The Processing of Linguistic Prominence

Heather Kember
The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia;
ARC Centre of Excellence for the Dynamics of Language, Australia

Jiyoun Choi
ARC Centre of Excellence for the Dynamics of Language, Australia; Sookmyung Women's University, Seoul, Korea

Jenny Yu
Anne Cutler [iD]
The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia;
ARC Centre of Excellence for the Dynamics of Language, Australia

Abstract

Prominence, the expression of informational weight within utterances, can be signaled by prosodic highlighting (*head-prominence*, as in English) or by position (as in Korean *edge-prominence*). Prominence confers processing advantages, even if conveyed only by discourse manipulations. Here we compared processing of prominence in English and Korean, using a task that indexes processing success, namely recognition memory. In each language, participants' memory was tested for target words heard in sentences in which they were prominent due to prosody, position, both or neither. Prominence produced recall advantage, but the relative effects differed across language. For Korean listeners the positional advantage was greater, but for English listeners prosodic and syntactic prominence had equivalent and additive effects. In a further experiment semantic and phonological foils tested depth of processing of the recall targets. Both foil types were correctly rejected, suggesting that semantic processing had not reached the level at which word form was no longer available. Together the results suggest that prominence processing is primarily driven by universal effects of information structure; but language-specific differences in frequency of experience prompt different relative advantages of prominence signal types. Processing efficiency increases in each case, however, creating more accurate and more rapidly contactable memory representations.

Corresponding author:
Anne Cutler, The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Locked Bag 1797, Penrith, NSW 2751, Australia.
Email: a.cutler@westernsydney.edu.au

Introduction

It is rare that an utterance stands alone and makes no reference to context; discourse is the communicative norm, and what we term *information structure* is the expression of how an utterance relates to the discourse to which it contributes. Listeners need to recognize not only the meaning of utterance parts, but also each part's relative importance in the ongoing discourse. Some parts of an utterance will be new and will carry greater informational weight; this is signaled by prominence within the utterance.

Prosodic highlighting is one expression of prominence, and indeed the marking of informational salience has been claimed to be a prosodic universal (Bolinger, 1978). It is how prominence is typically realized in English (*He sent a TWEET about it?*). But it is not the only means; for instance, syntactic devices such as clefting can also make a word prominent (*It was a tweet that caused all the fuss*). Further, prominence can be conveyed by semantic means such as a preceding question or other arrangement of the discourse context.

Language-specific phonology interacts differently with prominence signals. Thus, prominence influences vowel articulation in French, German and English in subtly different ways (Hay, Sato, Coren, Moran, & Diehl, 2006), and language-specific rhythmic structure affects which phonetic cues are used for prominence in English and also in other—typologically unrelated—languages (Burdin, Phillips-Bourass, Turnbull, Yasavul, Clopper, & Tonhauser, 2015). Burdin et al.'s finding of rhythmic structure effects orthogonal to the languages' prominence typology classifications is consistent with a proposal of Calhoun (2010) that prosodic prominence is realized by how words are placed in metrical structure, such that metrical structure itself effectively expresses information structure—inevitably, in language-specific realizations.

The point of each of these methods of highlighting is always the same: to call listeners' attention to the most important part of a message. And indeed, prominence does attract listeners' attention by whatever method it is signaled. Moreover, listeners benefit in that they process the utterance more efficiently. Prominence on new information makes utterances easier to process in both English (Bock & Mazzella, 1983) and Dutch (Terken & Nooteboom, 1987). Prosodically prominent words are recalled more accurately (Fraundorf, Watson, & Benjamin, 2010), and elicit more rapid responses in detection tasks, again in English and Dutch (Akker & Cutler, 2003; Cutler & Foss, 1977). Syntactic prominence also elicits better recall (Birch & Garnsey, 1995) and faster response times (Birch, Albrecht, & Myers, 2000). Semantic prominence (conveyed by a preceding question) also induces faster responses to prominent words (Cutler & Fodor, 1979).

Phoneme detection studies showed that the processing advantage of prosodic prominence encourages listeners to actively seek it. Identical neutrally produced words were detected more rapidly when they replaced words spoken with prominence than when they replaced words in a non-prominent position, showing that listeners had predicted where prominence would occur by entraining to the prosodic contour (Cutler, 1976). When semantic (prior question) and prosodic prominence were examined together, their effects were not additive (Akker & Cutler, 2003), suggesting that prominence is a unitary (semantic) property and listeners do not differentiate sources of information about it; finding prominence is important, but either type of signal will do.

In the present study one of the languages we examine is English, and indeed most of the existing processing evidence for prominence comes from English, or from closely related West Germanic languages (e.g., Dutch, German). These West Germanic languages are very similar in their prosodic structure and in how listeners process them. They are all stress languages, with stress placement and metrical structure not only forming the basis for the representation of information structure, but also for perceptual segmentation of speech (Cutler & Norris, 1988; Vroomen,

Van Zon, & De Gelder, 1996). The Germanic pattern for expression of prominence is called "head-prominence."

Korean is a language that differs from English in all the ways just listed, suggesting that a direct English–Korean comparison of prominence processing should provide an informative view of whether the currently available prominence processing literature has been limited by having been conducted largely in West Germanic languages. The perceptual segmentation of spoken Korean resembles that of French: in both languages listeners have been shown to segment speech at the level of the syllable (Kim, Davis, & Cutler, 2008, for Korean; Mehler, Dommergues, Frauenfelder, & Segui, 1981, for French) and at the level of the accentual phrase (Jeon & Nolan, 2010, for Korean; Michelas & D'Imperio, 2010, for French). The Korean pattern for prominence expression is called "edge-prominence."

Edge-prominence languages mark prominence by arrangement of utterance phrasing, such that prominent material appears at the designated edge of a particular phonological unit. The unit and placement decisions for prominence are language specific. French, for example, places prominent words at the end of an accentual phrase (AP; D'Imperio, German, & Michelas, 2012; Jun & Fougeron, 2002). In Korean, the designated phonological unit is also the AP, but the designated edge for prominence is the beginning. The AP is the unit directly above a prosodic word. It is demarcated by a pitch contour, with an underlying rising tonal pattern at its edge (Jun, 1998, 2000; see also the proposal for a further unit, Intermediate Phrase (ip) to describe prominence-induced pitch reset: Jun, 2011; Jun & Cha, 2015).

Prominent elements in Korean are also prosodically strengthened, with the strengthening being domain-initial, that is, at the prominent AP's left edge (Cho, Lee, & Kim, 2011). This leads to following material being shorter and produced with a smaller pitch range than in a neutral phrase (Jun & Lee, 1998). Further phonetic effects in AP realization are subject to debate. Jun and Lee found a tendency for sequences preceding a prominent AP to be shorter, but Lee and Xu (2010) observed no difference pre-prominence, only post-prominence. Phrase-final lengthening has been reported at the end of larger (intonational) phrases, yet it is inconsistent at the AP level, arguably due to an interaction between prominence, syntax, and utterance length (Cho & Keating, 2001).

There is little literature on the processing of prominence in Korean. In one recent study (Lee, Wang, Chen, Adda-Decker, Amelot, Nambu, & Liberman, 2015), speakers of English, Mandarin Chinese, and Korean produced number strings read as connected groups in the style of an American phone number (XXX-XXX-XXXX), placing contrastive prominence in turn on each of the 10 digits. For English and Mandarin speakers, consistent acoustic differences appeared in the production of numbers with versus without prominence: prominent digits exhibited greater duration, F0 and intensity, as well as triggering post-prominence compression. Listeners of both languages were highly accurate at identifying which number within such a string had contrastive prominence (94–97%). For speakers of Korean, however, the design meant that prosody and position were decoupled; as predicted, then, no clear differences between numbers with versus without prominence were produced, and perceptual accuracy was also low (44–55%).

The experiments reported here are aimed at providing a more comprehensive picture of the role of prominence in processing of Korean and English, with their differing systems of prominence realization. Not only do we compare across these languages, we also compare, within each language, across the two principal prominence markers used in the literature we found: marking by prosodic emphasis and by syntactic placement. Of the studies reviewed above, none conducted such a within-experiment comparison.

We use a recognition memory task, which as described above has previously shown robust benefits for prominent words in English. Our study uses an auditory version of Birch and Garnsey's (1995) visually based technique, in which participants first read sentences in which target words

were rendered prominent by means of constructions such as "It was the. . ." or "There was this. . .", then judged for printed target words whether or not each such word had appeared in the previously presented sentences; target words that had been prominent were recognized faster than those without prominence, even when target testing occurred only after every 10 sentences. While memory testing necessarily occurs after rather than during sentence presentation and is therefore less sensitive to online attentional fluctuations than for instance phoneme detection, the task reliably reveals differences in the successful achievement of word or utterance processing.

Furthermore, the method allows examination of the stage of processing achieved, by the inclusion of foil, or distractor, targets related in different ways (in sound or in meaning) to the actually presented word targets. The processing of spoken words in sentences involves phonological activation that precedes semantic activation (Norris, Cutler, McQueen, & Butterfield, 2006). Foil targets that did not actually occur in the stimuli, but overlap phonologically with words that did occur, can be more difficult to reject than foils without such resemblance because of residues of competitor inhibition that occurred during the earlier target recognition (Goldinger, Luce, & Pisoni, 1989). Semantically related foils are also hard to reject; indeed, more than a half century of memory research has addressed the remarkable ease with which false memories for words can be induced in the laboratory (Deese 1959; Roediger & McDermott, 1995; Underwood, 1965). If the auditory realization of prominence affects the (earlier) phonological activation stage of word recognition, then we should observe differing ease of rejection for foils related to targets with versus without prominence. Likewise, if prominence alters processing (also, or only) at the (later) semantic stage of word recognition, then memory for the precise word that conveyed a concept may vary, and hence semantically related foil rejection may also vary, as a function of whether or not a target word had been prominent. Experiment 2 incorporates such a test for phonological versus semantic level of lexical processing in this memory task.

