Listener adjustment of stress cue use to fit language vocabulary structure

Laurence Bruggeman[1], *Jenny Yu*[1] *and Anne Cutler*[1]

[1]The MARCS Institute for Brain, Behaviour & Development, Western Sydney University, Australia
l.bruggeman/jenny.yu/a.cutler@westernsydney.edu.au

Abstract

In lexical stress languages, phonemically identical syllables can differ suprasegmentally (in duration, amplitude, F0). Such stress cues allow listeners to speed spoken-word recognition by rejecting mismatching competitors (e.g., unstressed *set-* in *settee* rules out stressed *set-* in *setting, setter, settle*). Such processing effects have indeed been observed in Spanish, Dutch and German, but English listeners are known to largely ignore stress cues. Dutch and German listeners even outdo English listeners in distinguishing stressed versus unstressed English syllables. This has been attributed to the relative frequency across the stress languages of unstressed syllables with full vowels; in English most unstressed syllables contain schwa, instead, and stress cues on full vowels are thus least often informative in this language. If only informativeness matters, would English listeners who encounter situations where such cues would pay off for them (e.g., learning one of those other stress languages) then shift to using stress cues? Likewise, would stress cue users with English as L2, if mainly using English, shift away from using the cues in English? Here we report tests of these two questions, with each receiving a yes answer. We propose that English listeners' disregard of stress cues is purely pragmatic.

Index Terms: lexical stress, suprasegmental cues, spoken-word recognition, English, German, Dutch

1. Introduction

Recognising speech requires listeners to parse an incoming continuous stream of sound into component parts that correspond to stored forms – spoken words – ordered in an expected pattern (a meaningful and grammatically regular utterance). Every part of this operation involves the listener in mind-bendingly complicated consideration of multiple options to arrive at decisions at each necessary (phonological, syntactic, semantic) descriptive level. As the research reported here will show, the listener's calculations include vocabulary-specific estimates of the likely payoff involved in invoking individual sources of potential information.

The stored word forms in any user's vocabulary contrast in vowels and consonants (*pot/got/goat/goad*), but also, most importantly for the present convention, in their prosody. In lexical stress languages, for example, one syllable in each word bears the primary stress. In principle, two words can share all their phonemes and contrast only in stress (such as *GOAty/goaTEE*; N.B., upper case signals primary stress). In practice, though, such pairs are quite uncommon in all lexical stress languages, which might suggest an asymmetry in the contribution of phonemic and prosodic information to lexical identity – perhaps unsurprising in that such asymmetry also follows from the fact that prosodic contrasts require more than one syllable, so that they are irrelevant to a large part of any vocabulary. Not only do many vocabularies contain large numbers of monosyllabic words, a high proportion of multi- (especially bi-) syllabic words also consist of a stem plus a morphological affix, the latter hardly ever carrying stress. Consequently, for sets such as *potted/potter/potting*, lack of stress on the suffix may not be crucial for identifying each word.

Lexical recognition studies have revealed that the role of lexical stress varies in importance across languages. Spanish listeners, for example, can distinguish *PRINcipe* vs. *prinCIpio* before the end of the first syllable [1]; similar results appear with Dutch, German, Turkish, and Italian [2-5] (and possibly many more languages). English, though, is not a language in which listeners make use of stress cues. Decades of research have shown English listeners to pay little attention to the cues that differentiate a syllable with primary versus secondary stress [6-9]. For instance, they have difficulty with (and even underperform Dutch or German listeners in) classifying an isolated syllable (e.g. *goa-* from *goatee*) as stressed or not [10, 11], and find sets of words such as *autoMAtion, auTOMata* and *AUtumn* equally acceptable (a) as originally uttered or (b) with the initial syllables interchanged [12].

Lexical statistics reported at an earlier Speech Prosody meeting [13] have provided a potential explanation for the different results across languages (even across closely related and phonologically similar languages, as in the case of Dutch, German and English). This explanation involves the relative number of potential competitor words (i.e., words that cannot yet be ruled out by the listener engaged in lexical recognition). When suprasegmental cues to stress position are considered, this number significantly declines in Spanish and in Dutch and in German (by at least half, often more). By comparison, the reduction for English is much smaller (no more than one-third). In other words, there is not a significant reward for English listeners which would repay them for the increased processing load of attending to suprasegmental information!

