

Lexical stress and spoken word recognition: Dutch vs. English

Klaske van Leyden and Vincent J. van Heuven

0. Introduction

One of the current issues in auditory word recognition concerns the role of lexical stress.¹ In stress-accent languages such as Dutch and English, stress does not occur in fixed position with respect to word boundaries and is therefore available as a potential determinant of word identity. Studies investigating to what extent lexical stress narrows down the cohort of possible word candidates have so far produced a conflicting pattern of results. Cutler & Clifton (1984) found that prior knowledge of stress pattern does not facilitate lexical decision responses. They also reported that the strong correspondences between grammatical category and stress pattern in disyllabic English words (strong-weak stress being associated primarily with nouns, weak-strong with verbs) are not exploited in the recognition of isolated words. This pattern of results suggests that lexical stress information is not used to narrow down the cohort of potential word candidates and thereby speed word recognition. However, van Heuven (1984) found that Dutch listeners performing a gating task with isolated words only need the first syllable of the target word to know whether this syllable is lexically stressed or not. Yet, subjects were biased to respond with initially stressed words when segmental information was poor. Van Heuven (1988) reported evidence that stressed versus unstressed realisations of otherwise identical word-initial full syllables effectively narrowed down rhythmically different cohorts of word candidates. These findings indicate that lexical stress information may facilitate word recognition.

Several studies have investigated to what extent the word recognition process is impaired when words are incorrectly stressed. The rationale behind this is that the more incorrectly placed stress impairs the recognition of words, the more important the role of stress is in the word recognition process. In other words, if lexical stress information is functional, then its distortion should impair spoken word recognition. Cutler & Clifton (1984), van Heuven (1985) and Slowiaczek (1986) showed that when words are deliberately mis-stressed, word recognition is

¹ Thanks are due to Maarten Hijzelendoorn for multiple consultations of the CELEX database in order to establish the cohorts for various rhythmical types of words in Dutch and English. Also, we gratefully acknowledge the support and hospitality offered by the German Department of Edinburgh University, where the gating tests with English listeners were carried out. Finally, we thank Hugo Quené and Mariette Koster for comments on an earlier version of this manuscript.

delayed. This strongly suggests that the stress pattern is part of the lexical representation accessed in word recognition. When there is a clash between stored and perceived information, word recognition suffers.

The studies investigating the effect of mis-stressing on the recognition of words have, however, yielded contradictory results for English and Dutch. Cutler & Clifton (1984) found that recognition of English disyllabic words that were incorrectly stressed on the final syllable (e.g. **classic*), was severely delayed (up to 200 ms delay in a semantic category decision task), while the recognition of words with incorrect stress on the initial syllable (e.g. **typhoon*) hardly suffered. They offer the following explanation for this asymmetry: English listeners are familiar with "incorrect" stress on the initial syllable because this kind of stress shift regularly occurs in spoken English, namely when words with stressed (heavy) final syllables are used attributively, for example, *thirteen*, but *thirteen MEN*. This so-called iambic reversal occurs in Dutch as well, e.g. *katholiek*, but *katholieke Eredienst* 'Catholic worship'. However, van Heuven (1985), using a gating task (see below), reports that for Dutch, mis-stressing has the opposite effect: stress front-shift significantly impairs recognition (e.g. **Kapitein* 'captain'), while word recognition hardly suffers from stress back-shift (e.g. **paprika* 'green pepper').

The question now arises whether this discrepancy between the results for Dutch and English is an artifact of the different experimental methods that were used, gating with synthetic speech for Dutch versus a category monitoring (reaction-time) experiment with natural speech for English, or whether it can be explained in terms of structural differences between the two languages. To address this question we investigated the effect of mis-stressing on the recognition of spoken words in two comparative experiments, one for Dutch and one for English. The experiments were set up to be exactly the same: the recognition of a similar set of words (matched across the two languages in terms of word length and stress position and absence of vowel reduction in unstressed syllables) spoken in fixed carrier phrases with the same variation of correct and incorrect stress patterns was tested using (Dutch and English) university students as subjects. If the earlier results for Dutch and English are corroborated, we will accept the conclusion that the discrepancy noted above did indeed originate from structural differences between Dutch and English. However, should we find similar results for both languages, we will conclude that the earlier conflicting results can be ascribed to a difference in experimental techniques.