First, however, Experiment 1 assesses, in both Korean and English, recognition memory for words that had been prominent (either syntactically, prosodically, or both) in comparison to words without prominence. We record both recognition accuracy (whether processing was successfully achieved) and reaction time (RT) to express recognition (how rapidly it was achieved). We expect that, in line with prior studies, there will be robust processing advantages for either type of prominence in English, and that the two types will possibly run parallel but will not be additive, copying the demonstrated equivalence of prominence effects expressed prosodically versus semantically (Akker & Cutler, 2003). In Korean, processing benefits should be conferred as well, given that prominence expression is a universal element of information structure. Metalinguistic judgments for Korean show that, as for other languages, listeners rate naturalness as higher if focused, but not given, discourse elements are highlighted (Lee, 2012). In the absence of online processing data for the language, however, we cannot motivate particular predictions regarding the relative results for the two types of prominence. In Experiment 2, we replicate this study and extend it by incorporating different target types to allow examination of processing level.

2 Experiment 1

2.1 Participants

For Experiment 1a, 32 native speakers of Korean (M_{age} = 25.53, SD = 7.89) were recruited from the greater Western Sydney region. All were born and raised in Korea (except one who was born in Australia but returned as an infant to Korea). For Experiment 1b, 32 native speakers of Australian English (M_{age} = 29.63, SD = 13.32) were recruited either in the same manner from the Western Sydney community or through the Western Sydney University undergraduate psychology pool; all

had acquired English first and had two English-speaker parents though some had also learned another language (not as simultaneous bilinguals). No participants reported any speech, reading, or hearing problems. Students were reimbursed with course credit for their participation, and community participants were reimbursed for travel expenses.

2.2 Korean materials

Experimental sentences were constructed, 140 in all, each containing two high-frequency target words, one with and one without syntactic prominence. The topic-comment structure ". . .것 은/*kʌs-in*/ ('thing'-topic marker). . ." was used for syntactic prominence: the marker immediately precedes the syntactically prominent target (see 'A' sentences in Table 1). Mean word frequency was 168 per million for targets with syntactic prominence and 173 for targets without (Kim & Kang, 2008). No target words were repeated across sentences and every listener heard every target word just once (thus avoiding effects of repetition, which can be strong in memory tasks).

A female native speaker of Korean recorded two versions of each sentence, with prosodic prominence on the first and the second (syntactically prominent) target respectively. To induce the intended prominence placement, two prompt questions were created for each sentence (Table1); each question contained a word (e.g., 공급 /koŋkip/ 'supply', 술 /sul/ 'alcohol') that semantically contrasted with an intended target (수요 /sujo/ 'demand', 담배 /tampe/ 'tobacco') thus prompting corrective contrastive prominence on the target. This design produced four separate target-word conditions: no prominence (N), prosodic prominence (PR), syntactic prominence (ST), or both prosodic and syntactic prominence (PS). The speaker read the question–answer sentences silently first and then read the answer sentence aloud.

A native speaker of Korean (the second author) checked that prosodically prominent targets were produced in AP-initial position. For each target word, word duration, F0 peak and intensity peak values were obtained using PRAAT scripts (Boersma & Weenink, 2014) and then manually checked based on visual inspection of the F0 and intensity contours along with the waveform and spectrogram for each token. The effect of prosodic prominence was tested in paired-sample *t*-tests on each measure, separately for targets with and without syntactic prominence. In all cases acoustic values were significantly higher for prosodically prominent words (see Table 2). Figure 1 and

Table 1. Example question–answer dialogue for Korean stimuli. Underlining signifies target words; bold signifies prosodically prominent item. PR: prosodic prominence; ST: syntactic prominence; N: no prominence; PS: prosodic and syntactic prominence.

Version 1: Prosodic prominence on a target without syntactic prominence

        Q: 빠르게 **공급**이 증가한 것은 담배야?
        p*alike **koŋkip**-i ʧiŋkahan kʌs-in tampe-ja?
        "Is what showed a fast increase in **supply** the tobacco?"

        A: 빠르게 **수요**(PR)가 증가한 것은 <u>담배</u>(ST)야.
        p*alike **sujo** (PR)-ka ʧiŋkahan kʌs-in <u>tampe</u> (ST)-ja.
        "What showed a fast increase in **demand** (PR) is the <u>tobacco</u> (ST)."

Version 2: Prosodic prominence on a target with syntactic prominence

        Q: 빠르게 수요가 증가한 것은 **술**이야?
        p*alike sujo-ka ʧiŋkahan kʌs-in **sul**-ija?
        "Is what showed a fast increase in demand the **alcohol**?"

        A: 빠르게 **수요**(N)가 증가한 것은 **담배**(PS)야.
        p*alike <u>sujo</u> (N)-ka ʧiŋkahan kʌs-in **tampe** (PS)-ja.
        "What showed a fast increase in <u>demand</u> (N) is the **tobacco** (PS)."

Table 2. Korean stimuli: descriptive statistics and paired-sample *t*-test values for acoustic measures.

	First target: no syntactic prominence			Second target: with syntactic prominence		
	without prosodic prominence	with prosodic prominence	*p* values	without prosodic prominence	with prosodic prominence	*p* values
Duration (ms)	302.0 (5.5)	341.7 (5.4)	< 0.001	302.9 (5.0)	371.5 (5.6)	< 0.001
Peak F0 (Hz)	252.8 (6.7)	327.2 (7.0)	< 0.001	250.5 (2.7)	310.6 (9.4)	< 0.001
Peak intensity (dB)	68.2 (0.2)	73.2 (0.3)	< 0.001	66.8 (0.2)	70.6 (0.2)	< 0.001

Tables 1 and 2 illustrate examples of the sentences and the recording contexts, and summarize the acoustic properties of the targets.

Korean syntax entails that words with syntactic prominence were always the second target word in the sentence. To distract from this pattern, the materials also included 20 control sentences with target words in early or late positions, but neither syntactic nor prosodic prominence. Forty filler sentences (with foil words at test) were also constructed, likewise without prominence. Control and filler sentences were recorded by the same speaker who recorded the experimental sentences. The materials are listed in full in the online Supplementary Information.

2.3 English materials

Here too 140 experimental sentences were constructed. Again, each sentence contained two high-frequency (in CELEX: Baayen, Piepenbrock, & Van Rijn, 1995) target words, one with and one without syntactic prominence. Two cleft structures were used to create the syntactic prominence: "it was the. . ." and "there was this. . ." All syntactically prominent words were thus the first target word (the opposite pattern to that of the Korean stimuli). A female native speaker of Australian English recorded two versions of each sentence, one version with prosodic prominence on the first (syntactically prominent) target, and the other with prosodic prominence on the second target. The recording procedure was as for the Korean stimuli except that prompt questions for English stimuli were wh-questions (Table 3). As in Korean, this induced narrow prominence on the target, realized as a high tone pitch accent. The conditions were as in the Korean experiment: no prominence (N), prosodic prominence (PR), syntactic prominence (ST), or both (PS). Control stimuli and filler sentences were created as in the Korean experiment, and again, participants only heard one version of each experimental sentence. Figure 2 and Tables 3 and 4 again illustrate examples of the sentences and recording contexts, and summarize the acoustic properties of the targets.

Target word duration, F0 peak and intensity peak values were ascertained in PRAAT by a native speaker of Australian English (the third author). Paired-sample *t*-tests on each acoustic measure, separately for targets with or without syntactic prominence, again gave significantly higher acoustic values for prosodically prominent words in all comparisons (see Table 4).

