Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Cognition 104 (2007) 631–637
COGNITION
www.elsevier.com/locate/COGNIT

Discussion

Do we have innate knowledge about phonological markedness? Comments on Berent, Steriade, Lennertz, and Vaknin ☆, ☆☆

Sharon Peperkamp *

Laboratoire de Sciences Cognitives et Psycholinguistique (EHESS, DEC-ENS, CNRS) and Université de Paris 8, France

Received 18 September 2006; accepted 6 December 2006

1. Introduction

In their article 'What we know about what we have never heard: Evidence from perceptual illusions', Berent, Steriade, Lennertz, and Vaknin (2007) address the question of whether listeners have innate knowledge of phonological markedness. Focusing on sonority and syllable structure, they answer this question positively. Specifically, they argue that English listeners' perception of onset clusters is reflected by the markedness of these clusters. Their reasoning is as follows. On the basis of typological facts, onset clusters with a sonority rise (such as /bn/, /b/ being less sonorous than /n/) are said to be universally less marked than those with a sonority plateau (such as /bd/), which are in turn less marked than those with a sonority fall (such as /lb/). Using three different tasks – syllable judgment, discrimination, and lexical decision – the authors show that the more an onset cluster is marked, the more likely it is perceived with an illusory epenthetic vowel by English listeners. English words do not contain any of the onset cluster types used in the experiments; hence, this

☆ This manuscript was accepted under the editorship of Jacques Mehler.

☆☆ I would like to thank Emmanuel Dupoux and Katrin Skoruppa for comments and discussion. Financial support was provided by a grant from the Agence Nationale de la Recherche (ANR-05-BLAN-0065-01).

* Tel.: +33 1 44 32 23 58; fax: +33 1 44 32 23 60.
E-mail address: Sharon.Peperkamp@ens.fr.

0010-0277/$ - see front matter
doi:10.1016/j.cognition.2006.12.009

markedness effect cannot be the result of phonological learning. Moreover, the authors rule out two other sources of the effect within the native listeners' language experience, i.e. statistical properties of the English lexicon and the process of fast speech vowel deletion. Finally, they argue that the effect is not due to phonetic differences among clusters with varying sonority profiles either. They therefore conclude that knowledge about the markedness of onset clusters is innate.

I argue that this conclusion is not justified for two reasons. First, all experiments test whether the different types of clusters are perceived with an epenthetic schwa or not. Epenthesis, however, is not the only possible perceptual repair. Hence, we cannot rule out that those cluster types that tend not to undergo epenthesis are more often subject to some other perceptual repair, thus undermining the authors' conclusion that they are perceived more faithfully. Second, the arguments against a phonetic explanation of the results are inconclusive.

2. Epenthesis versus other perceptual repairs

Concerning the type of perceptual repair that is investigated, the authors assume that if illegal clusters are repaired during perception, it is by means of the insertion of an epenthetic schwa. All experiments are designed accordingly. Thus, participants judge the number of syllables of non-word items (Experiments 1 and 2), they try to discriminate monosyllables from their disyllabic counterparts containing a schwa (Experiments 3 and 4), and they do a double lexical decision task on pairs of items that differ at most in the presence of schwa (Experiments 5 and 6). But what if some of the clusters under scrutiny undergo a perceptual repair other than epenthesis? For two of them, i.e. /tl/ and /dl/, a number of articles indeed report different repairs in English listeners. Pitt (1998) observed that /tl/ is perceptually confused not only with /tl/ but also with /tr/, a finding reported earlier by Massaro and Cohen (1983), while Hallé and Best (to appear) found confusion of /tl/ and /dl/ with /kl/ and /gl/, respectively. Hence, it appears that as far as the perception of /tl/ and /dl/ is concerned, English listeners apply at least two or three different repair strategies. The same might of course be true for other clusters that were used in Berent et al.'s experiments and that have not been the topic of such extensive investigation in the literature.

