Research Article

UNIVERSALITY VERSUS LANGUAGE-SPECIFICITY IN LISTENING TO RUNNING SPEECH

Anne Cutler,[1] Katherine Demuth,[2] and James M. McQueen[1]

1Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, and
Department of Cognitive and Linguistic Sciences, Brown University

Abstract—*Recognizing spoken language involves automatic activation of multiple candidate words. The process of selection between candidates is made more efficient by inhibition of embedded words (like* egg *in* beg) *that leave a portion of the input stranded (here, b). Results from European languages suggest that this inhibition occurs when consonants arc stranded but not when syllables are stranded. The reason why leftover syllables do not lead to inhibition could be that in principle they might themselves be words; in European languages, a syllable can be a word. In Sesotho (a Bantu language), however, a single syllable cannot be a word. We report that in Sesotho. word recognition is inhibited by stranded consonants, but stranded monosyllables produce no more difficulty than stranded bisyllables i which could be Sesotho words). This finding suggests that the viability constraint which inhibits spurious embedded word candidates is not sensitive to language-specific word structure, but is universal.*

Listening to spoken language is one of the easiest things people do. Yet this ease is fully apparent only in the native language, because listening can be sensitive to language-specific structure—discriminating normative speech sounds is often hard, for instance. But listeners everywhere are born with the same mental apparatus, and use it to perform the same operations, and for this reason listening must also be to a large extent universal.

Current models of spoken-word recognition (e.g.. (Gaskell & Marslen-Wilson. 1997; Luce & Pisoni, 1998; McClelland & Elman. 1986; Norris. 1994) agree that spoken language activates multiple alternative word candidates. The candidates need not be aligned with one another, each neatly beginning where a previous one ended; thus. the utterance "we stop begging" might activate *east, top,* and *egg,* as well as the intended words. Activated words must compete with one another for recognition, and the outcome of the competition is ultimately the correct sequence of words in the input. The combination *we stop begging* accounts for the whole input and will therefore triumph over *east, top,* and *egg* because none of the latter can muster support from other words (*w* and *weace* and *b* are not words).

Spuriously present words can also be gotten rid of by other means. Norris. McQueen, Cutler, and Butlerfield (1997) discovered that in word-spotting experiments—in which the task was to spot embedded real words in nonsense items—listeners often failed to spot, for instance, *egg* in *fegg.* or, when they did detect it. their response times were slow. *Egg* was much easier to spot in *maffegg.* Norris et al. proposed that it is in general hard for listeners to extract words from contexts consisting only of consonants (e.g.. *top* in *stop, egg* in *beg),* and explained this difficulty as a reflection of a basic viability constraint in spoken-word recognition. Neither *f* nor *muff* is a word of English, but *maff* would be viable as a word *(map, muff,* and *gaff are* all words). whereas a single consonant like *f* would not. By leaving stranded nonviable residues of the input, candidate words like *egg* in *beg* (or *fegg*) forfeit their own credibility. The constraint could thus eliminate such spuriously active candidates. Norris et al. called this procedure the possible-word constraint (PWC) and implemented it in the Shortlist model of spoken-word recognition (Norris. 1994). Shortlist is a coneclionisl model in which an initial activation stage produces multiple candidate words, which may or may not be aligned with one other: in a subsequent stage, these shortlisted candidates compete for recogniiion via interword competition. In Shortlist, the PWC operates at the competition stage; activated words are penalized (their level of activation is reduced) if the effect of accepting them would be to strand a nonviable residue of the input between the putative word and the nearest known boundary. "Viable" was specified in this implementation as "containing a vowel"; a syllable was thus a viable residue, but a consonant alone was not.

Although activation and competition may characterize speech recognition in all languages, language-specitic features can also obtain. The nearest known boundary might be determined, for instance, by phoneme sequence restrictions, which are language-specific. If English listeners hear the sound sequence |vr|. they know that there must be a syllable boundary between the two sounds (as. for instance, in "ev'ry" or "have run"), because no syllables in English begin or end with |vr|. This conclusion is language-specific because in other languages (e.g.. German, Dutch) syllables can begin with |vr|. McQueen (1998) showed that words are easier to spot if their edges are aligned with such necessary syllable boundaries than if they are not so aligned.

