Published in final edited form as:

J Mem Lang. 2011 February 1; 64(2): 153–170. doi:10.1016/j.jml.2010.11.001.

Stress Matters: Effects of Anticipated Lexical Stress on Silent Reading

Mara Breen and Charles Clifton Jr.

University of Massachusetts Amherst

Abstract

This paper presents findings from two eye-tracking studies designed to investigate the role of metrical prosody in silent reading. In Experiment 1, participants read stress-alternating noun-verb or noun-adjective homographs (e.g. *PREsent, preSENT*) embedded in limericks, such that the lexical stress of the homograph, as determined by context, either matched or mismatched the metrical pattern of the limerick. The results demonstrated a reading cost when readers encountered a mismatch between the predicted and actual stress pattern of the word. Experiment 2 demonstrated a similar cost of a mismatch in stress patterns in a context where the metrical constraint was mediated by lexical category rather than by explicit meter. Both experiments demonstrated that readers are slower to read words when their stress pattern does not conform to expectations. The data from these two eye-tracking experiments provide some of the first on-line evidence that metrical information is part of the default representation of a word during silent reading.

Keywords

eyetracking; lexical stress; silent reading; implicit prosody; ambiguity resolution

As adult readers of English, the vast majority of our reading is silent. However, we share the impression that, even during silent reading, a voice is speaking inside our heads. This phenomenon can be particularly salient when we encounter foreign words or names in texts, and find that we are ‘tripped up’ by their pronunciation, even in the absence of overt production. This ‘little voice’ is of particular interest if it can be shown to play a functional role in reading comprehension. In the present paper, we briefly review some of the roles that the activation of phonological representations during silent reading has been shown to play, and present research showing that one aspect of a phonological representation – its metrical structure – affects eye movements during silent reading.

Rayner and Pollatsek (1989) chose the term phonological coding to describe “the mental representations of speech that can give rise to the experience of hearing sounds” (p. 189). They argued that the role of phonological coding is to strengthen memory for words during text processing. Slowiaczek & Clifton (1980) made a similar claim, demonstrating that suppressing subvocalization by having subjects constantly vocalize during reading impaired memory for inferences and for propositions that required integration across distinct sentences, but not memory for specifically stated propositions. They argued that phonological coding strengthens the memory representation of the material that was read, which facilitates conceptual integration.

There is ample evidence that phonological coding demonstrably influences word-level processing during silent reading: Homophones interfere with category decisions (it's hard to reject *rows* as a flower; Van Orden, 1987); phonological differences associated with spelling similarity slow processing (it's harder to choose a continuation for the phrase *nasty hasty* than the synonymous *mean rash*; Treiman, Freyd, & Baron, 1983); tongue twisters are read slowly (McCutchen & Perfetti, 1982), and with more errors (Dell & Repka, 1992) than control sentences; and phonological information (but apparently not word identity, except in cases of word skipping) is extracted from words in the parafovea (Pollatsek, Lesch, Morris, & Rayner, 1992; Ashby, Treiman, Kessler, & Rayner, 2006; Henderson, Dixon, Petersen, Twilley, & Ferreira, 1995).

There is some disagreement about when phonological information becomes available. Kennison and colleagues (Kennison, 2004; Kennison, Sieck, & Briesch, 2003), for example, demonstrated late effects of phonemic similarity in silent self-paced reading, such that slow-downs in reading due to repeated phonemes were later than (and did not interact with) those observed for syntactic garden-paths. In contrast, however, Warren and Morris (2009) found evidence of phonemic similarity affecting earlier processing when eye movements were measured.

There is also a smaller body of research showing effects of phonological coding on higher levels of sentence comprehension. This research has focused primarily on the role of suprasegmental information (i.e. prosody) during silent reading. Suprasegmental phonology describes the acoustic properties of speech that convey information beyond which segments make up a word, including tone, intonation, stress, and phrasing. There is some evidence that implicit prosodic phrasing can affect how silently-read sentences are parsed and interpreted. Fodor (1998) proposed a "same-size sister constraint", which suggests that speakers tend to produce intonational phrases of the same size, and readers implicitly do the same. She argued, for example, that speakers and readers may treat the string *The divorced bishop's daughter* as a single intonational phrase, and, due to some heuristic like late closure (Frazier, 1979), interpret it as referring to a divorced bishop. However, when the string is lengthened to *The recently divorced bishop's daughter*, Fodor argues that, in an effort to produce balanced phrases, readers will impose a phrase boundary between *divorced* and *bishop*. Because listeners are unlikely to make attachments over phrase boundaries (Kjelgaard & Speer, 1999; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Schafer, Speer, Warren, & White, 2000), readers will say it is the daughter who is divorced, not the bishop. Supporting evidence for implicit prosodic phrasing comes from Augursky (2008), Bader (1998), Hirose (2003), Hwang and Schafer (2008), and Swets, Desmet, Hambrick, & Ferreira, (2007).

Another aspect of suprasegmental phonology which has been argued to influence silent reading is metrical structure, which describes both the structure of syllables and how they relate to one another. There is some initial evidence that metrical structure is accessed during silent reading. Ashby and Martin (2008) showed that information about a word's syllable structure (whether it had a CV or a CVC initial syllable) was extracted from the word in the visual parafovea, which they took to be evidence that some aspects of the metrical phonology of a word are processed before the word is fixated. Ashby and Clifton (2005) demonstrated that readers look longer at four-syllable words with two stressed syllables (e.g. *RAdiAtion*) than one stressed syllable (e.g. *inTENSity*), while also showing that the sheer number of syllables in a word does not matter (two-syllable words were pronounced more slowly than one-syllable words but were not silently read more slowly). Following Sternberg, et al. (1978), they suggested that the time to prepare an implicit pronunciation of a word was determined by the number of stressed syllables, and that this time affected eye movements.

We turn our attention to another aspect of metrical phonology, namely, the pattern of stressed and unstressed syllables (referred to as S and W, for *strong* and *weak*). Stressed syllables have been shown to hold a privileged position in auditory language comprehension: Listeners are faster to detect phonemes in stressed syllables (Shields, McHugh, & Martin, 1974); lexical access is more disrupted by the mispronunciation of stressed syllables than unstressed syllables (Mattys & Samuel, 1997); and listeners interpret stressed syllables as word onsets (Cutler & Norris, 1987; cf. Cutler, Dahan, & van Donselaar, 1997).

Unlike languages like Polish and Finnish, which have regular and predictable stress patterns in words, stress in English words is not fully predictable, and is often lexically determined. Of interest for the experiments we report here, English has pairs of semantically related nouns and verbs that differ systematically in stress¹: nouns have initial syllable stress, verbs have non-initial stress (consider the noun *CONduct* and the verb *conDUCT*; the noun *ABstract* and the verb *abSTRACT*; etc.) (we follow the convention of putting stressed syllables in uppercase). We report two experiments that, in very different ways, create the anticipation of a word with one stress pattern or the other (SW or WS, i.e., strong-then-weak or weak-then-strong syllable structure), which is then confirmed or disconfirmed by the content of the sentence in which the word appears. If readers are in fact imposing an implicit metrical structure on what they read, and if there is measurable cost in revising an anticipated metrical structure to conform to what is actually required by the sentence, their reading may be disrupted by a metrically nonconforming word.