The authors briefly consider the possibility of repairs other than epenthesis. In a footnote, they report on a transcription task in which English speakers are found to transcribe the vast majority of the monosyllabic stimuli used for the first two tasks (number of syllable judgment and discrimination) as disyllabic. Interestingly, repairs other than epenthesis, such as replacement or deletion of one of the consonants, were more frequent for less marked clusters than for more marked clusters, hence showing a reverse markedness effect. This effect was significant in the comparison between sonority rises on the one hand and either plateaus or falls on the other hand. If the same pattern is found with a more reliable on-line perception task, this would be quite problematic for Berent et al.'s view. It should also be noted that the same transcription pretest was not carried out with the stimuli used for the lexical decision task, despite the fact that they have rather different phonetic properties. Indeed, whereas the first set of stimuli was recorded by a Russian speaker (all the onset clusters being legal in Russian), the second set was obtained by removing schwa out of the initial syllable of disyllabic tokens produced by an English speaker.[1]

The assumption that all illegal onset clusters are perceptually repaired by means of epenthesis appears to be inspired by the hypothesis that a single grammar handles both perception and production (Smolensky, 1996). According to the authors, then, the presence of epenthesis in the phonological grammar of English predicts the presence of epenthesis as a perceptual repair strategy. English indeed uses epenthesis to break up illegal clusters, such as /z/ in *the brushes* (cf. *the combs*) and *she pushes* (cf. *she pulls*). However, the inference that epenthesis should likewise apply in perception is mistaken. We can distinguish at least two stages in perception. One of them consists of the matching of surface forms against the more abstract forms that are stored in the lexicon. During this stage, the effects of phonological processes that apply during production are undone; that is, the processes that apply during this stage are the *reverse* of those that apply in production (for experimental evidence see, among others, Gaskell & Marslen-Wilson, 1996; Lahiri & Marslen-Wilson, 1991). Hence, listeners of a language whose phonological grammar contains a productive epenthesis process must delete the epenthetic segments for the purposes of word recognition. This stage taking legal surface forms as its input, it is irrelevant for the issue at hand. Rather, the perceptual repair of illegal onset clusters is arguably due to a previous processing stage, during which universal acoustic structures are mapped onto the closest language-specific surface structures. Non-native sounds and sound structures that are illegal in the listener's language get thus projected onto the closest native one (see, for instance, Best, 1994). Crucially, the processes that apply in order to repair illegal sounds and sound structures can be different from the phonological processes that apply during production. For instance, the perception of an epenthetic /u/-like vowel within consonant clusters by Japanese listeners (Dupoux et al., 1999) contrasts with the process of /i/-epenthesis within such clusters in the Japanese production grammar (Itô & Mester, 1999).

To sum up, there are no theoretical arguments to assume that illegal onset clusters are always repaired by means of epenthesis rather than by some other modification, and the empirical argument is not fully developed. Therefore, we cannot exclude that the clusters that were less often confused with their epenthetic counterpart in Berent et al.'s experiments are subject to some other perceptual repair. In other words, the authors have not demonstrated that universally less marked illegal clusters are perceived more faithfully than more marked ones; rather, all we can conclude is that an epenthetic vowel is perceived less often in the former than in the latter. Hence, the extent to which perceptual epenthesis applies in illegal onset clusters depends on their sonority profile. This is of course an interesting finding. The next question is whether these results show that English listeners have innate knowledge about the

[1] As it happens, we can expect that the onset clusters in the latter stimuli would overall elicit at least the same amount of transcriptions with an epenthetic vowel as those in the first set. This is because they were likely to contain co-articulatory cues of the removed vowel, cues to which listeners have been shown to be sensitive (Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999).

markedness of sonority profiles. Below, I examine one of the alternative explanations rejected by Berent et al.

3. Innate phonological knowledge versus phonetic properties of the stimuli

A possible account of the reported markedness effect that makes no appeal to innate knowledge lies with phonetic differences across onset clusters with varying sonority profiles. In particular, Berent et al. mention that clusters with falling sonority might be acoustically closer to their disyllabic counterparts, due to the fact that their initial consonant, a sonorant, shares spectral properties with vowels. They nevertheless argue against a phonetic explanation of their results, based on two arguments.