1. Method

The gating paradigm was first introduced by Grosjean (1980) and entails the repeated presentation of a word of which the presentation time, measured from the onset of the word, increases with each successive presentation. After each

presentation, subjects are asked to guess the word being presented. Since gating provides information about the narrowing-in process employed by listeners in the recognition of words, we used this method in both experiments. In the responses of a gating task we can also determine the length of the initial stimulus proportion that is necessary for correct recognition of the target.

Gating is an efficient and easily administered off-line word recognition task which has been advanced to simulate certain aspects of the on-line recognition process. We take the view that the on-line recognition process is adequately covered by gating as long as it does not rely on semantic and/or syntactic top-down information streams, i.e. as long as word recognition solely depends on properties of the input signal and lexical constraints (cf. Jongenburger 1996). In the present experiment, word recognition of single targets is studied in a semantically and syntactically non-constraining context, so that gating is an admissible choice of method. Note, moreover, that the choice of method is largely immaterial for the present study: as long as the same method is used in both languages, the results will always be conclusive. If the discrepancy between the two languages disappears, we know that the earlier results were caused by a difference in experimental task. In that case, a subsequent decision will have to be made whether the English on-line data or the Dutch off-line data are more credible.

1.1 Materials. The CELEX database (Burnage 1990) was employed to retrieve Dutch and English monomorphemic nouns. Stimuli for Dutch were 16 disyllabic and 27 trisyllabic monomorphemic nouns of low frequency of occurrence. In order to shift stress from the syllable that normally receives lexical stress to another syllable without affecting vowel quality, all words that were selected had a full vowel (i.e. no schwa) in the unstressed syllable(s). Of the disyllabic words, 8 had stress on the first syllable (Sw), the other 8 on the second (wS). The 27 trisyllabic words were evenly distributed over types with initial (Sww), medial (wSw) and final (wwS) stress.

As English words often have a schwa in their unstressed syllable(s), our choice of stimuli for the English version of the experiment was rather limited. The 43 monomorphemic nouns we selected had at least one unstressed syllable with a full vowel. There were ten instances of Sw, wS, Sww and wwS stress patterns, but only three wSw words (there are simply no more suitable words in this category). The full set of stimuli for both languages is included in the appendix.

Context sentences were constructed such that each word was pronounced once with a contrastive pitch accent on the syllable that normally carries lexical stress and also once (or twice, in trisyllabic words) with a contrastive accent on a lexically unstressed syllable. The same context sentence was used for each of the target words. Below are examples for Dutch as well as for English.

- (1) a Zei je BAZON? Nee, ik zei BIZON
 'Did you say bazon? No, I said bison'
 b Zei je bIZAN? Nee, ik zei bIZON
- (2) a Do you say shamPEE? No, I say shamPOO
 b Do you say SHIMpoo? No, I say SHAMPoo

The Dutch sentences were digitally recorded by a male native speaker of standard Dutch, the English sentences by a male native speaker of standard British English. The recordings were downsampled to 16 kHz and stored on computer disk.

The target words, together with the neutral carrier sentence *ik zei* or *I say* were digitally excerpted from the context sentence. Outside the original context sentence, words that are pronounced with a pitch accent on a lexically unstressed syllable, will be incorrectly stressed. Using a digital waveform editor, the utterances were cut into fragments of increasing length, under visual and auditory control. The first gate consisted of the preceding context plus the initial phoneme of the target word. Each next fragment contained one phoneme more, until the whole word was made audible. The total number of gates depended on the length of the individual target word.

For each experiment three experimental tapes were created such that each lexical word occurred only once per tape, with correctly stressed and incorrectly stressed words in random order. Thus, a target word with correct stress on one tape was presented with incorrect stress on the other tapes.² The tapes for the experiment contained 258 stimuli (gates) each for the Dutch version and 254 stimuli each for the English version. Both versions contained 43 test words with between 3 and 8 gates per word. A control tape, to be played to a fourth group of listeners contained correctly stressed words only, in order to check whether alternation of correctly and incorrectly stressed words negatively affects the subjecttask performance. The interstimulus interval was 5 seconds; an alert tone was recorded 1 second prior to each stimulus onset.

