LISTENING EXPERIENCE AND PHONETIC-TO-LEXICAL MAPPING IN L2

Anne Cutler[1] and Andrea Weber[2]

[1]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, and MARCS Auditory Laboratories, University of Western Sydney, Australia; [2]Saarland University, Saarbrücken, Germany
anne.cutler@mpi.nl; aweber@coli.uni-sb.de

ABSTRACT

In contrast to initial L1 vocabularies, which of necessity depend largely on heard exemplars, L2 vocabulary construction can draw on a variety of knowledge sources. This can lead to richer stored knowledge about the phonology of the L2 than the listener's prelexical phonetic processing capacity can support, and thus to mismatch between the level of detail required for accurate lexical mapping and the level of detail delivered by the prelexical processor. Experiments on spoken word recognition in L2 have shown that phonetic contrasts which are not reliably perceived are represented in the lexicon nonetheless. This lexical representation of contrast must be based on abstract knowledge, not on veridical representation of heard exemplars. New experiments confirm that provision of abstract knowledge (in the form of spelling) can induce lexical representation of a contrast which is not reliably perceived; but also that experience (in the form of frequency of occurrence) modulates the mismatch of phonetic and lexical processing. We conclude that a correct account of word recognition in L2 (as indeed in L1) requires consideration of both abstract and episodic information.

Keywords: Phonetic processing, word recognition, abstract, episodic, L2.

1. INTRODUCTION

There are many differences between acquisition of a first language in childhood (L1) and of a second language later in life (L2). One such difference is, without question, the richness of the information sources on which the learner can draw.

The infant learner of an L1 is not provided with metalinguistic descriptions, helpful hints about the learning task, or feedback on the correctness, or otherwise, of categorisation decisions. The input an infant receives is in many ways well adapted to the task to be accomplished [5,7], but the successful construction of a phonetic repertoire (one of the great achievements of the first year of life [12]) is, clearly, based on evaluation of heard input alone.

This situation contrasts starkly with that of the L2 learner. Even L2 learners who receive no explicit teaching of any kind – e.g., the most abjectly exploited migrant worker – can draw upon rich resources far beyond what is available to the infant. First, by having already acquired an L1 they know how language works; they know about words, they appreciate that vocabularies of many words are constructed from a selection of relatively few speech sounds, and they know a lot about how words may be used. Second, they know about learning; they know how to store knowledge for re-use, how to practise learned competences to improve facility with them, and how to ask for useful information from others, or in some other way elicit it. They know about social interaction and the role and importance of language in this, and in most cases about orthography and its use.

Explicit teaching of any kind – in a classroom, by individual tutoring, self-instruction texts, web-based courses or whatever – provides a wide range of further resources. These can be of many types: dictionary translations of words and constructions between the L1 and the L2, structured introduction of increasing levels of complexity, reading and writing practice in the L2, even attention to the precise nature of contrasts. The effect of all these options, however, is that the L2 learner, in whatever learning situation, is in possession of knowledge about the new target language drawn from many sources other than the spoken input of the available listening experience. We proceed from the assumption that communicative goals are sufficiently motivating that each such potentially useful information source will indeed be drawn on.

With respect to the phonology of the L2, then, information can be gleaned not only directly from listening, but also from spelling, from observing native speakers' perceptual behaviour, and from incidental relevant information such as song texts or jokes. This leads to a potentially interesting situation when such information suggests a difference between two sounds which are both within the range of variation of a single category in

the listener's L1. This is known to be the most difficult type of contrast for an L2 listener to distinguish [2]. Well-attested cases include /r/ versus /l/ for Chinese or Japanese learners of English. English /ae/ versus /ɛ/ is equally difficult for listeners with Dutch or German as L1. In each case, the L1 has only one phonemic category covering the range of the two English categories. Famously, the listeners confuse the two English categories in both perception and production.

