DOES LEXICAL STRESS OR METRICAL STRESS BETTER PREDICT WORD BOUNDARIES IN DUTCH?

David van Kuijk

Department of Language and Speech, University of Nijmegen, The Netherlands /
Max-PIanck-Institute for Psycholinguistics, Nijmegen, The Netherlands
Snail-mail: P.O. Box 310, NL-6500 AH Nijmegen, The Netherlands
E-mail: kuijk@mpi.nl

ABSTRACT

For both human and automatic speech recognizers it is difficult to segment continuous speech into discrete units such as words. Word segmentation is so hard because there seem to be no self-evident cues for word boundaries in the speech stream. However, it has been suggested that English listeners can profit from the occurrence of full vowels (i.e. vowels with metrical stress) in the speech stream to make a first good guess about the location of word boundaries. The CELEX-database study described here investigates whether such a strategy is also feasible for Dutch, and whether the occurrence of full vowels or the occurrence of vowels with primary word stress (i.e. vowels with lexical stress) is a better cue for word boundaries. The CELEX-counts suggest that for Dutch metrical stress seems to be a better predictor of word boundaries than lexical stress.

1. INTRODUCTION

Writers put white spaces between words, but speakers normally do not separate words by silences. This absence of pauses or any other deterministic cue to word boundaries makes it hard to segment fluent speech into discrete words. Lack of word boundary information implies that automatic speech recognizers have to assume that at each moment a new word may begin. This assumption results in many superfluous word hypotheses, and therefore in a huge search space in which many hypotheses have to be considered simultaneously. The computation of the probabilities for every path through this large search space slows down the processing of speech. If one could find reliable information about word boundaries in the speech stream, it might be possible to speed up the recognition process.

In this paper I investigate whether a word segmentation strategy which is proposed for English listeners could be used in Dutch automatic speech recognizers, but I will not yet suggest any concrete implementation of this segmentation strategy in a state-of-the-art recognizer. First I go into findings for the English language which support the idea that humans use a strategy of segmentation which is based on metrical stress (the occurrence of full vowels). Then I describe a CELEX-database study which investigates whether such a stress-based segmentation strategy is also feasible for Dutch, and whether metrical stress or lexical stress (the primary word stress) is a better cue for word boundaries.

2. BACKGROUND

Cutler and Norris (1988) found that for humans words were more difficult to spot at the beginning of bisyllabic nonsense strings when the second syllable had a full vowel (such as *mint* in *mintayve*) than when the second syllable had a reduced vowel (*mint* in *mint@f*). They proposed the Metrical Segmentation Strategy (MSS), which states that listeners try to segment the speech signal at the onset of strong syllables (i.e. syllables with a full vowel). Segmentation would occur at the second syllable of *mintayve*, leaving the listener with the nonwords *min* and *tayve*. Detection of *mint* would then ask for a second parse of the input, in which this putative boundary is dismissed. In the case of *mint@f* there is no segmentation at the second syllable, because this syllable is weak. So no second parse is necessary and detection of *mint* can be faster than in the *mintayve* case.

But why would listeners employ such a strategy? The answer is probably that it helps them to make good guesses about word boundaries in the speech stream. For that to be the case there should be relatively many words which start with strong syllables, and there should not be too many strong syllables in other positions in words. In other words: The probability p(S|F) that a syllable is strong when it is the first syllable of a word should be high, and the probability p(F|S) that a syllable is the first syllable of a word when it is strong should also be high.

For English these probabilities were estimated by Cutler and Carter (1987). They distinguished between lexical and function words because it was assumed that humans process these words in different ways, and that the MSS is mainly used to find the onsets of lexical words. Their counts in a corpus of 190,000 words of spontaneous British conversation showed that for English a strategy postulating word boundaries at the onset of strong syllables would be highly successful for finding the word boundaries of lexical words, because 74% of the strong syllables encountered were indeed initial syllables of lexical words (so p(F|S) = .74), while 11% were initial syllables of function words, and 15% were non-word-initial syllables. The MSS would find many of the word boundaries in fluent speech, because 90% of the lexical words began with strong syllables (so p(S|F) = .90). Of the weak syllables, 69% were initial syllables of function words, 26% were not word-initial, and 5% are initial syllables of lexical words.