2.4 Procedure

Participants were tested individually either in a sound attenuated room or at a convenient quiet location, seated at a computer running the experimental software E-Prime (Psychology Software Tools, 2016). The audio was presented through noise-attenuating Sennheiser HD 280 Pro
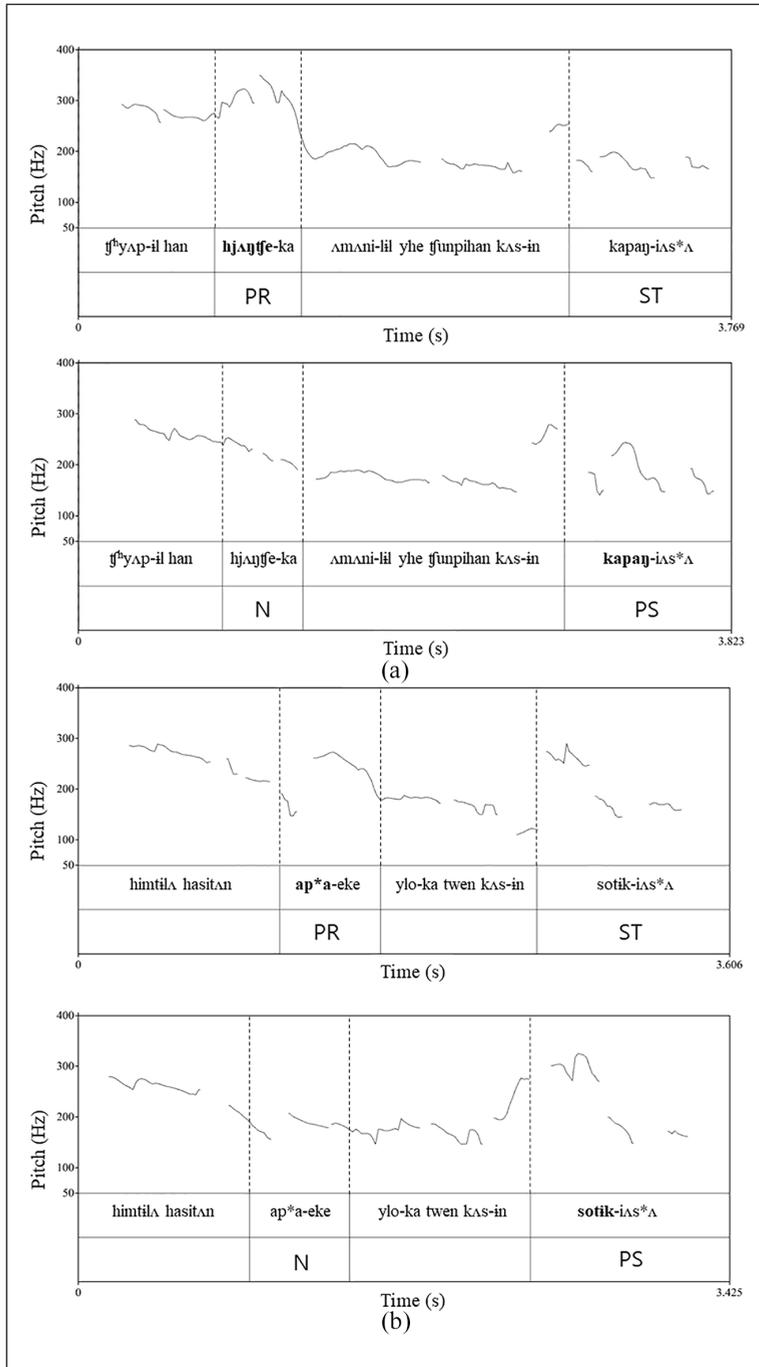
Figure 1. Example pitch tracks for two Korean sentences: (a) 취업을 한 형제가 어머니를 위해 준비한 것은 가방이었어. *"What the brothers who got jobs prepared for their mother was a bag."* (b) 힘들어하시던 아빠에게 위로가 된 것은 소득이었어. *"What comforted a dad who had a hard time was income."* Bold signifies prosodically prominent item. PR: prosodic prominence; ST: syntactic prominence; N: no prominence; PS: prosodic and syntactic prominence.

Table 3. Example question–answer dialogue for English stimuli. Underlining signifies target words; bold signifies prosodically prominent item. PR: prosodic prominence; ST: syntactic prominence; N: no prominence; PS: prosodic and syntactic prominence.

| |
|---|
| Version 1: Prosodic prominence on a target without syntactic prominence |
| Q: Who was the noise keeping awake at night? |
| A: It was the noise (ST) that was keeping the **kids** (PR) awake at night. |
| Version 2: Prosodic prominence on a target with syntactic prominence |
| Q: What was keeping the kids awake at night? |
| A: It was the **noise** (PS) that was keeping the kids (N) awake at night. |

headphones with volume levels adjusted individually for comfortable audibility. To encourage participants to attend to the sentences, they were accurately informed that they were participating in a recognition memory test. One block of 10 practice items preceded the experimental items. As noted above, participants only heard one version of each sentence.

Sentences were pseudo-randomized so that all sentence types were evenly dispersed across the experiment and presented in 20 blocks of 10 sentences (each containing seven experimental, one control and two fillers). On-screen instructions explained that participants would hear sentences via their headphones; after each block of sentences, words would appear on the screen and they should respond YES versus NO as to whether they had heard them in the preceding block, signaling their answer by pressing the M versus Z key on the computer keyboard (with the key used for YES being that on the side of their dominant hand). After instructions, participants were told to press the space bar when they were ready to start; this cleared the screen apart from a central fixation point while sentence audio was being presented. Participants again pressed the space bar to play each sentence. During the recognition memory test participants were presented with 10 words sequentially on screen (one from each sentence) and asked to respond as quickly and accurately as possible. RT was measured from visual display onset and trials timed out after five seconds. Order of sentence and words presented in the recognition memory block from those sentences was kept the same (rather than scrambling) to ensure an equal time delay between presentation and testing for each word.

2.5 Results

2.5.1 Experiment 1a (Korean). No participants were excluded. Only experimental trials were analyzed for accuracy and RT analyses. Due to experimental error, one word was presented twice and was removed from all analyses (so 64 of 4480 trials were deleted). One prominence condition per person thus had two items fewer than the others, but counterbalancing ensured that all items and conditions were tested equally. Across all participants, just 19 responses (0.5%) were missing due to time-out (RT above five seconds). These were removed from all analyses, as well as any responses where RT was below 150 ms (2 trials). The final dataset for analysis consisted of 4395 trials.

Recognition accuracy and RT analyses were carried out using mixed-effects regression models in R 3.5.1 (R Core Team, 2018) with the package *lme4* (Bates, Maechler, Bolker, & Walker, 2015). The initial model included the fixed factors and *subjects* and *items* as crossed random effects. Random slopes for fixed factors were then added and retained if the model converged, and fit improved based on LogLikelihood (Bates, Kliegl, Vasishth, & Baayen, 2015; Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). The BOBYQA optimizer was used for all models to increase the likelihood of convergence with random effects. For logistic mixed-effects
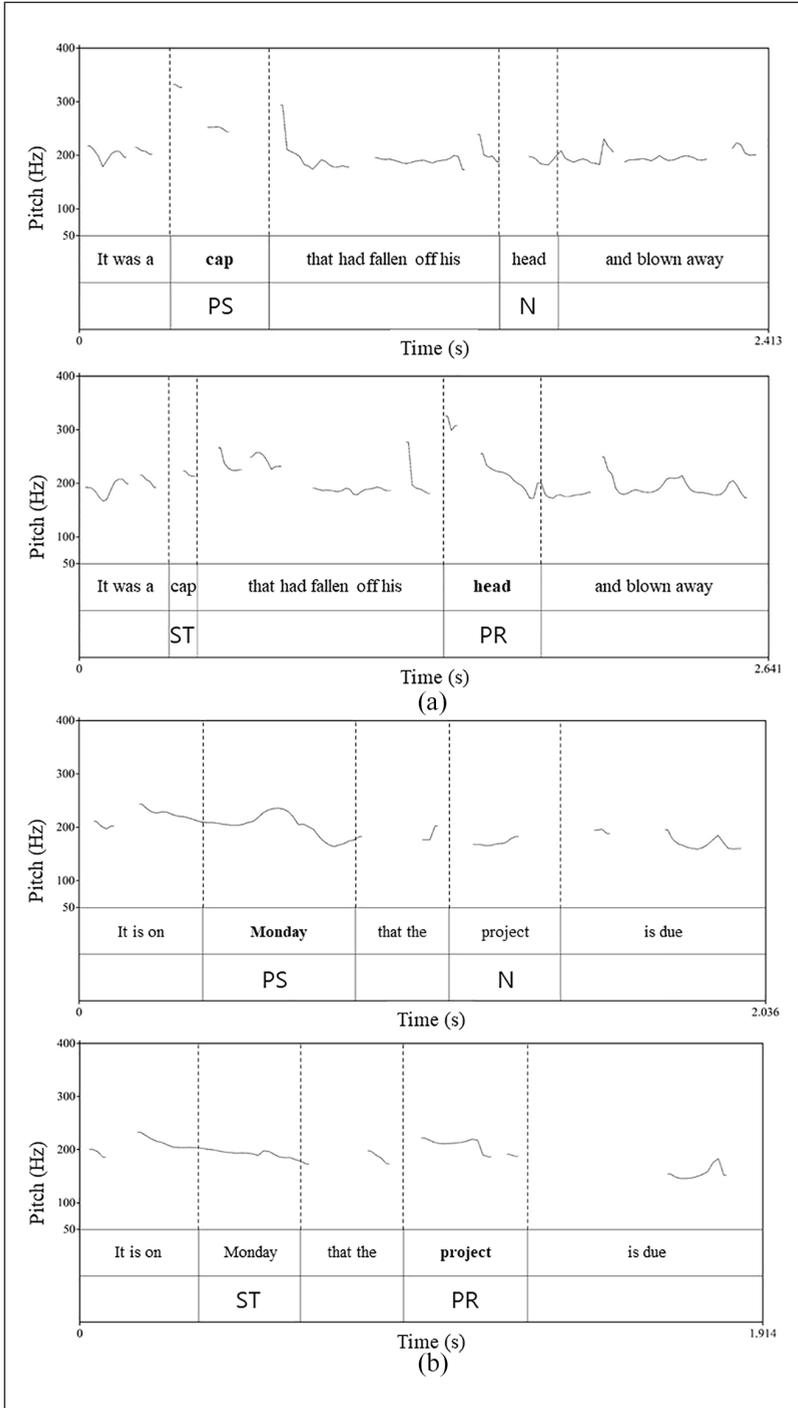
Figure 2. Example pitch tracks for two English sentences: (a) *It was a cap that had fallen off his head and blown away*; (b) *It is on Monday that the project is due*. Bold signifies prosodically prominent item. PR: prosodic prominence; ST: syntactic prominence; N: no prominence; PS: prosodic and syntactic prominence.