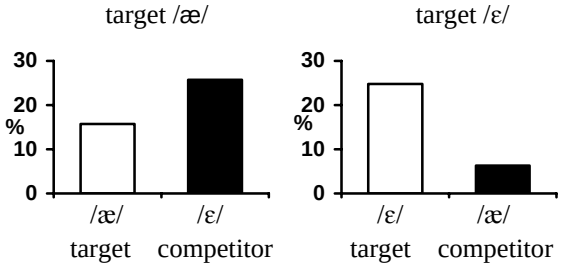
But equally famously, they often know that they are supposed to distinguish them. English speakers tell them so, or make fun of them for their errors (try a Google search on *flied lice*). There are many websites devoted to such errors or to training exercises for inducing perception of the contrast (at least for /r-l/). The knowledge of the existence of the contrast is meta-knowledge, not knowledge derived from analysis of the input. The input is uninformative with respect to the contrast because listeners confuse it; their stored knowledge about the L2, however, includes this distinction. This has quite drastic effects on their recognition of spoken words, as a series of recent studies has shown.

2. THE EYETRACKING TASK AND PHONETIC-TO-LEXICAL MAPPING

Psycholinguists have expended a lot of effort on devising ways to look at spoken-word recognition as it happens. The eyetracking task, which has only been in use in psycholinguistics for a dozen or so years, has proven itself one of the most informative methods we have. In an eyetracking experiment, listeners wear a miniature camera on the head, which allows the direction of their gaze to be continuously recorded. Typically, they face a display (e.g., on a computer screen) featuring a number of objects. The spoken-word recognition stimuli are usually presented as an instruction, e.g., to click on one of the pictured objects.

Weber and Cutler [14] used the eyetracking technique to examine the mapping of phonetic information to lexical entries in L2. Although the literature on L2 speech perception is huge, there has actually been little attention to how phonetic perception relates to lexical storage. Perhaps it was always assumed that phonetic perception fully determined lexical storage. If an L2 listener cannot hear the difference between *rice* and *lice*, the two words might be stored in the lexicon as fully homophonic, just as *mail* and *male* must be. However, because the eyetracking technique allows the mapping of phonetic information to lexical entries to be investigated with great sensitivity over time, it enabled Weber and Cutler to give the lie to this assumption.

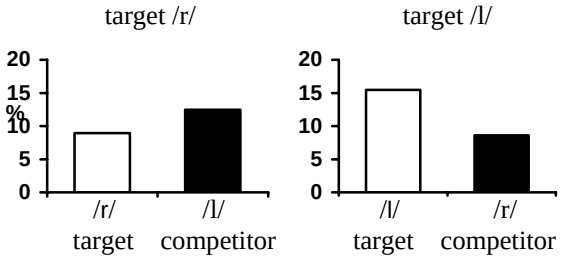
Figure 1: Competition effects in Weber & Cutler [14]; mean advantage (in percentage of looks) for the target picture and the competitor picture over the (averaged) distractor pictures in the time window 300-500 ms from target word onset, roughly covering eye movements driven by the first syllable of that word.



In their initial study Dutch listeners were shown a four-picture display including, for example, a panda and a pencil, plus two distractor pictures (e.g., a duck and a strawberry). Instructions (in English) asked them to click on the panda. The Dutch listeners (but not British native control listeners, of course) were likely to look first at a picture of a pencil when instructed to look at a panda. However, the reverse did not apply - instructions to click on the pencil did not induce many looks to the panda. Figure 1 summarises the relative attractiveness of the target and the competitor in each target condition in this experiment. The figure shows how much more atttractive the target picture and the competitor picture are than the two distractor pictures in the display, averaged across the time window which best captures eye movements influenced by just the first syllable of the target. When the listeners hear a target with /æ/, such as *panda* (left graph), there is significant competition from the /ɛ/ word (e.g., *pencil*). But the /ɛ/ target (right graph) receives very much less competition from the /æ/ word.

The asymmetry in these results clearly indicates that the two initial syllables were not represented as interchangeable homophones in the lexicon. If they had been, then the effects would have been similar in each condition – whatever the target, both names would have been equally possible, and both pictures should have received equivalent proportions of looks. But although both targets were apparently perceived as containing a vowel appropriate for the /ɛ/ words, this vowel was not

Figure 2: Competition effects in Cutler, Weber & Otake [3]; mean advantage for the target picture and the competitor picture over the (averaged) distractor pictures, computed as in Figure 1.



mapped to the lexical representations of the /æ/ words. Therefore, these listeners apparently maintained a distinction between English words containing [æ] and containing [ɛ] in their lexical representations, even though they could not successfully make this distinction in their acoustic-phonetic processing of English speech.