A similar investigation for Dutch is not easy, since there is no transcribed Dutch corpus of spontaneous speech available. But based on the CELEX-database (Baayen, Piepenbrock, and van

Rijn, 1993) one can try to estimate the benefits of the MSS for lexical segmentation in Dutch. Vroomen and de Gelder (1995) did such a study for a subset of the lexical words in CELEX. In their counts they had to compensate for the fact that in Dutch reduction of unstressed short vowels is not as compulsory as in English. Van Bergem (1993) gave as an example of this potential reduction the word *banaan* (banana), which may be realized as *baNAAN* or *b@NAAN*. Similar words are not forced to similar reduction, as was illustrated with the word *banier* (banner), which normally can not be realized as *b@NIER*. In CELEX these examples are all coded with their full vowel and there is no reasonable way to estimate how often the reductions are likely to take place in real speech. Therefore Vroomen and de Gelder formulated two criteria for the weakness of a vowel. The "Dutch criterion" was that a vowel was counted as reduced only if the vowel in CELEX was a schwa. According to the other criterion, which was called the "English criterion", a short vowel followed by a syllable with lexical stress was also reduced. The two criteria can be seen as extremes, so that the number of weak vowels for spontaneous Dutch speech will be somewhere in between. Vroomen and de Gelder found that for 30690 Dutch lexical words 87.7% have a strong initial syllable according to the Dutch criterion, and 81.4% according to the English criterion. This result suggests that p(S|F) lies somewhere between .814 and .877 for Dutch lexical words.

The previous study for Dutch does not tell us how far an implementation of a stress-based segmentation strategy would be able to correctly identify word boundaries, and whether such a strategy could be better based on lexical stress or metrical stress. In the next section I will define the concepts metrical stress and lexical stress more precisely, and formulate a criterion on which both proposals of the segmentation strategy (based on either metrical stress or lexical stress) will be judged.

3. METHOD

In order to generalize the CELEX-counts to fluent speech as much as possible, the following three decisions are made. First, the counts are weighted for word frequency, as listed in CELEX, so as to estimate the frequency of occurrence of each word in fluent speech.

Second, the lexical stress as listed in CELEX is adapted. In the database all words are listed with their citation form stress, so that for instance even monosyllables containing a schwa have lexical stress. Function words are also listed with lexical stress, but this stress is seldom realized in fluent speech. A reasonable assumption seems to be that monosyllables which have a schwa, as well as articles, pronouns and conjunctions do not have word stress. This gives a definition of lexical stress, but in order to see how strongly this definition influences the final results for the probabilities, two more extreme (and more unlikely!) definitions are used, thus giving a total of three definitions of lexical stress:

- Definition 1: Monosyllables which have a schwa do not have lexical stress, all other words have lexical stress as listed in CELEX.
- Definition 2: Monosyllables which have a schwa, and articles, pronouns and conjunctions do not have lexical stress. All other words have lexical stress as listed in CELEX.
- Definition 3: Monosyllables which have a schwa and all function words do not have lexical stress. Content words (nouns, verbs, and adjectives) have lexical stress as listed in CELEX.

Going from definition 1 to 3 the number of stressed words decreases.

Third, I adopted the two definitions of metrical stress which were formulated by Vroomen and de Gelder (1995), because it is necessary to consider the possibility that some vowels are listed as full vowels, but will often be pronounced as reduced vowels:

- Dutch criterion: A vowel is counted as reduced only if the vowel is a schwa in CELEX.
- English criterion: A vowel is counted as reduced if the vowel is a schwa in CELEX, or if it is a short vowel followed by a stressed syllable.

There is no interaction between the definitions of lexical stress and those of metrical stress.

For all these definitions the probabilities p(S|F) and p(F|S) were computed, as well as p(U|F) and p(F|U), where U stands for unstressed. The calculations were:

p(S|F)= n(S&F)/n(F) (and so p(U|F)= n(U&F)/n(F))

p(F|S)= n(S&F)/n(S) (and so p(F|U)= n(U&F)/n(U))

where n(S&F) is the number of word-initial stressed syllables, n(F) is the number of first syllables (or the number of words), and n(S) is the number of stressed syllables.