1.2 Subjects and procedure. Forty native speakers of Dutch participated in the experiment with Dutch stimuli (students of Leyden University) and forty native speakers of British English (students of Edinburgh University) took part in the English edition of the experiment. The subjects were tested in small groups, ten per experimental tape, in a language laboratory in experimental sessions lasting

² The design was, in fact, slightly more complicated so as to balance the distribution of the 16 (correct or incorrectly stressed) disyllabic words evenly over the three tapes. In order to solve this problem the "third" tape was produced in two versions, which were presented to two subgroups of listeners.

approximately 45 minutes. The stimuli were presented over headphones at a comfortable listening level. Subjects were instructed that they were going to listen to polysyllabic words or word fragments and that their task was to write down the complete word they believed was being presented, with an unlimited choice from the Dutch (or English) lexicon. Subjects were required always to write down a word, even if they had to guess. They also had to indicate on a 10-point scale how confident they were as to the eventual correctness of their response. Preceding the experiment there was a short practice session.

2. Results

A total of 40 (subjects) \times 258 (stimuli, i.e. gates, per list) = 10,320 responses for Dutch and 40 \times 254 = 10,160 responses for English were collected. With the exception of a few cases where a subject apparently did not know a particular word, all target words, irrespective of stress condition, were recognised at or before the last gate.

In order to be able to compare results across words, gate length (i.e. the duration of the audible word fragment) was expressed as percentage of the total word duration. For each word a subject-individual Isolation Point (IP) was then defined as the relative duration of the gate (in percent of word duration) where the subject correctly completed the word for the first time and did not change his response at any later gate for the same word.

Results for the confidence ratings were analysed but will not be reported here *in extenso*. Confidence ratings increased monotonically with the position of the isolation point. Clearly then, confidence increases as the listener completes the word from a larger word-initial fragment. The effects of all other factors (correct versus incorrect stress position, type of word, mixed versus correct-only stimulus lists) were negligible and statistically insignificant. Therefore, in this experiment, confidence ratings afford no insight additive to what we may learn from the analysis of the IP data.

Figure 1 presents mean isolation point for correctly and incorrectly stressed words, collapsed over di- and trisyllabic words and broken down by language.³

³ Our presentation of results is based on the three groups of listeners that responded to the tapes with mixed correct and incorrect stress patterns. Somewhat to our surprise, the results obtained for the control tape with correct stress patterns only did not deviate in any systematic way (in mean isolation points nor in confidence ratings) from the mixed tapes.

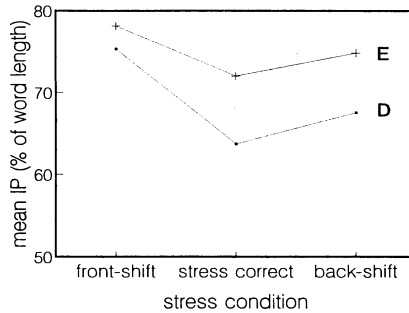


Figure 1: Mean relative isolation point (in percent of total word duration) for correctly and incorrectly stressed (front-shift and back-shift) words, collapsed over di- and trisyllabic words, broken down by language (English vs. Dutch).

As can be seen in figure 1, there is large main effect of stress condition. Stress front-shift (FS), when compared to the average isolation point for correctly stressed stimuli, delays the IP by 6.1 percentage points for English and by as much as 11.6 percentage points for Dutch. The effect of stress back-shift (BS) is smaller: the delay is 2.8 and 3.8 percentage points for English and Dutch, respectively. The main effect of stress condition is significant by separate one-way analyses of variance for Dutch and English with stress condition as a fixed factor, $F(2,1264)=37.3$ ($p < .001$) and $F(2,1250)=10.3$ ($p < .001$). Post hoc analyses for contrasts (Newman-Keuls procedure) showed that all three stress conditions differ from each other at the .05-level for both Dutch and English. Crucially, for both languages alike, FS increases the delay of the IP more than BS.

