The lexical statistics of word recognition problems caused by L2 phonetic confusion

Anne Cutler

Max Planck Institute for Psycholinguistics
Nijmegen, The Netherlands
anne.cutler@mpi.nl

Abstract

Phonemic confusions in L2 listening lead to three types of problem at the lexical level: inability to distinguish minimal pairs (e.g. *write*, *light*), spurious activation of embedded words (e.g. *write* in *delighted*) and delay in resolution of ambiguity (e.g. distinction between *register* and *legislate* at the sixth instead of the first phoneme) The statistics of each of these, computed from a 70,000+ word English lexicon backed by frequency statistics from a 17.9 million word corpus, establish that each causes substantial addded difficulty for L2 listeners.

1. Introduction

Listening in the mother tongue (L1) is effortless; but listening in a second language (L2) can prove hard. Unfamiliar words, unknown idioms, and hitherto unencountered accents present new challenges. Speech can seem too fast, if it mismatches L1 prosodic [1] or phonotactic expectations [2]. Segmentation procedures which work well for the L1 may be applied to the L2, even when they work inefficiently there [1,3]. Syntactic processing is less efficient [4], prosodic cues to idioms are less efficiently processed [5], as is the exploitation of prosody for information structure [6]. All together, these effects cause significant disadvantage for L2 in comparison to L1 listening.

The most well-studied problem in L2 listening is the inability to distinguish between L2 phonemes which map to a single L1 category – e.g. Japanese problems with English /l/-/r/ [7,8]. What is not so well studied is the consequences of phonemic misidentifications for recognizing L2 words.

Word recognition involves two important processes: multiple activation and competition. Incoming speech calls up an array of potential word candidates which form at least temporarily a partial match to the input. This is an inevitable result of the fact that vocabularies are constructed from only on average 30 phonetic categories [9]. Moreover, languages prefer short words to long ones. This results in large numbers of minimal pairs of shorter words, and longer words with shorter words embedded within them. In fact, only about 2% of English words do not contain some other word [10]. And because many of these embeddings occur at the beginning of the matrix word, the first full word a listener hears may not be the intended word, but only a spuriously embedded form. Thus *star* may turn into *start* or *stark* or *starve* or *starling* as more speech input arrives; *start* may become *starch* or *startle*; *starch* may turn out to have been *star chart* after all.

Experimental evidence for multiple simultaneous lexical activation is plentiful [11,12], as also for active competition between activated word candidates [13,14]. The more active candidates, and hence the more competition, the slower the recognition proceeds [15].

There will be competition in L2 just as in L1 listening. The question at issue is not whether such competition occurs, but how much of it occurs. Inability to distinguish phoneme categories in the L2 can impact in at least three ways on lexical processing: (1) it can result in pseudo-homophones, i.e. the inability to distinguish minimal pairs such as English *write, light* or *cattle*, *kettle*; (2) it can cause spurious word activation, whereby nonwords such as *daf*, *lem* in *daffodil*, *lemon* may actually be heard as real words (*deaf, lamb*); and (3) it can induce temporary ambiguity, so that a larger number of competitor words remains active for a longer time for the L2 listener in comparison with the L1 listener (as when inability to distinguish /r/ from /l/ makes *register* distinct from *legislate* only at the sixth rather than the first phoneme). We calculate the lexical statistics concerning each of these in turn.

2. Method

The severity of the problem of course depends on the number of phonemic confusions an individual listener makes. It also depends on vocabulary size. Exact computations are therefore impossible; but we can at least establish the potential addition to the competition process due to a given phonemic confusion. Here it makes a difference whether the confusion we choose is vocalic or consonantal. The contribution of vowels and consonants to vocabulary structure is not symmetric; for English words from two to 15 phonemes in length, for instance, there are about 2.2 times as many lexical neighbours if a consonant is replaced (*cat* becoming *mat*, etc) as if a vowel is replaced (*cat* becoming *cot* etc.). For the calculations we therefore chose one vowel and one consonant confusion, in each case choosing L2 contrasts which collapse to a single L1 category: English /r/-/l/ (for, e.g., Japanese listeners) and /æ/-/ɛ/ (for, e.g., Dutch listeners).

