Processing of Prosody and Semantics in Sepedi and L2 English

Giuseppina Turco, Sabine Zerbian

Please cite as: Turco, G., Zerbian, S. Processing of Prosody and Semantics in Sepedi and L2 English. *J Psycholinguist Res* 50, 681–706 (2021).

https://doi.org/10.1007/s10936-020-09746-z

Processing of Prosody and Semantics in Sepedi and L2 English

A phoneme-detection task shows that listeners of Sepedi use semantic information in processing but not prosody (Experiment 1). Sepedi is a language with no grammaticalised prosodic expression of focus. Sepedi listeners detected phoneme targets faster when the phoneme-bearing words were focussed (as opposed to unfocussed) but not when occurring in a context conducive to prosodic emphasis (as opposed to non-conducive). Experiment 2 tested the role of semantic focus and prosody in processing by Sepedi L1/English L2 listeners (English being a language with systematic focus-to-accent mapping). Non-native listeners detected phoneme-bearing words faster in focussed condition (as opposed to unfocussed) and in accented condition (as opposed to deaccented). The results suggest that the L2 prosodic structure is exploited by Black South African English listeners even if this feature is not present in their L1. Our experiments replicate the pattern of results found in Akker and Cutler's (Biling Lang Cogn 6:81–96, 2003) experiment for Dutch L1/ English L2 listeners, even with listeners whose L1 does not use prosody the way English does.

1. Introduction

Prosodic salience plays an important role in sentence comprehension (Cutler 1976, Terken and Nooteboom 1987; Watson, Tanenhaus, and Gunlogson 2008; Ito and Speer 2008, Braun and Tagliapietra 2010, among others). A phoneme-monitoring study by Cutler (1976) provided a first demonstration of the processing advantage of (contrastive) accents. The author tested the contribution of prosodic context in allocating listeners' attention to a word in a specific position. Participants had to detect target phonemes in acoustically identical words which had been spliced into contexts that contained prosodic cues to either accented or deaccented versions of that word. The experiment showed that the acoustic cues present in the respective prosodic contexts guided listeners' attention towards the phoneme-bearing word on which an accent had originally been realized, thus detecting the target phoneme much faster.

The reason for the facilitation effect found in Cutler (1976) might also have been due to the fact that the accent fell on a semantically focussed word; that is, on a word carrying the most important information. The similarity between prosodic processing and focus processing was tested in a follow-up phoneme-detection study by Cutler and Fodor (1979). Here, the authors manipulated the question preceding the answer (instead of the prosodic context, as in Cutler 1976). It was shown that phoneme-detection responses changed with the preceding question: detection of *[b]* in *blue* was faster when *blue* was the focus induced by the preceding question; that is, it was faster when preceded by the question *Which hat was the man on the corner wearing?* than when preceded by *Which man was wearing the hat?*. Thus, prosodic prominence (e.g. predictable by the surrounding prosodic context) and also semantic focus provide a processing advantage (as e.g. induced by a preceding question). According to Cutler and Fodor (1979), both prosodic and semantic context work towards the identification of semantic focus as an essential component in sentence processing.

Since both preceding prosodic context (Cutler 1976) and semantic context (Cutler and Fodor 1979) equally facilitate processing, Akker and Cutler (2003) tested the prediction that these two factors (prosody and focus) should not be independent of each other if manipulated in the same experiment. Therefore, their phoneme-detection task manipulated both prosodic and semantic context in the same experiment. It involved a question-induced semantic focus condition for which it is predicted that a processing advantage (i.e. faster detection of a target phoneme) occurs with words that provide an answer to a preceding question (i.e. focussed words) rather than by words that do not provide such an answer (i.e. unfocussed words). It also contained a 'predicted-accent' condition, in which processing advantages are expected to occur due to the prosodic contour preceding an originally accented target as opposed to an originally deaccented target. Both semantic processing and prosodic processing are seen as serving the same goal of identifying the semantic focus of a sentence, which is the most relevant information in discourse. In statistical terms, this shared goal is mirrored in an interaction between the factors of semantic focus and prosody. The results from English and Dutch native speakers (Akker and Cutler 2003, figure 1 and 2, p. 86-87) indeed show that prosody does not have an extra effect in the focussed condition, as would be expected if both processing advantages added up.

Evidence that listeners process (or entrain to) prosodic information of an upcoming accent comes mostly from studies testing West Germanic languages like English, German and Dutch (e.g. Dahan, Tanenhaus and Chambers, 2002; Akker and Cutler, 2003; Weber, Braun and Crocker, 2006; Ito and Speer, 2008; Kurumada, Brown, Bibyk, Pontillo and Tanenhaus 2014). In English and Dutch, prosodic focus is marked via pitch accent (e.g., Terken, 1984, Hirschberg, 1993), More recently, prosodic entrainment has been investigated in Mandarin Chinese, a tone language (Ip & Cutler 2017). In Mandarin Chinese, focus is prosodically marked via pitch range expansion (e.g., Xu, 1999). The results of these studies have been interpreted to suggest a universal capacity of native listeners to entrain to acoustic cues that are informative in their own native language (Ip & Cutler 2017: 1220; see also Cutler 2012, ch. 7, for a review and for discussion on universality in prosodic processing). Compared to English, a language in which semantic focus is systematically realized via pitch accent placement and concomitant prosodic modulations, the results on Mandarin Chinese are particularly remarkable. Being a tone language, Mandarin Chinese uses prosody, more specifically F0, primarily for signalling tonal contrast. However, given that Mandarin Chinese also makes systematic use of prosody for focus, native listeners still take advantage of prosodic changes for other processing purposes, namely for prosodic entrainment.

If systematic use of prosody for marking semantic focus is a necessary precondition for prosodic entrainment, then native listeners of a language in which there is no such use are predicted to show no evidence of prosodic entrainment. By using similar experimental conditions than previous studies (Cutler 1976, Cutler and Fodor 1979, Akker and Cutler 2003, Ip and Cutler 2017), the first goal of the current study is to test prosody and semantic focus in Sepedi, a South African Bantu tone language. For this language, neither accent nor phrasing seem to be exploited for the prosodic marking of focus in production (Zerbian 2006, 2007).[1] Previous research showed that in Sepedi, minimal sentence pairs which were elicited under different focus conditions, did not contain any systematic acoustic cues to semantic focus (Zerbian 2006, 2007). When presented with auditory stimuli elicited under different focus conditions, Sepedi listeners did not provide the intended semantic meanings, as tested in an offline perception experiment (Zerbian 2006). However, based on production alone and on rather coarse-grained offline interpretation experiments, it cannot be excluded that there are fine acoustic cues in the signal in connection with focus and/or focal emphasis that listeners could entrain to in an online task. In the current study, Sepedi listeners therefore performed a phoneme-monitoring task (Experiment 1) largely comparable to the one developed in Akker and Cutler's (2003) study on English and Dutch.

