Lexical manipulation as a discovery tool for psycholinguistic research

Laurence Bruggeman[1,2] *and Anne Cutler*[2]
[1]Department of Linguistics, Macquarie University, Australia
[2]The MARCS Institute, Western Sydney University, Australia
laurence.bruggeman@mq.edu.au; a.cutler@westernsydney.edu.au

Abstract

Consultation of machine-readable dictionaries has advanced understanding of language processing; but these resources also allow examination of processing consequences if the lexicon changes. To recognise speech, listeners must rapidly evaluate spoken input as matching or mismatching candidate words. Listeners use any speech cues that help this process, whereby identical cues across languages may be used in one language but not in another. Suprasegmental stress cues, for example, are similar in Dutch and English, but used only in Dutch. This asymmetry has been explained as due to vowel reduction in English; lexical manipulation here tests this proposal and suggests a refinement.

Index Terms: lexicon,stress,word recognition, English, Dutch

1. Introduction

Electronic versions of complete dictionaries have now been available to language researchers for more than three decades. Their availability revolutionised both the design of automatic speech recognition systems and the modelling of language processing by human speakers and listeners. For example, an entire class of early spoken-word recognition models [1,2] that was based on strictly sequential recognition of words in their order of arrival was rendered untenable by the discovery that most English words cannot be uniquely recognised until at or after their final sounds [3]; only the automatically searchable dictionary resources made the latter discovery possible.

In the intervening decades, psycholinguists have made good use of such dictionary tools, which have become available for a steadily increasing number of languages. The structure of vocabularies is well understood and differences between vocabularies have been easy to measure, enabling their processing implications to be derived and tested. Thus the makeup of language phoneme repertoires determines average word length in dictionaries of that language [4] and predicts that this will carry through to everyday speech experience, as is confirmed in standard spoken samples [5].

However, electronic dictionaries offer researchers the opportunity of going beyond the usual activities involved in dictionary use – consulting re a single item, searching for the full extent of classes of items, analysing overall structure. Electronic dictionaries can also be deliberately altered.

It may seem as if altering a dictionary would be a pointless exercise; it can, after all, have no effect on how users of a language actually deploy their vocabulary resources. But lexical manipulation in fact provides a tractable and effective way of testing psycholinguistic explanations or predictions in which vocabulary structure plays an operational role. Section 2 describes two case studies of research in which the lexical manipulation method was used to illuminate differing questions, and in Section 3 we apply the technique to a new case involving a cross-language processing asymmetry.

2. Lexical Manipulation

2.1. Phoneme substitution

Lexical manipulation can shed light on effects of phonemic confusion, in particular distinctions that prove intractable for second-language listeners (such as the English contrast in, e.g., *write*, *light* for many listeners with an Asian native language, or in *cattle*, *kettle* for listeners whose first language is Dutch or German). These two phoneme confusions were at issue in [6], in which the English lexicon was examined from the point of view (or hearing) of second-language users for whom these distinctions were not perceptible in word-initial or word-medial position.

The manipulation in [6] thus involved treating pairs like *write*/*light* or *cattle*/*kettle* as homophones, and replacing [r] by [l] and [æ] by [ɛ], consistent with independent evidence of the direction of these mergers [7,8]. The results revealed the greatest effect of phonemic confusion to be located not in whole-word substitutions (despite their salience to language users) but in increase in the two major measures of inter-word competition: embedding (spuriously, *egg* in *agriculture*, *let* in *reticent*) and overlap (*lemon* and *remedy* or *matter* and *metal* are heard as starting with the same syllable). Moreover, the effect of a consonantal confusion was significantly greater than that of a vowel confusion, and confusion direction was asymmetric: listeners with [r]/[l] difficulty would actually have less trouble collapsing English [r] and [l] to [r], rather than to [l]! These findings suggest lines of speech perception research and also possible avenues for listening training.