The examination of toy lexica (see Figure 1) can illustrate how the distribution of strong syllables over the lexicon influences the likely success rates of segmentation strategies based on lexical versus metrical stress. These toy lexica are clearly far removed from any realistic lexicon of a human language, but suffice for this goal. For each lexicon the various probabilities are given in the figure. For lexicon 1a a stress-based segmentation strategy would clearly work well. Not only would the chance that a putative word boundary is really there be 80% (because p(F|S)=.80), but the segmentation strategy would also find 80% of the boundaries (because p(S|F) = .80). For lexicon 1b the boundaries found by the strategy correspond to real word boundaries in 66.6% of the cases (because p(F|S) = .66), but only 40% of the word boundaries are found (because p(S|F) = .40). For lexicon 1c the stress-based strategy would be a bad idea; both probabilities are below chance level. A strategy where segmentation is attempted at the onset of weak syllables would have more success, because p(U|F)>p(S|F) and p(F|U)>p(F|S). For lexicon 1d it is less clear whether to hold on to a stress-based strategy or not; many type II errors (proposing a word boundary when it is not there) will be made because p(F|S) is only .40, but most of the actual boundaries will be detected because p(S|F) is .80. This dilemma is even more clearly illustrated in lexicon 2. What if the distribution of syllables with lexical stress and those with metrical stress is similar to this distribution in

Toy lexicon 1a	Toy lexicon 1b	Toy lexicon 1c
sww	sww	sww
s	w	w
sww	sww	ssw
wwsw	wwsw	wssw
s	w	w
p(S|F)= 4/5, p(F|S)= 4/5, p(U|F)= 1/5, p(F|U)= 1/7	p(S|F)= 2/5, p(F|S)= 2/3, p(U|F)= 3/5, p(F|U)= 3/9	p(S|F)= 2/5, p(F|S)=2/5, p(U|F)= 3/5, p(F|U)= 3/7
Toy lexicon 1d	**Toy lexicon 2 (Lexical stress)**	**Toy lexicon 2 (Metrical stress)**
sss	Suu	sww
s	S	s
sss	Suu	sww
wssw	uuSu	wssw
s	u	s
p(S|F)= 4/5, p(F|S)= 4/10, p(U|F)= 1/5, p(F|U)= 1/2	p(S|F)= 3/5, p(F|S)= 3/4, p(U|F)= 2/5, p(F|U)= 2/8	p(S|F)= 4/5, p(F|S)=4/6, p(U|F)= 1/5, p(F|U)= 1/6

Figure 1: Several toy lexica are shown which illustrate the effect of the distribution of stressed syllables over the lexicon on the predictive power of a stress-based segmentation strategy. Every 's' represents a strong syllable and every 'w' a weak syllable. In toy lexicon 2 the 'S' stands for a syllable with lexical stress, and the 'u' for a syllable without lexical stress.

lexicon 2? The p(S|F) is lower for lexical stress than for metrical stress, but for p(F|S) it is the other way around. We then have to decide whether we want to minimize the type I error (failing to propose a word boundary when it is there) or the type II error.

To clarify this problem we can imagine two recognizers which use a stress-based segmentation strategy in a different manner. The first recognizer would only generate word hypotheses at the onset of stressed syllables. The second recognizer is more like the state-of-the-art recognizers and tries lexical access at each time-frame, but the probability of the word hypotheses is influenced by the probability of a word-boundary, either during the Viterbi search or in a second parse afterwards. Word hypotheses which are not aligned with the suggested boundaries would be penalized. For the first recognizer it would be very important to find as many word boundaries as possible (i.e. to minimize the type I error), because for every word boundary which is not found there is also no word hypothesis. For the second recognizer it is more important to make sure that the proposed word boundaries are reliable (i.e. to minimize the type II error), because for any erroneously proposed word boundary potentially good word hypotheses are penalized. Since it is my goal to try and implement the strategy in a modern recognizer, I will first evaluate the two forms of stress on their type II error, so p(F|S) should be high. However, should the p(F|S) for lexical stress be similar to that of metrical stress, then p(S|F) can be used to decide between the two strategies.

4. THE EXPERIMENT

For the counts all 311,291 different Dutch word forms listed in CELEX were used. These word forms contained 18,482 split words (e.g. *aapte na*). These were treated as separate words, thus leading to a total of 329,773 words. The majority of these words were not used in the weighted condition, because 206,482 of them had a frequency of zero. If the remaining words are weighted for their frequency this makes 39,803,130 words, which is 93,9 % of the original 42,380,000 words of the corpus on which the word frequencies in CELEX were based.