Cutler, Weber and Otake [3] undertook another eyetracking study to examine this phenomenon once again; their new study involved the English /r-l/ distinction and Japanese listeners, who were instructed to click on pictures in a display containing, for example, a rocket and a locker. The Japanese listeners also showed asymmetric lexical confusions. Instructed to look at the rocket they significantly often looked at the locker, but instructions to look at the locker did not induce them to look so much at the rocket. Figure 2 summarises these results in the same way that Figure 1 did for the experiment with Dutch listeners. It is clear that this phonetic-to-lexical asymmetry is a general possibility in L2 processing, and is specific neither to the highly proficient Dutch English-speakers tested by Weber and Cutler, nor to vowel processing such as in the distinctions examined in that earlier study.

3. LEXICAL GENESIS IN L2

In both the Dutch and the Japanese cases, the results from these eyetracking experiments indicate that there must be a stored lexical distinction that cannot be heard. The preferred interpretation of the input is consistently in terms of one of the two L2 categories (for the Dutch /ɛ/, for the Japanese /l/) but the lexical representations of words containing /æ/ and /r/, which would be different from the lexical representations of /ɛ/- and /l/-words for L1 users, are also different for these L2 users. Clearly, the relation of phonetic processing to lexical storage in L2 is not as simple as one might have thought. L2 listeners maintain lexical distinctions even when success in mapping spoken input correctly to the two distinct categories is at best no more than a faint future hope.

But the important issue in the present case is the source of the phonological distinction represented in these L2 listeners' lexicons. Where a distinction can reliably be heard, a case can be made either for exploitation of an abstracted phonetic distinction, or for generalisation from episodic traces (or both!) in incorporating the distinction into lexical storage. But where a distinction cannot be reliably heard, the latter option is unavailable. Statistical learning studies [7,9] have shown that reliable bicategorical judgements in categorisation tests only result when listeners have been exposed to a bimodal training distribution; listeners trained with a unimodal distribution cannot respond with reliable decisions even if the absolute number of times they have heard the specific test items is identical to the experience of the bimodal group. Thus the L2 listener's exposure to what is apparently a unimodal distribution cannot support construction of lexical representations exactly matching those of native listeners: /r/ words in one form, /l/ words in another; or /æ/ words in one, /ɛ/ words in another.

A fundamental dispute in speech perception over the past years has been the contribution of abstract versus episodic knowledge in learning or recognising words of the L1 [6,10]; there is, by now, good evidence for the use of both. L2 clearly constitutes an excellent testbed for the evaluation of episodic vs abstractionist models, because the extra kinds of knowledge which the L2 learner brings to bear on the learning task are in some sense abstract. Note that we surely do not hold L2 and L1 learning to be different in kind (the same brain does it, after all!). Therefore it is probably the case that abstract and episodic knowledge both have a role in L2 learning too. The phonetic-to-lexical asymmetry, as it turns out, can form a specific test case, as the next two sections show.

4. ABSTRACT INFLUENCE ON LEXICAL REPRESENTATIONS

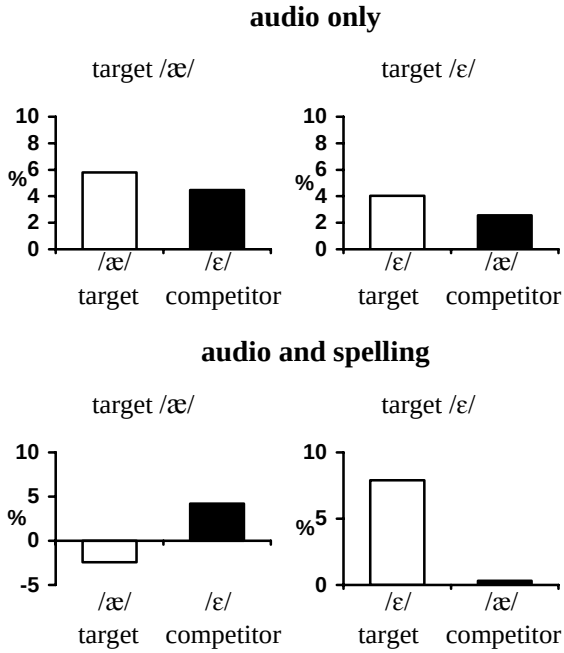
The recognition of newly-learned words is easily assessed with the eyetracking paradigm [13], and Escudero, Hayes-Harb and Mitterer [4] exploited this possibility, teaching novel object-name pairings to listeners from the same Dutch