Experiment 1

The first experiment used a “brute force” method of manipulating metrical expectations. Participants read limericks, which have an extremely salient metrical structure (described briefly below). The metrical structure of the first line entailed a particular metrical structure for the second line, but the part of speech that ended this line in our experiment forced a stress pattern that was either consistent or inconsistent with the required metrical structure. If readers impose a metrical structure during silent reading, a lexical requirement that this metrical structure be revised could disrupt reading. We explore the details of how and when this disruption might appear after presenting the data.

Method

Participants—30 undergraduates at the University of Massachusetts Amherst served as participants. All reported speaking English as their first language, and had normal or corrected-to-normal vision. Subjects received either extra course credit or \$8 for their participation. Data from two participants were removed due to a failure to track their eyes for the entire course of the experiment, such that 28 participants contributed data to the reported analyses.

Materials—Participants read limericks where a stress-alternating noun-verb homograph was the final word of the second line. Limericks, like (1), which served as a filler in the current experiment, are poetic devices with several constraints on both rhyme and meter.

- (1) While painting the church steeple gray,
The wind blew our brushes away.

¹The set of stress-alternating homographs in English includes both noun-verb homographs (*CONduct/conDUCT*) and noun-adjective homographs (*COMplex/comPLEX*). However, the vast majority of materials in Experiment 1, and all materials in Experiment 2, are composed of noun-verb homographs, so we focus on this subset.

We said to the pastor,
 “We’ve had a disaster.”
 He calmly replied, “Let us spray.”

With regard to rhyme scheme, lines one, two, and five (e.g. *gray, away, spray*), and lines three and four (e.g. *pastor, disaster*), respectively, rhyme. Each line is composed of a number of metrical feet, which are the smallest units of poetry, consisting of two-three syllables, one or more of which is stressed. With regard to meter, lines one, two, and five have three metrical feet, while lines three and four have two. These feet are usually of the form W-W-S (e.g. *let us SPRAY*), but there are two frequent exceptions: The first foot of each line can have the form W-S (e.g. *we SAID*); the last foot can have the form W-W-S-W (e.g. *to the PASTor*) (Zwicky & Zwicky, 1986). These constraints are not inviolable, and many limericks which are considered well-formed do not conform to them; however, limericks written for the experiment were designed to adhere to these constraints as tightly as possible.

Forty target stress-alternating noun-verb or noun-adjective homographs were selected from those used by Pitt & Samuel (1990), and from a list generated using the Celex database (Baayen, Piepenbrock, & van Rijn, 1993). Half of the items selected had a reduced vowel in the unstressed syllable of the verb form (e.g. *con* in *conVICT*), and half had an unreduced vowel in this position (e.g. *re* in *reFUND*) (according to the authors’ judgments; see Appendix 1). (We note that this factor never had a significant effect on the measures we analyzed, and therefore will be disregarded.) A list of all homographs used is found in Appendix A.

For each critical homograph, the authors wrote four limericks, like those in Table 1, with the following properties: (a) the lexical stress of the critical word was strong-weak (*PREsent* in Lines A and B of Table 1) or weak-strong (*preSENT* in lines C and D), and (b) the lexical stress of the critical word was consistent with the metrical context of the limerick (Lines A and C) or inconsistent (Lines B and D). In all cases, the first lines of the strong-weak consistent and weak-strong inconsistent limericks were the same, as were the first lines of the strong-weak inconsistent and weak-strong consistent limericks. The second lines necessarily differed among all four conditions, to force the final word to be a noun or to be a verb, and to maintain plausibility. The remaining three lines were written to be reasonably coherent and to maintain the pattern of the first line, but were not of experimental interest. The resulting 160 experimental limericks can be found at <http://people.umass.edu/mbreen/pubs/LimericksItems.pdf>

An additional 60 standard limericks were gathered from internet sources to serve as fillers. All the fillers adhered closely to the constraints outlined above, and were altered in cases where they violated the constraints. This extra care was taken to ensure that participants would see canonical limericks 80% of the time, so that those limericks intended to be metrically inconsistent would be highly salient.

In order to ensure that participants were semantically processing the limericks, but without drawing attention to the experimental manipulation, participants’ task was to decide whether each limerick they read was ‘dirty’ or not. This task was chosen because a successful ‘dirty’ determination required that readers fully attend to, and comprehend, the entire limerick. Twenty of the filler limericks (e.g. 2) were chosen to be ‘dirty.’

- (2) There was a young lady named Frances,
 Who suffered embarrassing trances.
 She stripped to the skin,

Before Father Flynn,
And made him indecent advances.

None of the ‘dirty’ limericks were sexually explicit or contained obscenities, but contained either mild sexual or scatological reference. Participants were all informed of the task in advance, and given the opportunity to withdraw from participation without penalty.

The 160 experimental limericks were distributed across four lists, according to a latin-square design. The sixty fillers (40 clean, 20 dirty) were the same on each of the four lists, resulting in 100 limericks on each list. Each participant was randomly assigned to one of the four lists, which was presented in an individually randomized order.

Procedure—Participants were tested individually. After providing written informed consent, participants were taken into the testing room, shown the eyetracker, and given instructions to read limericks and decide whether they were ‘dirty’ or not. Participants were seated 55 cm from a CRT monitor that displayed the sentences. Their eye movements were recorded using an EyeLink 1000 (SR Research, Toronto, Ontario, Canada) eyetracker controlled by a PC running the University of Massachusetts EyeTrack software (<http://www.psych.umass.edu/eyelab>). Eye position was sampled at 1kHz. Although participants viewed the stimuli with both eyes, only data from the right eye were collected.

After the tracker was calibrated, the experiment began. A trial was initiated when the participant fixated on a box located where the first letter of the limerick would appear. All limericks were presented in five lines, just as they appear in (3), in a fixed-width font. Each character subtended a visual angle of .36 degrees. When they had finished reading a limerick, participants were instructed to look to a colored square affixed to the right edge of the monitor and to pull a trigger on a game controller. The question “Dirty?” appeared on the screen, and participants answered “yes” or “no” by pulling one of two triggers on a game controller. Responses to these questions were recorded. Calibration was checked on each trial by checking the alignment with a single centered target. Minor errors were automatically corrected; unacceptably large errors led to recalibration. The entire experiment took 45 minutes or less.

Results

Participants’ mean question-answering accuracy was 88%. Because the questions were subjective (i.e., whether the limerick was ‘dirty’) we would not expect all participants to share our judgments; however, the high observed accuracy indicates that participants were engaged in the task of reading the limericks.

