – phonemes, morphemes, words – can be seen as idealizations of patterns that evolved for communicative efficiency that can serve the purposes of orthographic (and linguistic) description, rather than psychologically 'real' elements that are essential to language processing.

1. Introduction

Many linguists and laymen believe that natural languages are made up of discrete units of various kinds: (1) *words,* which can be easier or harder to identify (e.g., *dog, table*, *dunno*); (2) *phonemes,* roughly the acoustic/psychological version of letters; and (3) *sentences*. Another somewhat technical unit is (4) the *morpheme*, a fraction of speech that has a consistent meaning and cannot be divided into smaller such pieces (e.g., *tree-house, rain-bow-s, walk-ing* = 7 morphemes total). Linguists generally consider these four units to be widespread across human languages, because in many languages utterances appear to be composed out of sequences of words, words appear to be composed out of one or more morphemes and morphemes in turn appear to be composed out of phonemes. These units are often considered to be universal and to be essential

* Corresponding author.
E-mail address: port@indiana.edu (R.F. Port).

http://dx.doi.org/10.1016/j.langsci.2015.08.002
0388-0001/

components of language. The semantic interpretations of utterances are thought to be constructed along similar lines, by combining the meanings of the morphemes that comprise words that in turn may comprise a sentence. This view of language is usually called *compositionality*, and virtually all of modern linguistic theory is predicated on this principle.

The twin ideas of units and combinatorics have powerful intuitive appeal. However, most writing systems overlook many of the important contrasts that are realized in speech (such as intonation, stress or the different acoustics of the final sibilant in *pots* and *dogs)*, and of course, people who ponder the formal nature of language invariably have a high level of literacy, which raises the possibility that lifelong reading practice biases the theoretical intuitions of researchers. However, when critical attention is focused on the units of spoken language, they cannot be specified with sufficient clarity to play the central role that these linguistic theories attribute to them.

In this paper, we focus on the problem of speech perception and comprehension by adults leaving the important issue of the production of speech to another day. We will describe some of the many empirical and theoretical reasons that cast doubt on the idea that human speech comprehension relies on discrete phonemes, words, sentences, or any other units. We then sketch an alternative account of the workings of language based on learning theory. Critically, our account does not depend on language users being in possession of inventories of discrete components. Rather, from a learning perspective, language and linguistic knowledge can be seen to be fundamentally continuous in time and systematic in nature. This account treats languages as cultural systems that have been shaped by the constraints imposed by learning and communication. Indeed, the dynamic nature of language systems at both the communal and individual levels help explain why it is that theories of language based on units have inevitably proven so unsatisfactory when it comes to accounting for and explaining the facts of linguistic communication.

Why units don't work. While it cannot be denied that each language comprises a system of contrasts that yield patterns that do resemble all four of the units we listed above, research has shown that none of these units can be given clear definitions, and none of them can be consistently applied to real speech. None can play the role of discrete symbol tokens analogous to letters as required by formal theories of language (Chomsky and Miller, 1963; Chomsky and Halle, 1968). On the other hand, it is also clear that many cultures have developed discrete *orthographic systems*. However, although letters, words and sentences (or their equivalents) can be identified in most writing systems, attempts to specify corresponding units in speech that are reliable have foundered.

In the view we will sketch in this paper, language perception proceeds by means of a continuous-time process involving a learned system: a listener's current knowledge of a language and its relationship to the world, which comprises a series of relationships between semantic distinctions and acoustic phonetic information. Learning is a discriminative, predictive process (Ramscar et al., 2010a,b, for a review) that reduces a learner's uncertainty about the world. Comprehension involves a continuous series of predictions about upcoming semantic and acoustic patterns that reduce a listener's uncertainty about an intended meaning by a process of elimination. A listener's understanding of the intended semantic message of the speaker is gradually revealed as alternative messages are successively rejected. The important thing is the continuous interaction between semantic and acoustic phonetic aspects of the language system.

To give a concrete example, suppose someone says:

A. *Wanna nother drink?* (Do you want another drink?)
B. *What's 'at?* (What is that you said)?)
A. (Picks up a soda can.)
B. *Yeah, I'd love one*.

The context is presumably an occasion where A, the hostess, has offered a drink to B, a guest. B, saying *What's that?* (with stress on *what,* not on *that)* reveals that he is uncertain what she said, so instead of repeating herself, she raises a can of drink into B's visual field, thereby reducing his uncertainty about both what it was that she said and her intended meaning.

Let's look at B's response in more detail. B sees the soda can and this reduces his uncertainty about what A had just said sufficiently to allow him to say that he would actually like another drink. The phonetic pattern [jæ:] by B indicates acceptance of the offer and [aɪdlʌv …] (*I would love …*) accepts politely by implying that the host is still free to retract the offer. A third person overhearing this utterance can use the [jæ:] to partially predict B's *I'd* (rejecting *No thank you*) and using *I'd* to partially predict *love* (and reject *rather not…* etc.*)* and *love* to predict *another (*and reject *a couple of 'em*, as well as *Fresca* and *I'd love to but …* etc.). The point is that both acoustic phonetic information (acoustic and visual) and semantic (and social) context work together from moment to moment to reduce a listener's uncertainty about a speaker's intended message.

Although real-time predictions do not generally lead to certainty before the word begins, simply narrowing the range of possibilities is very valuable to a listener. Since the phonetic and semantic sides of the language system work continuously together, people in some situations may be quite unaware of complete interruptions of the speech signal. If someone scrapes a chair noisily across the floor to mask the acoustics of the entire word *love*, listeners will normally believe they heard the word *love* because their language system fully predicts it based on the previous (and following) acoustic and semantic context (Warren and Obusek, 1971). From the perspective we describe here, "speech perception" can be seen as a similar kind of continuous process in which the phonetic and semantic aspects of speech are inseparable. Speech understanding is the product of continuous-time learning that constructs a system of phonetic and semantic contrasts that enable a speaker's intended message (which is, itself, situated within an acoustic phonetic and semantic system) to be discriminated from other possible messages a speaker might have uttered. We put quotes around *speech perception* above because the view of speech