First, the two priming experiments, which differ only in whether or not participants are implicitly incited to pay attention to phonetic detail, show different results. According to the authors, this state of affairs is incompatible with a phonetic explanation. In the priming experiments, listeners perform a double lexical decision task: they listen to two items and have to indicate whether they are both words or not. The first item serves as a prime for the second one. In the crucial test conditions, prime and target are either identical (e.g. *bdif–bdif*) or they differ from one another in that the former contains schwa (*bedif–bdif*). Two types of illegal clusters are tested, containing a sonority plateau and a sonority fall, respectively. The results of the first experiment (Experiment 5) show a priming effect for clusters with a sonority fall in the schwa-related trials (participants are equally fast in *lbif–lbif* and *lebif–lbif*) but not for clusters with a sonority plateau (participants are faster in *bdif–bdif* than in *bedif–bdif*), suggesting that the most marked clusters, those with falling sonority, are perceived with an epenthetic vowel. Crucially, this priming effect disappears in the second experiment (Experiment 6), in which listeners are implicitly encouraged to pay attention to phonetic detail. This experiment contains a contingency between the presence of schwa in the prime and the correct response, in that the number of filler trials that consist of two schwa-related words (e.g. *polite–plight*) is doubled; as a consequence, there are twice as many schwa-related trials that require a yes response than there are requiring a no response. According to the authors, the absence of a priming effect (participants are always faster in the identity condition, regardless of the type of cluster) is unexplainable under a phonetic account. That is, they argue that if perceptual confusion in illegal clusters is due to their phonetic properties, then the effect should be persistent, even when listeners are encouraged to pay attention to phonetic detail. In their view, the difference between the results of the two experiments can only be explained by invoking different processing levels. In particular, whereas the first experiment would involve a phonological level, where grammatical repair takes place, the contingency in the second experiment would incite participants to perform the task at a grammar-free phonetic level.

I have argued above that the role of the grammar in phonological perception is not to repair phonologically illegal structure but rather to undo the effect of native phonological processes, and that perceptual repairs take place at a lower, phonetic, processing level. Contrary to what is argued by the authors, the finding that the added contingency changes the participants' performance is not in disagreement with the premise that the two experiments tap the same phonetic processing level. Phonetic perception is indeed modulated by several factors, including listeners' attention to phonetic cues. For instance, Norris, McQueen, and Cutler (2003) administered a categorization task and found that when listening to a continuum between the consonants /f/ and /s/ in a vowel context that always signals /f/, listeners tend to ignore the formant-transition cues between the vowel and the following consonant, due to the fact that they are uninformative. By contrast, when the most ambiguous part of the same continuum is presented after a lexical decision experiment containing an ambiguous sound between /f/ and /s/ in many different vocalic contexts, the formant-transition cues signaling /f/ are taken into account. Thus, participants who first do the lexical decision experiment show a bias towards /f/ in the following categorization task compared to those participants who only participate in the latter task. These results show that with the same (phonetic) task and stimuli, listeners can perform differently according to whether they are incited to either pay attention to or discard certain phonetic cues.

The authors' second argument against a phonetic explanation is based on the results of the syllable judgment task (Experiments 1 and 2). With this task, a reverse markedness effect was found in both the American participants and the Russian controls for disyllabic stimuli containing schwa, such as *benif*. That is, disyllables were more likely to be perceived as monosyllables if their monosyllabic counterparts had a less marked cluster. For both groups of participants this effect was significant in the comparison between sonority rises and plateaus: stimuli like *benif* were more often erroneously judged to be monosyllabic than stimuli like *bedif*. Whereas one might interpret the presence of the effect in the Russian controls as an indication that phonetic properties of the stimuli are at play (since both *bnif* and *bdif* are legal in Russian), Berent et al. argue in favor of a phonological account. In their view, the presence of innate phonological knowledge about sonority implies that less marked clusters are preferred to more marked clusters even in the Russian grammar, that allows for both. Moreover, they explicitly argue against a phonetic explanation, by showing that the duration of schwa does not differ across the two types of stimuli. Note, however, that there can be other phonetic cues that make disyllables more confusable with their monosyllabic counterparts if the relevant onset clusters have a sonority rise as opposed to a plateau. For instance, schwa might be more co-articulated with a following nasal consonant, as in *benif*, than with a following stop, as in *bedif*.[2] More in general, it can be difficult to determine which phonetic properties could influence a given perception task,