The limericks were divided for analysis purposes into six regions, as illustrated by the subscripts in (3):

- (3) There once was a clever young gent₁|
 Who had a nice talk to₂| present₃|
 But then he got scared₄|
 His speech was impaired₅|
 And nobody knew what he meant₆|

Note that the material is different across conditions for all regions but Region 3, the critical region. Since this variation introduces uncontrolled variability into eye movement measures, we will concentrate on the critical region, Region 3. However, because of the possibility that

difficulty reading Region 2 could influence the time to read Region 3, we will also present some data for Region 2.

Prior to analysis, the eye movement data were cleaned of track losses, blinks, and long fixations over 800ms using the program EyeDoctor (<http://www.psych.umass.edu/eyelab>). Following Rayner, Sereno, Morris, Schmauder, & Clifton (1989), we assume that short fixations, i.e., those under 80 ms, do not contribute useful information to the reader. Thus, short fixations were incorporated into the nearest neighboring fixation within three characters; otherwise, they were deleted. Fixations greater than 800 ms were also eliminated, on the assumption that these did not represent normal acquisition of information from the text (Rayner et al., 1989). Finally, trials with blinks or track losses on the critical word were eliminated. In total, 91 trials out of 896 were excluded from analysis, representing 10% of the data.

We present the following standard eyetracking measures (Rayner et al., 1989; Rayner, 1998): First-pass time (sum of all fixation durations made from first entering to first leaving a region, eliminating trials on which no such fixations occurred), go-past time (sum of all fixation durations made from first entering a region to first leaving it to the right), the probability of fixating a region, and the probability of regressing out of a region given that it was fixated during the first pass. We also examined first fixation duration for Region 3, but since no effects approached significance, these data are not reported here.

The mean values (and standard errors) of the eyetracking measures for Regions 2 and 3 appear in Table 2. Individual trial time measures were analyzed using a mixed model, as were the binary events of fixating a region or regressing out of a region. The R statistical programming language was used for all analyses (R Development Core Team, 2007). The fixed factors included in the model were (1) the metrical pattern (strong-weak, weak-strong, or SW/WS) of the critical word (region 3) and (2) the critical word's consistency with metrical context (whether the context required a word with the same metrical pattern of the critical word or a different pattern), and (3) the interaction of these two factors. Additional fixed factors were evaluated during model exploration (log frequency of occurrence in the Celex norms, the ratio of frequency of use as noun and as verb in these norms, and whether or not the vowel in the unstressed syllable was reduced) to be included in the model if they significantly improved the model fit. Improvement of model fit was quantified by a likelihood ratio test comparing models with and without the additional factor. If the comparison was significant, the additional factor was included. (We note that these additional factors never did improve model fits, so can be disregarded.) Treatment coding was used, with effects being evaluated with respect to the consistent, SW baseline (thus, inconsistent SW was compared to consistent SW, as was consistent WS; differential effects of consistency appear as the interaction between the two fixed-effect factors). Models were fit with subject and item intercepts as the random factors, as well as with the individual subject and item slopes for metrical pattern and consistency as random factors. The model fits were compared following Baayen, 2008, such that the more complex model with random slopes was used only if it improved the model fit. (See Baayen, 2008, for discussion.) Significance levels were estimated using Markov Chain Monte Carlo sampling when possible (Baayen, Davidson, & Bates, 2008) (which is not currently possible when random slopes are used). Occurrences of fixations and regressions out were analyzed in a similar way but using logistic regression (Jaeger, 2008). The results of the model fits for Region 3 (the critical word, the primary region of interest), including significance levels, appear in Table 3.

We first discuss the results of analyses on the critical region (Region 3), then turn to the results for Region 2. The critical word of Region 3 was more likely to be skipped (not

fixated) when it was consistent with the preceding context than when it was inconsistent. However, the higher frequency of fixating a metrically inconsistent word appeared only when the lexically-required stress of the word that actually appeared was WS (i.e., the word was a verb) (interaction $p < .02$). Previous work has demonstrated that a word is likely to be skipped if it is identified before the eye leaves the previous word (Drieghe, 2008; Rayner, 1998; Rayner & Duffy, 1988; see Reichle, Pollatsek, Fisher, & Rayner, 1998, for a theoretical analysis). Similarly, given that the critical word in Region 3 was fixated, there was a marginally higher probability of regressing out of it when it was inconsistent, but again, only when the lexical critical word was a WS word (interaction $p = .08$). First pass reading time in Region 3 gave little indication of difficulty reading the critical region when it was metrically inconsistent with context. However, go-past times were significantly slowed in the WS inconsistent condition of Region 3 (interaction $t = 2.43$, $p < .05$), providing clear evidence for disruption of reading in this condition (Figure 1).

This finding of disruption in Region 3 may be qualified by the fact that Region 2 was read at different speeds in the different conditions. First pass reading times for Region 2 were somewhat faster in the inconsistent SW condition compared to the consistent SW baseline ($t = 1.90$, $p < .10$) and significantly slower in the inconsistent-WS condition (interaction $t = 3.52$, $p < .01$). Go-past times showed a similar pattern of effects (t 's of 1.57, n.s., and 3.86, $p < .01$, respectively) (pMCMC values are not reported in these cases because random slopes improved the model fit, and it is not currently possible to compute pMCMC with random slopes). These differences are difficult to interpret, given that different words, different structures, and different content appeared in the material preceding the critical word in the different conditions (necessarily so, to induce the different metrical and syntactic expectations). They cannot, however, be attributed to 'preview' effects of Region 3, which might have resulted in long fixations at the end of Region 2 in the inconsistent WS condition, as analyses in which the material was resegmented to move the last 7 characters of Region 2 (where preview might have been expected) into Region 3 left the same pattern of results for both regions.

Importantly, the observed Region 3 differences cannot be attributed to confounding lexical differences, as was the case in Region 2. In Region 3, all items had the same spelling, and the verb/noun or WS/SW difference was orthogonal to the interaction effect. However, the observed effect of slowest reading times on inconsistent WS targets in Region 3 could in principle be a spillover of difficulty from Region 2. Reading difficulty in one analysis region is commonly observed to appear on the next region as well (Rayner, 1998). However, three pieces of evidence lead us to believe that spillover was not the basis of the long go-past times we observed in Region 3. First, spillover presumably happens on a trial-by-trial basis: a slow-down in reading in one region spills over to the next region. However, the trial-by-trial correlation of go-past time between Region 2 and Region 3 was a negligibly-small $r = 0.024$. Therefore, the trials on which the target word was read with difficulty were not trials on which the previous region was read slowly. Second, an initial analysis of Region 3 go-past times was conducted using Region 2 go-past times as the sole fixed effect predictor, and the residual Region 3 go-past times that remained after any effect of Region 2 times was removed were analyzed in the same way as the raw Region 3 go-past times had been (with stress and consistency as fixed effect factors). These residual reading times have any systematic effect of Region 2 reading difficulty statistically removed. Nonetheless, the interaction of stress and consistency remained significant (see Table 3), indicating difficulty in reading inconsistent WS words (verbs) over and above any persisting difficulty of reading earlier material on the line. Third, variations in Region 2 difficulty did not result in different Region 3 landing positions, which could have affected the time to read that region. Specifically, landing position in Region 3 ranged between character 4.84 and 5.0 in the different conditions, and a linear mixed model analysis of these data indicated that neither

the main effects nor an interaction were significant ($t < 1.0$). For these reasons, therefore, we conclude that the observed effects on Region 3 were not due to spillover, but rather to readers' local difficulty with the WS inconsistent condition.