Table 4. English stimuli: Descriptive statistics and paired-sample *t*-test values for acoustic measures.

	First target: no syntactic prominence			Second target: with syntactic prominence		
	without prosodic prominence	with prosodic prominence	*p* values	without prosodic prominence	with prosodic prominence	*p* values
Duration (ms)	306.0 (9.6)	433.9 (10.9)	< 0.001	297.1 (7.8)	408.7 (8.4)	< 0.005
Peak F0 (Hz)	327.0 (17.8)	420.4 (18.9)	< 0.001	355.8 (15.5)	434.2 (18.4)	< 0.001
Peak intensity (dB)	56.0 (0.2)	60.5 (0.2)	< 0.001	59.1 (1.8)	61.8 (0.2)	< 0.001

regression models (glmers) for which BOBYQA was not the only default optimizer, the function allFit() was used to ensure model parameters did not differ significantly with other optimizers (Bates et al., 2015).

Contrasts for between-condition pairwise comparisons were extracted using the package *emmeans* (Lenth, 2018) with Tukey-adjusted *p*-values for multiple comparisons. Tables 5 and 6 summarize the model, model specification, and contrasts for accuracy and RT analyses. ANOVA summary tables for the final logistic regression models were calculated using the Anova() function from the package *car* (Fox & Weisberg, 2011). For linear regression models the anova() function from *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2017) was used.

2.5.1.1 Recognition accuracy. Accuracy was analyzed using a logistic mixed-effects regression model with *condition* as a four-level fixed factor (N, PR, ST, PS), *subjects* and *items* as random intercepts and *condition* as a random slope for *items*. There was a highly significant effect of condition, Wald χ^2 (3, *n* =32) = 41.49, *p* < 0.001, such that all prominence conditions were more accurate than the no-prominence control (N: 66.1%; PR: 73.0%, *p* < 0.01; ST: 77.5%, *p* < 0.001; PS: 82.4%, *p* < 0.001; see Figure 3). Responses in the prosodic + syntactic prominence condition were significantly more accurate than in the prosodic condition, on average by 9.3% (*p* < 0.01), as well as in the syntactic condition (mean difference = 4.9%, *p* < 0.01). Responses in the PR and ST conditions did not significantly differ (mean difference: 4.5%, *p* = 0.75); see Table 5.

2.5.1.2 Reaction times. RT of correct responses was analyzed using a linear mixed-effects regression model (lmer) with *condition* as a fixed factor, *subjects* and *items* as random intercepts and *condition* as a random slope for *items*, see Table 6. The overall model was significant, *F*(3, 135.6) = 10.25, *p* < 0.001, and responses in all prominence conditions (ST: 898 ms; PR: 957 ms; PS: 869 ms) were faster than in the control condition (N: 1025 ms; *p* < 0.001 for ST and PS, *p* = 0.026 for PR), see Figure 3 and Table 6. Among the prominence conditions, PS condition responses were faster than PR (*p* = 0.012), but not than ST (*p* = 0.638). There was no difference between ST and PR responses (*p* = 0.116); see Table 6.

2.5.1.3 Control analyses. As recency effects in memory—faster responses for remembered elements that were presented more recently as opposed to less recently—have long been known to be particularly strong in the auditory modality (Conrad & Hull, 1968; Nairne, 1988), we assessed the effect of prominence with target position held constant. To do this, we compared RT for words with syntactic prominence, that is, the late targets in the experimental conditions, against late targets in the control sentences, in a linear mixed-effects regression model with Tukey-adjusted

Table 5. Recognition accuracy analyses for Experiment 1a (Korean, left) and Experiment 1b (English, right). Part A reports the analysis of deviance tables (using Type II Wald χ^2 tests) of the final models. Part B reports all pairwise contrasts between levels of prominence condition (Tukey-adjusted). Estimates are given on the log odds ratio (not the response) scale. Part C gives the specification of the final model.

	Korean				English			
Part A.	χ^2	Df	p value		χ^2	Df	p value	
condition	41.49	3	< 0.001		48.66	3	< 0.001	
Part B.	Estimate	Std error	z value	p value	Estimate	Std error	z value	p value
N – PR	–0.418	0.126	–3.305	0.005	–0.493	0.121	–4.061	< 0.001
N – PS	–1.007	0.160	–6.287	< 0.001	–0.871	0.148	–5.900	< 0.001
N – ST	–0.567	0.140	–4.045	< 0.001	–0.227	0.136	–1.668	0.341
PR – PS	–0.589	0.169	–3.475	0.003	–0.379	0.156	–2.429	0.072
PR – ST	–0.149	0.149	–1.005	0.747	0.265	0.129	2.051	0.170
PS – ST	0.439	0.129	3.415	0.004	0.644	0.135	4.778	< 0.001
Part C. Model specification	glmer(accuracy ~ condition + (1|subject) + (1+condition|item), data = K1, family = "binomial", control = glmerControl(optimizer = "bobyqa"))				glmer(accuracy ~ condition + (1|subject) + (1+condition|item), data = E1, family = "binomial", control = glmerControl(optimizer = "bobyqa"))			

Table 6. Reaction time analyses for Experiment 1a (Korean, left) and Experiment 1b (English, right). Part A reports the analysis of deviance tables (Satterthwaite's method) for the final models. Part B reports all pairwise contrasts between levels of prominence condition (Tukey-adjusted). Part C gives the specification of the final model.

	Korean				English			
Part A.	NumDF	DenDF	F value	p value	NumDF	DenDF	F value	p value
condition	3	135.6	10.25	< 0.001	3	137.6	9.337	< 0.001
Part B.	Estimate	Std error	z value	p value	Estimate	Std error	z value	p value
N – PR	66.946	23.905	2.801	0.026	53.008	23.625	2.244	0.112
N – PS	154.955	29.233	5.301	< 0.001	122.693	24.350	5.039	< 0.001
N – ST	126.491	27.292	4.635	< 0.001	51.772	26.927	1.923	0.218
PR – PS	88.008	28.764	3.060	0.012	69.685	25.669	2.715	0.034
PR – ST	59.544	26.755	2.226	0.116	–1.236	27.145	–0.046	1.000
PS – ST	–28.464	24.064	–1.183	0.638	–70.921	22.402	–3.166	0.008
Part C. Model specification	lmer(RT ~ condition + (1|subject) + (1+condition|item), data = K1)				lmer(RT ~ condition + (1|subject) + (1+condition|item), data = E1)			

pairwise comparisons. If there is no difference between the syntactically prominent versus the control targets in this pair, then the faster RTs for syntactic prominence overall could have arisen from a recency effect. The dependent variable was RT and the independent variable, *condition* (3 levels: ST, PS, and late control). The overall model was significant, $F(2, 45.7) = 5.53$, $p = 0.007$, and crucially, words in each experimental condition were recognized significantly faster than
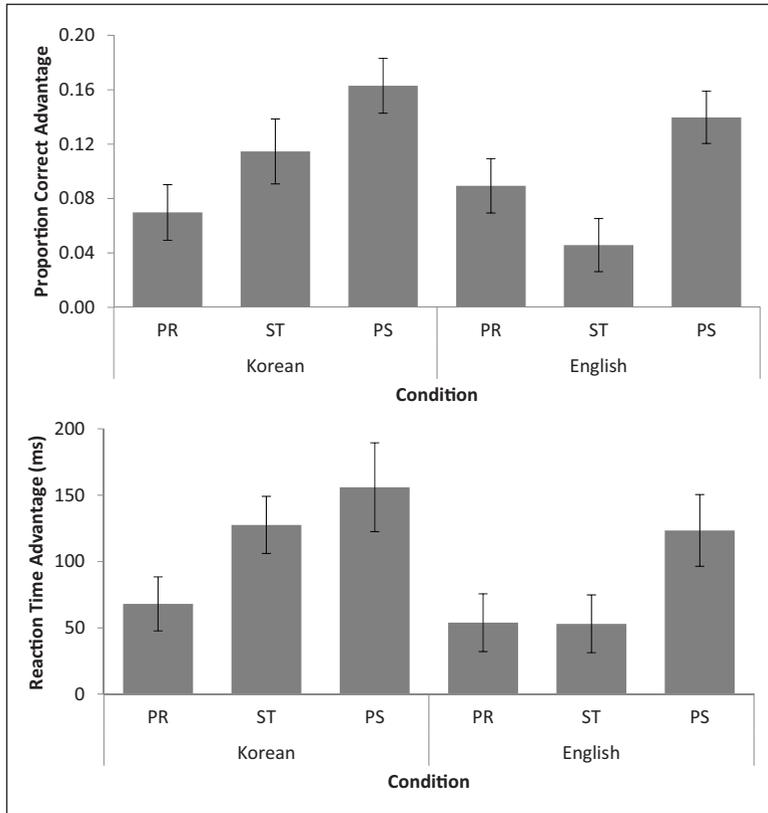
Figure 3. Experiment 1: Mean accuracy advantage (above, calculated as prominence condition mean proportion correct responses minus baseline mean proportion correct), and mean response time advantage (below, calculated as mean baseline condition RT minus prominence condition RT, in ms) of each prominence condition in each language. Error bars show standard error of the mean.

control-condition words (ST vs. control: Mean difference $= 105.20$, $p = 0.03$; PS vs. control: Mean difference $= 133.46$, $p = 0.006$). We therefore discard the possibility of a recency effect underlying the Korean syntactic prominence finding.