2.2. Phonological rule substitution

The lexical manipulation technique has also been deployed to elucidate different patterns of embedded-word location across languages [9]. These analyses (without lexical manipulation) showed that the predominance of word-initial over word-final location of embedded words (such as *cat* or *log* in *catalogue*), long known for English, was replicated even more strongly in the related languages Dutch and German, but did not appear in Japanese. Japanese has neither suffixes nor stress, while all the other languages have both. The French vocabulary, which has one of those two features (suffixes) but not the other (stress), fell in between the Japanese and English values, thus suggesting that both factors played a role. This was tested in [10] by a radical form of lexical manipulation, which has been termed "lexical miscegenation", since it effectively creates a hybrid of separate varieties. The French lexicon was augmented with pronounced schwa in all possible legal positions (giving *fille* two syllables and *petite* three). Its embedding pattern then came to closely resemble that of English and other stress languages. The contribution of stress to embedding patterns is thus based on schwa distributions.

3. Explaining asymmetry in speech cue use

3.1. Speech cue use

It is a remarkable fact that listeners do not always make optimal use of speech cues in recognizing spoken utterances. A striking example in phoneme recognition concerns the use of transitional cues to identify fricative sounds, which only happens when fricatives must be distinguished from highly similar alternatives [11,12]. If a sound has no close competitor sounds, available information in the signal may prove redundant and thus be ignored. The explanation of this pattern thus invokes language-specific phoneme repertoires.

At the word recognition level, similar asymmetries are found. The process of word recognition involves sorting out the intended words in a heard utterance from among the alternative candidate words that are embedded (i.e., fully supported) or overlapping (i.e., partially supported) in the speech signal. If there is one thing that is well known about spoken-word recognition, it is that listeners do not wait around to hear utterances or words or even syllables in full before attempting to recognize them; they constantly consider alternative interpretations and weigh the continually incoming evidence in the signal in terms of whether it offers support for, or counts against, the current options. Spoken *Speech Science and Technology* may activate the embedded words *pea/bee*, *peach/beach, each*, *sigh*, and *knowledge*, the cross-word embeddings *sand* or *antic*, and many temporarily supported candidates such as *speed*, *sign*, *text*, or *echo*. Such words are effortlessly discarded by listeners, though traces of their temporary presence can be discerned in psycholinguistic experiments. For instance, in the cross-modal priming task, where listeners make yes-no lexical decisions about visually presented words while hearing speech, words are recognised more slowly when they partially match the auditory input than when the auditory input is totally unrelated (e.g., responses to visual *feel* are slower after spoken *feed* than after spoken *name*; [13]). This response inhibition represents the temporary availability but later rejection of an alternative interpretation (*feel*) of the speech input. Because the word has been rejected it is momentarily less available to the lexical decision process.

With this same task, cross-language differences have appeared in the use of stress cues in spoken-word recognition. In English, for instance, there are minimal pairs of words that differ only in stress, such as *INsight*/*inCITE* (henceforth, upper case denotes a primary-stressed syllable). Dutch also has such minimal pairs, e.g., *VOORnaam* 'first name' vs. *voorNAAM* 'respectable'. English and Dutch are closely related languages, both having free lexical stress realised under highly similar rules [14]. Yet in cross-modal priming, if the mismatch involves only stress, English listeners do not show the inhibition that indicates a word candidate's rejection. Thus the spoken initial part of *admiRAtion*, *admi-*, does not inhibit responses to visual *admiral* (initial stress: *ADmiral*; [15]). In Dutch, in contrast, such mismatch indeed leads to inhibition – e.g., responses to visually presented *dominee* 'pastor' (pronounced *DOminee*) are slowed after *domi-* from *domiNANT* 'dominant'[16]. It seems that in these two very similar languages, in which stress is realised with virtually identical suprasegmental cues, Dutch listeners use the cues but English listeners don't (Fig. 1). The English result is only one of many demonstrations (in multiple tasks: recognition in noise [17]; acceptability judgment [18]; phoneme detection [19]; goodness rating [20]) of English listeners ignoring suprasegmental cues to word identity.