Figure 2 presents mean IP for correct and incorrect (BS and FS) stress patterns broken down by the individual stress types, for Dutch and for English.

Examining the results for Dutch, we observe that, irrespective of lexical stress type, the effect of FS (relative to the IP for correctly stressed items) is considerably larger than the effect of BS. To give a few examples: when *olifant* 'elephant' is pronounced as **oLifant* or **oliFANT* the IP is delayed by less than 2 percentage points; yet, when *kapiTEIN* 'captain' is incorrectly stressed as *wSw* or *Sww*, mean IP is delayed by some 16 percentage points.

For English, mean isolation points of incorrectly stressed stimuli vary from one word type to the next.⁴ We observe that, with respect to disyllabic words, FS as well as BS cause a delay of about 10 percentage points (relative to their

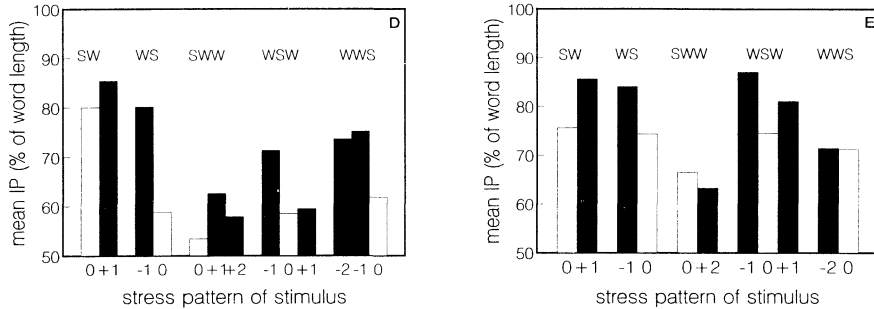


Figure 2: Mean isolation point broken down by lexical stress type and stress condition, for Dutch (left) and English (right). 0: stress correct; -1 and -2: stress front-shifted by 1 or 2 syllables, respectively; +1 and +2: stress back-shifted by 1 or 2 syllables (further, see figure 1).

correctly stressed counterparts). In the case of the trisyllabic words, both FS and BS delay isolation of wSw words; the recognition process hardly suffers when wWS words are realised incorrectly as Sww; what is more, when Sww words are wrongly pronounced as wwS, they are isolated even earlier, on average, than their correctly stressed counterparts. For example, **porcupine* or **suicide* are isolated by more than 10 percentage points earlier than *porcupine* or *suicide*. Summarising, we can say that, regarding English, the recognition process suffers slightly - but significantly - more from FS than from BS.

In order to investigate to what extent lexical stress helps the listener to narrow down the cohort of potential word candidates, an analysis of metrical properties was made of the error responses to the first syllable, i.e. accumulated over between maximally 4 gates, depending on the individual word. Monosyllabic content words were considered initially stressed, monosyllabic function words as initially unstressed; ambiguous responses (less than 1% of the total) were discarded.

Figure 3 (below) presents the results of the error response analysis for Dutch and for English.

With respect to both Dutch and English, it appears that, regardless of the lexical stress position, when words are correctly or incorrectly stressed on the

⁴ Not all theoretically possible stress conditions have been exploited. The reason for this is the following. Mis-stressing a trisyllabic word on the medial syllable is not possible in English because there are apparently no English Sww or wwS nouns that have an unreduced vowel in the medial syllable.

first syllable, about 80% of the responses are words with initial stress. When words receive non-initial stress this figure drops by more than 30 percentage points for Dutch, while in the case of English this decrease, at some 15 percentage points on average, is considerably smaller. So, on the whole, the bias towards perceiving stress on the first syllable, regardless whether this syllable receives stress or not is stronger in English than in Dutch.