5. RESULTS

The results are summarized in Tables I and II. As was expected, we can see that the definition of lexical stress is very important for the results, because the function words are used very often. For metrical stress the difference in results for the two definitions is marginal. Table I shows that more word-initial syllables are stressed in the metrical stress conditions, which is what I expected because the lexically stressed syllables are a subset of the metrically stressed syllables. But therefore also the number of full syllables which are *not* word-initial is higher, so one would expect that the predictive power of metrical stress is lower than that of lexical stress, which appears from Table II. We can also see that for Definition 3 both p(F|U) and p(F|S) are high. This seems strange at first sight, but can be explained by the fact that p(F|U) and (F|S) also reflect the average word-length in terms of syllables in the language. If all words were monosyllabic, these probabilities would always be 1; word segmentation is then of course completely independent of syllable-stress.

Based on the criterion of minimizing the Type II error, lexical stress seems to be the better choice, because P(F|S) is the highest for Definitions 1 and 2. Since Definition 1 (according to which, for instance, the high-frequent word *ik* (I) has lexical stress) is considered a very unlikely definition for the occurrence of stress in fluent speech, I will not discuss it here anymore. Also for

Probabilities	Lexical stress Definition 1	Lexical stress Definition 2	Lexical stress Definition 3	Metrical stress Dutch criterion	Metrical stress English criterion
P(S|F)	.701	.543	.370	.788	.759
P(U|F)	.299	.457	.630	.212	.241

Table I: The probability that a syllable is stressed or unstressed when it is the first syllable of a word, given for various definitions of stress and weighted for frequency.

Probabilities	Lexical stress Definition 1	Lexical stress Definition 2	Lexical stress Definition 3	Metrical stress Dutch criterion	Metrical stress English criterion
P(F|S)	.822	.781	.709	.729	.725
P(F|U)	.397	.501	.581	.405	.431

Table II: The probability that a syllable is the first syllable of a word when it is either stressed or unstressed, given for various definitions of stress and weighted for frequency.

Definition 2 some reserves have to be made. In the light of the assumptions made with respect to the generalization of lexical stress and word frequencies to fluent speech, the difference between p(F|S) for Definition 2 and p(F|S) for metrical stress is small (about 5%), which invites us to look at some other factors which determine the success of a stress-based segmentation strategy. First, the type I error is much (about 20% to 25%) lower for metrical stress, as can be seen by comparing p(S|F) for Definition 2 with that of metrical stress. Second, the reliable detection of stress is a factor which also must be taken into account, because the chance that there is a word boundary at the onset of a stressed syllable p(WB|S) will also be dependent on the quality of the detection:

$$p(WB|S) = p(stress) * p(F|S)$$

Taking the quality of the stress-detection itself into account gives an advantage to metrical stress, because metrical stress is probably easier to detect than lexical stress. The first is based on local characteristics of the spectrum at a certain time, while lexical stress is a relational phenomenon which has to take the context into account.

6. CONCLUSION

In this study the success of a stress-based segmentation strategy for words based on either lexical stress or metrical stress was investigated. The task of this strategy in our recognizer would be to indicate places in the speech signal where word boundaries are very likely, so that word hypotheses which are not aligned with these word boundaries can be disfavoured. Therefore it is important that the detection of word boundaries is maximally reliable. For the present study this means that the chance that a syllable is word-initial when it is stressed p(F|S) should be high. For lexical stress p(F|S) was 5% higher than for metrical stress, but this difference is small in the light of the assumptions made in this study. Based on the arguments that a segmentation strategy based on metrical stress would find 25% more word boundaries, and that metrical stress is probably easier to detect, I conclude that metrical stress is the better candidate for a stress-based segmentation strategy.

7. REFERENCES

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

Bergem, D. R. van (1993). "On the perception of acoustic and lexical vowel reduction". *Proceedings of the 3rd European Conference on Speech Communication and Technology*, Berlin, 689-692.

Cutler, A. (1986). "Forbear is a homophone: Lexical prosody does not constrain lexical access", *Language & Speech*, 29 (3), 201-220.

Cutler, A., and Carter, D.M. (1987). "The predominance of strong initial syllables in the English vocabulary". *Computer, Speech and Language*, 2, 133-142.

Cutler, A., and Norris, D. (1988). "The role of strong syllables in segmentation for lexical access". *Journal of Experimental Psychology: Human Perception and Performance*, 14(1), 113-121.

Vroomen, J., and de Gelder, B. (1995). "Metrical segmentation and lexical inhibition in spoken word recognition". *Journal of Experimental Psychology: Human Perception and Performance*, 21, 98-108.