Similar to Akker & Cutler's (2003) study, our experiment tests whether prosodic processing parallels semantic processing in Sepedi. Given the absence of grammaticalised prosodic cues to focus in Sepedi, we re-interpret the prosody condition to be one which is conducive to prosodic emphasis. In particular, we test a) whether semantic focus as induced by a preceding question facilitates processing; b) whether prosodic context conducive to emphasis facilitates processing; c) whether both semantic focus and prosody serve the same processing function, that is, they search for the semantically most important part of the utterance. Based on previous work, both on Sepedi and on semantic and prosodic processing (cf. above, Cutler 2012 for a review), we anticipate the following: Given that Sepedi does not have grammaticalised prosodic cues to focus, prosodic context will not have a comparable effect on processing as it has in English, Dutch or Mandarin. Should we nevertheless find evidence for prosodic entrainment in this online study, this would suggest that there are indeed cues in the larger prosodic context related to focus that listeners use in processing and which have escaped previous analyses.

Given the cross-linguistic differences in focus marking, the second goal of the study is to investigate whether (and if so, to what extent) language-specific modulations of prosody for focus marking are attained in L2 listening. Akker and Cutler's (2003) study on proficient L2 English listeners (with Dutch as mother tongue) shows that both prosodic and semantic context is processed in non-native listening even though both levels are not mapped native-like (see also Braun & Tagliapietra, 2011 for similar findings on focus-accent mapping in German proficient speakers of L2 Dutch). In other words, the study shows that non-native listeners benefit from the presence of an upcoming accent on the basis of acoustic information available in the prosodic context both in the focussed and unfocussed conditions. However, the reduced effect of accent in the focussed condition shown by native listeners is not observed in non-native listeners. That study, however, investigated a language pairing (namely Dutch and English) that shares a grammaticalised prosodic marking for semantic focus. It is hence an open question whether prosodic entrainment also occurs across typologically dissimilar language pairings. Dutch L2 listeners of English may have benefitted from similarities in prosodic processing between their two languages by means of transfer rather than by experience with the input. By using the same methodology (phoneme-monitoring in L2 English: Experiment 2), we test listeners whose languages differ from each other structurally in the crucial aspect of prosodic focus marking, namely Sepedi as a mother tongue (L1 henceforth) and English as a further language (L2 henceforth).

Speakers of Sepedi are multilingual, with at least South African English as a further language. Despite ongoing change, South Africa is still largely a heterogeneous society with recognizable varieties of English which are still being referred to along ethnic lines. In this study, we investigate a recognizable variety of English by educated black speakers (with Sepedi as a first language), which is comprised under the umbrella term 'Black South African English' (BlSAfE, henceforth), in line with existing literature. Sepedi listeners, in contrast to Dutch listeners, cannot necessarily fall back on the prosodic entrainment that they use in their L1 when listening to English. These listeners need to acquire using prosodic cues that lead to an accented word in English for processing. Previous studies on BlSAfE listening have shown difficulties in the semantic interpretation of intonational differences signalling focus (see also Gut, 2005 on Nigerian English), suggesting that the semantic interpretation of the prosodic cues has not been attained by these listeners. These studies involved the reconstruction of discourse (Zerbian, 2012) and question-answer judgements (Zerbian, 2015). However, these offline tasks required considerable meta-linguistic analysis.

Considering that for Sepedi speakers exposure to English starts in their primary education at the latest, one could argue that such early experience with English plays a crucial role in the attainment of prosodic entrainment (Flege, 2009). According to Flege (2009), input matters as long as there is regular use of and exposition to the L2. The importance of input has been shown for other L2 settings too: for L2 speakers being constantly exposed to the target language (once immigrating in the target-speaking country at an early age, e.g., Flege et al., 1995, Flege 2009) and for other language contact situations like heritage speakers (e.g., Benmamoun et al., 2013). As yet, the overwhelming majority of these studies have focussed on speech production (see Saito, 2018, for a review), whereas still little is known on prosodic processing in L2 listening, and even less is known on the mapping between prosody and semantics when the two languages involved differ in this respect.

In experiment 2, the added value of the phoneme-monitoring task is to provide information concerning the processing of prosody in an online task in order to investigate to what extent participants use semantic and prosodic context, and in how far these two levels are mapped in these L2 listeners.

2. Experiment 1

2.1 Participants

Forty-six Sepedi speakers between the age of 19 and 25 were recruited for the study (17 male, 29 female, age average= 21, SD= 2.9). All of them reported Sepedi as mother tongue. All were students at the University of Witwatersrand, Johannesburg, and did not report any history of vision/hearing impairment. The participants received a reimbursement for their participation.

2.2 Materials

Twenty-four semantically unrelated experimental sentences (see Appendix A for a full list) were constructed of the structure provided in (1). Sentences were rendered in Sepedi orthography.[2]

(1)	*O*	*apeela*	*monna*	*yo*	***b**ohlale*	*nama*	*ya*	*go*	***t**ura*	*mosegare.*
	SM1	cook_for	1.man	AGR1	intelligent	9.meat	POSS9	AGR15	expensive	afternoon

“She cooks expensive meat for the intelligent man in the afternoon.”

Each experimental sentence contained two target phonemes (given in bold). One target occurred in the post-nominal modifier of the first object noun (‘early’ target in the sentence, [b] in (1)), the other on the post-nominal modifier of the second object noun (‘late’ target, [t] in (1)). The word containing the late target phoneme was always followed by an adverb.[3]

Target phonemes consisted of unaspirated ejectives or aspirated plosives /t’ t^h p’ p^h k’ k^h/. Aspirated and unaspirated stops are contrastive in Sepedi (Ziervogel 1967: 122) and are also differentiated in orthography (<t>-<th>, <p>-<ph>, <k>-<kh>). These phonemes never co-occurred in the target sentences. There were at least five syllables before the early target, which made sure that responses from listeners were based on sufficient prosodic information (see also Akker and Cutler, 2003: 84, A&C henceforth). Furthermore, at least five syllables occurred between the early and late target phoneme. The syllables containing the target phonemes were all high-toned (according to the tone marking in the dictionaries of Ziervogel & Mokgokong, 1979).

In phoneme monitoring tasks, target phonemes customarily occur word-initially. For Sepedi, however, this was not possible.[4] Therefore, the target phonemes always occurred in the stem-initial syllable of the noun modifiers.[5] Modifiers were adjectival (2a), verbal (2b), relative (2c), or possessive (2d) constructions (which, in contrast to English, are all highly frequent modifying constructions in Sepedi).

2a. Adjectival construction

*Re nyadiša monna yo mo***t***elele mosadi wa go bina išago.*

SM1PL marry 1.man AGR1 1.tall 1.woman POSS1 AGR15 dance next_year

"We are marrying the tall man (with) the woman who dances next year."