But is what counts as a viable residue also language-specific? In the experiments by Norris et al. (1997). syllables were always viable residues (word spotting was easier) and consonants never were (word spotting was harder). Culler (1997) and Cutler. McQueen. Norris, and Somejuan (2001) therefore argued that syllables play a fundamental role in the segmentation of speech. No language allows words consisting only of single consonants; syllables are the smallest units that function as words. According to this account, the viability tested by the constraint of Norris et al. is the hypothetical potential of the residue—a syllable might conceivably itself be a word, but a consonant would not. Note that the constraint ignores actual lexical status—*maff* is not in fact an English word, though it could be. The viability constraint concerns only the smallest form (hat a word can possibly take; across languages in total, this is a syllable.

However, it is not true that in every language the smallest possible word is a syllable. Many Bantu languages, including Sesotho (Demuth. 1996; Doke & Mofokeng. 1957). Shona (Myers. 1987), and Chichewa (Kanerva. 1990), prohibit one-syllable stand-alone words. This prohibition obeys a larger rule, observed in many languages. whereby well-formed stand-alone words cannot consist of only one

Address correspondence to Anne Cutler. Max Planck Institute for Psycholinguistics, P.O. Box 310. 6500 All Nijmegen, The Netherlands; e-mail: anne.cutler@mpi.nl

mora (Broselow. 19X2; McCarthy & Prince. 1995). A mora is (a) a short vowel, (b) a short vowel plus onset, or (c) a coda (end pari) ol a Syllable. A long vowel is bimoraic (i.e., counts as two short vowels). The rule is thus minimally satisfied by words of two syllables, or of one syllable with two moras (Hayes. 1995).

English obeys this rule—*see* and *my*. with long vowels, are permissible monosyllabic English words, and *get* and *mat.* with short vowels plus codas, are also line, but *se* (with the vowel *of gel)* and *ma* (with the vowel *of mat)* are impossible. Bantu languages like Sesotho also obey the rule. Bui Sesotho has no coda consonants, and therefore no words like English *get* or mat. and it also has no long vowels or diphthongs. and thus no words like English *see* or *my.* Thus, the only way ol" satisfying the rule in Sesotho is not to allow monosyllabic words. So although the stem ja ("eat") is monosyllabic, the shortest stand-alone word with this stem is the imperative ("eat!"), alternatively realized as *eja* or *jaa* (pronounced "ja-a").

Many other languages (e.g.. French. Japanese) do allow monomoraic words. This rule for minimal words is thus not universal. Hence, languages that obey it. and especially those that, like Sesotho. can obey il only by creating bisyllables, provide an opportunity to examine the universality of the proposed viability constraint on word recognition. Across languages, the smallest form a word can take is a syllable. If viability always simply requires at least a syllable as residue, the operation of the constraint may be universal. But in Sesotho. the smallest form a word can take is two syllables. If viability in Sesotho requires a two-syllable residue, then the operation of the constraint is language specific (i.e.. sensitive to lexical potential within a Specific vocabulary).

To test this question, we compared the effect of different adjoined contexts on the recognition of spoken words in Sesotho. which is spoken in Lesotho and parts of South Africa. We conducted our experiments in Lesotho, at the National University of Lesotho (NUL). In experiments like those conducted by Norris et al. (1997), we examined the relative difficulty of spotting words adjoined to three types of short nonsense contexts. First, in the *pseudoword* context, the context could be (but was not) a stand-alone Sesotho word (e.g.. a bisyllabic nonword such as *hapi).* Second, in the *syllabic* context, the context could not be a Sesotho word, but could be a word of some other language (e.g.. a monosyllabic nonword such as *w).* Third, in the *consonant* context, the context was a single consonant (e.g.. *b)* and could not be a word of Sesotho or of other languages. We expected word spotting to prove very hard in the consonant context, in line with the results of Norris et al. 11997). The crucial question concerned the pseudoword versus the syllable context. If the syllable context was significantly harder than the pseudoword context, this would suggest that the constraint is sensitive to what may and may not be a word in the listener's own language. But if the two contexts did not contrast in difficulty, this would suggest that the constraint operates similarly across languages.