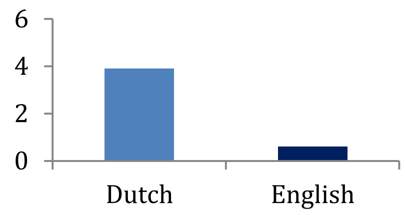

Figure 1 (data from [16] and [15]): *Dutch and English cross-modal priming: Inhibition due to mismatching stress (RT given mismatching prime minus RT given control prime, as % of control condition RT). The effect is highly significant in Dutch, insignificant in English.*

3.2. Lexical asymmetry as explanation

Lexical statistics suggest why English listeners should behave in this way. An estimate of the strength of competition from co-activated candidate words in speech recognition can be derived by tallying word embedding and overlap in the vocabulary, and this can be done taking only segmental structure into account, or stress pattern as well. In *enterprise*, for instance, *enter* and *prize* are embedded, and *settee* has *set* and *tea*. But if we require primary stress location to match also, *ENterprise* contains only *enter*, and *setTEE* contains only *tea* (neither *set* nor *prize* have primary stress there). Dutch examples are *OUderdom* 'old age' (*ouder* 'parent' and *dom* 'stupid', or only *ouder*) and *karPET* 'rug' (*kar* 'cart' and *pet* 'cap', or just *pet*). Dutch embedding competition reduces by more than 50% if stress match is indeed required in this computation [21]. There is a smaller (but also significant) reduction in English. But if the numbers are weighted by carrier word frequency to estimate actual competition in natural speech, a Dutch segments-only count gives 1.52 competitors per word of speech on average, while a segments-plus-stress count reduces this to .74. The equivalent numbers for English are .94 and .59 [21]; see Fig. 2. This is a quantum improvement for Dutch (from more than one to under one) but not for English (under one to a bit further under one).

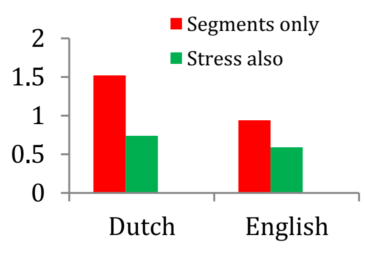

Figure 2 (data from [21]): *Dutch and English: Estimated number of competing words per word of speech, given match in segments only, or in segments and stress.*

One possible explanation lies in the fact that English syllables *without* primary stress often have a particular vowel, namely schwa. English listeners need attend only to segments in computing mismatch to competitors, as suprasegmental cues do not reduce competition sufficiently to make recourse to suprasegmental processing mechanisms worth the effort.

3.3. Dutch and English statistics

We test the hypothesis that schwa distribution might be the cause of this asymmetry (as it was the cause of the embedding asymmetry examined in [10]), by analysing the CELEX database [22]. This contains both English (British version) and Dutch lexicons. These closely related languages share many phonological attributes (e.g., both allow complex onsets and codas and have short vowels as well as long vowels and diphthongs). The phoneme inventory of the two languages differs somewhat; English has more consonants than Dutch and also more vowels, but each is a member of the class of languages with a rather high number of both phoneme types.

CELEX's lexicon of Dutch, as it happens, is much larger than its English lexicon. The Dutch lexicon contains more than 124,000 lemmas, and frequency statistics based on a 42 million word corpus; the English lexicon contains 52,000 lemmas, and frequencies from a 17.9 million word corpus. These differences – in the size of the frequency count corpora especially – largely arise from differences in the original sources from which the database was compiled. However, the differences in lexicon size also in part reflect cross-language morphological differences. Because of the size differences, absolute totals are obviously not directly comparable. Therefore our calculations are based on proportions.