2b. Infinitive (preceded by possessor)

O apeela malome wa go ***b****efelwa nama ya go tura mosegare.*

SM3SG cook_for 1.uncle POSS1 AGR15 angry 9.meat POSS9 AGR15 expensive afternoon

"She cooks expensive meat for the angry uncle in the afternoon."

2c. Relative clause structure

Ba nyakela lesogana leo le ***t****imanago monyana yoo a phagamilego legaeng.*

SM2 search_for 5.boy REL5 AGR5 stingy 1.girl REL1 AGR1 high_esteem village

"They are searching a highly esteemed girl for the stingy boy in the village."

2d. Possessive construction

O fa ngwana wa ***Kh****olofelo senepe sa go pentwa mošomong.*

SM1 give 1.child POSS1 PROPNAME 7.picture POSS7 AGR7 paint work

"He gives Kholofelo's child a painted picture at work."

Though they occurred in words belonging to different syntactic categories (namely adjective, verb or noun), the target phonemes occurred in structurally comparable positions, namely medial in the phonological word, which is the strong stem-initial position.[6] The phoneme-bearing words were not controlled for frequency since this factor was not found to have an effect in previous phoneme-monitoring tasks (see A&C: 84).

A native speaker of Sepedi (male, 21 years old at the time of recording) recorded each of the 24 sentences in three different contexts: (1) as an answer (e.g., A1 below) to a question (Q1) that narrowly focussed the early target-bearing modifier (e.g., *khuma* in A1; focus constituent marked by subscript *F*); (2) as an answer (A2) to a question (Q2) that narrowly focussed the late target-bearing modifier (e.g., *tura*), (3) as an answer (A3) to a broad focus question (Q3) in which all words were equally important and which showed the default prosody for declarative sentences.

Q1: *O apeela malome [ofe]$_F$ nama mosegare?*

SM3SG cook_for 1.uncle 1.which 9.meat afternoon

"Which uncle does she cook meat for in the afternoon?"

A1: *O apeela malome [wa go khuma]$_F$ nama ya go tura mosegare.*

SM3SG cook_for 1.uncle POSS1 AGR15 rich 9.meat POSS9 AGR15 expensive afternoon

"She cooks expensive meat for the rich uncle in the afternoon."

Q2: *O apeela malome nama [efe]$_F$ mosegare?*

SM3SG cook_for 1.uncle 9.meat 9.which afternoon

"Which meat does she cook for the uncle in the afternoon?"

A2: *O apeela malome wa go khuma nama [ya go tura]$_F$ mosegare.*

SM3SG cook_for 1.uncle POSS1 AGR15 rich 9.meat POSS9 AGR15 expensive afternoon

"She cooks expensive meat for the rich uncle in the afternoon."

Q3: *Go direga eng?*

SM17 happen what

"What is going on?"

A3: *O apeela malome wa go khuma nama ya go tura mosegare.*

SM3SG cook_for 1.uncle POSS1 AGR15 rich 9.meat POSS9 AGR15 expensive afternoon

"She cooks expensive meat for the rich uncle in the afternoon."

While in English the difference between accentuation and deaccentuation is acoustically discernible and the prosodic surrounding guides listeners to allocate attention to an upcoming de/accented target-bearing word, in Sepedi there is no consistent, salient prosodic difference between the three contexts shown above (cf. section 2, see also footnote 3). However, it is conceivable that there are prosodic cues (such as intensity, duration, or articulatory strength), which might indicate an upcoming focussed constituent to native listeners and which might have escaped qualitative and acoustic analyses so far. To ensure the realization of a prosodic structure in line with an upcoming focussed target-bearing word (i.e. *khuma* in A1) and of a prosodic structure cueing a non-focussed version of the same target-bearing word (i.e. *khuma* in A2, where *tura* is focussed), the Sepedi informant was instructed to read first the question and then the answer in a way that is semantically congruent. Question-answer pairs were read in a block-wise fashion and in the following order: broad focus, narrow focus on the early target-bearing modifier, narrow focus on the late target-bearing modifier.

For each (narrowly) focussed sentence (A1 and A2), one focussed modifier was spliced out (e.g. *khuma* in A1) and replaced by its counterpart from the broad focus context (e.g. *khuma* in A3), thereby leaving the rest of the prosodic structure intact (following the cross-splicing procedure as described in A&C). This operation led to the realization of four sentences, with one of the modifiers (early or late) acting as the target-bearing word (and being always cross-spliced with its broad-focus counterpart). These four sentences were alternatively combined with the two narrowly focusing questions (Q1 and Q2); all possible combinations of early/late target, focussed/non-focussed and two prosodic contexts resulted in eight different versions of each target sentence. Target-bearing modifiers were spliced at the zero-crossing of the release of the (unaspirated or aspirated) stop burst. The potential presence of disturbances and pops in the spliced part of the signal was further checked by the Sepedi informant who was instructed to listen to the whole set of spliced sentences and report anything unnatural for each cross-spliced sentence.

The twenty-four filler sentences (see Appendix B for a full list) had the same syntactic structure as the experimental sentences but contained target-bearing words differing in phoneme type (e.g., nasals, fricatives) and position. Of the 12 fillers with late phonemes, six target-bearing words were represented by a sentence-final adverb (yielding a *when* or *where* question) and six by a sentence-final modifier (the final adverb was then left out altogether; yielding a *which* question). Of the 12 fillers with early phonemes, target-bearing words were represented by a noun followed or not by a modifier (yielding a *what* question). Fillers were not controlled for tone and were recorded in two contexts: in broad focus and in narrow focus (i.e. in half of the cases focus was on the target-bearing word and in the other half on some other word). As with the experimental sentences, the target-bearing word from the broad focus condition was spliced into the narrow-focus sentence.

2.3 Procedure

All materials were recorded onto a digital audiotape in a soundproof cabin of the Radio Academy at the University of the Witwatersrand, Johannesburg.

Subjects were tested individually in a quiet room at the University of the Witwatersrand, Johannesburg. The entire experiment session was conducted in Sepedi to ensure that participants were in their Sepedi language mode (Grosjean, 2012). Stimuli were presented on a portable laptop and binaurally over headphones.

Subjects were instructed to pay attention to the content of each question-answer pair, listen for a specified target sound in the answer sentence and press the button as soon as they heard that sound. For each trial, participants first saw the target phoneme (appearing for 1 sec.) on a portable laptop screen, and then heard the question (after 2 sec.), followed by the answer (after 2 sec.).

The experiment was designed by using *E-prime* software (version 2.0, https://www.pstnet.com/eprime.cfm). Response times were recorded by means of a button box linked to the portable laptop and calculated in relation to the timing interval between the beginning of the sentence and the onset of the target-bearing word.

