Phonologically determined asymmetries in vocabulary structure across languages

Anne Cutler[a)]
MARCS Institute, University of Western Sydney, Penrith South, New South Wales 2751, Australia
Anne.Cutler@mpi.nl

Takashi Otake
E-listening Laboratory, 3-35-20 Higashi Tokorozawa, Tokorozawa 359 0021, Japan
otake@e-listeninglab.com

Laurence Bruggeman
Max Planck Institute for Psycholinguistics, POB 310, 6500AH Nijmegen, The Netherlands
Laurence.Bruggeman@mpi.nl

Abstract: Studies of spoken-word recognition have revealed that competition from embedded words differs in strength as a function of where in the carrier word the embedded word is found and have further shown embedding patterns to be skewed such that embeddings in initial position in carriers outnumber embeddings in final position. Lexico-statistical analyses show that this skew is highly attenuated in Japanese, a noninflectional language. Comparison of the extent of the asymmetry in the three Germanic languages English, Dutch, and German allows the source to be traced to a combination of suffixal morphology and vowel reduction in unstressed syllables.

PACS numbers: 43.71.Hw, 43.71.An [SGS]
Date Received: May 18, 2012 **Date Accepted:** July 3, 2012

1. Introduction

The recognition of words in continuous speech is based on concurrent consideration of multiple interpretations and competition between candidate words.[1,2] In all languages, the vocabulary contains tens to hundreds of thousands of words, which necessarily overlap and resemble each other because they draw on very few phonemes (the average inventory size is around 30). Speech is thus always temporarily compatible with more words than those intended. Luce[3] was the first to point out the significance of this for models of spoken-word recognition; his lexico-statistical proof that the first word encountered in many speech strings is not the intended word, due to the extent of word-initial embedding of shorter words in longer ones, put paid to strictly sequential word-by-word lexical access models.

In Luce's phoneme-based analyses, as well as in later embedding analyses with matching of syllable boundaries,[4,5] English carrier words consistently proved to have more shorter words embedded in their initial than in their final portions. This is relevant for models of word recognition because the number of competitors affects recognition[6–9] and the effects of competition differ at different points: Words with common onsets are easier to extract from preceding speech contexts than words with rare onsets[10] but at the same time are recognized less rapidly than otherwise similar

---

[a)] Author to whom correspondence should be addressed.

rare-onset words,[11] early- versus late-embedded words in spondees[12] and in idioms and compounds[13] differ in recognition effects.

Further, listener response to competition at different points is asymmetric. Relative strength of competition can be assessed with eye-tracking, which records what people look at as they hear speech. In such studies, competitors for word onsets attract more looks than competitors for offsets—e.g., hearing *candle* can induce early looks to a candy, or to a handle, but the former tendency is stronger.[14,15] This asymmetry suggests a greater role of initial than of final competition. However, the impact of mismatch between input and lexical candidate is modifiable; listeners adjust their tolerance of mismatch when listening conditions are not ideal, and this affects the initial/final asymmetry. Even words that are themselves clearly pronounced are more likely to induce looks to offset competitors and less likely to induce looks to onset competitors if occasional background crackle (as with a poorly tuned AM radio) interrupts other utterances[16] or if some words in other utterances are spoken casually with consequent speech reductions such as *ornry* instead of *ordinary*.[17]

Accurate modeling of these findings requires a complete understanding of competitor availability in the vocabulary. One possible reason for the skew in initial/final embedding patterns that has been reported for the English vocabulary, as described, is that English words predominantly have initial stress[18] and that unstressed syllables are often weak with reduced vowels. Every citation form pronunciation must contain at least one strong syllable, but polysyllables can end with successive weak syllables, so there is less opportunity for final embedding (e.g., in *deodorant* or *manufacturer*, the first two or three syllables might have been words, but the final two syllables could not be words).

The initial/final asymmetry could be solely due to stress patterning; but then it should not appear in languages that do not resemble English in this feature, such as Spanish, which has largely penultimate stress and no vowel reduction in unstressed syllables. An English-Spanish comparison,[4] however, showed a skew in both cases. The asymmetry could also be due to suffixing morphology. Across languages (including both English and Spanish), affixes, especially of person, tense and number, tend to be suffixes more often than prefixes or infixes.[19] Suffixes reduce the likelihood of final embedding (both *become* and *succumb* contain *come*; *became* contains *came*; but *succumbed* contains no final embedding).

2. Evidence from Japanese

To test this factor, we examine initial/final embedding ratios in a noninflectional language, Japanese. In Japanese, many tensed verbs have null affix realization. Thus the person distinction in English *I walk* versus *he walks* is signaled in Japanese only in the pronoun; the verb form *aruku* is identical in each case. Japanese past tense (e.g., *ita* for *aruku*) or nominalization markers (e.g., *koto*) are bimoraic elements that themselves can be words (*ita* is also “board”; *koto* is also “harp”). Moreover, a large proportion of the Japanese lexicon is Chinese-origin words with a two-part structure, each part also a stand-alone word.