Structural comparisons in the lexicon showed that the amount of influence on competitor numbers was primarily due to the relative frequency of syllables with full vowels but without primary stress; there are significantly fewer such syllables in English than in the other languages tested [14].

This picture suggests that English listeners' behaviour is rationally based, a view that we test further in the present study. If indeed informativeness is essentially the sole factor discouraging suprasegmental processing of English words, then with other words which would deliver a higher payoff (e.g., German words), such processing by English listeners may be quite possible. In Experiment 1 we test this question on a listener population of English-native users of L2 German.

Another view of the same issue is offered by users of L2 English whose L1 is one of the stress languages in which using stress cues pays off (e.g., Dutch). Do long-term users of L2 English ever come to realise that those cues (so handy in their L1) here deliver no rewards, and therefore actually stop using them? We test this second question in Experiment 2.

2. Experiments

2.1 Experiment 1

2.1.1 Participants

Twenty-one native English speakers with L2 German (M_{age} = 36.5 years; 12 females), recruited in Sydney, took part in return for a small payment. All provided written informed consent; none reported any language use issues. They took the German LexTALE test [15], a short lexical decision task designed to provide a measure of language proficiency, and received an average score of 69.27% (*SD* = 10.20), which is significantly above chance but well below L1 average scores.

2.1.2 Materials and Procedure.

This experiment used the tokens presented to German listeners in [11] which had been selected as follows. First, 36 bisyllabic German word pairs such as *Konto-Konzept* were selected from the CELEX database [16]. Both words in each pair had phonemically identical first syllables (e.g., *Kon-*) but differed in primary stress location: one member was stressed on the first syllable (e.g., *KONto),* the other on the second syllable (e.g., *KonZEPT).* In a pretest, these 72 words, with 48 other German words varying in frequency, were rated for familiarity by 11 native German speakers, using a 7-point scale; 24 pairs (listed in the Appendix) were selected, with mean familiarity rating for first and second syllable stress words of respectively 6.22 (*SD* = 0.58) and 6.19 (*SD* = 0.47). Mean log frequency was 1.07 (*SD* = 0.66) and 1.12 (*SD* = 0.53) respectively. Two instances of each word in isolation were recorded by a female native speaker of German with no knowledge of the study. Each word was then truncated at the offset of the first syllable (based on visual inspection of formant boundaries in spectrograms). These initial syllables (96 fragments in all: 24 pairs x 2 source words x 2 productions) became the tokens presented to listeners in Experiment 1.

Presentation software [17] was used to present stimuli and to record responses, with two occurrences of each fragment included (giving 192 tokens in all). Participants were tested individually in a quiet room. They listened to a word fragment over headphones; immediately following fragment offset, a response pair appeared in the centre of the screen in capital letters (e.g., KONTO-KONZEPT); participants were asked to select the matching fragment using the left and right “Shift” keys. A new trial commenced immediately after a response or after 5 sec without response. Items were pseudo-randomised per participant (the same word pair never twice in a row), and left and right positioning of the word options were counterbalanced. Two practice trials (and the option of asking questions) preceded the start of experimental trials.

2.1.3 Results and discussion

Seventy-four trials without response were excluded from the analysis. One word pair (*damals-damit*) was also excluded due to the possibility of alternative pronunciations. The remaining dataset comprised 3790 responses.

Mean proportion accuracy across stress conditions is shown in Figure 1. Fragments from both first- and second-syllable stress words were correctly judged significantly above chance (first: mean accuracy 66.17, $p < .001$, second: mean accuracy 65.55, $p < .001$). In a separate analysis we directly compared the performance of these English listeners to that of the German listeners tested on the same stimuli in [11]. Those earlier results are also included for comparison in Figure 1. This analysis showed no main effect of group (Wald $\chi 2$ (1, N=42) = 3.09, p = 0.08), i.e., accuracy did not significantly differ across groups, and as the figure shows, both groups scored above 50% on each stimulus type. Thus our English listeners, learners of German as L2, achieved a success rate comparable to that of the native German listeners in [11].