Sentence comprehension was tested via administration of a post-experiment test in the form of a multiple-choice test. This comprised 24 of the 48 target sentences heard during the phoneme-detection task with a gap for one of the modifiers. Subjects had to choose from four suggested words which one had been the (early or late) target-bearing word.

2.4 Statistical analyses

Statistical analyses were performed by using linear mixed effects models in R software (R core team 2016) as implemented in the *lme4* package. SPEAKER and ITEM were entered as crossed-random factors, allowing for random adjustments of by-participant and by-item intercepts and slopes for the predictors (e.g., Cunnings, 2012).

First, to test whether semantic focus and prosody affected the distribution of missing responses (i.e. no button press), we ran a binomial logistic regression analysis (Baayen, 2008, Pinheiro and Bates, 2000) containing BUTTON PRESS as a binary dependent variable (yes vs. no) and SEMANTIC STATUS (focussed vs. unfocussed) and PROSODIC CONTEXT (conducive vs. non-conducive to emphasis) as predictors.[7]

Values below 100 ms were removed since these points are more likely to reflect a simple guessing (resulting from anticipations) of the target-bearing word in phoneme-monitoring tasks (largely following the procedure applied in A&C; see also Cutler, 1976). The cleaned data was inspected again to check for normality violation by creating quantile-quantile plots for log-transformed reaction times, using a diagnostic procedure. The fixed structure of the linear mixed effects model contained (LOG-TRANSFORMED) RTS as dependent variable and SEMANTIC STATUS, PROSODIC CONTEXT and TARGET POSITION as fixed factors. Specifically, phoneme-detection experiments have shown that targets near the end of the sentences (i.e. late targets) are spotted faster than targets near the beginning of the sentence (i.e. early targets) because the likelihood of the target word to occur increases towards the end of a sentence (cf. Cutler & Fodor, 1979). *P*-values were calculated on the basis of the *Satterthwaite* approximation of degrees-of-freedom in the *lmerTest* package (Kuznetsova, 2013). The main effects of each predictor and of interactions were tested by comparing the model that contains a certain factor (or interaction) with a model that lacks that particular factor (or interaction) by using the Likelihood ratio test as implemented in the *anova*()-function. Predictors that did not significantly improve the fit of a model relative to the alternative without them were dropped ('backward elimination' procedure, cf. Baayen, 2008). Finally, R^2 values associated to each model were calculated as implemented in the *r.squaredGLMM*()-function (Nakagawa & Schielzeth, 2013), whose output provides marginal R^2 values (R^2m) associated with fixed effects of the model and conditional R^2 values (R^2c), thus accounting for the proportion of variance explained by both fixed and random effects of the model. In what follows, coefficients and fitted latencies of the model are reported in the transformed scale, whereas data is visualised in the original millisecond scale.

2.5 Results

The overall mean score of correct answers to the post-experiment test was 54%.

Five participants out of the 46 tested subjects were excluded because in more than half of the trials (i.e. 24 target sentences) they did not press the button and scored low in the post-experiment test (less than 6 correct responses). The overall number of data points was 984 (41 participants x 24 experimental sentences) with 99 missing responses (i.e. no button presses, see Appendix C for more details). The binomial logistic regression model (cf. section 2.4) revealed that there was no effect of semantic status (focussed: N= 49 vs. unfocussed: N= 50 out of 984 responses; p= .8, on average 91% percent chance) and of prosodic context (conducive: N= 55 vs. non-conducive: N= 44 out of 984, p= .2, 93%) on the distribution of the missing responses. Furthermore, there was no interaction between the two factors (focussed/conducive: N= 28, focussed/non-conducive: N= 21, unfocussed/conducive: N= 27; unfocussed/non-conducive: N= 23; p= .7, 92% on average). Missing responses were subtracted from the 984 data points, thereby leaving 885 reaction time responses for data screening.

The RT dataset comprised 844 points after removing data points below 100 ms and above 1500 ms (N= 41 data points removed; accounting for 4.6% of the data). The linear mixed-effects model (R^2m= .014, R^2c= .526; fitted as described in section 2.4.) revealed a main effect of semantic status (*Likelihood ratio test*: $\chi^2(1)$= 4.21, *p*= .04) in that phonemes in focussed target-bearing words were detected faster (mean RT= 603 ms) than in unfocussed target-bearing words (mean RT= 633 ms; difference, $\beta_{unfocussed}$= 0.04, standard error (*SE*)= 0.02, *t*= 2.05, *p*= .04). It revealed a further main effect of target position ($\chi^2(1)$= 15.8, *p*< .0001) in that targets in late position were detected faster (mean RT= 643 ms) than targets in early position (mean RT= 591 ms; β_{early}= 0.09, *SE*= 0.02, *t*= 3.99, *p*< .0001). The model revealed no main effect of prosodic context (conducive: mean RT= 614 ms vs. non-conducive: mean RT= 622 ms; *p*= .4), and no interaction between any of the factors (all *p*-values> .2, see Figure 1).

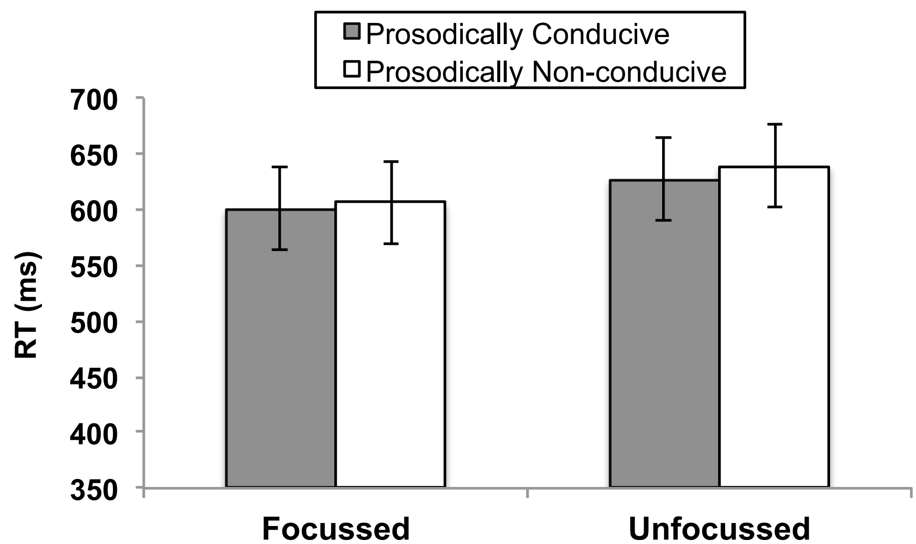

Figure 1. Mean response times (ms) in Experiment 1 (41 Sepedi participants tested on Sepedi materials) as function of Semantic status (focussed vs. unfocussed) and Prosodic context (conducive vs. non-conducive to emphasis). Whiskers represent standard errors as calculated by the model.