We first assessed whether Dutch and English in fact differ in the overall frequency of schwa. Both languages have stress, both have vowel reduction, and both have extensive affixation with affixes (both suffixes, especially inflectional suffixes, and prefixes) typically realised by weak syllables containg schwa. Here the morphological differences between the two languages would in fact tend to bias the frequency of schwa towards more in Dutch, since Dutch word formation both uses prefixing more extensively than English (e.g., *gevoel* 'feeling', *geval* 'case', *geluk* 'luck' and many more, with the syllable *ge-* containing schwa) and includes a larger range of verbal inflexions including both prefixes and suffixes (thus where English *walk* adds the forms *walks*, *walking* and *walked*, Dutch *lopen* 'walk' adds *loopt*, *loop*, *liep, liepen*, *gelopen, lopend, geloop*). We thus initially computed the proportion of full vowels versus schwa in different word positions in each lexicon. For this count we included lemmas (base word entries) only, since including all word forms, especially in the case of verbs, leads to a greater size asymmetry between the lexicons and, for the reasons listed above, hugely increases the number of syllables with schwa, particularly in final position. Using lemmas still leaves morphologically caused asymmetries, since in fact all verb infinitives of Dutch (e.g., *lopen*) are marked as such by a final *–en* (with schwa). Similarly, a very large number of Dutch nouns derived from verbs begin with the prefix *ge-* which is the past participle inflection (e.g., *gebouw* 'building').

Indeed, even this lemma-based count revealed that Dutch final syllables are somewhat more likely to contain schwa than English final syllables, and while the proportions for initial syllables were fairly close for the two languages, there was a slight predominance of the schwa count in Dutch there too. Thus neither in initial nor in final syllables did we find support for greater frequency of schwa in English.

Medial syllables, however, patterned differently. Here the languages tend towards opposite patterns, as Figure 3 shows. In all word lengths with medial unstressed syllables (3 syllables and above), Dutch has more full vowels in such syllables (and English has more schwa).

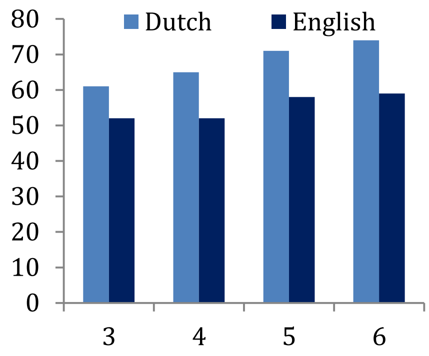

Figure 3: *Dutch and English: Percent word-medial syllables containing a full vowel, for words of 3 to 6 syllables. The equivalent figure for word-medial syllables containing schwa is the reverse of this (English always the higher value).*

We next assessed whether Dutch and English differ in the relative amount of competition offered by the makeup of their respective vocabularies. We tallied for each phoneme, the number of words beginning with that phoneme; for each syllable, the number of words beginning with that syllable; and for each bisyllabic string, the number of words beginning with that bisyllable. Given the size asymmetry between the two CELEX lexicons, it was necessary to adjust these results for the relative number of words involved. For the phoneme measure and the first-syllable measure we found no significnt cross-vocabulary difference. Thus there is (as expected) no tendency for either language to have, in principle, more inter-word similarity and hence a greater amount of competition. The two-syllable measure revealed that English in fact had more such overlap in its vocabulary. This was indubitably due to the prevalence of schwa; consider that English *coral*, *correlate*, *corridor*, *coroner*, *corrugated* and *coryphee*, despite being all spelled differently, begin in each case with the same CVCV sequence, the constant second V being schwa.

3.4. Stress cue substitution

The calculations so far do not suggest that Dutch listeners use suprasegmental cues to recognise words because their lexicon confronts them with more onset-overlapping competitor pairs; on the contrary, English presents the more difficult listening task in this respect. This renders the next step in this project even more interesting. We now undertake some lexical manipulation in order to shed further light on this issue; the manipulation involves moving one of these two lexicons in the direction of the other (a form of miscegenation, as in [10]).