The results of Experiment 1 thus reveal that the ability to distinguish between segmentally identical syllables differing only in whether they bear primary or secondary stress remains fully intact in English-native listeners, even though they do not in general make use of this ability when hearing their L1.

2.2 Experiment 2

2.2.1 Participants

Twenty participants were recruited from the Dutch emigrant community in Sydney (M_{age} = 48.8 years; 14 females). All were native Dutch speakers, had grown up in the Netherlands and had migrated to Australia as adults ($M_{age\ at\ migration}$ = 28.4 years; $M_{time\ since\ arrival}$ = 20.5 years). All reported regular use of both Dutch (L1) and English (L2). None reported any language use issues. All provided written informed consent and received a small payment for participating.

2.2.2 Materials and Procedure

The stimulus materials used (also listed in the Appendix) were from Experiment 3 of [10]. They consisted of recordings of 21 pairs of bisyllabic English words, spoken by a male native speaker of Australian English. As with the German word pairs of Experiment 1, words in each pair had phonemically identical first syllables, always with full vowels, but differed in either first or second syllable primary stress (e.g., *RObot, roBUST*). Mean log word frequencies from [16] as reported in [10] were 2.18 for first-syllable stress words, 1.88 for second-syllable stress words. Each word was recorded twice; all spoken tokens were truncated at the end of the first syllable, giving 84 fragments. These were each presented twice (168 trials), with fragments from a given word pair never occurring in successive trials. As in Experiment 1, a separate pseudo-randomised stimulus list was created for each participant.

Procedure was as in Experiment 1 except that the testing space was a sound-attenuated booth, instructions were orally clarified (in Dutch) by the first author, the response words were displayed already while the word fragment was presented, and all trials started 500 ms after a response to the preceding trial had been received.

2.2.3 Results and discussion

One trial had a response time of less than 100 ms and was excluded from all analyses. Overall, the emigrants correctly identified the source word for 61.9% of truncated fragments, with first-syllable stress fragments receiving higher scores (72.3%) than second-syllable stress fragments (51.5%). Note that first-syllable stress is the most frequently occurring stress pattern in English [18, 19], so that this asymmetry may reflect an overall tendency to select the response option with first-syllable stress more often than the alternative. Indeed, a word with first-syllable stress was chosen in 60.4% of all trials here, similar to the 62.9% of all choices by the L1 listeners tested by [10] with these same materials.

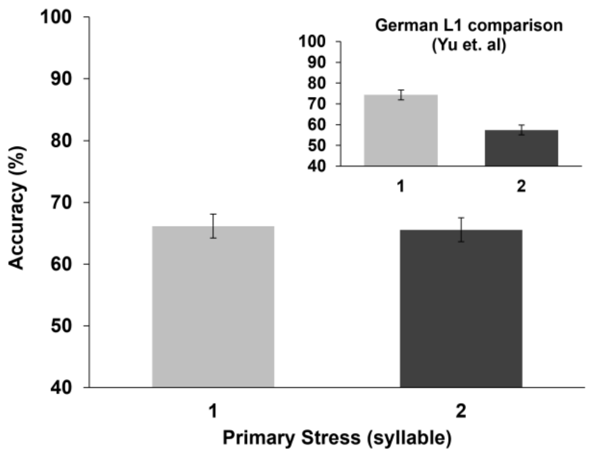

Figure 1. *Experiment. 1: Mean accuracy of identification of initial fragments from German words with first- and second-syllable primary stress by English-native L2 German listeners. Mean accuracy of native German listeners on the same task (from [11]) is included above right.*