To test whether specific prosodic cues preceding the focussed target word were present, we conducted an auditory analysis of the stimulus recordings used in the experiment. Two linguistically-trained labellers with a good knowledge of Sepedi listened for prosodic differences in the utterances preceding the focussed modifiers. Each labeller compared the BF utterance to its rendering in the narrow focus conditions (cf. section 2.2). Parameters which might be relevant for a perceptual impression of emphasis in Sepedi include loudness, pitch, duration, pauses and tonal variations (see also Raborife et al. 2016 for more details). In the stretch preceding the phoneme-bearing word, narrow focus sentences were realized with a raised pitch in 13 cases out of 48, as compared to the broad focus condition. In a linear mixed effects model, we implemented reaction times as function of RAISED PITCH (coded as a *yes-no* binary variable) and of TARGET POSITION (early vs. late) and SPEAKER and ITEM as random factors allowing for random intercepts and slopes. The model revealed no effect of RAISED PITCH (RTs were not significantly faster in raised pitch sentences compared to non-raised pitch ones; mean RTs for group of stimuli with preceding raised pitch= 615.8 ms; SD= 313.6; mean RTs for group without preceding raised pitch= 620.2 ms, SD= 309.9, p= .9) and no interaction with target position (mean RTs for early targets with raised pitch= 663.9 ms; SD= 356.4, mean RTs for late targets with raised pitch= 573.5 ms, SD= 266.1, mean RTs for early targets without raised pitch= 637.4 ms, SD= 330.9, mean RTs for late targets without raised pitch= 599.6 ms, SD= 282.4, p= .5). [8] Hence Sepedi listeners did not exploit the presence of raised pitch preceding the focussed word in processing.

Experiment 1 confirms that in Sepedi, just as in English, the semantic information cued by a preceding question (a question-induced effect) facilitates processing in directing attention to the target word (as evidenced by faster reaction times). At the same time, the lack of an effect of prosodic context shows that Sepedi listeners do not exploit prosody in processing, at least not in connection with focus, as processing is not facilitated by prosodic cues that could be considered conducive for the signaling of focus, emphasis or highlighting. This finding suggests that there are indeed no prosodic cues in the stretch preceding the target word to assist the listeners' search for semantic focus, as also corroborated by the auditory analysis on the sentence material produced by our informant. Finally, there is no interaction of the two factors as found for English and Dutch native listeners (no difference between prosodically conducive and non-conducive under focus vs. a marked difference between them under unfocussed condition, cf. Introduction). Hence, in Sepedi there is no relation between semantic status and prosodic context. If predicting sentence focus on the basis of the preceding semantic information occurs, it is independent of the processing of prosody as there is no focus-related prosody that directs Sepedi listeners' attention to the focus of the sentence. The semantic notion of focus, however, serves to facilitate processing across languages, as also shown in previous studies (Cutler & Fodor 1979, Ip & Cutler 2017).

On the basis of the cross-linguistic differences concerning the prosody-semantic interface between English and Sepedi, Experiment 2 tested whether L2 English speakers with Sepedi as L1 make use of the processing advantage of prosodic information in English and thus exploit the form-to-function mapping that exists in their L2 but not in their L1.

3. Experiment 2

3.1 Participants

Forty-one speakers of Black South African English (22 male, 19 female, age average= 20.5, *SD*= 1.5) with no impaired vision and hearing took part in the study. All were students at the University of the Witwatersrand, Johannesburg, where English is the language of teaching and learning. All participants of the study had already been exposed to and had had tuition in English since their primary education. Participants reported Sepedi as a mother tongue. They were tested in English and the entire experiment session was conducted in English to ensure that participants were in their English language mode. The participants received a small reimbursement for taking part in the study.

3.2 Materials

The materials were those tested in A&C[9] on English native and non-native participants with Dutch as their mother tongue. They comprised of twenty-four experimental sentences (e.g., *The young man on the corner was wearing a blue hat*) and 24 fillers. Early and late target-phonemes occurred in two prosodic contexts (accented vs. deaccented), two question-induced contexts (focussed vs. unfocussed) and two target positions in the sentence (early vs. late). Specifically, the question-induced context either focussed the target-bearing word of the answer (e.g., *The man on the corner was wearing a [BLUE]$_F$ hat* following after *Which hat was the man wearing*?) or not (*The man on the corner was wearing a [BLUE]$_F$ hat* following after *Which man was wearing the hat?*). The prosodic contour preceding the target-bearing word predicted either an accented or a deaccented target-bearing word. Each experimental sentence contained one of the target phonemes – either /d/ or /k/ or /b/ – in the initial position of the bearing words, one located in the early part of the sentence (e.g. /k/ of *corner* in the example above) and another in the late part of the sentence (e.g. /b/ of *blue*). On the basis of the cross-splicing procedure described above (cf. section 2), A&C created eight versions of the same experimental sentence (2 question-induced contexts x 2 prosodic contexts x 2 target positions). These eight versions were rotated across the 8 lists; participants were randomly assigned to one of them.

Like in Experiment 1 (cf. section 2.3), the post-experiment test comprised 24 sentences of the 48 experimental sentences heard during the task with a gap for one of the target-bearing words (in the English material, the word was mostly represented by a noun). Subjects had to decide among four words and choose the one that had been the (early or late) target-bearing word.

3.3 Procedure and statistical analyses

Procedure and statistical analyses were the same as the ones described in section 2.3 and 2.4 respectively.

3.4 Results and discussion

The mean score on the post-experiment test was 58%.

With the same procedure as in Experiment 1, two speakers were excluded because they did not press the button in more than half of the cases. Following A&C (p. 85), one critical item (see A&C, the *bond* experimental item) was excluded from the analysis (39 subjects x 23 experimental items= 897 data points). Missing responses (N= 118) were checked for statistical significance and then subtracted of the 897 data points, thus leaving 779 reaction time responses for data screening (see Appendix C for more details). The binomial logistic regression model revealed no effect of semantic status (focussed: N= 59 vs. unfocussed: N= 59 out of 897 responses; p= .6, on average 94% of chance) and of prosodic context (accented: N= 52 vs. deaccented: N= 66 out of 897; p= .4, on average 92%) on the distribution of missing responses. Furthermore, no interaction was found (focussed/accented: N= 27, focussed/deaccented: N= 32, unfocussed/accented: N= 25; unfocussed/deaccented: N= 34; $p > .5$, on average 90%).