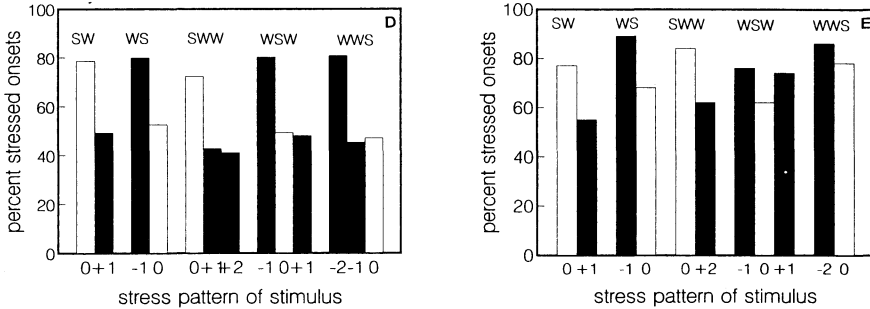


Figure 3: Percentage of stressed word onsets in error responses to first syllables, broken down by lexical stress type and stress condition, for Dutch (left) and English (right) (further, see figure 2).

To gain more insight into the temporal development of the perception of stress, we also performed a rhythmic analysis of the error responses to the first two phonemes of the stimuli. Specifically, this procedure will allow us to determine the individual contribution of acoustic information supplied to the identification of the stressed/unstressed nature of the initial syllable by the onset consonant versus that of the vowel. It has generally been claimed in the literature that the perceptual cues for stress (duration, intensity and spectral quality) are located in the vocalic nuclei of syllables, rather than in the consonants. We predict from this that no effect of stress pattern in the first gate will be found. This would be in contrast to a claim made by Cutler and co-workers that the presence of an upcoming stress can be predicted by the listener from the prosody of the preceding context (Cutler, 1976; Cutler & Darwin, 1981). In this case the listeners should be able to determine the stressed nature of our target word's onset at - or even before - the first (onset consonant) gate.

The results of the rhythmical analysis carried out to choose between these competing predictions are presented in figure 4.

It is apparent from figure 4 that the first consonant of a particular stimulus does not provide the listener with any useful prosodic information: listeners are biased towards initial stress, and again this bias is larger for English than for Dutch. When the first consonant as well as the following vowel of an unstressed initial syllable have been made audible, however, the number of initially stressed responses drops by nearly 20 percentage points on average.

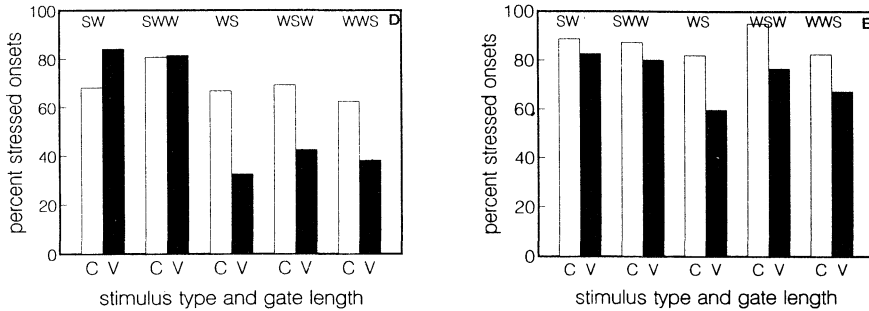


Figure 4: Percentage of stressed word onsets in error responses to gates 1 (C: first consonant audible) and 2 (V: first consonant plus following vowel audible), broken down by lexical stress type and stress condition, for Dutch (left) and English (right).

3. Discussion

Two comparative gating experiments were carried out to investigate whether an observed discrepancy between the effect of mis-stressing on the recognition of spoken words in Dutch and English originates from structural differences between the two languages or can be ascribed to different experimental techniques employed in earlier studies. It was found that, firstly, deliberate mis-stressing impairs word recognition; yet the recognition process suffers more from stress front-shift than from stress back-shift and this effect is larger for Dutch than for English. Secondly, there is a strong bias towards perceiving stress on the first syllable, irrespective of the presence or absence of a prosodically marked stress; this bias is especially strong in English. Finally, prosodic information only becomes available when the first vowel has been made audible; the preceding consonant does not contribute to such information.