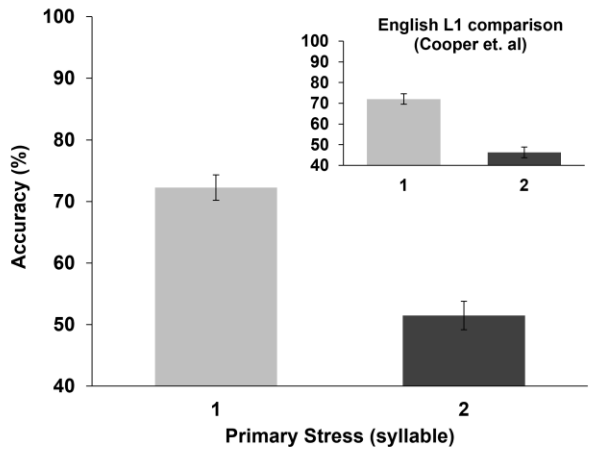

Figure 2. *Experiment 2: Mean accuracy of identification of initial fragments from English words with first- and second-syllable primary stress by Dutch emigrants. Mean accuracy of native English listeners on the same task (from [10]) is included above right.*

The emigrants' results are pictured in Figure 2, with, again for comparison, the results of that earlier group of Australian L1 listeners of English [10]. The identification accuracy of each of the two groups was statistically compared by fitting a generalised linear mixed-effects model to the combined data, carried out in R [20] using family 'binomial' and the logit-link function of the lme4 package [21]. This analysis showed that the emigrants' accuracy and the accuracy of the earlier English-speaking group did not significantly differ (p = .58).

We then compared the emigrants' response accuracy to chance level (i.e., 50%) with a two-sided binomial test. Since the potential bias towards first-syllable-stress responses prevents a meaningful interpretation of participants' accuracy for fragments with that stress pattern, this comparison was only carried out with participants' judgments for items with second-syllable stress. While Dutch listeners perform this task significantly better than chance [10], this was not the case for the emigrants, who performed neither better nor worse than chance level (z = 1.34, p = .181).

The Experiment 2 results show that the Dutch emigrants do not exploit suprasegmental information when listening to English in the way that is typical of Dutch L2 listeners living in the Netherlands. Their use of this information is essentially indistinguishable from that of native Australian listeners. Thus extended daily L2 use appears to have enabled the emigrants to adjust the way they listen to fit the properties of the English lexicon, presumably to optimise processing efficiency.

3. Conclusions

Each of our two predictions has received a positive answer. Experiment 1 showed that English listeners (a population who, in many studies, have been shown to leave stress cues unprocessed when listening to their native language) nevertheless have not lost the ability to exploit suprasegmental information if they are so inclined; yes was the answer to the question of whether they can use the cues in German when German is their L2. Confronted with a lexicon in which competitor numbers prove to be usefully reduced by including stress cue analysis in the recognition process, these non-native listeners indeed adjust their L2 recognition accordingly.

Likewise, in Experiment 2 we have received a yes answer to our question for Dutch listeners, who are accustomed to using suprasegmental stress cues when they are listening to speech because this is a useful strategy in processing input in their L1. We now know that such listeners will, with sufficient experience of English on a regular basis, come to realise that the use of such cues in English is not really furthering their goal of understanding English as fast as possible. That realisation appears to prompt them to drop stress-cue analysis from their recognition schedule, at least with input in this L2.

Each of these findings supports the proposed explanation of the puzzle confronting prosody research since the 1980s; why do English listeners not use the suprasegmental cues which clearly mark stress location in English words [6-9], while essentially the same cues that signal stress location in Dutch [3], German [4], Spanish [1], and Turkish [5] are fully exploited in word recognition by native speakers of those languages? It is not the cues themselves that somehow are less effective in the English vocabulary; L2 users whose native language is Dutch [10] or German [11] presented with the present experimental task in English succeed in exploiting the information in the English cues very efficiently.