Using the CELEX lexicon of over 70,000 English words [16], we can calculate how much pseudo-homophony results from a given phonemic confusion, and how much spurious embedding and temporary ambiguity the same confusion causes. To do this we counted (a) for every word containing one such phoneme whether replacing it with its confusable alternative produces another real word. For instance, *and* becomes *end* which is a word, so a pseudo-homophone would be created; but *as* becomes *ez* which is not a word. Next, we checked (b) whether words with substituted vowels (e.g. /dæf/, /ɛz/) occur embedded in other words (e.g. /ɛz/ in *residue* or *mezzanine* or *esoteric*). This assesses the possibility for spurious lexical activation, e.g. of *as* in *residue*. Finally, we checked (c) the number of temporary competitors which would be added by a substitution, i.e. how many words begin /əbɛ/ if the third phoneme of *abandon* is misheard as /ɛ/.

3. Pseudo-homophony

3.1. Evidence from listening

Lexical decision experiments, in which listeners hear spoken forms and decide for each item whether or not it is a real word., show that L2 listeners really do suffer from pseudo-homophony. A robust effect in lexical decision is “repetition priming” - responses are faster to items which have been presented before. Spanish-Catalan bilinguals dominant in Spanish experience repetition priming for some Catalan word pairs differing only in a distinction found in Catalan but not in Spanish [17]. The same occurs for Dutch and Japanese listeners to English presented with English /æ/-/ɛ/ minimal pairs such as *cattle/kettle, flesh/flash* and /r/-/l/ minimal pairs such as *right/light, glass/grass* [18]. The Dutch listeners respond significantly faster to one member of a *cattle/kettle* pair after having heard the other member earlier in the list (compared with having heard a control word), suggesting that both words are activated whichever is heard. Japanese listeners, however, show no such priming for *cattle/kettle* words, but do show priming across /r/-/l/ pairs, e.g. *right/light* or *glass/grass*. Thus confusable contrasts (/r/-/l/ for Japanese, /æ/-/ɛ/ for Dutch) increase homophony for L2 listeners.

True homophones (e.g. *meet/meat* or *sale/sail*) must be interpreted by reference to context. L2 listeners will thus have to perform this operation more often than L1 listeners do.

3.2. Lexical statistics

Pseudo-homophony turns out not to be an extensive problem. Confusion between /æ/ and /ɛ/, for example, adds less than 150 cases of homophony to the English vocabulary. *Bland* and *blend*, *cattle* and *kettle* become homophones for Dutch listeners or for others who cannot distinguish between these vowels; whereas of course they are not homophones for English listeners. The exact number depends on the direction of substitution. Replacing /æ/ by /ɛ/ adds 137 homophones to the lexicon. Substitution in the opposite direction does not give exactly the same number because some words contain instances of each phoneme. Thus *access* will in the above count become *excess*. But if we replace /ɛ/ by /æ/, we would replace both tokens of /ɛ/ in *excess*, giving a nonword. Replacement of /ɛ/ by /æ/ adds 135 cases of homophony.

The situation for confusion of /r/ and /l/ is a little worse – around 300 homophones are added. *Glass* and *grass*, *parrot* and *palate* are homophones for Japanese or any other listener who cannot distinguish /r/ from /l/, where English listeners have no trouble keeping these pairs apart. Again the count is asymmetric (*rightly* becomes *lightly* if /r/ goes to /l/, but *lightly* becomes a nonword if /l/ goes to /r/; *celebration* becomes *cerebration* if /l/ goes to /r/ but *cerebration* becomes a nonword if /r/ goes to /l/); replacement of /l/ by /r/ adds 287 cases, replacement of /r/ by /l/ 311 cases. Note that the r/l counts do not include words with syllable-final /r/. Japanese listeners would be unlikely to confuse American English *peer* with *peel* since they can generally make this distinction syllable-finally [8]. It was not necessary to exclude such cases from our count, however, since we used the CELEX British transcriptions; in British English, final r is not pronounced.

Given that listening in any language involves a certain amount of homophony, and actually there are hundreds more real homophones in English, the additional load caused by any one such phonemic confusion may seem relatively minor. In fact, computing the exact size of the additional load is probably not possible, because we need a precise definition of what counts as homophony. We can count the number of words in CELEX for which there is another word with the same pronunciation but different spelling - like *meet* and *meat* or *bury* and *berry*. There are 660 such words in CELEX. But many homophone pairs exist which have the same spelling - the bank of the river and the bank which administers money, for instance. Those are different meanings; but is the mouth of the river and the mouth of the human different, or just a metaphorical derivation? However one draws the line, it is clear that dealing with different meanings of a single auditory sequence is something listeners do a great deal of. Each phonetic confusion further complicates the L2 listener's task.