The counts were based on the electronic version of the NHK Pronunciation Dictionary.[20] Japanese non-kanji orthography and dictionary entries are based on the mora (vowel, onset + vowel, lengthening element or syllable coda). Stand-alone words must (as in English) be minimally bimoraic. The “special moras” (nasal coda: *tempura* te-m-pu-ra; geminate marker: *Nissan* ni-s-sa-n; long vowel: *Tokyo* to-o-kyo-o) cannot occur alone or initially, reducing the number of embedding options (the last two moras of *Nintendo* could never be a word *ndo*). The most common word form in Japanese has four moras. We therefore analyzed words with four or five moras (the largest proportion of the vocabulary, 55.65% of the total database). Note that moras may, but need not, correspond to syllables (*Tokyo*, with two long vowels, has four moras, just as the loan word *panorama*).

Table 1. Percentage of words having shorter words embedded within them in initial versus final position, as a function of length of the carrier word and length of the embedded word, across four languages (Japanese, English, Dutch, German).

(a) Japanese			
Carrier length	Embedded word length	Initial	Final
4 moras ($N = 25\,974$)	2 moras	84.53	80.84
	3 moras	25.01	17.46
5 moras ($N = 11\,326$)	2 moras	79.25	65.99
	3 moras	42.95	46.32
	4 moras	12.25	7.32
(b) English			
2 syllables ($N = 44\,880$)	1 syllable	60.69	42.16
3 syllables ($N = 29\,544$)	1 syllable	43.28	33.68
	2 syllables	30.16	15.92
4 syllables ($N = 13\,454$)	1 syllable	29.22	23.28
	2 syllables	10.32	6.93
	3 syllables	22.26	13.05
(c) Dutch			
2 syllables ($N = 70\,731$)	1 syllable	73.7	64.6
3 syllables ($N = 115\,133$)	1 syllable	70.52	49.18
	2 syllables	43.79	15.43
4 syllables ($N = 87\,486$)	1 syllable	70.16	45.9
	2 syllables	34.62	8.42
	3 syllables	30.9	13.08
(d) German			
2 syllables ($N = 51\,647$)	1 syllable	60.02	37.1
3 syllables ($N = 104\,818$)	1 syllable	60.07	11.69
	2 syllables	35.23	16.27
4 syllables ($N = 91\,060$)	1 syllable	56.34	4.67
	2 syllables	30.15	6.2
	3 syllables	44.08	7.49

For each word, each possible embedding site was checked as to whether it indeed corresponded to at least one real word. For a four-mora word such as *panorama*, we thus checked *pano*, *panora*, *nora*, *norama*, and *rama*. For a five-mora word such as *amerikan*, there were nine possibilities (*ame*, *ameri*, *amerika*, *meri*, *merika*, *merikan*, *rika*, *rikan*, and *kan*). Special moras were not allowed in initial position, including long vowels (e.g., in *Kyoto* kyo-o-to, *kyoo* "today" would be counted but not *oto* "sound"). Table 1(a) shows the results. It can be seen that in Japanese, too, initial embeddings in carrier words slightly outnumber final embeddings, but the overall initial:final ratio (1.118:1) is not very asymmetric.

3. Evidence from English

Because the Japanese count had focused on the most frequently occurring word lengths, we undertook an analogous count for English (the two, three, and four syllable words in the CELEX English database,[21] forming 38%, 25%, and 11.4% of the total). Syllable boundaries were respected (thus in *scandalous* we counted *scan* and *scandal* but not *can*, *candle*, or *scanned*). Diphthongs were treated as unitary; so *plate* [pleɪt] does not contain *it* [ɪt]). The results [Table 1(b)] show that the pattern previously reported for English is again observed: More embedded words are found at the onset than at the offset of carriers, with an overall initial:final ratio of 1.454:1. Embedding patterns in Japanese and English thus differ.

One further reason for this could also be an accident of orthography. In English, most compound words are separated in writing (*nail file*, *can opener*), while similar Japanese compounds are unitary and thus potential carriers within which their components can be found. We test the effect of this by comparing English with Dutch; phonologically and morphologically, these related languages are highly similar (Dutch also has largely initial stress, vowel reduction, and suffixing inflectional morphology), but orthographically they differ including in that compounds are written as one word in Dutch (*nagelvijl, blikopener*).

4. Evidence from Dutch

Again we used the CELEX database and the same range of word lengths and constraints as for English (respectively, 20.4%, 38.25%, and 25.25% of the Dutch total). The results [Table 1(c)] reveal that the skew seen in English is even stronger in Dutch, with an overall initial: final ratio of 1.69:1.