Here we can either choose to make Dutch more English-like (having a greater proportion of schwa, especially in medial position), or English more Dutch-like (having a greater proportion of full vowels in medial position) – or indeed both, though such duplication would of course deliver no additional information. The replacement of full vowels by schwa, a many-to-one mapping, is the simpler calculation. The replacement of schwa by full vowels, in contrast, is a one-to-many mapping, so inherently more difficult; furthermore, it raises difficult issues of phonological principle in the choice of mapping (e.g., should the schwa in *coral, correlate*, *corridor*, etc. be replaced by the same full vowel, or by different ones?). For these reasons we chose the simpler option and undertook an alteration of the CELEX Dutch lexicon.

To obtain an estimate of embedding frequency, we altered the proportions of schwa/full vowel in polysyllabic Dutch words to bring the schwa distribution in this lexicon in line with that in the English one, and computed embedding statistics analogous to those reported by [21] and summarised in Figure 2. Compared with the unmodified Dutch lexicon, embedding in this altered lexicon was reduced by almost 10%. To further obtain a measure of overlap competition, we selected the two-syllable overlap set computed for Dutch and described above (section 3.3), and compared it across the original unmodified lexicon and the altered lexicon. Again, this led to a reduction of competition in the latter, albeit this time only by 3% if stress location was ignored and by 2.5% if competitors had to match in the location of primary stress.

These two consistent outcomes indicate that the Dutch lexicon would contain fewer competitors if more use was made of the optional reduction to schwa provided by the phonological rules of the language; by implication there is thus support for the use of suprasegmental stress cues where they can assist in narrowing lexical competitor sets. This is in line with one part of the proposed explanation of why Dutch listeners use such stress cues in spoken-word recognition while English listeners ignore them, though it does not directly address English listeners' choices. The English lexicon, in part because of the predominance of schwa in medial syllables, confronts listeners with more competition. Schwa in English or Dutch cannot bear stress, and as a result of this, English listeners actually have very few opportunities to profit from suprasegmental stress cues to resolve competition.

4. Conclusions

The structure of vocabularies determines both what listeners have to recognise to understand spoken utterances (stand-alone words, agglutinative particles, polysynthetic sequences?) and also what speech cues are necessary and worthwhile to attend to in the task. In the case of suprasegmental cues to the distinction between primary and secondary stress, abundant empirical evidence shows that Dutch listeners make good use of the information they provide, but English listeners do not. The cues are acoustically similar in both languages, and Dutch listeners are capable of using the English cues even to the point of outdoing native listeners in an identification task [15], but English listeners fail to attend to them [17-20].

Lexical manipulation has supported the suggestion that English vowel reduction plays a role in this pattern. However, our results overall counter-indicate a larger claim that vowel reduction alone is responsible, because overall, and in initial or final syllables, Dutch has a greater proportion of syllables with schwa than English. More can be discovered here, and in particular we note again the higher proportion of morphological affixes in the Dutch lexicon. CELEX does not provide sufficient information for us to calculate exactly the relative proportion of affixes with schwa, but it does indicate an upper and a lower bound: English has a lower proportion of affixes with schwa in the lexicon than Dutch does (75% of the Dutch level at most, but at least 23.5%). Future work could address the processing implications of this asymmetry and its potential role in influencing word recognition processes.

We strongly encourage other researchers to join us in this view of dictionary resources: dictionaries are not holy untouchables, they are alterable objects that can serve as tools to discover novel potential patterns and hence to prompt new and innovative empirical studies of language processing.

5. Acknowledgements

Support is acknowledged from the ARC Centres of Excellence in Cognition and its Disorders (LB) and for the Dynamics of Language (AC) as well as from the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands (AC, LB).

