Comprehension of patterns of phonetic simplification in English by non-native speakers: A cognitive account

Eiichi Yubune

1. Introduction

Non-native speakers of English are much less familiar with how English phonemes are realized in connected speech, let alone their simplified forms. Japanese speakers, as a whole, are more familiar with, what Brown (1990) calls, ‘ideal sounds’ compared to sounds in normal speech (This is partly because teaching has been more focused on the comprehension of written text, and partly because the learners have often been exposed to English spoken relatively slowly and clearly in the context of formal teaching.). Although there are of course many examples of phonetic simplification of both vowels and consonants in Japanese, they must be permissible under the system of the syllable structure of Japanese, which is drastically different from that of English.

English words, on the other hand, have a lexical stress, due to which stressed syllables and unstressed syllables are manifested. Stressed syllables are likely to retain their ideal or canonical sound realization whereas unstressed syllables are more likely to be simplified or reduced both qualitatively and quantitatively. These phenomena may occur within a word and even in word boundaries. Besides, English is a stress-timed language, where elements in between stresses are squeezed into one bunch of reduced sequence so that the utterance as a whole can hold the stress-timed rhythm. A sequence of grammatical words such as prepositions and articles tend to appear in such intervals of stresses and thus reduced in normal speech whereas lexical words tend to have a sentential stress. For example:

(1) Michael must have been in the station.

In (1), the content words Michael and station may well have a stress in a normal situation while those items in between them are likely to be pronounced very fast and
reduced both in vowels and consonants. However, Japanese speakers who are not familiar with spoken form of English would have great difficulty to identify the sequence of these reduced items and eventually to comprehend the whole utterance.

In this essay, I want to discuss general problems concerning the cognitive process in comprehension of patterns of phonetic simplification in English by non-native speakers, particularly EFL Japanese learners in comparison with native English speakers from the cognitive standpoint.

2. Examples of simplification of consonants and their interpretation process

Let us look at examples of a rather weak type of simplification for consonants. Although the citation form of the word *hand* is [hænd], when it occurs in a context such as *Hand me the book*, the word may be produced as [hæm] because the final [d] is deleted and the preceding [n] assimilates with the initial sound of the following word in place of articulation. In another context, *hand* may be just [hæn] in a phrase like *Hand Tony the book* with deletion of [d] but without the subsequent assimilation, due to the matching of the place of articulation in /n/ and /t/. These phonological processes of coarticulation, and eventually phonetic simplification, are widely common and pervasive in fluent speech, but nonetheless do not usually deteriorate the native listeners’ perceptual interpretation.

How, then, is the native listener actually doing this rather sophisticated and systematic interpretation? What source of information is crucial for this mental process to be available? Furthermore, do non-native speakers also use the same kind of analysis as the native speakers do? With regard to the native speakers, Marslen-Wilson et al (1995) provide some empirical strands to these questions. One is the evidence that lexical representations are abstract and probably underspecified (e.g., the underspecification theory as in Archanegli (1988) and Pulman and Hepple (1993)). Another is that the treatment of simplification is highly context-sensitive.

In normal speech, potential ambiguities due to simplification can be resolved by different levels of constraints—phonological, lexical and sentential (Marslen-Wilson et al, 1995). Let us take the ambiguous velar /k/, for example: In [steik] as in *steak dinner*, the ambiguity is resolved phonologically because the following segment /d/
does not have velar place. In [swik] as in *sweet girl*, the ambiguity is resolved by the absence of a lexical item such as *sweek, sweak* and so on. In [eik] as in *There were eight kids at the party*, the ambiguity is resolved through the syntactic and semantic knowledge that *ache kids* is not acceptable and *eight kids* is most appropriate. Thus, these assimilation and deletion process may well be discussed in the framework of abstract phoneme representation, and in some cases an ambiguous segment due to coarticulation will be disambiguated only through the phonological sensitivity to coarticulation, but in other cases it may only be resolved by way of higher levels of contexts.