[2] The Russian controls also show a small but significant markedness effect with monosyllabic stimuli, where they make more errors in the case of a sonority fall (e.g. *lbif*) than in the case of a sonority plateau (e.g. *bdif*). A very similar question thus arises: is the effect due to a dispreference in the Russian grammar for the marked clusters with a sonority fall, as Berent et al. argue, or are there phonetic properties that make the consonant transition in these clusters be more schwa-like?

and, moreover, certain phonetic properties are hard to quantify. Fortunately, in this case an experiment can be designed that directly pits the hypothesis based on innate phonological knowledge against the phonetic one. Consider two pairs of schwa-related non-words such as *abenif–abnif* and *abedif–abdif*. The items without schwa contain a syllable boundary in between the two adjacent consonants. The presence of this syllable boundary reverses the markedness pattern, for the larger the difference in sonority between two consonants on either side of a syllable boundary, the better the syllable contact (Vennemann, 1988). In other words, whereas *bdif* is more marked than *bnif*, the reverse holds for their vowel-initial counterparts: *abdif* is less marked than *abnif*. Now suppose that we compare listeners' syllable judgments of *benif* and *bedif* to those of *abenif* and *abedif* while keeping the phonetic properties constant (which can be done by recording the larger non-words and splicing out the initial vowels). If perceptual confusion is influenced by phonological markedness, then listeners should make more errors with *benif* than with *bedif* – as found before – but less with *abenif* than with *abedif*. Conversely, if perceptual confusion is due to phonetic properties of the stimuli, then the error pattern should be the same, regardless of the presence or absence of an initial vowel; that is, listeners should make more errors with *(a)benif* than with *(a)bedif*.

4. Conclusion

It appears that despite Berent et al.'s laudable efforts, we do not know yet whether we have innate knowledge of the markedness of onset clusters with varying sonority profiles. I would like to conclude with a remark about the inherent difficulty of their enterprise. There is broad consensus that phonological markedness has its roots in phonetic properties of sounds and sound structures (for a review, see Hayes & Steriade, 2004). Concerning sonority, Wright (2004) specifically argues that the markedness of consonant clusters as a function of their sonority profile is entirely rooted in the robustness of phonetic cues for perception. For this reason, it is almost impossible to find out whether listeners rely on innate knowledge over and above their sensitivity to phonetic stimuli properties in perception experiments. I have argued that Berent et al. have not been convincing in ruling out a phonetic explanation of the observed markedness effect, and I have proposed an improved design of one of their experiments that factors out the influence of phonetics. If, in addition, this experiment is restricted to those clusters that are unambiguously repaired by means of epenthesis (as opposed to some other process), its results should allow us to settle the issue and gain important insight into the possible role of innate phonological knowledge in speech perception.

References

Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition, 104*, 591–630.

Best, C. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. Goodman & H. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 67–224). Cambridge, MA: MIT Press.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: a perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance, 25*, 1568–1578.

Gaskell, M., & Marslen-Wilson, W. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology. Human Perception and Performance, 22*, 144–158.

Hallé, P. & C. Best (to appear). Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop+/l/ clusters. *Journal of the Acoustical Society of America*.

Hayes, B., & Steriade, D. (2004). Introduction: the phonetic bases of phonological Markedness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically Based Phonology* (pp. 1–33). Cambridge: Cambridge University Press.

Itô, J., & Mester, A. (1999). The structure of the phonological lexicon. In N. Tsujimura (Ed.), *A Handbook of Japanese Linguistics* (pp. 2–100). Blackwell: Oxford.

Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form: a phonological approach to the recognition lexicon. *Cognition, 38*, 245–294.

Massaro, D., & Cohen, M. (1983). Phonological context in speech perception. *Perception and Psychophysics, 34*, 338–348.

Norris, D., McQueen, J., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*, 204–238.

Pitt, M. (1998). Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception and Psychophysics, 60*, 941–951.

Smolensky, P. (1996). On the comprehension/production dilemma in child language. *Linguistic Inquiry, 27*, 720–731.

Vennemann, T. (1988). *Preference laws for syllable structure and the explanation of sound change*. Berlin: Mouton de Gruyter.

Wright, R. (2004). A review of perceptual cues and cue robustness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 34–57). Cambridge: Cambridge University Press.