METHOD

The materials were based on 57 existing bisyllabic and trisyllabic Sesotho words. There were 24 bisyllables. all with CVCV (consonant-vowel-consonant-vowel) structure, such as rora ("to roar"), and 33 trisyllables. all V*CVCV*. such as *alafa* ("to prescribe"). All items, like these two examples, are verbs in common use. (Note that verbs and nouns are equally easy for listeners in word-spotting experiments; McQueen & Cutler. 1998).

For each trisyllable, we constructed three different nonsense conlexts: one bisyllabic (pseudoword). one monosyllabic (syllable), and One consisting of a single consonant (consonant). Thus, *alafa* could be preceded by *pafo* in *pafoalafa.* by *ro* in *roalafa.* or by h in *halafa.* For the bisyllables. there were only pseudoword and syllable contexts: *Rora.* for example, was preceded by *hapi* in *hapi rora* and by *ji* in *jirora.* We could not use consonant contexts in this case because Sesotho prohibits consonant clusters. We also constructed 114 filler items that were not words and contained no embedded words. The filler items, constructed to resemble the experimental items, comprised 66 legal VCVCV nonwords. 22 combined with each type of context used with the trisyllables, and 48 legal CVCV nonwords, 24 with each type of context used with the bisyllables.

All items were recorded by a female native speaker of Sesotho. The recording was digitized and used to make three tapes for use in the word-spotting experiment. All fillers and each of the 33 trisyllabic experimental words occurred in the same pseudorandom order on each tape, and the 24 bisyllabic words were divided such that 16 occurred on each tape. Context was counterbalanced across tapes, with each tape containing 11 trisyllabic words with a pseudoword context, 11 with a syllable context, and 11 with a consonant context, plus 8 bisyllabic words with a pseudoword context and 8 with a syllable context. Each tape began with a practice set that also contained nonwords both With and without embedded words.

We then constructed three further tapes for use in a lexical decision experiment. Words spoken in different contexts may form better or worse approximations to canonical pronunciations. If so. listeners may recognize the embedded words more rapidly or more slowly simply because of how close these are to their ideal forms, rather than because of the difficulty of extracting them from their context. Likewise. word-spotting speed and accuracy could be influenced by word frequency; although we tried lo avoid rare words, we could not know in advance whether all words would be recognized by our subject population. It was therefore important to determine how recognizable the individual tokens were by themselves. We did this using a standard control procedure: a simple lexical decision task on the embedded words, using the actual recording from the word-spotting experiment but without the contexts, these having been edited out of the recording. The three tapes tor our lexical decision task therefore consisted of the same test words and nonword tillers on the word-spotting tapes. but without the accompanying contexts.

One hundred four listeners (2nd- and 3rd-year NUL undergraduates. plus a few NUL lecturers and staff) participated in the study and received a small payment. Twenty-one participants heard each word-spotting tape. and 14, 14, and 13 heard the three lexical decision tapes. respectively. All were nalive speakers of Sesotho without known hearing impairment.

Listeners were tested *one* at a lime in a quiet room. They received instructions in Sesotho from a nalive Sesotho-speaking assistant. For word spotting, listeners were told that they would hear a list of nonsense words, that they should try lo spot any real words embedded in the nonsense words, and that when they spoiled a word they should press a single response key as quickly as possible, and then say aloud the word that they had detected. For lexical decision, listeners were instructed to listen lo a list of words and nonwords. and to press the response key as quickly as possible whenever they heard a real word. and then to say aloud that word. In both experiments, the materials were presented over headphones from digital audiotape (with an inter-stimulus interval of 3.5 s). and vocal responses were recorded onto another digital audiotape. Key-press response limes (RTs) were recorded by a personal computer. Only those accompanied by correct vocal responses were analyzed. Statistical analyses were conducted first on the lexical decision data, then on the word-spotting data: means from each task are shown in Table 1.