4. Spurious word activation

4.1. Evidence from listening

The problems causing pseudo-homophony could also lead L2 listeners to recognise words which are not there at all. Thus *phantom* may activate *fan* for any listener; *chastise* may activate *chess* for Dutch listeners and *regular* may activate *leg* for Japanese listeners. There are no real English words *reg* or *chass*, so for the native listener no such competition arises.

Does this happen? Is *chass* perceived as a token of *chess* by a Dutch listener? Broersma [19] conducted a lexical decision experiment in which Dutch and English listeners heard, among real English words and clear nonwords, "near words" formed by replacing a confusable phoneme in a word (e.g. *chess* became *chass* and *gang* became *geng*). Dutch listeners responded YES to the near words in 66% of the cases - they heard “words” where native English listeners did not. Although it is true that speakers do not, in general, say non-words, nevertheless this spurious word activation does constitute a real problem for L2 listening, because, as we have seen, such strings can occur as embeddings: *daf* in *daffodil*, *lem* in *lemon*, *stemp* in *The Last Emperor*, etc. Broersma [19] tested this with cross-modal priming: visual lexical decision preceded by spoken primes. This method measures multiple activation and competition in word recognition. Broersma presented fragments of spoken English words in a short context (e.g. *She looked at the daf-* from *She looked at the daffodil*) to Dutch and English listeners. A word appeared on screen for lexical decision while they were hearing the fragment. In this example, the visual word might be DEAF.

As expected, English listeners’ responses were faster after matching primes (*def-* DEAF) but slower after minimally mismatching primes (*daf-* DEAF). Dutch listeners, however, showed significant facilitation both in the matching condition and for the minimally different fragments. In other words, such fragments occurring embedded in real words do result in spurious word activation for L2 listeners; their listening thus involves competitors which are not competing for L1 listeners’ attention, since the L1 listeners hear only nonwords in the first syllables of *daffodil*, *lemon* or *chastise*. The more competition there is, the slower words are recognised [15]; thus this effect too will slow the recognition of spoken language by L2 listeners.

4.2. Lexical statistics

The spurious activation of pseudo-embedded words is far more serious. *Chess* in *chastise*, *deaf* in *daffodil*, *testicle* in *fantastical*, *as* in *residue*, *rag* in *regular*, *flag* in *phlegmatic* - there are scores of such cases for the /æ/-/ɛ/ confusion. We examined these in detail. One question we asked was whether embeddings like *as* in *residue* or *pen* in *span* are less of a problem than embeddings which preserve the syllabic structure of the matrix word (*as* in *esoteric*, *pen* in *panda*).

The /æ/-/ɛ/ embeddings are not equally common in either direction. 7090 spurious embeddings arise if /æ/ is perceived as /ɛ/ (*egg* in *fag* or *agriculture*, *stem* in *stamp* or *stampede*), while perception of /ɛ/ as /æ/ (*as* in *residue* or *esoteric*, *lass* in *bless* or *lesson*) yields nearly twice as many, namely 13658 cases. It might be argued that embeddings occur more often, for obvious reasons, in longer words, and longer words are encountered less frequently than shorter words so that perhaps embeddings do not cause much of a listening problem in reality. The extent of the problem can, however, be estimated by taking into account the frequency of occurrence statistics for the words which contain embeddings. CELEX contains frequency of occurrence statistics based on a corpus of 17.9 million words. These are statistics for frequency of occurrence in written text, so that they provide no direct estimate of the problem confronting the listener, but do suggest an upper bound. For misperception of /æ/ as /ɛ/ the frequency-adjusted count suggests 25631 spurious embeddings per million words, and for misperception of /ɛ/ as /æ/ 92284 per million.

These numbers are considerably reduced, as expected, if syllable boundaries of matrix and embedded word have to match. Embeddings which match reduce to 3636 cases for misperception of /æ/ as /ɛ/ (*egg* in *agriculture* but not *fag*) and 8054 cases for misperception of /ɛ/ as /a/e (*as* in *esoteric* but not *residue*). Frequency-adjusted estimates then suggest 14193 spurious embeddings per million words for misperception of /æ/ as /ɛ/ and 64198 per million words for misperception of /ɛ/ as /æ/.

We next computed the same statistics for the consonantal confusion /r/-/l/, again distinguishing embeddings which mismatch (*crow* in *clone*, *let* in *pretzel*) from embeddings which match the syllabic structure of the matrix (*crow* in *clothing*, *let* in *reticent*). Again the extent of the embedding problem is not symmetrically distributed in each direction. The spurious embeddings which would arise from perception of /l/ as /r/ (*crow* in *clone* or *clothing*) yield 15381 cases, while if /r/ is perceived as /l/ (*let* in *pretzel* or *reticent*) 1.65 times as many cases result, namely 25470. Frequency-adjusted estimates suggest 59079 spurious embeddings in every million words due to misperception of /l/ as /r/, and 108873 per million due to misperception of /r/ as /l/.