In addition to co-articulation processes, there are many other types of phonetic realization in which a phoneme is altered in terms of its voicing feature. It is well described in the literature of phonetics that voiced consonants tend to be voiceless when they occur in word-initial or word-final position. Specifically for word-initial plosives, however, the voicing cue is provided as the VOT value so that it is relatively easy to perceive the voice opposition. For word-final plosives as well as fricatives, however, the possible cues for the voicing may be the quality and the quantity of the preceding vowel. In other words, the final voiceless consonants are distinguished by a comparatively short vowel with tight voicing and glottalization; and a relatively long vowel with full voicing preceding a voiced consonant (Brown, 1974). Although these voicing cues may be reliable in isolated words, they may not be so evident in various phonological contexts in fluent speech. Thus, for example, *peace* and *peas* are less distinguishable in fast speech and other contextual information becomes reliable cues; *peace* is more likely in the context *keep the peace* and *peas* in the context *eat the peas* (Yubune, 1999). In this case, too, an abstract phoneme representation is being used to restore the altered phonemes by utilizing a higher-level context.

The question, then, is whether non-native speakers will do the same kind of analysis to interpret these assimilated or deleted sounds. Non-native speakers, whatever language they speak, also have to use top-down context such as lexical or semantic, or even discoursal, to interpret such difficult sounds. However, the speakers of a language in which the same kind of simplification often occur will have less difficulty in interpretation of such sounds than the speakers whose native language does
not see such simplification processes. Japanese is a CV language, as is well known, where consonants never occur in coda and consonant clusters never occur within and across words except for gemination. In what follows, in Japanese, consonants always occur between vowels and thus adjacent consonant assimilation or deletion cannot, in principle, occur. If Japanese speakers transfer their way of decoding Japanese into their interpretation of normal English speech, they have to be puzzled at a lot of phonetic change and absence at segmental level, since they expect each segment to be present even in normal continuous speech. Thus, Japanese need to be able to what kind of elements are likely to be changed or deleted in what phonological context, and what elements are likely to remain.

I argue, however, Japanese speakers should be, at the same time, encouraged to use as much top-down information as possible, because I consider that such knowledge as patterns and process of simplification are hard to utilize in listening if the speakers are not so familiar with them, especially at their early stage of learning.

3. Examples of simplification of vowels and their interpretation process

English vowels are typically centralized in unstressed syllables, particularly in non-lexical words such as I [ɪ] or can [kən]. In another case, they are completely deleted in an unstressed syllable in a longer word such as interest [ɪntrəst] or different [dɪfrənt]. However, Japanese vowels are not usually centralized or deleted because such simplifications will break the Japanese syllable system, where each syllable tends to be the same length. Besides, the standard Japanese does not have weak vowels such as /ə/. Consequently Japanese learners tend to expect English vowels to be always ideal sounds (not centralized) even when they occur in unstressed syllables. Thus, such a simplification pattern is considered to be one of the most difficult things for Japanese speakers to handle. In what follows, Japanese speakers need to know that unstressed vowels are typically centralized or sometimes deleted to realize the English stress-timed rhythm. However, in this case, top-down processing may also play an important role for the learners to interpret such weakened items, learners activating a cognitive metrical template for words and sometimes using higher level context such as syntactic or semantic context.
4. General discussion

A general idea of the effect of simplification patterns on the comprehension of English by non-native speakers entails some cognitive accounts, since simplified items, whether they are vowels or consonants, are not their canonical ones which vary in terms of at least one distinctive features such as place or manner of articulation and voicing. As a result, the listener must restore the ideal sounds in their minds in comprehension using the abstract notion of phoneme.

However, there is a case in which such restoration process may not apparently occur for native speakers. For example, contracted forms such as gonna or wanna are considered possible to be represented as going to and want to without taking any recourse to such restoration process since these instances are well-established and so common. In this case, the native listener may exclusively use the bottom-up cues to interpret the sequence of these simplified sounds. However, as Brown (1990) states, the native speakers, in general, are only partially reliant on the acoustic signal in forming an interpretation, suggesting that they rely much on top-down information to decode the sounds in normal speech.