RESULTS

Lexical Decision Analyses

We excluded from further analysis 6 bisyllabic and 19 trisyllabic words not correctly identified by at least two thirds of the subjects who heard them. The word-spotting data for these items may reflect differences in the recognizability of the target words themselves (either because of acoustic factors or because the words were unfamiliar to the subjects), rather than differences due to the contexts. We also excluded from further analysis 2 other trisyllabic words missed by all word-spotting subjects in at least one condition (these items could not be included in RT analyses).

Across the 30 remaining words, the mean RT. from word offset, was 413 ms for items from pseudoword contexts. 404 ms for items from syllable contexts, and 384 ms for items from consonant contexts. Mean error rates were quite high for these words (note that none of the subjects had ever participated in a psycholinguistic experiment before): 17% for items from pseudoword contexts, 15% for items from syllable contexts, and 16% for items from consonant contexts. Analyses of variance (ANOVAs) across subjects (F_1) and items (F_2) were conducted separately for the bisyllabic and trisyllabic words. For the bisyllables (e.g.. *rora*). there was no difference in errors as a function of original context but a marginal difference in RTs: Words from syllable contexts *(e.g., jirora)* were detected more slowly than words from pseudoword contexts (e.g.. *hapimra*). (F_1) (1.38) = 6.30, $p < .01$. and $p < .01$ and $F_2(1, 17) = 3.11$, $p < .1$. For the trisyllables (e.g.. *alafa*). there was again no effect of context in error rates and a marginal effect of context in RTs. $F_1(2.\ 76) = 2.90$. $p = .06$. and $F_2(2.\ 22) = 2.55$. $p = .1$. Post hoc tests showed the latter effect to be due solely to longer RTs to words from pseudoword contexts (e.g., *pafoalafa)* than to words from syllable contexts (e.g.. *roalafa).*

These results suggest that there were at least some differences in goodness of these tokens, favoring pseudoword over syllable contexts for the bisyllabic words and syllable over pseudoword contexts for the trisyllabic words. Note, though, that the words taken from consonant contexts were not less easy to recognize than words taken from either of the other contexts.

Word-Spotting Analyses

Figure 1 shows the mean RT and mean error rate across the three context types. As expected, words were hardest to spot in the consonant context—RTs were longer and error rates higher than in the other two conditions. Because the same tokens had not been harder to recognize in the lexical decision experiment, we assume that these items were harder because the consonant context is not a viable residue. This is exactly as expected from the finding in English (Norris et al., 1997: McQueen & Culler. 1998. and McQueen. Otake. & Cutler. 2001, have observed the same effect in Dutch and Japanese, respectively).

The crucial comparison concerns the pseudoword and syllable contexts. As Figure 1 shows, there was little difference in the results for these contexts, across all items. The separate ANOVA for the bisyllables (e.g.. *rora)* showed no significant difference in either RTs or error rates: analyses of covariance (ANCOVAs) by items, taking both lexical decision RTs and error rates as covariates. confirmed that there was no difference in listeners' ability to spot (for example) *rora* in *hapirora* versus *jirora.* In the separate analysis for the trisyllables (e.g., *alafa).* however, there was a marginal effect of context in RTs. $F_1(2,\ 120) = 48.40$. $p < \mathbf{.001}$, and F_2 n.s., and a fully reliable effect in errors. $F_1(2,\ 120) = 42.40$. $p < .001$, and $F_2(2.\ 22) = 7.52$, $p < .005$. Post hoc tests showed a disadvantage for pseudoword over syllable contexts, significant by both subjects and items for errors, though significant only by subjects for RTs. In an ANCOVA by items, however. taking both lexical decision RTs and error rates as covariates, this difference in error rates was no longer significant: that is. this pseudoword-syllable difference arose from the differences in token goodness discovered in lexical decision. The same post hoc tests showed a large difference between the consonant and syllable eontexts, again significant by both subjects and items for errors, and by subjects only for RTs. This difference was not due to token goodness: In by-items ANCOVAs taking lexical decision speed and accuracy as covariates, the effect in errors remained significant and the effect in RTs became significant.