Application of a syllabic match constraint, however, again greatly reduces these numbers: by about 25% to 11458 cases (*crow* in *clothing* but not *clone*) for misperception of /r/ as /l/, and by nearly 40% to 15428 cases for misperception of /l/ as /r/ (*let* in *reticent* but not *pretzel*). Frequency-adjusted estimates then suggest 49508 spurious embeddings per million words for misperception of /l/ as /r/ and 69923 per million words for misperception of /r/ as /l/.

Although L2 listeners may be able to escape being troubled by embeddings which mismatch syllables of the input (*crow* in *clone, egg* in *fag*), a substantial number of spurious activated words clearly remain. Phoneme confusions can thus cause a substantial increase in lexical competition in L2 listening via spurious activation of embedded words.

5. Temporary ambiguity

5.1. Evidence from listening

A third problem that phoneme indiscriminability can cause for L2 listeners is temporary ambiguity. Even where no spurious embeddings are involved, a phonemic minimal pair which is confusable for L2 listeners could keep two alternative word candidates activated, where for L1 listeners the phonemic information is unambiguous, so that only one candidate remains. Thus if a listener hears *mal-* it is clear that this fragment could become *mallet, malady, malcontent,* so all of these words will be temporarily activated. Likewise, *mel-* could become *melody, mellow, melancholy*. But for a listener who cannot distinguish *mal-* from *mel-*, all six of these words may stay active at once.

Weber and Cutler [20] addressed this issue via a methodology eminently suited to the investigation of competitor activation, namely listening plus the recording of gaze. Since eye movements can be continuously recorded, it is possible to monitor comprehension as speech is heard, and hence evaluate relative competitor activation over time. Participants instructed to move objects in a display look to the correct object (e.g., *candy*) more slowly when there is another object with a phonologically similar name (e.g., *candle*) than when no such object appears [10]. Weber and Cutler's study again used the English vowel contrasts that are difficult for Dutch listeners. The experiments were conducted completely in English, and participants were unaware that their native language was relevant. They were instructed in spoken English to click on pictures of objects. The name of one distractor picture shared initial segments with the name of the target picture (for example, target *panda*, competitor *pencil*). The Dutch listeners showed clear evidence of interference; when they heard *pan-*, they were likely to look at the pencil. Native listeners presented with the same materials did not show such interference. This clearly demonstrates that temporary ambiguity is indeed another source of added competition for L2 listeners. They need to choose between words which for L1 listeners cause no competition, because they are clearly dissimilar.

5.2. Lexical statistics

To compute the lexical statistics underlying this problem we tallied the number of extra competitors remaining active at the point of a substituted phoneme. For instance, for every word with an /æ/ in it, if that /æ/ is substituted by an /ɛ/, how many words in the lexicon are there which are the same up to that point? Thus *abandon* would be truncated to /əbæ/ which would be retranscribed as /əbɛ/; the search would then deliver words beginning /əbɛ/, such as *abet*, *abed*.

For this calculation there is little point in computing the contribution of syllable boundary matches, since there is no evidence that syllabic structure plays a role in initial activation. The earlier the point of substitution, of course, the more competitors remain under potential consideration, and thus the averages across all words are quite high. Across the 7926 CELEX English words containing /ae/, replacement by /ɛ/ adds an average of 138.03 competitors per word. Replacement of /ɛ/ by /ae/ adds on average 135.77 competitors. The consonantal tallies are, as predicted, even higher: replacement of /l/ by /r/ adds on average 264.85 competitors, while replacement of /r/ by /l/ adds on average 305.04. These totals are, of course, heavily weighted by large numbers of competitors arising from replacements in early positions in the word – first or second phoneme. However, even later in the word there is still a substantial increase in competition, as can be seen from Table 1 which shows the time course of the added competition effect. Note again that this is added competition; the legitimate competitors (for *abandon*: *aback*, *abash*) should also play their role, just as for the L1 listener. The difference is that the L1 listener is not subject to the added competition.

Table 1: Mean number of added competitors for phoneme substitutions as a function of position in the word.