Dutch in fact has more inflections than English, not only on verbs; plurals are also often an extra syllable (*boeken* = *books*). But if it is the weak syllables of inflections that really underlies the skew, we might expect an even stronger skew in German. Like Dutch and English, German has suffixes, syllables with schwa, and predominantly initial stress; but German also has many monomorphemic schwa-final words. Consider such triplets of English-Dutch-German cognates as: *cat kat Katze*; *cigar sigaar Zigarre*; *role rol Rolle*. There are many more of these. The final syllable in each German case contains schwa.

5. Evidence from German

Again we used the CELEX database and the same range of word lengths and constraints as for English and Dutch (respectively, 16%, 32.7%, 28.4% of the German database). As Table 1(d) shows, the skew is far more pronounced in German than in English or Dutch; the overall initial: final embedding ratio is now 6.04:1.

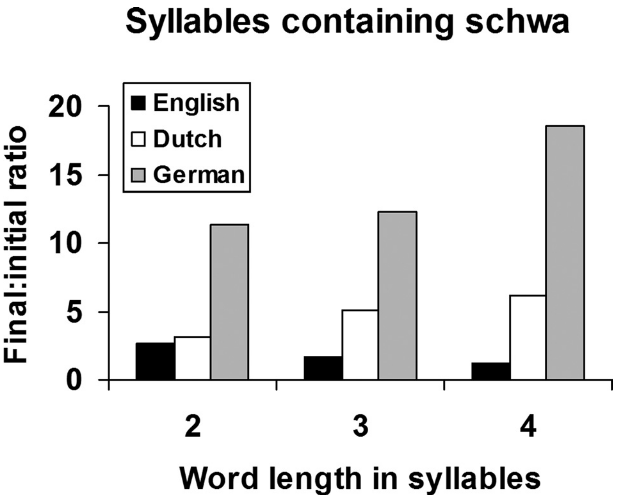

Fig. 1. The ratio of words with final syllable containing schwa to words with initial syllable containing schwa in the three Germanic languages English, Dutch, and German, separately for words of two, three, or four syllables in length. In English, the greater likelihood of affixes being suffixal rather than prefixal means that the final:initial schwa ratio is indeed always positive but reduces in longer words because the longer the word, the more likely it is to have a prefix. Both Dutch and German have more inflectional suffixes (on verbs and as noun plurals) than English, so their ratios are higher than the English ratio; German, however, has in addition very many monomorphemic words with final schwa.

To further examine the role of schwa in this clear difference of patterning across three closely related languages, we calculated the occurrence of schwa in initial versus final syllables in the three vocabularies. Figure 1 shows the asymmetry in the likelihood of final- versus initial-syllable schwa in each word length category in each language. The German asymmetry dwarfs the asymmetries in English or Dutch.

6. Conclusion

Language-specific phonology determines not only what words compete with one another but also where they compete. In Japanese, which has little affixation and no weak syllables, there is little difference between the number of lexical competitors for initial versus for final portions of words. (Note that our special-mora restriction, barring, e.g., *oto* in *Kyoto*, may have acted to reduce the Japanese final-embedding tally and hence increase initial-final differences.) In this, Japanese contrasts with English for which greater competition for initial than for final portions of words has consistently been demonstrated.

The cross-language difference is not due to the extensive compounding characteristic of Japanese because the initial-final asymmetry observed in English also appears in Dutch and German, which, like Japanese, have many compounds. We traced the asymmetry to a combination of extensive suffixing (present in English, Dutch, and German as well as in Spanish, where the asymmetry also appears), intensified by the presence of schwa in the phoneme repertoire, allowing these suffixes to be weak syllables. The more final weak syllables in the vocabulary, the greater the size of the asymmetry (so that it is largest in German, where many monomorphemic words also have a final syllable with schwa).

This asymmetry thus joins other language-specific phonological factors that significantly impact word structure and processing. Thus languages with fewer phonemes have longer words, with in consequence more embedding, than languages with more phonemes.[4] Languages with consonant clusters allow more within-syllable embedding (e.g., *ring* in *bring*, *bell* in *belt*) than languages that bar clusters, so that variation also arises in how often processes for rapidly removing such competitors[22,23] come into play. Embedding of words in other words, a major determinant of spoken-word recognition complexity,[8,9,24] thus varies in multiple ways across languages. Because competitors exercise differing effects when embedded at different word positions,[10–13] our results thus force the conclusion that competition in spoken-word recognition also fluctuates variably across languages.

Acknowledgments

The first two authors are also associated with the Max Planck Institute for Psycholinguistics, and the first and third authors are also associated with the Radboud University Nijmegen. Financial support was provided by the NWO-SPINOZA project "Native and Nonnative Listening" (A.C.).