Table 1. *Mean response time (in milliseconds from word offset) and mean percentage error in the word-spotting and lexical decision tasks*

Task and dependent variable	Bisyllabic words: Syllable	Bisyllabic words: Pseudoword	Trisyllabic words: Consonant	Trisyllabic words: Syllable	Trisyllabic words: Pseudoword
Lexical decision					
Response time	445	411	384	365	414
Percentage error	21	18	16	10	15
Word spotting					
Response time	912	923	1,028	817	910
Percentage error	33	33	73	41	61

Note. In the word-spotting task, words were presented in consonant, syllable, and pseudoword contexts. For example, the bisyllabic word *rora* appeared in *jirora* and *hapirora* (no consonant context was possible with bisyllables), and the trisyllabic word *alafa* appeared in *halafa*, *roalafa*, and *pafoalafa*. The same tokens were presented with their contexts edited out in the lexical decision task.

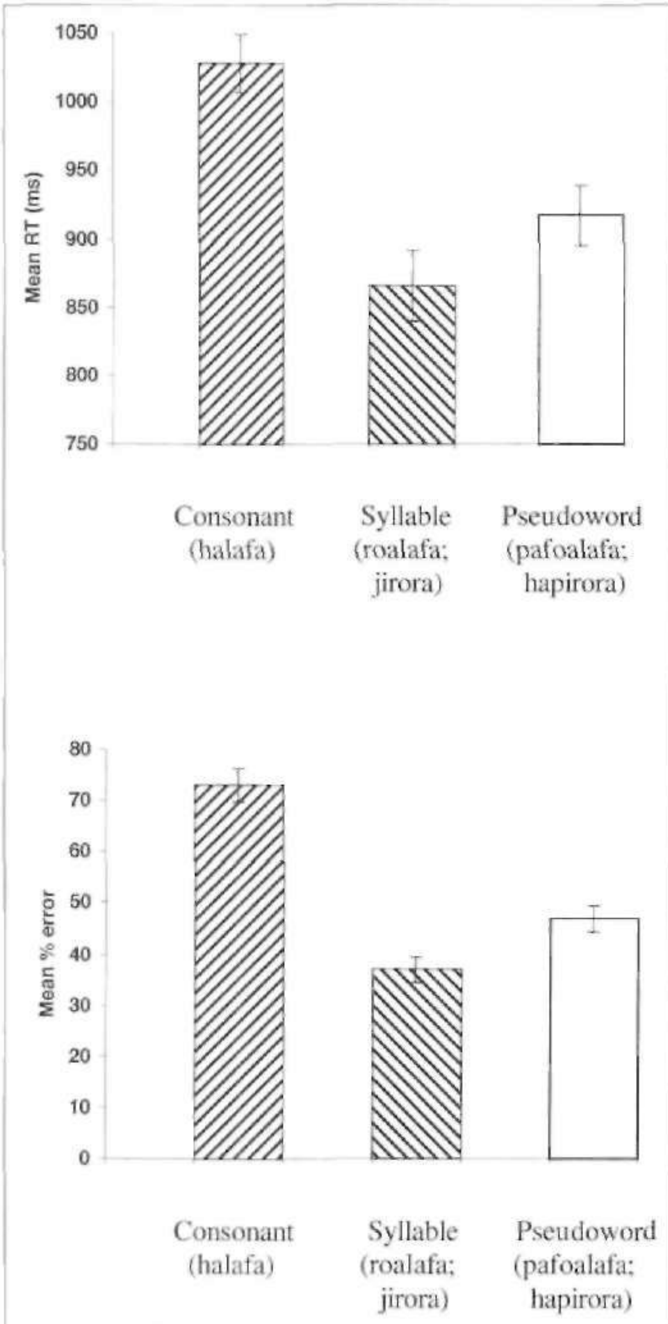

Fig. 1. Mean response time (RT) in milliseconds from word offset (top) and error percentage (bottom) in the word-spotting experiment. as a function of type of context. Error bars show standard errors.