Discussion

The first experiment provided some indication that readers experienced difficulty when they had to read a word whose stress pattern was inconsistent with context. The effect was apparent in a lowered probability of skipping the critical word, a marginally higher probability of regressing out of the critical word given that it was fixated, and longer times to move past the critical word after fixating it. However, this effect appeared only when this word was a WS word (a verb) in a SW metrical context, not when the word was a SW word (a noun) in a WS context.

We hypothesize that stress inconsistency was disruptive because readers activate metrical patterns of words during silent reading. We further hypothesize that the disruptive effect of metrical inconsistency reflects the role of implicit metrical coding in creating a representation of text that supports interpretation and integration of the text. Such a representation must be faithful to the material that is read if it is to be useful, and must in particular contain representations of lexical items that are accurate, including being metrically accurate.

The preceding interpretation of the role of implicit metrical stress assignment supports a coherent interpretation of the asymmetry of the disruptive effect of stress inconsistency. The metrical consistency of limericks provides a strong constraint on what the metrical pattern of the critical word should be. This constraint allows readers to predict the metrical pattern of the upcoming critical word, and impose this pattern on their 'inner speech' representation. When a reader encounters a mismatch between the metrical pattern predicted by context and the actual pattern of the critical word, the reader must shift the location of the metrically strong (i.e. stressed) syllable of the critical word. For example, when readers encounter the verb *preSENT* in a context that predicts a SW word like *PREsent*, they must shift stress on the critical word rightward from the predicted to the required stress; when they encounter the noun *PREsent* in a context that predicts a WS word like *preSENT*, they shift stress leftward. Only the former shift was found to be costly.

The fact that we only observe longer reading times when readers encounter a WS word in a context which predicts SW can be traced to the relative acceptability, and frequency, of left- and rightward stress shifts in speech. English speakers fairly frequently shift the stress of a word from a later syllable to the initial syllable, presumably to avoid a stress clash with a following word that has initial-syllable stress (e.g., *THIRteen MEN*) (Grabe & Warren, 1995; Shattuck-Hufnagel, Ostendorf, & Ross, 1994; Liberman & Prince, 1977). In fact, there is a diachronic shift in English toward stressing words on the initial syllable (e.g. *esCORT* to *EScort*) (Bolinger, 1965), perhaps reflecting the preponderance of English words with initial syllable stress (Cutler, Dahan, & van Daneselaar, 1997). We propose that when the second line of the limerick metrically induces the reader to anticipate a WS word, but syntactically induces the expectation of a noun (which is always a SW item in our materials), the reader has no difficulty shifting the implicit stress pattern to SW. Shifting stress from a later to an earlier syllable is a well-practiced skill for English speakers. Thus, an inner speech representation of a SW noun can be formed quite easily, even in a metrically inconsistent context. This leads to fairly fluent identification of the target word, promoting skipping and reducing the need to re-read earlier material. However, when the metrical anticipation is for a SW word but a WS verb is syntactically required, the reader must shift the implicit stress from the first to the second syllable. This is seldom observed in spoken English, and is not a skill that speakers are highly practiced in. Assuming, as we do, that a

full phonologically faithful mental representation of a word is generally made available before the eyes move on, having to shift from SW to WS stress will disrupt identifying the critical word. Since this word is the last word on the line and the last word of a clause, a position from which regressions are quite common, word identification difficulty could be expected to result in frequent regressive eye movements. This is especially so because material earlier on the line led to the metrical and syntactic expectations, and a reader might well be tempted to see whether anything about the line had been miscoded.

We are thus attributing the metrical inconsistency effect observed in Experiment 1 to the difficulty of producing an appropriate inner speech representation; that is, it is difficult to make a SW representation into a WS one, by shifting stress to the right. An attractive alternative explanation, which our data allow us to reject, would place the source of difficulty in a perceptual, rather than a production, process. Such an explanation assumes that the inner speech representation plays a role in actually identifying the word. The reader forms an inner speech representation based on the orthographic form plus contextual expectations, and then “listens” to this representation to identify the word. In listening experiments (Cutler & Clifton, 1984; Taft, 1984), auditory word identification seems to be disrupted when a normally SW word is pronounced as WS, but minimally disrupted when a normally WS word is pronounced SW. If our readers were identifying the visual word by “listening” to its inner speech representation, this asymmetry would lead one to expect difficulty when a normally SW word appeared in a context that encouraged a WS pronunciation, but not vice versa. The observed asymmetry was the reverse of this, suggesting that the difficulty arises from creating a faithful inner speech representation when it clashes with context, not from the difficulty of recognizing a word with stress on the wrong syllable.

Although the data from Experiment 1 provide some evidence for the activation of metrical stress patterns during silent reading, the result may not generalize to normal reading. These results may be due in part to the fact that metrical patterns are necessarily more salient in poetic contexts like limericks than in normal prose. In addition, the fact that the contexts in which the critical words appeared necessarily had to differ raises concerns about whether the observed effects were due only to the local requirement to shift the stress of the target word. Therefore, we designed a second experiment which manipulated metrical expectations in a different way, and also allowed for full control of the contexts in which the critical words appeared.

Experiment 2

Some readers may allow that Experiment 1 demonstrated effects of metrical stress pattern on the silent reading of limericks, but doubt that this demonstration would extend to the reading of prose. Limericks, after all, are designed to have a very salient metrical beat. For this reason, a second experiment was carried out to investigate whether inconsistencies between predicted and actual metrical patterns lead to disruption in silent reading; however, the materials and the manipulation were quite different. While the limerick context of Experiment 1 directly led readers to expect a specific metrical pattern, the second experiment led readers to expect a specific syntactic construction which entailed a particular metrical pattern. Experiment 2 was similar to Experiment 1 in that this metrical expectation was either confirmed or disconfirmed, depending on the part of speech of a word that could have SW or WS stress. The observation of disrupted reading in Experiment 2 when a metrical expectation is disconfirmed will provide evidence that the implicit generation of metrical structure is a general characteristic of silent reading.

Readers were induced to generate a syntactic expectation by the exploitation of a long-recognized noun-verb ambiguity, illustrated in *The old man the boats*. There is a strong tendency to take the adjective-noun homograph *old* to be an adjective and the noun-verb homograph *man* to be a noun, an analysis that must give way to a noun-verb analysis upon reading the following *the*. The critical words (see Table 4) were noun-verb homographs that did (e.g., *abstract*) or did not (e.g., *report*) alternate in stress pattern between their noun and verb uses. When the stress of a word alternated, the noun pattern was strong-weak (SW) (e.g., *ABstract*) while the verb pattern was weak-strong (WS) (e.g., *abSTRACT*). We compared these homographs to nearly-unambiguous nouns and verbs (e.g., *paper*, *suggest*). The adjective-noun homographs in our sentences (e.g., *brilliant*) were selected to be biased toward an adjective interpretation, in order to bias the noun-verb homograph towards a noun interpretation, which would have to be revised when the following material indicated that it had to be a verb. This revision should cause reading difficulty, compared to sentences with unambiguous nouns or verbs. If metrical stress was implicitly assigned to the critical word, and if a costly revision of metrical stress was required as part of the revision of grammatical category, reading difficulty should be greater when the noun-verb homographs differed in metrical pattern (e.g., *abstract*) than when they did not (e.g., *report*).