population as Weber and Cutler had tested. All the instructions in their experiment were, again, in English. The subjects were taught, for instance, that one novel object was called a *tenzer* while another was called a *tandik*. After training the listeners could reliably select the correct novel object's picture. Crucially, Escudero et al. had two training conditions. In one, the participants heard the picture names and never saw them written. In the other, they heard the names and also saw them.

The spelling should function as a clue to the existence of a contrast between the initial syllables *ten-* and *tan-*. If just having this single clue is sufficient to initiate the establishment of lexical representations contrasting in this property, then subjects in the audio+spelling condition should represent the novel names distinctly. Assuming that, like Weber and Cutler's subjects, they cannot reliably hear the English [æ]-[ɛ] distinction, they should then give evidence of the same mismatch of phonetic and lexical representations as Weber and Cutler observed. The audio-only subjects, in contrast, should only instantiate a distinction to the extent that they can hear it, which, with this group, is proposed to be hardly at all.

After their training, both groups of participants followed eyetracking instructions to click on pictures, in displays containing (in the critical trials) both members of the pairs such as a tenzer and a tandik. The results from their two groups are shown in Figure 3, in the same summary form as Figures 1 and 2 presented the earlier results from Dutch and Japanese listeners. The audio-only participants produced similar response patterns given (the early portions of) either input, suggesting that they had perceived no difference and that the initial syllables were represented similarly. The audio+spelling group, however, displayed exactly the same asymmetry as Weber and Cutler had found; when they heard *tan-* from *tandik*, they were more likely to look at the tenzer, whereas when they heard *ten-* from *tenzer* they looked at the tenzer but hardly looked at the tandik. This suggests that they are more likely to perceive either input as containing /ɛ/, but input containing /ɛ/ does not contact the stored representation of *tandik*. The additional information about the spelling was sufficient for them to store *tandik* with the same kind of lexical representation that they possess for known English words containing /æ/, such as *panda*.

Figure 3: Competition effects in Escudero, Hayes-Harb & Mitterer [4]; mean advantage for the target picture and the competitor picture over the (averaged) distractor pictures, computed as in Figure 1 except that the time window is 400-800 ms from target word onset. The upper pair of graphs shows the results when listeners were trained with the auditory input only, the lower pair shows the results when the training also included spelling information.



5. EXPERIENTIAL INFLUENCE ON LEXICAL REPRESENTATIONS

Experiential influence which is strong enough to alter the balance of phonetic and lexical processing is hard to achieve in a laboratory setting. However, it is possible that we can test for pre-existing modulations of the observed phonetic-to-lexical mismatch due to effects of experience. The classic example of experiential influence is, of course, the effect of frequency of occurrence. Most simply, we can examine whether the likely accrued total of encounters with known words can modulate the effects that have been observed in these studies, such as the low probability of competition from words containing /æ/ given input containing /ɛ/.

The word pairs tested by Weber and Cutler [14] were balanced overall for frequency in a deliberate attempt to exclude differential influence of experience. We addressed the present question by deliberately selecting unbalanced materials instead. Thus we contrasted pairs such as *hat* and *hedge*. In each of several corpora we had access to, the frequency count for *hat* was more than three times

that of *hedge*. This allows us to determine, for instance, whether very common words with /æ/ would provide competition when lower-frequency words with /ɛ/ were heard, so that we would find a pattern different from that in the right panel of Figure 1, where *panda* did not provide much competition when *pencil* was heard.