In control analyses of the data using several alternative criteria for inclusion/rejection of subjects, items, or both (e.g.. error analyses including high-error items for which RT analyses were impossible), the same pattern of results was consistently observed. The two major findings are therefore those that can be seen in Figure 1: Word-spotting performance was poorer in consonant contexts than in syllable contexts. but there were no differences in ease of spotting words in syllable versus pseudoword contexts.

DISCUSSION

Sesotho listeners find it as easy to spot Sesotho words in single-syllable contexts (which cannot themselves form Sesotho words) as in two-syllable contexts (which can form Sesotho words). This is not because any context is as easy (or hard) as any other—single-consonant contexts make it hard to find Sesotho words just as they make it hard to find English. Dutch, and Japanese words. This pattern shows that word recognition in Sesotho is subject to the same type of viability constraint as word recognition in languages that allow monosyllabic words. Even though neither a single syllable nor a single consonant could be a stand-alone Sesotho word, only the latter appears to constitute a nonviable residue in word recognition. This pattern, as we argued in the introduction, is consistent with a universal constraint, operating similarly across languages irrespective of their vocabularies.

The efficiency with which listeners recognize spoken language belies the complexity of the recognition process. In any language, lens of thousands of words are constructed from just a handful of speech sounds. Inevitably, words resemble one another, and shorter words may be embedded within longer words. Unintended words may thus be accidentally present in a spoken utterance. Many psycholinguistic studies have shown (hat multiple word candidates, including words that are only accidentally present, can be simultaneously activated in listening (Gow & Gordon. 1995: Tabossi. Burani. & Scott. 1995). But this does not mean that listening bogs down in a welter of indecision: active competition between the available word candidates ensues (Goldinger. Luce. Pisoni. & Marcario. 1992: McQueen. Norris. & Culler. 1994: Norris. McQueen. & Cutler. 1995). and the competition will be won by the sequence that fully accounts for the input. Moreover, ihe competition process is made even more efficient by procedures that enable some potential competitors to be quickly jettisoned. The PWC (Norris et al., 1997), which penalizes any activated word that Strands a nonviable residue of the input, is one such useful process.

Our present results suggest that the criterion for a viable residual chunk of the input is. under any circumstances, that it must be, minimally, a syllable. Even though in Sesotho the vocabulary contains only words of two or more syllables, vocabulary-membership restrictions do not affect the way word recognition is constrained for viability; the constraint has a universal form. Other recent results support this claim. Norris, McQueen. Cutler. Butterfield. and Kearns (2001) found that English listeners spot words such as *canal* more easily in [ze]*canal* *t*han in [s] canal, though [ze], an open syllable with a short vowel, is monomoraic and hence not a legal word of English.

The name that Norris et al. (1997) chose for the PWC implies a reason for the existence of the viability constraint: Lexical candidates can be rejected if they would spawn impossible words. Bui the name now appears in one sense misleading. Viability of a residue left by an activated word does not depend on whether it might turn out itself to be a word in (he utterance. Norris et al. pointed out that the constraint is insensitive to whether the residue actually is a word *{mqff* in *maf-fegg* is a viable residue though not a word): we now know further thai the constraint is insensitive to whether the residue might potentially be a word in the specific lexicon in question.

Instead, the sense in which a viable residue is a "possible word" is more abstract and truly universal. The syllable is a viable residue because across languages as a whole the syllable is the smallest thing that can be a word. Language-specific vocabulary characteristics play no role. Only the universal generalization holds: it is irrelevant whether a syllable can be a word in the actual language being listened to. or indeed whether it can operate in any way as a combinatorial clement. Combinatorial elements, after all. can bo very small, even just single consonants, in just about any language (consider English *it's three eighths copied*—it + s + three + eight + th + s + copy + d).