2.5.2 Experiment 1b (English). All participants and all items were included. Missing data (failure to respond) comprised only 18 out of 4480 trials (0.4%), and there were no trials where RT was less than 150 ms. Analyses were as for Experiment 1a.

2.5.2.1 Recognition accuracy. The final regression model consisted of *condition* as a fixed factor, *subjects* and *items* as random intercepts, and *condition* as a random slope for *items*. Again, a significant effect of *condition* on accuracy appeared, Wald χ^2 (3, $n = 32$) $= 48.66$, $p < 0.001$. Responses in the prosodic (PR: 75.2%) and prosodic + syntactic (PS: 80.2%) conditions were significantly more accurate than in the control (N: 66.2%) condition ($p < 0.001$), but there was no significant advantage for the syntactic prominence condition (ST: 70.9%; $p = 0.34$), see Figure 3. Accuracy was significantly greater in PS than ST ($p < 0.001$) but not than PR ($p = 0.07$). PR and ST did not significantly differ ($p = 0.17$), see Table 5.

2.5.2.2 Reaction times. The final regression model consisted of *condition* as a fixed factor, *subjects* and *items* as random intercepts, and *condition* as a random slope for *items*. There was again an overall effect of *condition*, $F(3, 137.58) = 9.34$, $p < 0.001$. Words with both prosodic + syntactic prominence (PS; mean RT: 916 ms) were responded to significantly faster than words in any other conditions: ST (986 ms; $p < 0.01$), PR (981 ms; $p = 0.034$), and control (N: 1033 ms; $p < 0.001$), see Figure 3. No other difference was significant (see Table 6).

A recency effect was not relevant to explanation of the English results pattern, so that no positional control analysis was undertaken.

2.6 Discussion

The experiment has revealed a clear effect of prominence in a heard utterance on memory for a heard word, yet the pattern of prominence effects was not identical across Korean and English. The difference can be clearly seen in the RT advantages (mean control condition correct RT minus mean experimental condition correct RT) shown in Figure 3. In each language the advantage for the PS condition is strongest, but the single-source prominence conditions produce different patterns across language: a consistently larger numerical advantage for ST than for PR in Korean, but not in English. Effectively, the principal asymmetry between the two languages lies in the effect size for ST, syntactic prominence only. Where position in the sentence is the standard criterion for the expression of prominence (Korean), position in the sentence exercises greater effect on listeners' responses in this task. Where both prosodic and syntactic options exist for expressing prominence (English), each exercises a similar effect.

In Experiment 2, we turn to the issue of the depth of these effects on processing and recognition. The course of phonological versus semantic activation in lexical recognition is well established (Norris et al., 2006) and we see no reason to postulate language-specificity in the processes involved. However, as described in the introduction, comparison across different types of foil recognition targets allows us to view the strength of the prominence manipulation during phonological and semantic processing respectively, potentially shedding light on the role of prominence in sentence processing. A strong effect of prominence on the processing of phonological foils will suggest that prominence processing may benefit from the acoustic expression of prominent versus non-prominent words. A strong effect of prominence on the processing of semantic foils will suggest that prominence effects may be associated with the computation of information structure.

3 Experiment 2

3.1 Participants

For Experiment 2a, 37 native Korean speakers (27 female: $M_{age} = 28.57$ years, $SD = 6.10$, range: 18–47 years), meeting the same criteria as for Experiment 1a, were recruited from the greater Western Sydney region. For Experiment 2b, 48 Western Sydney University undergraduate psychology students (37 female: $M_{age} = 21.29$ years, $SD = 3.93$, range: 17–37 years) took part; they met the same criteria as for Experiment 1b. Community participants were reimbursed for travel expenses, students received course credit. No Experiment 2 participant had taken part in Experiment 1.

3.2 Stimuli and procedure

The same recorded sentences were used as in Experiment 1. Again, the design was the same for Experiments 2a (Korean) and 2b (English). In each, 46 experimental sentences and their test words

remained exactly as in Experiment 1 (this enables a valuable direct replication of Experiment 1). For 47 sentences a phonologically related test word was displayed in place of the original sentence target (e.g., *charter* in place of *charcoal* in English) and for the remaining 47 sentences a semantically related target was selected (e.g., *fable* for *story*). For English, related words were chosen from a relatedness database (Nelson, McEvoy, & Schreiber, 1998); for Korean, words were selected using a relatedness task in which word pairs were rated on a 7-point scale, with 7 indicating highly related (chosen items had an average score of 6). Phonologically- and semantically-related test target words were matched to the prior targets on frequency and number of syllables. The procedure was as in Experiment 1, with each block containing at least one of each target type. Results (for accuracy and RT) were analyzed as for Experiment 1, with the addition of *target type* as a three-level fixed factor (Identical, Phonological, Semantic). The resulting models, model specifications, and contrasts for these analyses are summarized in Tables 7 and 8 below.

3.3 Results

3.3.1 Experiment 2a (Korean). No items or participants were excluded from analysis. Figures 4-6 show the results of Experiment 2 in the same way as for Experiment 1. In total, only 41 responses were missing across 5180 experimental trials (0.8%). No responses had a RT below 150 ms and the final dataset for analysis consisted of 5139 trials.

3.3.1.1 Accuracy. The final model consisted of *condition* and *target* as fixed factors with a *condition*target* interaction term, *subject* and *item* as random intercepts, *target* as a random slope for *subject*, and *condition* as a random slope for *item*. Analyses revealed a main effect of target type, Wald χ^2 (2, $n = 37$) = 30.98, $p < 0.001$, such that overall, phonological targets were more accurately rejected than identical targets were identified ($p < 0.001$). Phonological targets were also more accurately rejected than semantic targets ($p < 0.001$). There was also a significant interaction between prominence condition and target type, Wald χ^2 (6, $n = 37$) = 14.54, $p = 0.024$. We therefore examined between-condition contrasts by levels of target type.

For **identical** targets (i.e., the replication condition, in which correct answers were YES), accuracy was higher when targets were presented with PS compared to N ($p < 0.01$). No other pairwise comparisons of identical targets reached significance (see Table 7). For **phonological targets** (correct response NO), there were no significant pairwise comparisons, and likewise with **semantic targets** (correct response NO; see online Supplementary Information for detail).

3.3.1.2 Reaction times. The final model consisted of *condition* and *target* as fixed factors, *subject* and *item* as random intercepts, *target* as a random slope for *subject*, and *condition* as a random slope for *item*. Model comparison revealed that the *condition*target* interaction term did not improve the fit of the model based on LogLikelihood and it was therefore removed from the final model. Here our analyses revealed neither main effect to be significant (prominence condition: $p = 0.33$, target type: $p = 0.16$; see Table 8).

3.3.2 Experiment 2b (English). Again, no items or participants were excluded. There were only 10 missing responses (0.2%) and one response below 150 ms, resulting in 6661 trials for analysis.

3.3.2.1 Accuracy. The final model consisted of *condition* and *target* as fixed factors with a *condition*target* interaction term, *subject* and *item* as random intercepts, and *condition* as a random slope for *item.* The analyses again showed target type had a main effect on accuracy, Wald χ^2 (2, $n = 48$) = 15.30, $p < 0.001$, such that phonological and semantic targets were again more accurately rejected than identical targets were identified ($p < .001$ for phonological; $p = 0.047$ for semantic).

Table 7. Accuracy analyses for Experiments 2a (Korean, left) and 2b (English, right). Part A reports the analysis of deviance tables (using Type II Wald χ² tests) for the final models. Part B reports the Tukey-adjusted pairwise contrasts for the significant main effect, target type. Part C reports pairwise contrasts between prominence conditions for identical targets. Part D gives the specification of the final model.

	Korean				English			
Part A.		χ^2	Df	p value		χ^2	Df	p value
condition		6.14	3	0.105		5.62	3	0.132
target		30.98	2	< 0.001		15.30	2	< 0.001
condition*target		14.54	6	0.024		16.12	6	0.013
Part B. *target* contrasts	Estimate	Std error	z value	p value	Estimate	Std error	z value	p value
ident - phono	−1.305	0.327	−3.988	<.001	−0.481	0.132	−3.641	< 0.001
ident - seman	−0.455	0.306	−1.487	0.297	−0.304	0.128	−2.368	0.047
phono - seman	0.850	0.162	5.239	<.001	0.177	0.131	1.353	0.366
Part C. *identical* target contrasts	Estimate	Std error	z value	p value	Estimate	Std error	z value	p value
N − PR	−0.481	0.215	−2.241	0.112	−0.327	0.185	−1.772	0.287
N − PS	−0.789	0.227	−3.468	0.003	−0.803	0.235	−3.415	0.004
N − ST	−0.543	0.236	−2.300	0.098	−0.057	0.203	−0.281	0.992
PR − PS	−0.308	0.227	−1.356	0.527	−0.475	0.250	−1.897	0.229
PR − ST	−0.062	0.226	−0.274	0.993	0.270	0.239	1.133	0.669
PS − ST	0.246	0.197	1.251	0.594	0.745	0.206	3.615	0.002
Part D. Model specification	glmer(accuracy ~ condition*target + (1 + target|subject) + (1 + condition|item), data = K2, family = "binomial", control = glmerControl(optimizer = "bobyqa"))				glmer(accuracy ~ condition*target + (1|subject) + (1 + condition|item), data = E2, family = "binomial", control = glmerControl(optimizer = "bobyqa"))			

There was also an interaction between prominence condition and target type on accuracy (Wald χ^2 (6, $n = 48$) = 16.12, $p = 0.013$). For **identical** targets, accuracy was higher with PS than the baseline N condition ($p = 0.004$). Targets with PS also showed significantly higher accuracy than with ST ($p = 0.002$), but not compared with PR ($p = 0.23$) as in Experiment 1b. For **phonological and semantic targets**, there were no statistically significant pairwise comparisons (see online Supplementary Information).

