Available online at www.sciencedirect.com
ScienceDirect
System 36 (2008) 35–51
SYSTEM
www.elsevier.com/locate/system

Revising segmentation hypotheses in first and second language listening

John Field

Department of Applied Linguistics, University of Reading, Whiteknights, PO Box 218, Reading, RG6 6AA, UK

Received 15 June 2007; received in revised form 2 September 2007; accepted 31 October 2007

Abstract

Any on-line processing that takes place while an utterance is unfolding is extremely tentative, with early-formed hypotheses having to be revised as the utterance proceeds. The hypotheses in question relate not only to the words that are present but also to where their boundaries fall. This study examines how first and second language listeners adjust their segmentation assumptions as new perceptual evidence comes in. It employs a variant of the gating task in which subjects transcribe a short utterance presented in sections of gradually increasing length. The first two presentations were phonetically ambiguous and could be segmented in any one of three ways. The third and fourth presentations provided disambiguating input; subjects' responses were examined at these points to see how quickly they switched from a wrong interpretation to one that fitted the perceptual evidence. The results indicated a significant difference in the way in which first and second language listeners deal with incorrect segmentation hypotheses. Whereas native listeners are quick to change their interpretations on the basis of incoming evidence, non-native listeners are considerably more reluctant to do so.

Keywords: Perseveration; Gating; Second language listening; Spoken word recognition; Auditory processing; On-line processing; Lexical segmentation

E-mail address: j.c.field@reading.ac.uk

0346-251X/$ - see front matter © 2008 Published by Elsevier Ltd.
doi:10.1016/j.system.2007.10.003

1. Speech decoding as a tentative process

1.1. Decoding and the L1 listener

One way of viewing second language listening is as a form of expertise which the native listener possesses and the learner aims to acquire. The value of this perspective is that it provides us with a benchmark in the form of a model of expert listening against which the performance of learners can be measured. It also enables us to draw upon empirical evidence obtained by cognitive psychologists, phoneticians and others in order to identify distinct processes which can form the basis for listening practice (Field, in press).

But when we examine first language research, it becomes clear that it offers us two apparently contradictory accounts of how words, phrases and clauses are recognised. One view, widely held by psycholinguists, is that listeners do not wait until the end of a major syntactic constituent before attempting to assemble the words that a speaker has uttered. Instead, they analyse the speaker's utterances while they are still unfolding. Early studies by Marslen-Wilson (1973, 1975) suggest that listeners are capable of accurately repeating (or 'shadowing') what a speaker says at a delay of only about 200–250 milliseconds behind the speaker's voice. What is more, the shadowing process appears to entail matching at word level and not just the parroting of sounds, since incorrectly pronounced words are corrected in the process of reporting them. The delay is about the length of a syllable in English, suggesting that the syllable may be an important unit of analysis for the listener.

It is reasonable to express a few reservations about the original findings. The speakers were talking relatively slowly and the material was read-aloud rather than natural connected speech. In addition, just because listeners are *capable* of shadowing speakers in this way, it does not mean that they actually do so in practice when processing a piece of extended discourse. The demands upon attention may simply be too heavy. Nevertheless, the assumption that listening takes place 'on line' has become widely accepted, and is a tenet of many psycholinguistic models. It has also given rise to Marslen-Wilson's own Cohort Theory (1987) in which the opening sounds of an utterance activate a range of possible word matches, which is gradually narrowed down as the utterance continues and more perceptual and contextual evidence becomes available.

A rather different picture is painted by the evidence of researchers who have examined the intelligibility of small groups of sounds within stretches of connected speech. In another early landmark study, Pickett and Pollack (1963) found that only 55% of words excised from a piece of connected speech could be recognised without accompanying phonetic context. We can find an obvious explanation for this phenomenon in the accounts provided by phoneticians, (e.g. Cruttenden, 1986), who draw attention to the reductive effect of intonation. Intonation groups give prominence to one focally stressed syllable, at the expense of weakening the duration, loudness and/or vowel quality of those that surround it. It is thus unsurprising if it is difficult for listeners, even L1 listeners, to match excised syllables or pairs of syllables to known words. The word is in effect, subsidiary to the group in which it occurs. There are also, of course, factors such as assimilation, resyllabification and elision, which cause the word to vary from its citation form in ways that are heavily dependent upon the context in which the word occurs (Brown, 1990).

How, then, is one to reconcile these two accounts: one suggesting that listeners process spoken input as it is heard, the other indicating that they are incapable of matching short