A universal constraint of this type could be useful in language acquisition. Words spoken in isolation offer the infant excellent information about the rules governing what may be a word in the input language (e.g.. imperatives such as the forms of *eat!* occur often in Sesotho child-directed speech; Demuth. 1989). But even in the first year of life, infants can cope with words produced in running speech (Jusczyk. 1997). Thus, infants learn early to segment speech streams into the word forms that will stock their vocabulary. Clearly, they cannot use adultlike competition to segment speech, but they could use a universal process of distinguishing possible from impossible words. Indeed. Brent and Cartwright (1996) suggested just such a constraint on initial word learning. In computational simulations, they showed that a vocabulary-acquisition model performed better when vocabulary membership depended on presence of a vowel, so that no candidates consisting only of consonants were considered as possible words, than when the model incorporated no constraint on what strings might constitute words. The universality of the constraint is comprehensible if its source is in language acquisition.: Infants are not programmed in advance for any specific language.

Language-specificity in listening does exist. Word-boundary cues are provided by language-specific phoneme sequence restrictions (McQueen. 1998), and additionally in English by stress (Cutler & Norris. 1988). in French by syllable structure (Content. Kearns. & Frauenfelder. 2001; Cutler. Mehler. Norris. & Segui. 1986). and in Japanese by moras (McQueen et al.. 2001; Otake. Hatano. Cutler, & Mehler. 1993)- But the word recognition system operates in a universal manner: The aim is optimally rapid and efficient identification of the words making up a running speech signal. Words supported by the signal are automatically activated: spuriously present ones can often be identified at an early stage and eliminated as inherently unlikely. What makes them unlikely is that they leave an unusable residue between their edge and the nearest boundary. That boundary may have been set by language-specific factors (stress, sequence restrictions, etc.): but the viability of the residue is tested against a universal criterion whereby the residue must be. minimally, a syllable.

Acknowledgments—The authors contributed equally and order of authorship is alphabetical. Financial support was provided by a special grant from the Max Planck Society. Germany. We are very grateful to the Institute of Southern African Studies at the National University of Lesotho, and to its director, Gisela Prasad, for hospitality and assistance that enabled us to conduct the empirical work. We further thank Phokwane Mawasha for reCording the materials: Litsepiso Matlosa and Lehlohonolo Phafoli for experimental assistance in Lesotho; and 'Malillo Maehobane. Eraneina Moloi. James J. Jenkins. David E. Irwin, and two anonymous referees for helpful suggestions. The materials used in the experiment are listed in full at http://www.mpi.nl/world/persons/private/annne/materials.html

REFERENCES

Brent. M.R.. & Cartwright, t.A. (1996), Disiribuuonal regularity and phonocactic coif sininits arc useful lor segmentation. *Cognition, ftl,* 93-125.

Hioselow. E. 11982}. On the interaction of stress and epenthesis, *Ghssa, 16,* I IS 132,

Content, A.. Kearns. R.K.. & Frauenfcldcr, l.H. (2001), Boundaries versus onsets to sy> tabic segmentation. *Journaloj Memory and Language, -15.* 177 f99,

Culler, A. 11997). The syllable's role in the segmentation of stress languages. *Language ami Cognitivi Processes,* /2.839 845.

Cutler, A., McQueen. J.M.. Norris. IX. & Somejuan.A. (2001). The roll ol the silly ball. In E. Dupoux (Bel.). *Language, brain and cognitive development: Essays in honor of Jacques Mehlei* l|>p. 181-194). Cambridge. MA: MIT Press.

Cutter, A.. Mehler. J.. Norris. I).. & Segui. J. (1986). The syllable's differing role in the segmentation of French and English. *Journal "I Mental-\ and Language. 25.* 385-400

Cutler, A.. & Norris. D. i 19X8). The role of strong syllables in segmentation for lexical access. *Journal oj Experimental Psychology: Human Perception und Performance, 14,* 113-121

Demuth, K.G9N9). Maturation and the acquisition ol Scsothopassive. *Language, 65,* 56 80,

Denuitli. K. (1996). The prosodic structure of early words. In J. Morgan & K. Demuth 11 ils), *Signal to syntax: Booistrappingfwm speech to grammar in early acquisition* (pp. I7I-IX4), Mahwah. NJ: t-rlbaum.