Methods

Materials—Thirty sets of sentences were constructed, in six versions each as illustrated in Table 4 (where the ^ marks indicate the regions to be analyzed). All materials appear in Appendix B. Three versions of a sentence set began with a Determiner-Adjective-Noun sequence, and three with a Determiner-Noun-Verb sequence. The second word was initially ambiguous between adjective and noun, but strongly biased toward noun. The three versions of each of these types differed only at the third word of the sentence, which was the critical word. In the Ambiguous experimental sentences (conditions A–B, D–E), the critical word was ambiguously a noun or a verb (e.g., *abstract*, *report*). In the stress-alternating versions (e.g., *abstract* in conditions A & D), the critical word varied in stress pattern, such that it was SW as a noun and WS as a verb. In the Non-alternating versions (e.g., *report* in conditions B & E), stress pattern was the same for both noun and verb uses, although it varied between SW and WS among items (e.g., *rePORT* vs. *RALly*). Separate analyses of the non-alternating ambiguous items indicated that this factor did not affect reading times. In the Unambiguous sentences, the third word was nearly unambiguously a noun or a verb (e.g., *paper* in condition C, *suggest* in condition F, respectively). Although eight of the 60 ‘unambiguous’ nouns or verbs also appeared in their less frequent form—verb for noun and vice versa—in the Ku era-Francis (Francis & Ku era, 1968) norms, none occurred in their less frequent form more frequently than seven times per million.

The word or two that followed the critical third word disambiguated the sentence-initial sequence toward the adjective-noun analysis or toward the noun-verb analysis. We will refer to these two disambiguations as “Noun” and “Verb” respectively, reflecting the part of speech of the critical word (e.g., *abstract*). The manner of disambiguation varied among sentences. In most cases that were disambiguated toward the Noun interpretation, the first word in the disambiguating region was an auxiliary verb, but in some sentences it was a main verb or a preposition that modified the preceding noun. In most cases that were disambiguated toward the Verb interpretation, the first word was a determiner, but in some cases it was an unambiguous adjective or a bare plural noun.

The mean Ku era-Francis (Francis & Ku era, 1968) frequencies of the adjective-noun homograph that was the second word in each sentence, were 58 per million for use as adjective, and three per million for use noun (with no Ku era-Francis noun entries for 19 of the 30 entries.) There was thus a very clear bias in favor of the adjective use of the word, as

intended. The mean lengths of the critical third words, which served as a noun or a verb, and the Ku era-Francis frequencies of use of each word as a noun and as a verb appear in Table 5. For the stress-alternating words, the difference in frequency between noun (40 per million) and verb (44 per million) usage did not approach significance. Noun uses of the non-alternating items (97 per million) were marginally more frequent than verb uses (60 per million) ($p = .06$), but the frequency of each was fairly high. The overall frequencies of the unambiguous nouns (87 per million) and verbs (94 per million) did not differ significantly.

The 30 items were divided into three counterbalanced lists, each containing two sentences from each set of six versions of an item. A latin-square design was used, so that the two versions of a given lexical item were always different and always resolved the noun-verb homograph in different ways (in particular, if the Alternating Noun item (condition A) was assigned to a list, so was the Non-alternating Verb item (condition E), and similarly for Non-alternating Noun and Unambiguous Verb versions (conditions B and F), and the Unambiguous Noun and Alternating Verb versions (conditions C and D). Each list contained ten different items in each of these pairs of versions, and each item was tested in each version in one list. The resulting 60 sentences were combined with 72 unrelated items from other experiments plus eight practice items. Simple yes-no comprehension questions were made up for 12 of the 30 sets of experimental sentences (e.g., *Was the paper accepted?* and *Do the brilliant find good ideas?*).

Norming. Before reporting the results of Experiment 2, we report normative data collected after Experiment 2 was completed.

The prediction of increased reading times when metrical stress and syntactic category have to be simultaneously revised depends on two assumptions: The first assumption is that participants initially interpret a noun-verb homograph (stress-alternating or not) as a noun in the Experiment 2 sentence-initial contexts. This is a precondition for the assumed cost of syntactic reanalysis when the homograph is disambiguated as a verb. The second assumption is that participants pronounce the stress-alternating homographs with SW stress when they are used as nouns (e.g., *ABstract*), but with WS stress when they are used as verbs (e.g., *abSTRACT*). This is a precondition for the prediction that there will be a cost of revising their metrical assignment when they revise syntactically.²

To ensure that both of these assumptions were correct, we conducted two norming studies with the Experiment 2 materials. In the first, we used a fragment-completion technique to assess how often participants interpreted the noun-verb homographs as nouns in the same sentence-initial context in which they occurred in the experiment. Sixty-two participants who had not participated in the eye-tracking study participated in the paper-and-pencil completion study for course credit. The first three words of an item appeared as a prompt, as in (4), and participants were instructed to provide a coherent completion of the sentence.

- (4) a. The brilliant abstract _____
 b. The brilliant report _____
 c. The brilliant paper _____
 d. The brilliant suggest _____

Three counterbalanced lists were constructed in a similar way to how the lists were constructed for Experiment 2, as described above. Each list contained each initial *The adj/*

²The second assumption must also be made for Experiment 1. While the pronunciation norms to be reported did not test all the 40 homographs used in Experiment 1, they tested 20 of them. Any support the norms provide for making the assumption in Experiment 2 therefore provides some support for making the assumption in Experiment 1.

noun (e.g., *The brilliant*) twice: One-third of these initial strings were followed by an alternating homograph (4a) on one occurrence and a non-alternating homograph (4b) on the other occurrence; one-third were followed by a non-alternating homograph and an unambiguous verb (4d); and one-third were followed by an unambiguous noun (4c) and an alternating homograph. These assignments were counterbalanced across items in the three lists. In addition to the 60 experimental items, each list also contained 88 items from two unrelated experiments, which did not feature syntactically ambiguous materials. Each list was presented in one randomized order to all participants. Approximately equal numbers of participants received each list.