Rather, the explanation that was first proposed by [13] draws on the vocabulary as a whole and the competitive nature of the word recognition task. When we hear speech, potential words are activated in a constantly changing progression as the incoming speech input rules some candidates out or opens the list to more potential options. The input consists primarily of segmental information (concerning which phonemes have been heard) that acts to reject candidates. Note that aspects of the lexicon's structure will affect the competition process even considering the segmental stream alone. For instance, the size of a language's phoneme inventory is important, because with many phonemes in the inventory the average word length is shorter, while a more restricted phoneme inventory necessarily means that words have to be longer. Such phoneme repertoire differences then also affect competitor numbers [22].

But as we have shown in the case of lexical stress, competition issues can also reflect the vocabulary beyond the segmental sphere. At least in the languages named above, the input speech stream offers suprasegmental information about word identity, elaborating upon the segmental identifications.

Here is where there are major differences across the vocabularies; they vary in the likelihood of vowel reduction in unstressed syllables [14]. Where this is very common (i.e., in English), the inevitable result is fewer cases of competition between words containing sequences of full vowels, and, in consequence, fewer cases where attention to suprasegmental information could helpfully reduce competitor populations during spoken-word recognition. These resulting differences in informativeness suffice to lead listeners to use, or not to use, the signals of stress placement.

Note that the role of stress in English word recognition is not only less useful for listeners in that minimal pairs such as *discount* or *forbear* as noun vs. verb are extremely rare, but also by implementation of stress-sensitive phonetic effects. Stress placement not only alters vowel pronunciations but licenses other realisation effects and thus distinguishes words phonetically as well as prosodically, i.e., vowel reduction in unstressed syllables affects a word's consonants as well. Thus an intervocalic consonant followed by a fully reduced vowel can become ambisyllabic rather than syllable-final or syllable-initial; and the same position preceding an unstressed (though not necessarily fully reduced) vowel allows a consonant to be uttered in a casual-speech form (e.g., in US English, the inter-vocalic /t/ in *goaty*/*goatee* may be realised as a tap in the stress-initial but not in the stress-final word).

In summary: the recognition of speech is a complex task which must rate as one of the primary achievements of the human brain. It shows us our brains functioning in an optimal fashion. The use or not of suprasegmental information in lexical recognition is just one more way in which this system displays its perfection: the cues are used when they are useful. If they are not useful, they are not used.

4. Acknowledgements

Financial support was provided by the Australian Research Council Centre of Excellence for the Dynamics of Language. We thank Rob Mailhammer for guidance in the German study. LB and JY share equal first authorship.

5. Appendix

Test pairs for (a) Experiment 1, German and (b) Experiment 2, English. In each pair, the first member has primary stress on the first syllable.

(a) *Abschied, abstrakt; Aktie, Aktion; Arche, Archiv; Atlas, Athlet; Bruder, brutal; chemisch, Chemie; Chronik, Chronist; damals, damit; Dose, Dozent; Globus, global; Hupe, human; Kompass, kompakt; Konto, Konzept; Logik, lokal; Marke, markant; Masse, massiv; Moped, mobil; Motor, Motiv; Profi, Profit; Segel, Sequenz; tote, total; Tresen, Tresor; Turban, Turbine; Turner, Turnier.*

(b) *booking, bouquet; campus, campaign; carton, cartoon; cashew, cashier; convent, convex; distance, distinct; district, distress; diver, divert; harpist, harpoon; humid, humane; impact, impress; influence, inform; liquid, liqueur; massive, masseur; motive; motel; music, museum; mystic, mistake; robot, robust; ruler, roulette; typhus, typhoon; union, unique.*

6. References

[1] S. Soto-Faraco, N. Sebastián-Gallés, and A. Cutler, "Segmental and suprasegmental mismatch in lexical access," *Journal of Memory and Language*, vol. 45, pp. 412-432, 2001.

[2] L. Tagliapietra and P. Tabossi, "Lexical stress effects in Italian spoken word recognition," in *Proceedings of the XXVII Annual Conference of the Cognitive Science Society*, B. G. Bara, L. Barsalou, and M. Bucciarelli, Eds. Stresa, Italy: Lawrence Erlbaum, 2005, pp. 2140-2144.

[3] W. v. Donselaar, M. Koster, and A. Cutler, "Exploring the role of lexical stress in lexical recognition," *The Quarterly Journal of Experimental Psychology,* vol. 58A, no. 2, pp. 251-273, 2005.