Out of the 779 latency responses, reaction times shorter than 100 ms and longer than 1500 ms (N= 56; overall accounting for 7.1% of the data) were removed, thus leaving 723 data points for the RT analysis. The linear mixed-effects model (R^2m= .037, R^2c= .463) revealed an effect of semantic status on the RTs (*Likelihood ratio test*: $\chi^2(1)$= 10.17, p= .0001): Targets in focussed words (mean RT= 486 ms) were detected faster than targets in unfocussed words (mean RT= 528 ms; $\beta_{unfocussed}$= 0.08, *SE*= 0.02, *t*= 3.28, *p*= .0001). It also revealed an effect of prosodic context ($\chi^2(1)$= 13.6, *p*= .0002): Targets in accented words (mean RT= 485 ms) were detected faster than targets in deaccented words (mean RT= 531 ms; $\beta_{deaccented}$= 0.09, *SE*= 0.02, *t*= 3.70, *p*= .0002). Finally, it also revealed an effect of target position ($\chi^2(1)$= 20.6, *p*< .0001): Targets in late words (mean RT= 464 ms) were detected faster than targets in early words (mean RT= 541 ms; β_{early}= 0.11, *SE*= 0.02, *t*= 4.57, *p*< .0001). No interaction between any of the factors was found (all *p*-values> .3, see Figure 2).

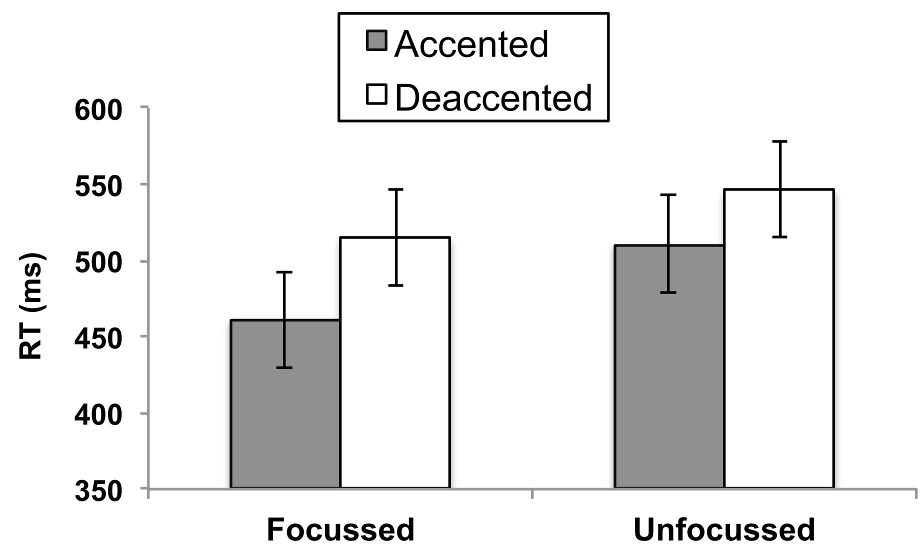

Figure 2. Mean response times (ms) in Experiment 2 (39 Black South African English speakers tested on British English materials taken from Akker and Cutler, 2003) as function of Semantic status (focussed vs. unfocussed) and Prosodic context (accented vs. unaccented). Whiskers represent standard errors as calculated by the model.

The results of Experiment 2 are in line with previous findings on non-native listening (cf. Introduction). The effect of semantic focus indicates that non-native listeners exploit focus for processing (i.e. faster RTs for targets in focussed than unfocussed words) through the semantic information driven by the preceding question, just like native listeners do. In addition, the effect of prosodic context on their RTs indicates that they allocate attention to the predicted sentence accent by using the prosodic information available from the preceding context (i.e. faster RTs for targets in accented than deaccented condition). Turning now to the relation between the two factors, the statistical analysis shows that there is no interaction between prosodic status and semantic status, which replicates the pattern of results found in Akker and Cutler's experiment 4 for L2 listeners of English with Dutch as their first language.

4. General Discussion

The first experiment aimed at testing processing advantages of semantic focus in Sepedi, a Bantu language without prosodic marking of focus (e.g. by means of accentuation and/or phrasing). On the basis of previous studies on sentence processing in other languages (cf. Introduction), we hypothesised that because there is no grammaticalised or otherwise systematic prosodic expression of focus in this language, listeners would only benefit from semantic context. This is so because the notion of focus is a universally shared notion in all languages, independent of its grammatical marking and despite language-specific preferences in the organisation of discourse structure (cf. Zimmerman & Onea, 2011, see Dimroth & Narasimhan, 2012 for discussion).

First of all, the phoneme-monitoring task confirms that Sepedi listeners actively search for and allocate their attention to the location of sentence focus using the information available in the preceding semantic context. When looking at the role of prosodic information, the results show prosodic context does not facilitate processing in Sepedi. The lack of an effect of prosody is independent of the fact that Sepedi is a tone language. Findings for Mandarin Chinese (Ip & Cutler,

2017) speak in favor of a double function of F0: this cue can serve for prosodic focus marking at the sentence level as well as for marking lexical contrasts at the word-level. For Sepedi, F0 also serves for marking distinctions at the word-level and sentence-level (for instance, sentence mode, see Zerbian, 2006 for Sepedi; Zerbian, 2017 for closely related Tswana). Crucially, however, it does not mark semantic prominence. In a few cases, the narrow focus sentences showed raised pitch before the target word. However, we could not show that listeners relied on such a prosodic cue in processing (i.e. reaction times were not faster for those words whose preceding context was characterised by raised pitch).

It is thus the systematic and grammaticalised use of prosody for focus marking that is relevant for predicting whether prosodic entrainment will occur in a language. This is fully in line with Cutler's (2012) central argument about what it means to be a native listener. Native listeners know which cues are useful in their native language, and which ones are not. They do not use the ones that are not useful. Our study thus corroborates the idea that prosodic entrainment correlates with systematic, measurable focus-related prosody. Follow-up studies should test by means of a memory task whether Sepedi listeners can benefit from other linguistic marking strategies (like word order) for focus processing, as Korean listeners do with syntax (Kember, Jiyoun, and Cutler 2016).

In lacking an effect in the prosodic condition, the results of the first experiment are also in line with previous findings based on production and offline perception studies that did not reveal a phonetic expression of focus and acoustic cues for focus interpretation in Sepedi (Zerbian, 2006, 2007).

Despite cross-linguistic research being extremely valuable in psycholinguistics (see Cutler 2012: 28), research on lesser-studied languages has been rare. It might therefore be interesting to have a look at the general difficulties our Sepedi listeners encountered in performing the task.

Compared to the English native group (mean RTs: 394 ms; scores: 60%) tested in Akker and Cutler (2003), they had a slower mean detection rate (605 ms on average) and scored lower in the post-recognition test (54%). It is worth recalling, however, that in order to make the closest comparison possible with the English (*wh-*)question-answer pair sentences, the target word in the Sepedi material was a modifier instead of a noun. Hence, general differences in dealing with the task observed between the two native groups may also be attributable to differences in the complexity of the sentence structure (more complex in Sepedi than in English). Furthermore, another aspect that may have played a role is that participants were unfamiliar with the use of Sepedi in an academic context, and/or that most of the Sepedi participants had never participated in a psycholinguistic experiment before (see also Cutler, Demuth and McQueen, 2002).