References and links

[1] J. L. McClelland and J. L. Elman, "The TRACE model of speech perception," Cognit. Psychol. **18**, 1–86 (1986).
[2] D. Norris and J. M. McQueen, "Shortlist B: A Bayesian model of continuous speech recognition," Psychol. Rev. **115**, 357–395 (2008).
[3] P. A. Luce, "A computational analysis of uniqueness points in auditory word recognition," Percept. Psychophys. **39**, 155–158 (1986).
[4] 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* (2004), Jeju Island, Korea, edited by S. H. Kin and M. Jin Bae (Sunjin Printing Co., Seoul, 2004), Vol. 1, pp. 65–68 [CD-ROM].
[5] J. M. McQueen and A. Cutler, "Words within words: Lexical statistics and lexical access," in *Proceedings of the International Conference on Spoken Language Processing'92*, Banff, Alberta, edited by

J. J. Ohala, T. M. Nearey, B. L. Derwing, M. M. Hodge, and G. E. Wiebe (University of Alberta, Edmonton, Canada, 1992), Vol. 1, pp. 221–224.

[6] J. A. Bashford, R. M. Warren, and P. W. Lenz, "Polling the effective neighborhoods of spoken words with the verbal transformation effect," J. Acoust. Soc. Am. **119**, EL55–EL59 (2006).

[7] J. A. Bashford, R. M. Warren, and P. W. Lenz, "Evoking biphone neighborhoods with verbal transformations: Illusory changes demonstrate both lexical competition and inhibition," J. Acoust. Soc. Am. **123**, EL32–EL38 (2008).

[8] S. D. Goldinger, P. A. Luce, and D. B. Pisoni, "Priming lexical neighbors of spoken words: Effects of competition and inhibition," J. Mem. Lang. **28**, 501–518 (1989).

[9] D. Norris, J. M. McQueen, and A. Cutler, "Competition and segmentation in spoken word recognition," J. Exp. Psychol. Learn. **21**, 1209–1228 (1995).

[10] A. H. van der Lugt, "The use of sequential probabilities in the segmentation of speech," Percept. Psychophys. **63**, 811–823 (2001).

[11] M. S. Vitevitch, "Influence of onset density on spoken-word recognition," J. Exp. Psychol. Hum. **28**, 270–278 (2002).

[12] M. S. Cluff and P. A. Luce, "Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation," J. Exp. Psychol. Hum. **18**, 551–563 (1990).

[13] D. A. Swinney, "Lexical processing during sentence comprehension: Effects of higher order constraints and implications for representation," in *The Cognitive Representation of Speech*, edited by T. Myers, J. Laver, and J. Anderson (North-Holland, Amsterdam, 1981), pp. 201–209.

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

[15] J. M. McQueen and M. Viebahn, "Tracking recognition of spoken words by tracking looks to printed words," Q. J. Exp. Psychol. **60**, 661–671 (2007).

[16] J. M. McQueen and F. Huettig, "Changing only the probability that spoken words will be distorted changes how they are recognized," J. Acoust. Soc. Am. **131**, 509–517 (2012).

[17] S. Brouwer, H. Mitterer, and F. Huettig, "Speech reductions change the dynamics in competition during spoken word recognition," Lang. Cognit. Processes **27**, 539–571 (2012).

[18] A. Cutler and D. M. Carter, "The predominance of strong initial syllables in the English vocabulary," Comput. Speech Lang. **2**, 133–142 (1987).

[19] J. A. Hawkins and G. Gilligan, "Prefixing and suffixing universals in relation to basic word order," Lingua **74**, 219–259 (1988).

[20] M. Sugito, "Osaka-Tokyo Akusento Onsei Jiten," *Osaka-Tokyo Accent Pronunciation Dictionary* (Maruzen, Tokyo, 1995) [CD-ROM].

[21] R. H. Baayen, R. Piepenbrock, and H. van Rijn, The CELEX lexical database (Linguistic Data Consortium, University of Pennsylvania, Philadelphia, 1993) [CD-ROM].

[22] D. Norris, J. M. McQueen, A. Cutler, and S. Butterfield, "The possible-word constraint in the segmentation of continuous speech," Cognit. Psychol. **34**, 191–243 (1997).

[23] A. Cutler, T. Otake, and J. M. McQueen, "Vowel devoicing and the perception of spoken Japanese words," J. Acoust. Soc. Am. **125**, 1693–1703 (2009).

[24] T. Otake, J. M. McQueen, and A. Cutler, "Competition in the perception of spoken Japanese words," in *Proceedings of the 11th Annual Conference of the International Speech Communication Association* (2010), Makuhari, edited by T. Kobayashi, K. Hirose, and S. Nakamura (Tokyo Institute of Technology, Tokyo), pp. 114–117.