Three native English speaking research assistants, who were naïve to the predictions of the experiment, coded the completions as indicating the participant's interpretation of the critical homograph as (a) a noun, (b) a verb, (c) an adjective, or (d) unclassifiable. Because there appeared to be little or no ambiguity about coding categories, we collected only one coding of each completion. That is, each assistant coded a different subset of the completed questionnaires, and, therefore, there was no basis for evaluating agreement among coders. The proportion of each completion type by condition is shown in Table 6. The important result is that participants most often interpreted the homograph as a noun when it appeared in the left context that was used in Experiment 2, and that this interpretation was quite similar for the stress-alternating (79%) and non-alternating (84%) homographs. Unambiguous nouns were almost always treated as nouns, while unambiguous verbs were most often treated as verbs, but on occasion as nouns or in ways that resulted in unclassifiable completions. All evidence points to the conclusion that the *The X* preamble was quite consistently taken as *The* adj, leading to the expectation that the next word would be a noun.

In a second norming study, we elicited productions of our experimental items from a new set of participants, who had not participated in either the eye-tracking study or the completion study. The experiment was conducted using Linger, a software platform for language processing experiments (Rohde, 2008). Fourteen pairs of naïve participants, a speaker and a listener, native English speakers who participated for course credit, sat at computers in the same room such that neither could see the other's screen. The "speakers" were instructed that they would be producing sentences for their partners (the "listeners"), and that the listeners would be required to answer a comprehension question about each sentence immediately after it was produced. Each trial began with the speaker being presented with a sentence on the computer screen to read silently until s/he understood it. The speaker then answered a true-false content question about the sentence, to ensure understanding, and then proceeded to produce the sentence out loud. Speakers' question answering accuracy ranged from 89–95% across conditions. The listener sat at another computer, and saw a blank screen while the speaker went through the procedure described above. After the speaker produced a sentence out loud for the listener, the listener would press the space bar on his/her computer. A true-false question about the content of the sentence just heard would then appear. Listeners were provided feedback when they answered a question incorrectly. Listeners' accuracy ranged from 88–92% across conditions.

Another native English speaking research assistant, who was also naïve to the experimental hypotheses, first pre-screened each production. Sixteen out of the 840 total productions were excluded due to speaker disfluency or recording failure. The research assistant then digitally spliced the critical word out of each useable production so that a listener could judge the stress pattern of the target word without being biased by his/her knowledge of the word's syntactic category. The authors independently listened to each target word and determined whether the stress pattern was S-W or W-S. Before independently classifying all the productions, the authors listened to a subset and agreed upon classification criteria (most

importantly, a reduced vowel indicated W stress; if neither vowel was reduced, the syllable marked by higher pitch, greater intensity, and longer duration was determined to be stressed (Fry, 1955; Lieberman, 1960). The authors' subsequent independent classification of the full dataset resulted in 98.6% agreement.

The proportion of WS productions for each condition is shown in Figure 2. The results demonstrate that the elicited stress patterns were consistent with the predicted patterns in three ways. First, stress-alternating homographs were produced with WS stress only 16% of the time when disambiguated as nouns but 77% of the time when disambiguated as verbs. Second, the stress pattern of the non-alternating homographs did not differ based on syntactic category, such that they were produced with WS stress 27% of the time when disambiguated as nouns, and 28% of the time when disambiguated as verbs. Moreover, these two subsets were the same, such that homographs produced with SW stress as nouns were also produced with SW stress as verbs, and homographs produced with WS stress as nouns were also produced with WS stress as verbs.³ Third, unambiguous nouns were never produced with WS stress, and unambiguous verbs were produced with WS stress 98% of the time. These data also indicated that the verbal use of *ally* was not in our subjects' vocabularies: they essentially never pronounced it with WS stress, and they very frequently stumbled over it when pronouncing a sentence that used *ally* as a verb. Therefore, in the present data and in all further analyses, we excluded sentences containing this item.

Participants—Forty-three University of Massachusetts undergraduates served as participants in Experiment 2. All reported speaking English as their first language, and all had normal or corrected-to-normal vision. All subjects received course credit for their participation. Five participant sessions were aborted due to an inability to maintain tracker calibration. Six more sessions were completed, but the data were excluded from subsequent analysis as excessive blinking and long saccades removed 25% or more trials. Therefore, 32 subjects contributed data to the analyses presented below.

Procedure—The procedure for Experiment 2 was the same as for Experiment 1 except that a single sentence was presented on each trial, on a single line, with each character subtending a visual angle of .32 degrees. When participants had finished reading a sentence, they were instructed to look to a colored square affixed to the right edge of the monitor and to pull a trigger on a game controller. Questions appeared as a whole immediately after this trigger-pull. Participants answered “yes” or “no” by pulling one of two triggers on the game controller. Responses to these questions were recorded. The experiment lasted approximately 45 minutes.

Results

Mean question answering accuracy was 86% (SD 12.8%) indicating that participants read the sentences with adequate comprehension.

Prior to analysis, the eye movement data were cleaned of track losses, blinks, and long fixations over 800ms using EyeDoctor. As in Experiment 1, short fixations were incorporated into the nearest neighboring fixation within 3 characters; otherwise, they were deleted. Two trials were deleted due to track loss in one of the critical regions (2, 3, or 4); 50 were deleted due to blinks in a critical region; 58 were deleted due to fixations on the critical regions in excess of 800ms. These exclusions totaled 110 trials, or six percent of all trials. Further, we discovered an error in one unambiguous item: we used *repair* as an

³There were three instances where this was not the case. On one occasion, a speaker produced the verb form of report as *REport*. There were also two cases where a speaker produced the verb form of outfit as *outFIT*.

unambiguous verb, which is not the case. Therefore, in analyzing the data, we eliminated sentences containing this item.

We report the five reading time measures that were reported in Experiment 1: first fixation duration, first pass time, go past time, proportion of regressions in, and proportion of regressions out (Rayner et al., 1989; Rayner, 1998). In addition, we report second pass time, which measures the time spent re-reading a region, once the region has been exited to the right, i.e., the time after a regression into a region that the reader spends re-reading.

The primary reading duration (first fixation, first pass time, and go past time) and probability measures (proportion of regressions out) were analyzed for three critical regions of the sentences, Regions 2, 3 and 4, as illustrated in the sample sentences of Table 4. Re-reading measures (second pass time, regressions in) were analyzed for Regions 1, 2, and 3. The mean values (and standard errors) of these measures appear in Table 7. The data (apart from second pass times) were analyzed on an individual trial basis using a mixed model with Lexical Condition (non-alternating ambiguous, alternating ambiguous and unambiguous) and Noun vs. Verb disambiguation as fixed factors. As in Experiment 1, subject and item slopes were treated as random factors if they improved the model fit over one which included only subject and item intercepts as random factors. If they did not result in a significantly better model (when tested by chi-square, using the anova function in R), we report the model with subject and item intercepts as random factors. When only intercepts were included, significance levels were estimated using Markov Chain Monte Carlo sampling (Baayen, Davidson, & Bates, 2008); as this method is not implemented when models include random slopes, a value of t greater than 2.0 was interpreted as significant when random slopes were included. The three-level Lexical Condition factor was analyzed as two treatment contrasts. The first contrasted Unambiguous vs. Non-alternating Ambiguous items, in order to identify any effects of the lexical ambiguity present in the Non-alternating condition; the second contrasted Non-alternating vs. Alternating Ambiguous items, to identify any additional effects of the stress alternation in the latter condition, which would presumably implicate implicit metrical stress. The best-fitting model parameters are given in Table 8.⁴

The first analyses to be reported reflect the initial reading of each region together with regressions out of the region. Analyses of re-reading will be presented later.