At the same time we tackled another difficult issue in this type of study. Setting up eyetracking experiments is never easy because in the classic design, the crucial words have to be accompanied by images. This makes all sorts of additional control studies necessary to ensure that the pictures used in the study are indeed typically named with the crucial experimental word, that the pictures are all equally attractive, and so on. When one adds the constraint that the words must be known to L2 learners, the selection of materials for such a study becomes a torment. Eyetracking researchers have therefore begun to explore whether printed words might not constitute acceptable target items. Indeed, it appears that patterns of competition given printed words the pronunciation of which begins in the same way (e.g., CANDY CANDLE) is just like the patterns observed when the stimuli are pictures with names that begin in the same way [11]. Using printed words means that the selection of items is not confined to picturable words; removal of this restriction would certainly help to overcome the limitations on materials selection in the L2 case. Our new frequency study was hence conducted with pictures and again with words (in the latter case we had twice as many items!).

The results are summarised in Figures 4 and 5. When listeners heard *hedge* they were indeed significantly likely to look at the hat instead; when frequency favours the /æ/ word, its representation is capable of being contacted by /ɛ/ input. There is still no indication that these Dutch listeners hear an /æ/-/ɛ/ difference, just as in Weber and Cutler's or Escudero et al.'s studies. The words (and their spellings) were known to these listeners, so that both the picture and the printed-word conditions in this case should have involved the same lexical representations. Indeed, the result is the same in both cases: the availability of a high-frequency name is such that significant competition is offered from an /æ/ competitor to a target with /ɛ/. Thus frequency effects do modulate the abstractly determined mismatches between the phonetic and lexical levels of processing.

Figure 4: Competition effects in Dutch listeners' responses to lower-frequency target words containing /**ɛ**/ given a higher-frequency competitor containing /**æ**/; mean advantage for the target picture and the competitor picture over the (averaged) distractor pictures, computed as in Figure 1.

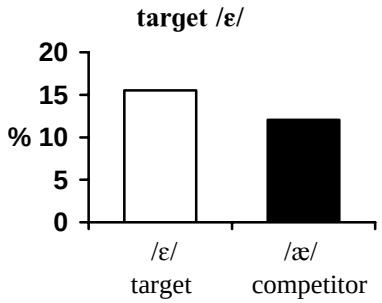
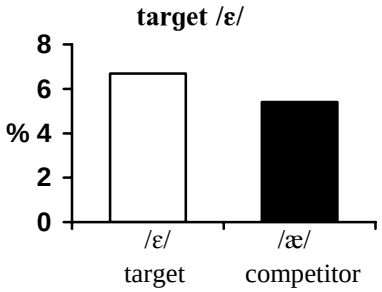


Figure 5: Competition effects in Dutch listeners' responses to lower-frequency target words containing /**ɛ**/ given a higher-frequency competitor containing /**æ**/; mean advantage for the printed target word form and the printed competitor word form over the (averaged) distractor words, computed as in Figure 1.



6. CONCLUSION

The results of these two series of new studies clearly lay out the challenge for modelling lexical genesis in L2. Both abstract knowledge about the linguistic system being acquired, and episodic experience with input in the new language, affect the construction of the lexical representations of new vocabulary items.

Spelling information is not encoded in auditory episodes (especially not when the target language is English, in which *laugh* rhymes with *half*, and *sane* with *reign*!). The effect of information about spelling overlaid upon pronunciation information in Escudero et al.'s [4] study was dramatic: similar performance irrespective of input was replaced by the highly asymmetric performance previously found for words which the listeners knew to have differing target pronunciations. The information about spelling sufficed to install this difference into the newly constructed representations.

Thus abstract knowledge is clearly effective in influencing the content of newly created lexical representations, such as are needed for the learning of an L2. Such an effect can of course be exercised only by way of representations (of the categories concerned) which are themselves abstract.