3.3.2.2 Reaction times. Again, the *condition*target* interaction term did not improve the fit of the model based on LogLikelihood and was removed from the final model. The final model consisted of *condition* and *target* as fixed factors, *subject* and *item* as random intercepts, *target* as a random slope for *subject*, and *condition* as a random slope for *item*. There was no significant main effect of prominence condition on RT ($p = 0.066$) and also no main effect of target type ($p = 0.17$), see Table 8.

In summary, Experiment 2 has suggested that prominence effects in memory do not rapidly attain processing depth. Accuracy in our experiment was high; the foil targets were correctly rejected. When responses were incorrect, RT was always slower (by a greater margin, as expected, in the identical condition with YES as correct response; however, even in the other two conditions,

Table 8. Reaction time analyses for Experiment 2a (Korean, left) and Experiment 2b (English, right). Part A reports the analysis of deviance tables (Satterthwaite's method) for the final models. Part B gives the specification of the final model.

	Korean				English			
Part A.	NumDF	DenDF	*F* value	*p* value	NumDF	DenDF	*F* value	*p* value
condition	3	156.126	1.154	0.329	3	123.88	2.462	0.066
target	2	49.551	1.921	0.157	2	107.34	1.801	0.170
Part B. Model specification	lmer(RT ~ condition + target + (1 + target|subject) + (1 + condition|item), data = K2)				lmer(RT ~ condition + target + (1 + target|subject) + (1 + condition|item), data = E2)			

there was no sign of foils attracting fast incorrect responses). A table of correct versus incorrect response measures can be found in the online Supplementary Information.

For **identical targets**, which formed a replication of Experiment 1, the replication was as predicted. Although most comparisons did not reach statistical significance, the overall patterns in the difference analyses depicted in Figure 4 strongly resemble the original pattern of Figure 3. Thus the pattern we observed in Experiment 1 appears again, in each language, with a new set of participants. The foil conditions receive further attention in Section 4.

One explanation of the non-significance in the Experiment 2 results could be that the experiment, in which the number of items in each condition was lower than in Experiment 1, simply had insufficient power to replicate the earlier study. To check this, we conducted post-hoc power analyses using the R package *simR* (Green & MacLeod, 2016), a robust tool for calculating power in mixed effects models (Brysbaert & Stevens, 2018). Power was estimated based on simulations using accuracy data from Experiment 1. Simulations were modelled on 100 random samples of 32 participants and 180 items (a third of the data) in each of the English and Korean datasets. The average power based on 20 simulations each was 93.4 for Korean and 84 for English, exceeding the recommended level of 80 (Cohen, 1988). Therefore, it is unlikely that the Experiment 2 results pattern reflects an insufficient number of items.

To gain a more detailed picture of participants' response patterns, we examined the effects of each factor across Experiment 2 quartile by quartile. Comparison against condition by quartile analyses for Experiment 1 revealed that patterns of response altered across the course of the experiment in the case of Experiment 2, while this had not happened in Experiment 1. (Condition x [target type x] quartile means may be found in Table S4 in the online Supplementary Information.)

The obvious candidate for such an effect is the addition of the target type factor. The proportion of correct YES responses in Experiment 1 was high (four in five, i.e., all trials bar the 40 fillers). In Experiment 2, in contrast, it was low (one in three, i.e., only the identical and control sentences), with the offered target words in each block being potentially confusing in requiring a NO response despite having a semantic or phonological similarity to a word that had indeed occurred. While the Experiment 1 pattern may have fulfilled the intention of the experimenters (both in prior studies and in the present case) by positively encouraging participants to attend to processing the sentence content in order to remember the words, the Experiment 2 pattern, in contrast, could actually have discouraged such sentence-level processing since the remembered meaning may then have been short on the detail required to reject the similar-sounding or similar-signifying foils.

This interpretation is supported by the nature of the changing effects. In English, an effect of prominence (faster RTs to words with than to controls without prominence) was present but only in Experiment 2b's first quartile (in Experiment 1 the prominence effect appeared in all quartiles).
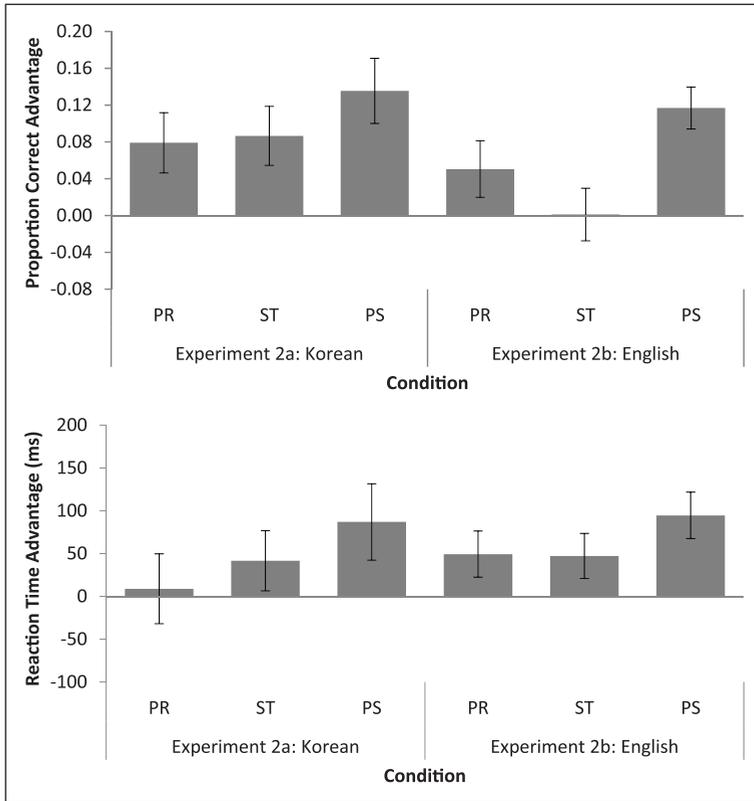
Figure 4. Experiment 2, Replication Condition: Mean accuracy advantage (above, calculated as prominence condition mean proportion correct responses minus baseline mean proportion correct), and mean response time advantage (below, calculated as mean baseline condition RT minus prominence condition RT, in ms) of each prominence condition in each language. Error bars show standard error of the mean.

In Korean, both the accuracy and RT results for Experiment 2a similarly resembled the Experiment 1 pattern (higher accuracy and faster RTs for PS and ST than for the other two conditions) in the first quartile only. Notably, response accuracy (which had stayed in the 70–80% range in each quartile in Experiments 1a and 1b) declined steadily across quartiles for the same responses, that is, in the replication condition, of both Experiment 2a and 2b (from 80 to 59% in Korean, and from 78 to 66% in English). All this suggests that the experiment itself induced a change of response strategy, with the YES response becoming harder to make when it had proved to be less likely. The differences between conditions (whereby prominent items had shown both RT and accuracy advantages over control items across Experiments 1a and 1b) also largely disappeared in Experiments 2a and 2b, possibly due to the strategy changes having resulted in lesser processing of the input sentence's information structure.