[4] C. K. Friedrich, S. A. Kotz, A. D. Friederici, and T. C. Gunter, "ERPs reflect lexical identification in word fragment priming," *Journal of Cognitive Neuroscience,* vol. 16, no. 4, pp. 541-552, 2004.

[5] H. Zora, M. Heldner, and I.-C. Schwarz, "Perceptual correlates of Turkish word stress and their contribution to automatic lexical access: Evidence from early ERP components," *Frontiers in Neuroscience,* vol. 10, no. 7, 2016.

[6] A. Cutler, "*Forbear* is a homophone: Lexical prosody does not constrain lexical access," *Language and Speech,* vol. 29, no. 3, pp. 201-220, 1986.

[7] L. M. Slowiaczek, "Effects of lexical stress in auditory word recognition," *Language and Speech,* vol. 33, no. 1, pp. 47-68, 1990.

[8] L. M. Slowiaczek, "Stress and context in auditory word recognition," *Journal of Psycholinguistic Research,* vol. 20, no. 6, pp. 465-481, 1991.

[9] L. H. Small, S. D. Simon, and J. S. Goldberg, "Lexical stress and lexical access: Homographs versus nonhomographs," *Perception & Psychophysics,* vol. 44, no. 3, pp. 272-280, 1988/05/01 1988.

[10] N. Cooper, A. Cutler, and R. Wales, "Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners," *Language and Speech,* vol. 45, no. 3, pp. 207-228, 2002.

[11] J. Yu, R. Mailhammer, and A. Cutler, "Vocabulary structure affects word recognition: Evidence from German listeners," in *Proceedings of the 10th International Conference on Speech Prosody (Speech Prosody 2020),* Tokyo, Japan, 2020, pp. 474-478.

[12] B. D. Fear, A. Cutler, and S. Butterfield, "The strong/weak syllable distinction in English," *The Journal of the Acoustical Society of America,* vol. 97, no. 3, pp. 1893-1904, 1995.

[13] A. Cutler and D. Pasveer, "Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition.," in *Proceedings of Speech Prosody 2006,* R. Hoff-mann and H. Mixdorff, Eds. Dresden, Germany: TUD Press, 2006.

[14] L. Bruggeman and A. Cutler, "Lexical manipulation as a discovery tool for psycholinguistic research," in *Proceedings of the 16th Australasian International Conference on Speech Science and Technology,* C. Carignan and M. D. Tyler, Eds. Parramatta, Australia, 2016, pp. 313-316.

[15] K. Lemhöfer and M. Broersma, " Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English," *Behavior Research Methods,* vol. 44, no. 2, pp. 325-343, 2012.

[16] R. H. Baayen, R. Piepenbrock, and L. Gulikers, "The Celex lexical database (CD-rom)," ed. Philadelphia: University of Pennsylvania, Linguistic Data Consortium, 1995.

[17] Neurobehavioral Systems Inc., "Presentation software," ed. Berkeley, CA.

[18] A. Cutler and D. M. Carter, "The predominance of strong initial syllables in the English vocabulary," *Computer Speech & Language,* vol. 2, no. 3–4, pp. 133-142, 1987.

[19] C. G. Clopper, "Frequency of stress patterns in English: A computational analysis," *Indiana University Linguistics Club Working Papers Online,* vol. 2, pp. 1-9, 2002.

[20] R Core Team, "R: A language and environment for statistical computing," 3.6.1 ed. Vienna, Austria: R Foundation for Statistical Computing, 2019.

[21] D. Bates, M. Maechler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software,* vol. 67, no. 1, pp. 1-48, 2015.

[22] A. Cutler, D. Norris, and N. Sebastián-Gallés, "Phonemic repertoire and similarity within the vocabulary," in *Proceedings of the 8th International Conference on Spoken Language Processing (Interspeech 2004-ICSLP),* S. Kim and M. J. Bae, Eds. Seoul: Sunjijn Printing Co, 2004, pp. 65-68.