However, the content of the same lexical entries is apparently also subject to modulation by effects of experience (here instantiated as frequency of occurrence). There are several possible alternative explanations for how such effects might arise. First, one might postulate that very common words containing the sound /æ/ would have been accessed very often from the lexicon, with the result that accumulated traces of access episodes would have created a strong veridical trace of the correct English vowel. Input containing this vowel would then be more likely to contact the stored representations of the word. This explanation can be tested by comparing the patterns in Figures 4 and 5 against results from a study with the same input but native L1 listeners. Second, experience of successful access may have altered the stored abstract representation itself, to include multiple possible realisations of the vowel. This would be the case if, for instance, the L2 listeners had heard the common words more often than rare words in accented realisations from other L2 speakers. This explanation could perhaps be tested via deliberate mispronunciations such as [hɛt] for *hat*. (Note that it is known that L2 listeners can find other L2 users' speech easy to perceive [1]). Third, the effect could involve a more complex interaction of factors. Consider that Cutler et al. [3] proposed that the phonetic-to-lexical asymmetry might be a graded phenomenon, with the two L2 categories coded in terms of goodness of fit to the single L1 category that appeared to dominate the perceptual processing. This explanation allows for *panda* to be activated when *pan-* is heard, for instance – just more weakly activated than *pencil*. The effect of frequency then would be orthogonal to the effect of goodness of fit, and additive.

Whatever the final explanation, these accounts all rely at some level on the effect of accrued episodes of auditory experience. Thus not only abstract knowledge forms L2 lexical instantiations; episodic experience also plays a role. Vocabulary acquisition in L2 thus draws on many sources, as we proposed in the introduction. Even for phonetic category learning, explicit comparison of lexical and statistical evidence has shown that effects of the former type of evidence are stronger [7]. It is, now, a challenge for models of vocabulary learning (for which, we note, the L2 has shown itself to be a superb testbed) also to capture the simultaneous effects of abstract and episodic influence.

7. ACKNOWLEDGEMENTS

Financial support is acknowledged from NWO-Spinoza (AC) and from SFB378 of the Deutsche Forschungsgemeinschaft (AW).

8. REFERENCES

[1] Bent, T., Bradlow, A.R. 2003. The interlanguage speech intelligibility benefit. *J.Acoust.Soc.Am*. 114, 1600-1609.

[2] Best, C.T. 1995. A direct realist view of cross-language speech perception. In: Strange, W. (ed.), *Speech Perception and Linguistic Experience*. Baltimore, MD: York Press, 171–204.

[3] Cutler, A., Weber, A., Otake, T. 2006. Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics* 34, 269-284.

[4] Escudero, P., Hayes-Harb, R., Mitterer, H. submitted. Novel second-language words and asymmetric lexical access.

[5] Fernald, A., Simon, T. 1984. Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology* 20, 104-113.

[6] Goldinger, S.D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psych. Rev.* 105, 251-279.

[7] Hayes-Harb, R. 2007. Lexical and statistical evidence in the acquisition of second-language phonemes. *Second Language Research* 23, 61-90.

[8] Kuhl, P.K., Andruski, J.E., Chistovich, I.A., Chistovich, L.A., Kkozhevnikova, E.V., Ryskina, V.L., Stolyarova, E.I., Sundberg, U., Lacerda, F. 1997. Cross-language analysis of phonetic units in language addressed to infants. *Science* 277, 684-686.

[9] Maye, J., Gerken, L. 2000. Learning phoneme categories without minimal pairs. *Proc. 24th Annual Boston Univ. Conference on Language Development*, 522-533.

[10] McQueen, J.M., Cutler, A., Norris, D. 2006. Phonological abstraction in the mental lexicon. *Cogn. Sc*. 30, 1113-1126.

[11] McQueen, J.M., Viebahn, M.C. 2007. Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology* 60, 661-671.

[12] Saffran, J.R., Werker, J.F.,Werner, L.A. 2006. The infant's auditory world: Hearing, speech, and the beginning of language. In: Siegler, R., Kuhn, D. (eds.), *Handbook of Child Development.* New York: Wiley, Vol. 6, 58-108.

[13] Shatzman, K.B., McQueen, J.M. 2006. Prosodic knowledge affects the recognition of newly-acquired words. *Psychological Science* 17, 372-377.

[14] Weber, A., Cutler, A. 2004. Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language* 50, 1-25.