   Change of listening strategy across an experiment is a sign that participants are behaving rationally and adopting a response pattern which will enable them to perform as well as possible within the experimental constraints. It is by no means a rare phenomenon. For instance, it is also known to occur when normally reliable speech cues become unreliable. When prosodic components (in F0, duration and amplitude) of focus marking in speech are manipulated such that they give
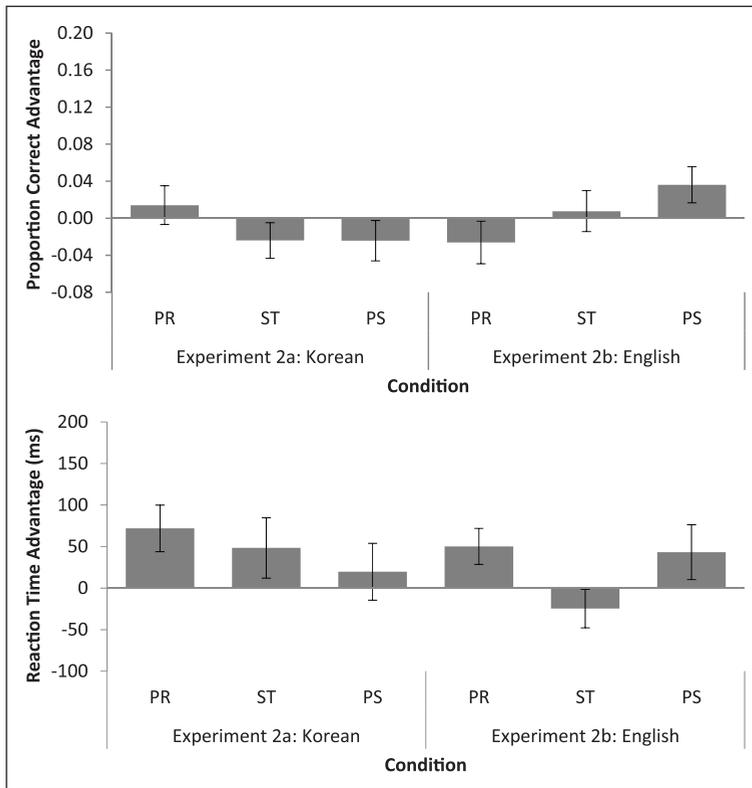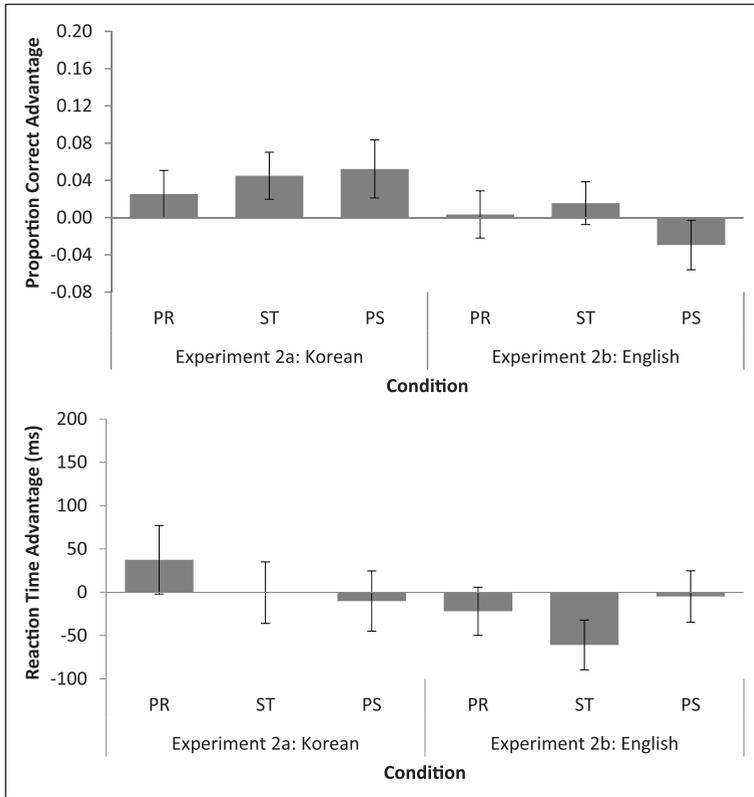
Figure 5. Experiment 2, Phonological Foil Condition: Mean accuracy advantage (above, calculated as prominence condition mean proportion correct responses minus baseline mean proportion correct), and mean response time advantage (below, calculated as mean baseline condition RT minus prominence condition RT, in ms) of each prominence condition in each language. Error bars show standard error of the mean.

conflicting information, listeners do not choose to attend to just one cue or another; instead, they stop processing all prosodic cues and attend to the segmental structure alone (Cutler & McQueen, 2014, an effect that there too was found to manifest itself increasingly across the course of individual experiments). Listener performance (in the laboratory, but presumably also in everyday listening situations) consists of many adjustable components by which the demands of a specific listening environment can be satisfied.

4 General discussion

Across languages, listeners need to process the information structure within utterances, and, as much earlier research has demonstrated, doing this involves, in particular, paying attention to where prominence is located. Our study has now shed a more intense light upon the manner in which prominence is processed, allowing us to propose tentative answers to some important questions (and thus suggest avenues for further investigations).

Our first question was whether improved memory for prominent elements, previously demonstrated in English and its close relative German, both languages with head prominence, would be observed in an unrelated language with a different prominence expression. Here the clear answer

Figure 6. Experiment 2, Semantic Foil Condition: Mean accuracy advantage (above, calculated as prominence condition mean proportion correct responses minus baseline mean proportion correct), and mean response time advantage (below, calculated as mean baseline condition RT minus prominence condition RT, in ms) of each prominence condition in each language. Error bars show standard error of the mean.

was, as expected, yes. Our study revealed broad equivalence of the effect of prominence on memory across our experiments. Language structure of prominence expression does not modulate listeners' processing of information structure; the most important elements need attention, and in consequence they are easier to recall.

If language structure does not affect *how* prominence is processed, does it then affect *where* and *when* it is processed? Here again the answer must be yes. In Korean, conditions where prominence was signaled by syntactic position showed greater processing benefit overall; in English, prominence signaled by prosodic highlighting had more benefit, and the condition with prominence signaled by syntax alone did not even significantly differ from the baseline condition. Note that the syntactic prominence was cued very clearly in both materials sets. In Korean, the ST-prominent word always followed the same morphosyntactic marker (은 /ɯn/, a grammatical particle indicating the topic of a sentence). In English, the ST element was always preceded by the copula and a determiner (*It was the. . . There is this. . .*). These cues should, and no doubt did, make the ST element stand out so that the listener's attention was drawn to it. However, this effect alone was only significant in Korean.

We suggest that differing frequency of exposure to the separate means of signaling prominence is most likely to underlie the contrasting patterns. The relative frequency with which users of English encounter prosodic highlighting versus cleft constructions has not, to our knowledge, been computed in spoken corpora, but it is probable that prosodic expression would be more frequent. In Korean, in contrast, a new phrase boundary is inserted before the location of prominence. In principle, prosodic and syntactic highlighting can be applied independently in Korean utterances, and if (as in our sentences) both are available and occur in a phonologically legal location at the onset of an accentual phrase, both can be used by listeners. But recall again here the findings of Lee et al. (2015) that when prosodic emphasis had to be realized in such a way as to violate these position constraints, neither production nor perception was fully accurate. In Korean, the fact that our listeners apparently benefited more from syntactic than from prosodic prominence signals (both correctly positioned with respect to accentual phrasing) may also reflect the relative frequency of the constructions in everyday Korean speech. Again, we have no spoken Korean corpus data on this, but once more it seems a likely scenario.

What we do know from corpus research is that there has been little success in automatic prominence detection using training corpora from another language, even one from the same language family (Rosenberg, Cooper, Levitan, & Hirschberg, 2012). This strongly suggests that language-specificity in prominence realization affects the perception of prominence, and further, that categories such as head- versus edge-prominence are not the sole determinant of where differences will be found. Note that our choice of Korean in this comparative study was driven not only by its status as an edge-prominence language, but also by the detail of how it realised the positional effect (at an initial edge). French also allows edge prominence, and as noted in the introduction resembles Korean both in the realization and perceptual implications of metrical structure; yet with the designated position for edge-prominence being a final phrasal edge, manipulations such as in our study are complicated, as French also has final accenting of lexical categories. Prominence judgements for French natural speech by humans and machines can mismatch due to this complexity (Goldman, Auchlin, Roekhaut, Simon, & Avanzi, 2010). Further, realization of prominence in French often differs in scope from the equivalent expressions of information structure in the Germanic languages (see for instance Destruel & Féry, issue), which would make it difficult to construct scope-matched materials such as those of the study we report. We conclude that language-specificity of prominence processing may extend beyond the level of family groupings such as prominence realization type.

These language-specific patterns also call for reinterpretation of prior findings. Akker and Cutler (2003) found *no* additive effects of combining semantic and prosodic prominence; in both English and Dutch, the two effects interacted such that the effect size for the one or both was the same. Akker and Cutler argued for a unitary prominence effect, equally communicable by whatever realization a speaker chose. That conclusion seems no longer tenable in the light of the additivity we have observed here for syntactic and prosodic highlighting in English. Prosodic prominence may be equivalent to semantic prominence, but its effect is separate from (and indeed additive to) that of syntactic prominence.

Substantial evidence confirms that processing of semantic structure can include discourse scenarios requiring potential alternatives to be ruled out. The active consideration of such alternatives was demonstrated in Dutch by Braun and Tagliapietra (2010) and in German by Gotzner, Spalek, and Wartenburger (2013). Interestingly, the latter team (Gotzner et al., 2013; Gotzner, Wartenburger, & Spalek, 2016; Spalek, Gotzner, & Wartenburger, 2014) also examined focusing particles (such as the German equivalents of *only* and *even*), which likewise trigger the activation of contextual alternatives. Enhanced memory traces for mentioned alternatives were again observed. When such

a particle and a prosodic cue were used in the same sentence, however, listeners' RTs to confirm that a probe word had occurred were slowed, indicating a processing cost in comparison to the trials with prosodic cues alone. This, like the present results, suggests that syntactic and prosodic signals of prominence function separately. Thus, in English, Dutch and German, the connection of prosodic highlighting to semantic prominence in the discourse context may be relatively more direct, while the action of syntactic devices in interpreting the discourse context may be less direct. This is clearly a difference that warrants further research attention.

In languages such as Korean, though, positional placement is significantly more important, and (our present results may suggest) more directly associated with discourse information structure. This could, in turn, motivate a Korean experiment in which semantic prominence was manipulated externally to the target-bearing sentence, such as by a preceding question, while syntactic prominence was manipulated within that sentence; the interaction observed by Akker and Cutler (2003) for semantic and prosodic prominence in Dutch and English might be observed in Korean for semantic and syntactic prominence.

The depth of processing manipulations incorporated in Experiment 2 were intended to address our next question: At what level(s) can we observe processing effects of prominence? Even at the beginning of Experiment 2 (where response patterns resembled those in Experiment 1), there was no indication that responses to semantically related targets differed across prominent versus control conditions, thus no evidence that prominence in an utterance is processed at a relatively deep or late level of processing. Nor did responses to phonologically related targets display effects of the prominence condition manipulation.