Cross-linguistic differences in prosodic focus marking provided the motivation to look at the processing of prosodic and semantic focus in non-native listening. In experiment 2, we tested whether the relation between focus and accentuation is exploited by Black South African English listeners with Sepedi as a mother tongue, that is, with an L1 with no focus-accent mapping. On the whole, findings show that these non-native listeners perform comparable to the Dutch L2 listeners of English tested in Akker and Cutler (2003). Their target detection rates were similar to the other L2 English group (Dutch L2 English 485 ms, Sepedi L2 English 489 ms) and so were their scores (58% vs. 60% respectively). As the original English stimuli from Akker and Cutler (2003) were used, the results should be comparable across the two studies (cf. footnote 8).

Turning to the core findings of Experiment 2, results show that BlSAfE listeners with Sepedi as L1 use the semantic information available from the preceding question (main effect of semantic status) and the prosodic information preceding an accented target word in processing their L2 (main effect of prosodic context). The results suggest that our L2 listeners track both sources of information, prosodic context and semantic focus, independent of each other (no interaction between the two factors, with a slight accent advantage effect under focus condition) unlike English native listeners. This pattern mirrors the results found in A&C's experiment 4 for Dutch L1/ English L2 listeners, even for listeners whose L1 does not use prosody as similarly to English as the L1 of listeners in A&C's study.

A&C (2003: 92ff.) discuss two possible reasons why their Dutch L1/English L2 listeners' responses did not yield the same interaction between semantic focus and prosodic context that native listening did. As we find the same result pattern with our BlSAfE listeners, we will briefly present A&C's discussion here.

The two main effects show that BlSAfE listeners process both prosodic structure and sentence semantics. However, the result pattern suggests that listeners do not map these levels of processing. This is actually less surprising for listeners of BlSAfE than for DutchL1/ English L2 listeners as their L1 Sepedi does not map prosody to semantics in a similar way. A&C offer two possible explanations: (1) semantic processing might be considerably slower in non-native listening than in native listening (due to e.g. lower lexical familiarity or less automatic processing of multi-word strings). In native listening, semantic processing could be seen to outrun prosodic processing, leading to no added effect of prosodic processing once semantic processing has allocated a listener's attention to the focussed constituent, resulting in the observed interaction in experiment 1 in A&C 2003. Due to being slower in non-native listening, semantic processing might not outrun prosodic processing in non-native listening. (2) The parallel processing of semantic and prosodic information providing the same information (active search for focus) can be considered a 'fail-safe, belt-and-braces' strategy in non-native processing (Cutler 2012: 352): Using two sources of information instead of one, even if one might be redundant, is a safer communicative strategy. Redundancy seems to be a common property in non-native speech, as e.g. mirrored in the tendency to overuse pitch accents on lexical items in L2 English productions (e.g., Jilka, 2000, Gut, 2009).

As a reviewer pointed out, the two explanations offered by A&C are crucially different: in (1), non-native semantic processing is not fast enough to parallel native semantic processing, in (2), sources of information for processing are strategically used differently in non-native listening as opposed to native listening. Just as A&C (2003)'s results could not decide between these two interpretations, neither can ours. They remain to be tested further in future research.

Another aspect that awaits further investigation is a possible influence of the British accent (at a segmental level) in the test materials on the BlSAfE listener group (see Zerbian and Turco, 2019).

Finally, the current online task has shown that the prosodic context of English is indeed used by Black South African English listeners even if this feature is not present in their L1 (cf. Experiment 1). This adds to previous offline perception data (discourse reconstruction: Zerbian, 2012, question-answer judgements: Zerbian, 2015) elicited by the same L2 variety. One could argue that compared to the online task presented here, the offline tasks proved difficult for non-native listeners since the requested judgements required considerable meta-linguistic analysis.

Acknowledgements

This research was funded by the German Research Foundation (DFG) grant to the project A7 in the SFB 732 "Incremental Specification in Context" at the University of Stuttgart. We thank all participants and the School of Languages, Literature and Media Studies at the University of the Witwatersrand, Johannesburg for their support in data collection, the Radio Academy at the same institution for assistance in the recording of the Sepedi stimuli, Anne Cutler for providing us with the English audio sentence material used in Akker and Cutler (2003), and several reviewers for constructive and helpful comments. Our heartfelt thanks go to Khtušo Matlou for his enthusiasm and help concerning the study. Without him, this would not have been possible. Re lebogile! . This research was also partially supported by the European Union Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 661530 awarded to the first author.

References

Akker, E., and Cutler, A. (2003). Prosodic cues to semantic structure in native and nonnative listening, *Bilingualism: Language and Cognition, 6*(2): 81–96.

Baayen, R. H. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*, Cambridge: Cambridge University Press.

Benmamoun, E., Montrul, S., and Polinsky, M. (2013). Heritage languages and their speakers: Opportunities and challenges for linguistics, *Theoretical Linguistics*, *39*: 129–181.

Braun, B., and Tagliapietra, L. (2010). The role of contrastive intonation contours in the retrieval of contextual alternatives. *Language and Cognitive Processes, 25*, 1024-1043.

Braun, B., and Tagliapietra, L. (2011). On-line interpretation of intonational meaning in L2, *Language and Cognitive Processes*, *26*(2): 224–235.

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers, *Second Language Research, 28*(3): 369–382.

Cutler, A. (1976). Phoneme monitoring and intonation contour, *Attention, Perception, & Psychophysics, 20*(1): 55–60.

Cutler, A., and Fodor, J. A. (1979). Semantic focus and sentence comprehension, *Cognition, 7*(1): 49–59.

Cutler, A., Demuth, K., and McQueen, J. M. (2002). Universality versus language-specificity in listening to running speech, *Psychological Science, 13*(3): 258–262.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*, Cambridge: MIT Press.

Dahan, D., Tanenhaus, M. K., and Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension, *Journal of Memory and Language, 47*(2): 292–314.

Dimroth, C., and Narasimhan, B. (2012). The acquisition of information structure. In M. Krifka & R. Musan (Eds.), *The expression of information structure* (pp. 319–362). Berlin: de Gruyter/ Mouton.

Downing, L. J. (2010). Accent in African Languages. In H. van der Hulst, R. Goedemans, and E. van Zanten (Eds.), *A survey of Word Accentual Patterns in the Languages of the World* (pp. 381–428). Berlin: Mouton de Gruyter.

Flege, J. E., Liu, S. (1995). The effect of experience on adults' acquisition of a second language, *Studies in Second Language Acquisition*, *23*(4): 527–552.

Flege, J. E. (2009). Give input a chance! In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 175–190), Clevedon: Multilingual Matters.