Phoneme:	1st	2nd	3rd	4th	5th
Substitution:					
/æ/->/ɛ/	583	174.99	27.39	3.09	1.39
/ɛ/->/æ/	903	177.99	32.75	2.13	1.01
/l/-> /r/	2412	349.12	9.80	2.61	0.76
/r/ -> /l/	1623	208.51	18.15	2.35	0.80

6. Conclusion

L2 listeners know that listening is harder in L2 than L1. Much of the problem occurs at the lexical level, when words which the L1 listener can exclude compete for recognition by the L2 listener because phonemic processing is not precise enough to exclude them. The statistics show that both vowel and consonant confusions result in an explosion of added competitors for the L2 listener.

7. Acknowledgements

This research was supported by the NWO-SPINOZA project "Native and non-native listening". Thanks to Dennis Pasveer for technical assistance.

8. References

[1] Cutler, A., Mehler, J., Norris, D. and Segui, J. "The syllable's differing role in the segmentation of French and English", *J. Mem. Lang.*, *Vol. 25*, 1986, 385-400.

[2] Weber, A. "Language-specific listening: The case of phonetic sequences", PhD dissertation. MPI Series in Psycholinguistics 16. University of Nijmegen, 2001.

[3] Cutler, A. and Otake, T. "Mora or phoneme? Further evidence for language-specific listening", *J. Mem. Lang., Vol. 33*, 1994, 824-844.

[4] Sorace, A. "Incomplete vs divergent representations of unaccusativity in near-native grammars of Italian", *Second Lang. Res., Vol. 9*, 1993, 22-47.

[5] Vanlancker-Sidtis, D. "Auditory recognition of idioms by first and second speakers of English", *Applied Psycholinguistics, Vol. 24*, 2003, 45-57.

[6] Akker, E. and Cutler, A. "Prosodic cues to semantic structure in native and nonnative listening", *Bilingualism: Language and Cognition, Vol. 6*, 2003, 81-96.

[7] Yamada, R. "Age and acquisition of second language speech sounds: Perception of American English /r/ and /l/ by native speakers of Japanese", in: *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*, ed. Strange, W. (pp. 305-320). York Press Inc., Timonium, MD, 1995.

[8] Sheldon, A. and Strange, W. "The acquisition of /r/ and /l/ by Japanese learners of English: evidence that speech production can precede speech perception", *Applied Psycholinguistics, Vol. 3*, 1982, 243-261 .

[9] Maddieson, I. (1984). *Patterns of sounds*, Cambridge University Press, Cambridge.

[10] Cutler, A., McQueen, J.M., Jansonius, M., and Bayerl, S. "The lexical statistics of competitor activation in spoken-word recognition", Proc. 9th Australian International Conference on Speech Science and Technology, Melbourne; pp. 40-45, 2002.

[11] Allopenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. "Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models", *J. Mem. Lang., Vol. 38*, 1998, 419-439.

[12] Zwitserlood, P. "The locus of the effects of sentential-semantic context in spoken-word processing", *Cogn., Vol. 32*, 1989, 25-64.

[13] Goldinger, S.D., Luce, P.A., Pisoni, D.B., and Marcario, J.K. "Form-based priming in spoken word recognition: The roles of competition and bias", *J. Exp. Psych.: Learning, Mem. & Cogn., Vol. 18*, 1992, 1211-1238.

[14] McQueen, J.M., Norris, D. and Cutler, A. "Competition in spoken word recognition, Spotting words in other words", *J. Exp. Psych.: Learning, Mem. & Cogn., Vol. 20*, 1994, 621-638.

[15] Norris, D., McQueen, J.M., and Cutler, A. "Competition and segmentation in spoken word recognition", *J. Exp. Psych.: Learning, Mem. & Cogn., Vol. 21*, 1995, 1209-1228.

[16] Baayen, H., Piepenbrock, R., & Van Rijn, H. "The CELEX lexical database" (CD-ROM), Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA, 1993.

[17] Pallier, C., Colomé, A., and Sebastián-Gallés, N. "The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries", *Psych. Science*, *Vol. 12,* 2001, 445-449.

[18] Cutler, A. and Otake, T. "Pseudo-homophony in non-native listening", *147th Meeting, Acoustical Society of America*, New York, May 2004.

[19] Broersma, M. "Phonetic and lexical processing in a second language", PhD dissertation. MPI Series in Psycholinguistics. University of Nijmegen, 2005.

[20] Weber, A. and Cutler, A. "Lexical competition in non-native spoken-word recognition", *J. Mem. Lang.*, *Vol. 50*, 2004, 1-25.