The demonstration that mis-stressing delays word recognition is strong evidence that lexical stress information indeed plays a role in word recognition. It appears that, although there is a bias for initially stressed responses, stressed versus unstressed realisations of word-initial syllables effectively narrow down rhythmically different cohorts of word candidates. Therefore, the role of stress and the observed bias should be explicitly accounted for in models of spoken word recognition.

The finding that the recognition process suffers more from stress front-shift than from stress back-shift was expected for Dutch and corresponds to earlier findings by van Heuven (1985). However, as regards English, our data contradict the results of the reaction time experiment reported in Cutler & Clifton (1984), who found that stress back-shift had a detrimental effect (on the recognition of

disyllabic words), while stress front-shift had little effect. Therefore, our results so far suggest that the discrepancy between the outcome of the experiments for Dutch by van Heuven (1985) and for English by Cutler & Clifton (1984) originates from a difference in experimental design. It is unclear at this time whether the discrepancy has been caused by a difference in experimental task (gating in Dutch versus semantic category detection in English) or in type of lexical materials (invariant stress patterns in Dutch versus stress-shift sensitive words in English). Follow-up experiments are needed to solve this issue.

The question now remains why stress front-shift has, on average, a more damaging effect on the recognition process than stress back-shift. An analysis of the individual isolation points for each of the individual test words revealed that, for Dutch as well as for English, the effect of mis-stressing differed considerably from one word to the next and this finding led to the following hypothesis. Deliberate mis-stressing impairs word recognition as soon as an NWP has been reached (Non Word Point, segmentally and prosodically; the earliest point at which the cohort of possible recognition candidates is empty). The later this point is reached, the greater the possibility that a mis-stressed word will be recognised despite an incorrect location of stress. Consequently, when stress is front-shifted so that words are mis-stressed on the first or second syllable, an NWP will be reached more frequently, as well as earlier, than when the final syllable of a word is incorrectly stressed. For example, in Dutch there are many words that begin with *MA* or *ma*; yet, there are no Dutch words that begin with *Maga* or *mAGa*. Thus, when *magAZIJN* 'warehouse' is incorrectly stressed on the initial or medial syllable, the NWP is reached as soon as the vowel of the second syllable becomes audible. Likewise, no word in British English begins with *Flan*, so, when *fIAnCee* is mis-stressed on the first syllable, the NWP is reached in the course of the medial syllable. Conversely, when stress is back-shifted so that a word like *Festival* 'id.' is incorrectly pronounced as *festIVAL*, the NWP occurs after the so-called uniqueness point (i.e., the place within the word where it is first uniquely distinguished from all other words in the lexicon, which, for *festival* is reached at the onset of the final vowel *a*), hence, after recognition of the word based on segmental information.

Apart from leading to an NWP, mis-stressing can also activate the wrong cohort of recognition candidates, which also has a damaging effect on the recognition process. For example, the fragment *baSI*, from the Dutch word *baSILIEK* 'basilica' incorrectly stressed on the medial syllable, prompts listeners to respond *basilicum* 'basil'. Only when the final consonant has been made audible, do listeners change their minds and respond *basiliek*.

At this stage it is not yet clear which has a more detrimental effect on the word recognition process: an incorrect stress which creates an NWP (cohort empty) or an incorrectly located stress which activates the wrong cohort. So far, little or nothing is known about the kind of mechanism that could be invoked

here. Moreover, it is highly unlikely that gating is an appropriate technique for probing these time-critical processes.

Exactly when mis-stressing leads to an NWP, or activates the wrong cohort of recognition candidates can be established on the basis of the lexicon. A pilot investigation based on the CELEX databases for Dutch and English has revealed that in those cases where an incorrectly located stress severely impaired word recognition in our experiment (typically occurring with stress front-shifts), mis-stressing indeed led to either an NWP early in the word or activated the wrong cohort.

Finally, the bias favouring initial stress, which was stronger in our English data than in the Dutch data, is most likely related to the distribution of stress patterns in the lexicons of the two languages. Both in Dutch and in English, primary stress generally falls on the initial syllable of a (compound) word: 66% for Dutch (van Heuven & Hagman 1988) and 61% for English (Cutler & Carter 1987). Note that these are lexical frequencies, which do not reflect frequency of occurrence in actual language use. A 80/20% token frequency distribution favouring primary stressed over secondary and unstressed word-initial syllables in English has been reported by Cutler & Carter (1987). No such token frequency count is available for Dutch at this time.⁵ We would predict, from our experimental results, that the proportion of stressed word-initial syllables is smaller in Dutch than in English.