Grosjean, F. (2012). Bilingual and Monolingual Language Modes. In *The Encyclopedia of Applied Linguistics*. doi: 10.1002/9781405198431.wbeal0090

Gut, U. (2005). Nigerian English prosody, *English World-Wide, 26*(2): 153–177.

Gut, U. (2009). *Non-native speech*, Frankfurt/M.: Peter Lang.

Hirschberg, J. (1993). Pitch accents in context: predicting intonational prominence from text, *Artificial Intelligence*, *63*: 305–340.

Hyman, L. M., and Katamba, F. X. (2005). The word in Luganda. In E. Voeltz (ed.), *Studies in African Linguistic Typology* (pp. 171–193). Amsterdam/Philadelphia: John Benjamins.

Ip, M. H. K., and Cutler, A. (2017). Intonation Facilitates Prediction of Focus even in the Presence of Lexical Tones, *Proceedings of Interspeech 2017* (pp. 1218–1222).

Ito, K., and Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search, *Journal of Memory and Language*, *85*(2): 541–573.

Jilka, M. (2000). *The contribution of intonation to foreign accent*. (Doctoral dissertation). University of Stuttgart, Stuttgart.

Kember, H., Jiyoun. C., and Cutler, A. (2016). Processing advantage for focused words in Korean, *Proceedings of Speech Prosody 8* (pp. 702–705). doi: 10.21437/SpeechProsody.2016-144

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. F., and Tanenhaus, M. K. (2014). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings, *Cognition*, *133*(2): 335–342.

Kuznetsova, A. (2013). "Lmertest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package)" *R package version 2.029*, 1–17. Retrieved from https://cran.r-project.org/web/ packages/lmerTest/lmerTest.pdf

Nakagawa, S., and Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models, *Methods in Ecology and Evolution*, *4*(2): 133–142.

Pinheiro, J. C., and Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*, New York: Springer.

Raborife, M., Turco, G., and Zerbian, S. (2016). The prosody of focus and emphasis in Sepedi. *Proceedings of PRASA* (Pattern Recognition Association of South Africa). Stellenbosch. Pp. 15-18.

Saito, K. (2018). Advanced segmental and suprasegmental acquisition. In P. Malovrh & A. Benati (Eds.)*, The handbook of advanced proficiency in second language acquisition*. Wiley Blackwell, pp. 282-303.

Terken, J. (1984). The distribution of pitch accents in instructions as a function of discourse structure, *Language and Speech, 27*: 269–289.

Terken, J. and Nooteboom, S. G. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and Cognitive Processes*, 2(3/4):145-163.

Watson, D.G., Tanenhaus, M.K., Gunlogson, C.A. (2008) Interpreting pitch accents in online comprehension: H* vs. L H*. *Cognitive Science 2008*, 32:1232-1244.

Weber, A., Braun, B., and Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents, *Language and Speech, 49*(3): 367–392.

Xu, Y. (1999). Effects of tone and focus on f0 contour formation, *Journal of Phonetics, 27*: 55–105.

Zerbian, S. (2006). Expression of information structure in the Bantu language Northern Sotho, *ZAS Papers in Linguistics, 45*. Berlin: ZAS.

Zerbian, S. (2007). Investigating the prosodic expression of focus in Northern Sotho. In E. Aboh, K. Hartmann & M. Zimmermann (Eds.), *Focus strategies: evidence from African languages* (pp. 55–79), Mouton de Gruyter, Berlin.

Zerbian, S. (2012). Perception and interpretation of intonational prominence in varieties of South African English. In K. Braunmüller & C. Gabriel (Eds.), *Multilingual Individuals and Multilingual Societies* (pp. 335–348). Amsterdam: John Benjamins.

Zerbian, S. (2015). Syntactic and prosodic focus in contact varieties of South African English, *English World-Wide, 36*(2): 228–258.

Zerbian, S. (2017) Intonation in Sotho-Tswana. In L.J. Downing & A. Rialland (Eds.), *Intonation in African Tone Languages* (pp. 393-433). Berlin: Mouton de Gruyter.

Zerbian, S., and Turco, G. (2019). Processing focus and accent across dialects. *Proceedings of ICPhS 2019, Melbourne*.

Ziervogel, D., and Mokgokong, P. C. (1979). *klein noord-sotho woordeboek* (3rd ed.), Pretoria: J. L. van Schaik.

Ziervogel, D. (1967). *Handbook of the Speech Sounds and Sound Changes of the Bantu Languages of South Africa,* Pretoria: University of South Africa.

Zimmermann, M., Onea, E. (2011). Focus marking and focus interpretation. In *Lingua, 121*: 1651–1670.

---

[1] The most natural option to express focus in Sepedi is through syntactic operations, such as inversion and clefting (see Zerbian, 2006 for details).

[2] The following abbreviations are used in the glosses: Roman numbers refer to nominal class and are given where relevant for agreement; SM= subject marker, AGR= agreement marker, REL= relative marker, POSS= possessive marker, SG= singular, PL= plural, PROPNAME= proper name.

[3] The penultimate syllable of a phrase is always lengthened in Sepedi (Zerbian, 2006). In order to avoid a target phoneme in a lengthened syllable (as would be the case for [t] in (1)), an adverb was added in sentence-final position.

4 Due to the agreement prefixes, this would give a cue to the position of the target phoneme even before the answer is heard.

[5] It can be argued that the stem-initial position is a phonologically strong position in the Bantu languages in which all phonemic and tonal contrasts occur (Downing 2010). More research comparing the processing advantages of different positions across typologically different languages is needed.

[6] Except for (2a), target phonemes occur initially in the orthographic word. However, orthography is not a reliable indicator of wordhood (see Hyman & Katamba, 2005 for a discussion of wordhood criteria in Bantu), especially not in the case of the Sepedi disjunctive orthography.

[7] The factor target position could not be integrated in the binomial model since the model failed to converge. Hence, only the impact of the two relevant factors (semantic status and prosodic context) was tested on the distribution of missing responses.

[8] Notice that the model reported in Raborife et al. (2016) showed an interaction of prosody and target position: compared to early and late targets *without* raised pitch, early targets *with* raised pitch were detected slower than late ones. However, early and late targets with raised pitch were not significantly faster than early and late targets without

prosody. If we compare the mean values of early/late targets with/without prosody reported in Raborife et al. (2016) with the mean values reported in the present paper, the direction of the results is similar. In the current model, however, this interaction is not significant. This is due to the fact that the outlier diagnostic of the two models is different: unlike here, Raborife et al.'s model included data points above 1500 ms and RTs values higher than this threshold (15 data points) were mainly found for early targets with prosody, which may explain the significant interaction. In the current model, removing data points above such a threshold improved the fit of the model.

[9] We are grateful to Anne Cutler for providing us with the original English audio material used in Akker & Cutler (2003).