6. References

[1] Cole, R. and Jakimik, J., "Understanding speech: How words are heard", in G. Underwood [Ed.], Strategies of information processing, 67-116, Academic Press, 1978.
[2] Marslen-Wilson, W. and Welsh, A., "Processing interactions and lexical access during word recognition in continuous speech", Cog. Psychol., 10: 29-63, 1978.
[3] Luce, P., "A computational analysis of uniqueness points in auditory word recognition", Perc. Psychophys., 39: 155-158, 1986.
[4] Cutler, A., Norris, D. and Sebastian-Galles, N., "Phonemic repertoire and similarity within the vocabulary. Proc. INTERSPEECH 2004, Jeju, Korea; vol. 1, 65-68, 2004.
[5] Maddieson, I., "Word length is (in part) predicted by phoneme inventory size and syllable structure", J. Acoust. Soc. Am., 132: 2218, 2016.
[6] Cutler, A., "The lexical statistics of word recognition problems caused by L2 phonetic confusion", Proc. INTERSPEECH 2005, Lisbon, 413-416, 2005.
[7] Iverson, P. et al., "A perceptual interference account of acquisition difficulties for non-native phonemes", Cogn., 87, B47–B57, 2003.
[8] Broersma, M., "*Kettle* hinders *cat*, *shadow* does not hinder *shed*: activation of 'almost embedded' words in nonnative listening", Proc. INTERSPEECH 2007, 1893-1896, 2007.
[9] Cutler, A., Otake, T. and Bruggeman, L., "Phonologically determined asymmetries in vocabulary structure across languages", J. Acoust. Soc. Am., 132: EL155-EL160, 2012.
[10] Cutler, A. and Bruggeman, L., "Vocabulary structure and spoken-word recognition: Evidence from French reveals the source of embedding asymmetry", Proc. INTERSPEECH 2013, Lyon, 2812-2816, 2013.
[11] Wagner, A., Ernestus, M., and Cutler, A., "Formant transitions in fricative identification: The role of native fricative inventory", J. Acoust. Soc. Am., 120: 2267–2277, 2006.
[12] Wagner, A., "Cross-language similarities and differences in the uptake of place information", J. Acoust. Soc. Am., 133: 4256-4267, 2013.
[13] Marslen-Wilson, W., "Activation, competition, and frequency in lexical access", in G.T.M. Altmann [Ed.], Cognitive Models of Speech Processing, 148–172, MIT Press, 1990.
[14] Trommelen M. and Zonneveld, W., "Word stress in the West Germanic languages", in H. van der Hulst [Ed.], Word Prosodic Systems in the Languages of Europe, 478-515, Mouton, 1999.
[15] Cooper, N., Cutler, A. and Wales R., "Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners", Lang. Speech, 45: 207–228, 2002.
[16] Van Donselaar W., Koster, M. and Cutler, A., "Exploring the role of lexical stress in lexical recognition", Quart. J. Exp. Psy., 58A: 251–273, 2005.
[17] Slowiaczek, L., "Effects of lexical stress in auditory word recognition", Lang. Speech, 33: 47-68, 1990.
[18] Slowiaczek, L., "Stress and context in auditory word recognition", J. Psycholing. Res., 20:465-481, 1991.
[19] Small, L., Simon, S. and Goldberg, J.S., "Lexical stress and lexical access: Homographs versus nonhomographs", Perc. Psychophys., 44:272-280, 1988.
[20] Fear, B., Cutler, A. and Butterfield, S., "The strong/weak syllable distinction in English", J. Acoust. Soc. Am., 97:1893-1904, 1995.
[21] Cutler A. and Pasveer D., "Cross-linguistic differences in effects of lexical stress on spoken-word recognition", Proc. 3rd Int. Conf. Speech Prosody, Dresden, 237–240, 2006.
[22] Baayen, H., Piepenbrock, R. and van Rijn, H., The CELEX lexical database. Philadelphia: Linguistic Data Consortium, University of Pennsylvania (CD-ROM), 1993.