However, as already explained, we now suspect that the required processing level was not in fact tapped throughout Experiment 2, in that participants may have learned to adopt a strategy of paying less attention to the very level of processing (sentence-level information structure) that we were interested in. The similar pattern of effects in the portion of Experiment 2 that served as a mini-replication of Experiment 1 is gratifying and useful, but Experiment 2 otherwise serves as a cautionary tale, suggesting that a majority of distracting foils in memory studies of this type may alter participants' strategic choices about level of processing.

The final question (one with a distinguished history: Bolinger, 1978) concerns prosodic prominence in particular; is prosodic highlighting indeed a universal prominence signal? There are two ways to interpret universality here. The first is: do all languages signal prominence by prosodic means? The answer to this version of the question is already known to be negative, however (e.g., from Rialland and Robert's (2001) analysis of the African non-tone language Wolof, which is held to eschew intonational marking of focus prominence in favour of morphosyntactic markers doing the same work). Another version of the universals question, though, is perceptual: if prosodic highlighting is applied, will listeners of all languages interpret it as a signal of prominence? To this our findings support a cautiously positive answer; in a language in which prominence is preferentially signaled by positioning, and effects of syntactic prominence were strongest, prosodic highlighting of words in a different position nevertheless resulted in improved memory, suggesting facilitated lexical processing. The final picture thus combines universal options with language-specific choices, even when the field of action for prominence processing is no greater than the processing of a prominent word.

Acknowledgements

Funding

ORCID iD

Anne Cutler ⓘD https://orcid.org/0000-0002-4203-0692

Supplemental material

Supplemental material for this article is available online.

References

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

Baayen, R., Piepenbrock, R., & Van Rijn, H. (1995). The CELEX database. *Nijmegen: Center for Lexical Information, Max Planck Institute for Psycholinguistics, CD-ROM*.

Bates, D. M., Kliegl, R., Vasishth, S., & Baayen, H. R. (2015). Parsimonious mixed models. *arXiv preprints*, arXiv:1506.04967.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Birch, S. L., Albrecht, J. E., & Myers, J. L. (2000). Syntactic focusing structures influence discourse processing. *Discourse Processes*, *30*, 285–304.

Birch, S. L., & Garnsey, S. M. (1995). The effect of prominence on memory for words in sentences. *Journal of Memory and Language*, *34*, 232–267.

Bock, J. K., & Mazzella, J. R. (1983). Intonational marking of given and new information: Some consequences for comprehension. *Memory & Cognition*, *11*, 64–76.

Boersma, P., & Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 5.3.62. Retrieved from http://www.praat.org/

Bolinger, D. (1978). Intonation across languages. *Universals of Human Language*, *2*, 471–524.

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

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*, 1–20.

Burdin, R.S., Phillips-Bourass, S.P., Turnbull, R., Yasavul, M., Clopper, C.G., & Tonhauser, J. (2015). Variation in the prosody of focus in head- and head/edge-prominence languages. *Lingua*, *165*, 254–276.

Calhoun, S. (2010). The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language*, *86*, 1–42.

Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, *29*, 155–190.

Cho, T., Lee, Y., & Kim, S. (2011). Communicatively driven versus prosodically driven hyper-articulation in Korean. *Journal of Phonetics*, *39*, 344–361.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

Conrad, R., & Hull, A. J. (1968). Input modality and the serial position curve in short-term memory. *Psychonomic Science*, *10*, 135–136.

Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics*, *20*, 55–60.

Cutler, A., & Fodor, J. A. (1979). Semantic prominence and sentence comprehension. *Cognition*, 7, 49–59.

Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech*, *20*, 1–10.

Cutler, A., & McQueen, J. M. (2014). How prosody is both mandatory and optional. In J. Caspers, Y. Chen, W. Heeren, J. Pacilly, N. Schiller & E. van Zanten (Eds), *Above and Beyond the Segments* (pp. 71–82). Amsterdam: John Benjamins.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 113–128.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17–22.

Destruel, E., & Féry, C. (this issue). Prominence in French dual focus. *Language and Speech*, *64*, 319–345.

D'Imperio, M., German, J., & Michelas, A. (2012). A multi-level approach to prominence, phrasing and intonation in French. In G. Elordieta and P. Prieto (Eds), *Prosody and Meaning* (pp. 11–34). Berlin: de Gruyter Mouton.

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory and Language*, *63*, 367–386.

Fox, J., & Weisberg, S. (2011). *An {R} Companion to Applied Regression*. 2nd ed. Thousand Oaks CA: SAGE.

Goldinger, S.D., Luce, P.A., & Pisoni, D.B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, *28*, 501–518.

Goldman, J.-P., Auchlin, A., Roekhaut, S., Simon, A.C., & Avanzi, M. (2010). Prominence perception and accent detection in French: A corpus-based account. *Speech Prosody 2010, 100575*.

Gotzner, N., Spalek, K., & Wartenburger, I. (2013). How pitch accents and prominence particles affect the recognition of contextual alternatives. *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, 36, 2434–2439.

Gotzner, N., Wartenburger, I., & Spalek, K. (2016). The impact of prominence particles on the recognition and rejection of contrastive alternatives. *Language and Cognition*, *8*, 59–95.

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*, 493–498.

Hay, J. F., Sato, M., Coren, A. E., Moran, C. L., & Diehl, R. L. (2006). Enhanced contrast for vowels in utterance prominence: A cross-language study. *Journal of the Acoustical Society of America*, *119*, 3022–3033.

Jeon, H.-S., & Nolan, F. (2010). Segmentation of the accentual phrase in Seoul Korean. *Speech Prosody 2010, 10023*.

Jun, S.-A. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology*, *15*, 189–226.

Jun, S.-A. (2000). K-ToBI (Korean ToBI) labelling conventions. *UCLA Working Papers in Phonetics*, 99, 149–173.

Jun, S.-A. (2011). Prosodic markings of complex NP prominence, syntax, and the pre-/post-prominence string. In M. B. Washburn, K. McKinney-Bock, E. Varis, A. Sawyer, & B. Tomaszewicz (Eds), *Proceedings of the 28th West Coast Conference on Formal Linguistics* (pp. 214–230). Somervilla, MA: Cascadilla Press.

Jun, S.-A., & Cha, J. (2015). High-toned [il] in Korean: Phonetics, intonational phonology, and sound change. *Journal of Phonetics*, 51, 93–108.

Jun, S.-A., & Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus*, *14*, 147–172.

Jun, S.-A., & Lee, H.-J. (1998). Phonetic and phonological markers of contrastive prominence in Korean. *Proceedings of the Fifth International Conference on Spoken Language Processing*, *4*, 1295–1298.

Kim, H.-G., & Kang, B.-M. (2008). *Frequency analysis of Korean morpheme and word usage*. Seoul, Korea: Institute of Korean Culture, Korean University.

Kim, J., Davis, C., & Cutler, A. (2008). Perceptual tests of rhythmic similarity: II. Syllable rhythm. *Language and Speech*, *51*, 343–359.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26.

Lee, Y.-C. (2012). Prosodic correlation between the focusing adverb *ozik* 'only' and focus/GIVENness in Korean. *Journal of Speech Sciences*, *2*, 85–111.

Lee, Y.-C., Wang, B., Chen, S., Adda-Decker, M., Amelot, A., Nambu, S., & Liberman, M. (2015). A crosslinguistic study of prosodic prominence. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 4754–4758. Brisbane: Curran Associates, Inc.

Lee, Y.-C., & Xu, Y. (2010). Phonetic realization of contrastive prominence in Korean. *Speech Prosody 2010*, 1000033.

Lenth, L. (2018). emmeans: Estimated marginal means, aka least-squares means. R package version 1.2.4. Retrieved from https://CRAN.R-project.org/package=emmeans.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H. R., & Bates, D. M. (2017). Balancing type 1 error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.

Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, *20*, 298–305.

Michelas, A., & D'Imperio, M. (2010). Durational cues and prosodic phrasing in French: evidence for the intermediate phrase. *Speech Prosody 2010*, 100118.

Nairne, J. S. (1988). A framework for interpreting recency effects in immediate serial recall. *Memory & Cognition*, *16*, 343–352.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. Retrieved from http://www.usf.edu/FreeAssociation.

Norris, D., Cutler, A., McQueen, J.M., & Butterfield, S. (2006). Phonological and conceptual activation in speech comprehension. *Cognitive Psychology*, *53*, 146–193.

Psychology Software Tools, Inc. [E-Prime 3.0]. (2016). Retrieved from https://www.pstnet.com.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/.

Rialland, A., & Robert, S. (2001). The intonational system of Wolof. *Linguistics*, *39*, 893–940.

Roediger, H.L., III, & McDermott, K.B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *21*, 803–814.

Rosenberg, A., Cooper, E., Levitan, R., & Hirschberg, J. (2012). Cross-language prominence detection. In Q. Ma, H. Ding, & D. Hirst (Eds), *Proceedings of the 6th International Conference on Speech Prosody 2012* (pp. 278–281). Shanghai, China: Tongji University Press.

Spalek, K., Gotzner, N., & Wartenburger, I. (2014). Not only the apples: Prominence sensitive particles improve memory for information-structural alternatives. *Journal of Memory and Language*, *70*, 68–84.

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

Underwood, B.J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*,*70*, 122–129.

Vroomen, J., Van Zon, M., & De Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory & Cognition*, *24*, 744–755.