Brown (1990) claims that this situation is different for non-native speakers; they are less sure of the general knowledge of English in terms of social and psychological complexity of the context of the language use. Besides, they are less sure of the forms of English, such as relevant vocabulary or syntactic structures. Thus, she concludes that non-native speakers are less able to bring to bear top-down processing to form an interpretation and thus rely more on bottom-up cues. Brown’s view of the interpretation of connected speech stresses the role of top-down information in comprehension of normal speech. However, if there are such ‘patterns’ of simplification as assimilation or elision, they should also be prestored more or less in the native speaker’s bottom-up processor, and then they may often be able to decode such simplified items only through the bottom-up cues, suggesting less reliance on top-down information.

In fact, some researchers argue against the prior effect of top-down context in spoken word recognition. For example, in Klatt’s (1979) Lexical Access From Spectra
model (LAFS), he assumes that spoken words are recognized directly from an analysis of the (sufficiently invariant) input spectrum using a larger unit called diphones (phoneme-phoneme sequences). He claims that by using the diphone, the LAFS model is capable of handling variability in spoken language such as speech rate, prosody and talker normalization. Furthermore, Stevens (1986) and Marslen-Wilson and Warren (1994) also stress the direct access of lexical entries on the basis of an acoustic analysis of the incoming speech signal. Stevens articulates the importance of linguistic binary phonetic features, whereas Marslen-Wilson and Warren highlight lexical representations specified in terms of phonological distinctive features such as [+voice] or [+nasal].

These models largely stress bottom-up processing, and deny a level of phonemic representation in spoken word recognition. However, these models may have their limits in coping with the degree of simplification. For example, how can the listener identify or restore the phrase “I am going to” from one of its possible simplified forms such as [ ] (Brown, 1999: personal communication)? In this circumstance, there appear to be no such segments that correspond to Klatt’s diphones. Furthermore, distinctive features such as [+voice] or [+nasal] would be no use to restore the original phrase. In these cases, higher information such as semantic or discoursal context would then be crucial for interpretation of such a simplified item.

It is noted that, among such information higher than segmental level, lexical context brings a robust feedback process in phoneme restoration or perception. One plausible reason is that listeners use pre-stored lexical knowledge in processing speech (Connine, 1990; Connine and Clifton, 1987; Tanenhaus and Lucas, 1987). It is claimed that the pre-stored phonological form of a lexical item contributes directly to the computation of the auditory input being processed.

An important point here is that although the native speakers may well utilize top-down information in interpretation of simplified speech, they also may often use very sophisticated and rather pre-stored, automatized bottom-up processing. However, the utilization of such bottom-up cues may be much dependent on the degree of simplification. Here I claim that the native speaker’s bottom-up processing is different in nature from that of non-native speakers. In other words, any slight simplification
will be a big problem for non-native speakers, especially at their early stage of learning, whereas most simplifications will be correctly and automatically interpreted by native speakers by attending to simply bottom-up acoustic cues (as is proposed by the above bottom-up priority researchers) if the simplifications take place only locally.

Thus, if the degree of simplification is small, the native speakers tend to cope with it by bottom-up confirmation to a great extent whereas non-native speakers must rely more on top-down information. In other words, the processing cost is much bigger for non-native speakers. This means that the capacity of the working memory (Baddeley, 1986; 1999 and others) of the native speakers is not as fully occupied by bottom-up processing load as that of the learners is. This follows that the native speakers afford to attend to and utilize much more semantic information at the same time than the learners can. At any rate, one learners’ goal is to increase the size of the processing unit they process at one time when listening in order to utilize the working memory more economically and effectively. I thus argue that, at the learning stage, learners must be exposed to a great amount of simplified data together with sufficient amount of context in a fully accessible way.

5. Conclusion

Native Japanese speakers are, as a matter of fact, less skillful in bottom-up processing in interpretation of normal English speech where many patterns of simplification are manifested because of the difference of the phonological systems between the two languages. They also less familiar with general information of the language use, resulting in less availability of top-down processing. I suggest, however, just as native English speakers often do, Japanese speakers also need to try to use as much top-down information as possible to interpret less familiar simplified items in normal English speech, because Japanese learners basically lack the amount of bottom-up data on simplified forms of spoken English. Thus I argue that the interaction of both these two types of cognitive processing manners will enable non-native speakers to learn most quickly how to understand normal English speech.

4. References