Appendix: stimulus words

Dutch:

Sw: altaar, armoe, bivak, kilo, koffie, koning, lichaam, pinda
 wS: balkon, copie, idee, kantoer, konijn, moeras, radijs, vulkaan
 Sww: bariton, festival, horizon, lucifer, marathon, olifant, pagina, paprika, pergola
 wSw: andijvie, bacterie, embargo, fiasco, kanarie, mitella, parochie, piano, vakantie
 wwS: amulet, basiliek, calorie, document, formulier, kapitein, legioen, magazijn, paradijs

English:

Sw: arrow, aspect, coffee, curfew, herring, impulse, rhubarb, termite, turmoil, virtue
 wS: antique, cartoon, cartoon, cigar, duet, guitar, hotel, pontoon, settee, shampoo
 Sww: alibi, anecdote, appetite, golliwog, imbecile, paradise, porcupine, restaurant, revenue, suicide
 wSw: fiancée, inferno, stiletto
 wwS: accolade, bagatelle, balustrade, carousel, cavalcade, fontanelle, jamboree, macaroon, personnel, tambourine

⁵ In the potentially relevant table I in Quené (1992: 350) primary and secondary stresses were lumped together due to the information structure of the CELEX database, which precludes proper comparison with the English frequencies.

References

- Burnage, G. (1990) *CELEX - A guide for users*, Centre for Lexical Information, University of Nijmegen, The Netherlands.
- Cutler, A. (1976) 'Phoneme monitoring reaction time as a function of preceding intonation contour', *Perception & Psychophysics* 20, 55-60.
- Cutler, A. and D.M. Carter (1987) 'The Predominance of strong initial syllables in English vocabulary', *Computer Speech and Language* 2, 133-42.
- Cutler, A. and C. Darwin (1981) 'Phoneme Monitoring reaction time and preceding prosody: Effects of stop closure duration and fundamental frequency', *Perception & Psychophysics* 29, 217-24.
- Cutler, A. and C. Clifton Jr. (1984) 'The Use of Prosodic Information in Word Recognition', in H. Bouma and D.G. Bouwhuis, eds., *Attention and Performance X*, Erlbaum, Hillsdale NJ.
- Grosjean, F. (1980) 'Spoken Word Recognition Processes and the Gating Paradigm', *Perception & Psychophysics* 28, 267-83.
- Heuven, V.J. van (1984) 'Segmentele versus Prosodische Effecten van Klemtoon op de Woordherkenning', *Verslagen van de Nederlandse Vereniging voor Fonetische Wetenschappen*, 159-162, 22-38.
- Heuven, V.J. van (1985) 'Perception of Stress Pattern and Word Recognition: Recognition of Dutch Words with Incorrect Stress Position', *Journal of the Acoustical Society of America*, 78, S21.
- Heuven, V.J. van (1988) 'Effects of Stress and Accent on the Human Recognition of Word Fragments in spoken Context: Gating and Shadowing', *Proceedings of the 7th FASE/Speech-88 Symposium, Edinburgh*, 811-18.
- Heuven, V.J. van and P.J. Hagman (1988) 'Lexical statistics and spoken word recognition in Dutch', in P. Coopmans and A. Hulk, eds., *Linguistics in the Netherlands 1988*, Foris, Dordrecht, 59-68.
- Jongenburger, W. (1996) *The role of lexical stress during spoken-word recognition*, PhD dissertation, Leiden University.
- Quené, H. (1992) 'Integration of Acoustic-Phonetic Cues in Word Segmentation', in M.E.H. Schouten, ed., *The Auditory Processing of Speech*, Mouton de Gruyter, Berlin, 349-56.
- Slowiaczek, L.M. (1986) *On the Role of Lexical Stress in Auditory Word Recognition*, PhD dissertation, Indiana University.