Region 2—This region contained the critical lexical item. The initial analyses compared Non-Alternating and Unambiguous words. First fixation durations were disproportionately long for Unambiguous Verbs (see Table 7 for means, and Table 8 for parameter estimates and significance of the relevant interactions). First pass times and go-past times showed the same effect; the interaction between Noun vs. Verb and Unambiguous vs. Non-Alternating for first pass times can be seen in Figure 2. These effects demonstrate that an unambiguous verb like *suggest* after *The brilliant* (see example in Table 4) is disruptive, because *suggest* requires *brilliant* to be analyzed as a noun, which presumably requires reanalysis from an initial adjective analysis.

The next set of analyses of Region 2 compared Non-alternating vs. Alternating lexical items. First fixation and first pass times were disproportionately long for Alternating Ambiguous

⁴We also conducted various analyses in which we included as predictors measures derived from the normative data presented earlier. For instance, we included a measure of how strongly each sentence context biased toward an expectation that the critical word would be a noun vs. a verb, and (in analyses of alternating items), a measure of how consistently the verb received a WS pronunciation. In no case did adding these predictors result in a better-fitting model. We believe that this is because there was a severe restriction of range of our predictor variables – that is, our manipulations were uniformly quite successful – and we note that few simple correlations of our predictor variables with measures of reading time were significant.

words that were to be disambiguated as verbs (see Table 7 and Figure 3). While the interaction between Noun vs. Verb and Alternating vs. Non-alternating words was significant or nearly significant for both of these measures, it was not significant for go-past time. This presumably reflected the relatively small increase from first pass to go-past time for Alternating verbs, which in turn presumably reflected the low frequency of regressions out of the verb in that condition (although the corresponding effect in the analysis of regression-out frequencies did not reach significance). (Frequencies of regressions into the region and second pass times will be discussed later, as they reflect re-reading.)

The greater effect of disambiguation of the critical word toward a Verb for Alternating than Non-alternating words is initially puzzling, since the verb-disambiguation and noun-disambiguation sentences are identical up through Region 2. Both verb and noun sentences begin (e.g.; see Table 4) *The brilliant abstract* where *abstract* is Region 2. However, detailed analysis of the data provides support for an interesting conjecture; namely, that the effect of type of disambiguation reflects the effect of parafoveal identification of the next word. This word was often a short function word (e.g., as in Table 4, *was* in the case of noun disambiguation, *the* in the case of verb disambiguation). We examined trials on which the initial, disambiguating part of the disambiguating region was skipped and compared them to trials on which this region was fixated. If a reader who is fixating in Region 2 (the critical word) obtains enough information parafoveally about the next region to skip it, this parafoveal information may well be enough to disambiguate the critical word. On the other hand, fixating the next region provides evidence that the reader did not identify the word it contained while fixating Region 2 (cf. Drieghe, Rayner, & Pollatsek, 2005; Drieghe, 2008, for an analysis of word-skipping during reading).

To do this examination, we re-segmented the materials, dividing Region 3 into the first word (as long as that word disambiguated toward a noun vs. verb use of the critical word) and the second word. Most typically, this meant that the new first part of Region 3 (which we will call Region 3') contained an auxiliary verb or a determiner. When Region 3 originally contained a single word, it was treated as the new Region 3' (except when the word was 'themselves', in which case only 'them' was treated as Region 3'). We then analyzed first fixation and first pass times conditional upon the subject fixating Region 3' on a trial, and conditional on the subject skipping Region 3'. This resulted in a total of 1136 observations when Region 3' was fixated, and 610 when it was skipped.

The resulting mean reading times appear in Table 9 (with first pass times in Figure 4). The increased reading times for Alternating items that were to be disambiguated as a Verb were substantially affected by skipping. First fixation durations showed a Verb-disambiguation cost of 9 ms when Region 3' was fixated, but a cost of 45 ms when it was skipped. First pass durations showed costs of 1 and 71 ms respectively. A different pattern of results was seen for Non-alternating items. They generally showed no clear added cost when the next region disambiguated toward a Verb as compared to a Noun, whether that region was skipped or not (first fixation, -6 ms when Region 3' was fixated, 10 ms when it was skipped; first pass, -5 ms when it was fixated, -24 ms when it was skipped). The interaction of Noun vs. Verb disambiguation and Non-alternating vs Alternating word was significant when Region 3 was skipped (pMCMC < .05 for first fixation, p MCMC < .001 for first pass) but not when Region 3 was fixated (pMCMC > .30 for both measures) (Figure 4).⁵

⁵A reviewer noted that alternating words were fixated longer than non-alternating words when they were fixated, and suggested that this might be a concern. We note that this difference did not interact with disambiguation toward noun vs. verb (24 vs. 30 ms effects, respectively). We conclude, therefore, that this difference is likely a lexical one, reflecting in part the somewhat higher frequency of occurrence of the non-alternating words.

The Unambiguous critical items showed a substantial cost of being disambiguated toward a Verb, but in contrast to the Alternating items, the cost was numerically larger when Region 3' was fixated than when it was skipped (a first fixation cost of 30 ms when Region 3' was fixated, 12 ms when it was skipped; and a first pass cost of 41 ms in the former case, 21 ms in the latter case). This pattern differed from the pattern for Non-alternating items. The interaction of Noun vs. Verb disambiguation and Non-alternating vs. Unambiguous critical word was significant when Region 3 was fixated ($pMCMC < .02$ for first fixation, $pMCMC < .04$ for first pass). When Region 3 was skipped, the interaction was not significant for first fixation ($pMCMC = .75$) but was significant for first pass ($pMCMC = .05$).

In short, the extra increase in Region 2 reading time when an Alternating item was to be disambiguated as a verb appeared only when the following (disambiguating) word was skipped, not when that following word was fixated. This pattern did not appear for the Non-alternating or the Unambiguous words. We will argue in the Discussion that this reflects an early cost of changing the implicit metrical pattern of the Alternating item.

Region 3—This region (the first word or two after the critical word) contained the syntactic disambiguation to a Noun vs. a Verb use of the critical word. The region contained at least 8 characters (including spaces) and generally contained an auxiliary or main verb when the critical word was disambiguated as a noun, or a determiner or adjective when the critical word was disambiguated as a verb. Slowed reading was expected when the region disambiguated the critical word toward a Verb.

The slow-down expected when disambiguation was toward the Verb did not appear in first fixation or first pass times. When tested at the Noun baseline, first fixation durations were longer following a Non-alternating Ambiguous word than following an Unambiguous word, and the interaction with disambiguation toward Verb vs. Noun was nonsignificant, indicating that the penalty for Non-alternating words held true for Verbs as well as Nouns. Neither the difference between Alternating vs. Non-alternating words, nor its interaction with disambiguation toward Verb or Noun, reached significance. First pass times were not significantly affected by either factor.