Doke, C.M., & Molokcng. S.M. (1957). *textbook <\i"Southern Sotlto grammar.* Cape Town, South Africa. Longman.

Gaskell. M.G.. & Marslen-Wilson, W.D, (1997). Integrating form and meaning: A disiribiilcd model of speech perception, *hmguage and Cognitive Processes, 12.* 613 656.

Goldinger. S.I).. Luce. I'.-V. Pisoni. D.B.. & Marcario. J.K. (1992). form-based priming in spoken word recognition: The roles of competition and bias. *Journal afExperimenuil I'svi Hology: Learning, Memory, ami Cognition. IN.* 1211—1238.

(iow. D.W., ;< Gordon. I'.c. 11995). Lexical and prelexieal influences on word segmentation: Evidence from priming. *Journal oj Experimental Psychology: Human Percvp' lion and Performance, 21.* 344 359.

Hayes, B. 11995). *Metrical* ttottv *theory: Principles and case studies.* Chicago: I iimcrsiiy of Chicago Press,

Jusc/yk. P.W. 11997). *The discovery oj spOke>\ language.* Cambridge. MA: MIT Press.

Kanerva. J. (1990). *Focus and phrasing in < 'hichena phonology* New York: Garland Press.

Luce. P.A.. & Pisoni, D.B. (1998). Recognizing spoken words: the neighborhood activation model. *Far and Hearing. 19,* 1-36.

McCarthy. J.. & Prince. A. 11995). Prosodic morphology. In J. Goldsmith tRd.i. *Handbook ofphonoiog)* (pp. 318-366). Oxford. England: Blaekwell.

McClelland. .1.1... & Hlman. J.I.. I 1986). The TRACK model of speech perception. *Cognitive Psychology, 18,* 1 86.

McQueen, *iM,* 11998), Segmentation of continuous speech using *phcmolaetics. Journal aj Memory and Language,* 39, 21-46,

McQueen. J.M.. & Cutler. A (I99NI. Spoiling (different kinds of) words in (different kinds of) context, In *Proceedings of the Filth International Conference on Spoken Language Processing* (Vol. 6, pp. 2791-2734). Sydney. Australia: Australian Speech Science and Technology Association.

McQueen, J.M.. Norris. t).. & Cutler. A. (1994) Competition in spoken word recognition: Spoiling words in other words. *Journal of Experimental Psychology: Learning, Memory, and ('Ognitfon, 2<>.* 621-638.

McQueen, J.M.. Olake. T. & Culler. A. 120011. Rhythmic cues anil possible-word constraints ill Japanese speech segmentation. *Journal of Memory and Language. -15.* 103 I 32.

Myers, s i 19X7). *Tone and the structure q) words in Shona.* Unpublished doctoral dissertation. University ol Massachusetts, Amherst.

Norris, I). (1994). Shortlist: A connecliQnisI model ≫i continuous speech recognition. *Cognition, 52.* IN9-234.

Norris. I).. McQueen. J.M.. & Cutler. A. i l'JOsi. Competition and segmentation in spoken word recognition. *Journal iff Experimental. Psychology: Learning, Memory, and Cognition,!!.* 1209 1228.

Norris. D.. McQueen. J.M.. Culler. A., $ Butlerlield. S. (11997), The possible-word constraint in the segmentation of continuous speech. (*ognltive Psychology, 34,* 191 243.

Norris, D,, McQueen. J.M. Cuder. A.. Butlerlield. S,, & Kearns, R. i20lil1. Language-universal constraints on speech segmentation, *Language and Cognitive Processes, l'6* 637-660,

Ot.ikc. T. Hatano. C... Culler. A.. & Mehler. J. 11993). Mora or s> liable.' Speech Segmentation in Japanese. *Journal of Memon and Language, 32,* 358—378.

Tabossi, P.. Buranl, C & Scott, D.R. 11995). Word identification in fluent *speech. Journal of Memory and language.* >/. 4-10 467.

(RECEIVED 2/21/01 REVISION ACCEPTED 7/25/011