In contrast, the effect of Noun vs. Verb was significant in the frequency of regressions out of Region 3: there were more regressions out of Region 3 following disambiguation of the critical word toward the Verb form than toward the Noun form. However, no effects of, or interactions involving, type of critical word were significant. Similarly, the expected disruption in the Verb condition was nearly significant in go-past times when tested at the baseline (Non-alternating condition) ($t = 1.97$). There was a tendency toward longer go-past times (at the Noun baseline) for Alternating than for Non-alternating items. The disruption in the Verb condition held consistently across different types of critical items: It did not interact with either Non-alternating vs. Unambiguous or with Non-alternating vs. Alternating items.

In sum, the Region 3 analyses did demonstrate the existence of a greater processing cost for disambiguation toward the Verb than the Noun form of the critical item, but showed only marginal effects of the nature of the critical item. First fixations were slowed for Non-alternating compared to Unambiguous words (and, although not tested explicitly, equally slowed for Alternating compared to Unambiguous words). However, contrary to our initial expectations, there was no suggestion in Region 3 that having to revise implicit metrical coding increased processing cost. We offer some speculations about why this was so in the following Discussion.

Region 4—The only effect that was significant in first fixation durations was an interaction between Noun vs. Verb disambiguation and the Unambiguous-Non-alternating ambiguous contrast. When disambiguation had been toward the Noun form, fixations were longer following an Unambiguous critical item than following a Non-alternating item, while the opposite held when disambiguation had been toward the Verb form. No effects approached significance in the analyses of first pass times, but disambiguation toward the Verb resulted in significantly longer Region 4 go-past times than disambiguation toward the Noun (at the Non-alternating baseline). Analysis of frequency of regressions out of Region 4 indicated two apparent effects. There were marginally more regressions following disambiguation toward Verb than toward Noun at the Non-alternating baseline, and there was a significant interaction of Noun vs. Verb and the Non-alternating-Unambiguous contrast, because the increased frequency of regressions in the Verb condition disappeared in the Unambiguous condition.

In short, the Region 4 data contained some evidence for greater disruption following disambiguation of the critical word toward a verb than toward a noun, limited to the Non-alternating items.

Re-reading effects—The percentages of regressions into Region 1–3, and the second pass times for each region, were analyzed to obtain information about readers' need to revisit the region in the different conditions (Table 7). Frequencies of regressions in were analyzed using a mixed effects logistic model. Because the frequent values of zero for individual trials violated normality assumptions, second pass times were averaged over items and averaged over subjects, and the resulting averages were analyzed in separate mixed model analyses with subjects and with items as random factors (including random slopes when they improved the model fit). The contrasts used for the previously-reported analyses were used in these analyses.

In Region 1, no effects reached significance in the analysis of regressions in. However, second pass times were longer following Verb than Noun disambiguation at the Non-alternating baseline, an effect that held true for the other critical item conditions since no interactions involving critical item approached significance (apart from a significant Noun-Verb x Non-alternating-Unambiguous interaction in the by-items analysis; the Verb cost was possibly larger in the Non-alternating than the Unambiguous condition).

In Region 2, regressions-in were more frequent following Verb than Noun disambiguation. They were also more frequent following disambiguation of a Non-alternating than an Unambiguous item. No interactions approached significance. A similar cost for Verb disambiguation was observed in second pass time. The apparent extra time spent in the Non-alternating condition compared to the Unambiguous condition only approached significance in the by-subjects analysis.

By Region 3, no effects of frequency of regressions-in were significant. However, second pass times were significantly longer at the baseline following disambiguation toward a Verb than toward a Noun. Further, the interaction of this contrast with the contrast of Non-alternating and Unambiguous was also significant. Reading times following Verb disambiguation were particularly long when the critical item had been a Non-alternating item compared to an Unambiguous item, a difference that held true without significantly changing in size for Alternating items.

Discussion

We will discuss the effects of the syntactic and metrical analyses separately. The effectiveness of the syntactic manipulation is not surprising, but demonstrating it is essential

to the experiment. An adjective-noun homograph that was biased toward the adjective reading (e.g. *brilliant*) led readers to interpret it as an adjective and then follow a noun interpretation of the next word. Readers incurred a reading cost on the target word when it was heavily biased as a verb (e.g. *suggest*) compared to when the target was heavily biased as a noun (e.g. *paper*), indicating that they were garden-pathed by the verb in a context which strongly predicted a noun. Second, when the target was an equi-biased noun-verb homograph (e.g. *abstract, report*), readers had longer go-past times on the subsequent region, and more frequent regressions out, when it disambiguated the noun-verb homograph as a verb (e.g. *the...*) compared to when it disambiguated the homograph as a noun (e.g. *was...*).

The novel contribution of the present research is its observation of apparent metrical reanalysis. Readers incurred a reading cost when a stress-alternating noun-verb homograph that was biased to have a noun interpretation (e.g. *abstract*) had to be disambiguated as a verb, as in *The brilliant abstract the...* Importantly, the reading cost in this condition was larger than when a non-stress-alternating noun-verb homograph was disambiguated as a verb, as in *The brilliant report the...*, indicating that readers incurred a reading cost for lexical stress reanalysis over and above that due to syntactic reanalysis.

Although this metrical cost occurred in the predicted condition, its timing was surprising. We initially predicted that metrical reanalysis would be concurrent with syntactic reanalysis, such that we would observe even longer reading times on *the* in *The brilliant abstract the* than on *The brilliant report the*. This result was expected under the assumption that both the metrical and syntactic representations of an Alternating critical word *abstract* must be revised at the same time, whereas only the syntactic representation of a Non-alternating word like *report* has to be revised. However, this is ultimately not what we observed. Rather, evidence of metrical reanalysis appeared earlier than syntactic reanalysis. That is, while syntactic reanalysis influenced reading times on the disambiguating region, metrical reanalysis influenced reading times *before* the disambiguating region. Specifically, when readers subsequently skipped the initial portion of the disambiguating region, they spent more time fixating the Alternating ambiguous homograph, *abstract*, than the Non-alternating ambiguous homograph, *report*.

We take this result as evidence that readers were sometimes able to identify the word in the subsequent disambiguating region while still fixating the target. When the disambiguating word (e.g., *the*) forced the homograph to be changed from a (presumably preferred) noun to a verb, they revised the metrical reanalysis of the target and then skipped the disambiguating word. This metrical reanalysis required a change from the SW pattern appropriate for the noun to the WS pattern appropriate for the verb, the direction of change that was found to be costly in Experiment 1.

The interpretation of our data analysis depends on the assumption that if a word is skipped, it is likely identified parafoveally. There is a good deal of support for this assumption, including the observation that short, frequent words, especially function words, are most often skipped (Rayner, 1998), presumably because they can be identified on the basis of the minimal information available parafoveally. Recent evidence from Drieghe and his colleagues (Drieghe et al., 2005; Drieghe, 2008) has also provided support for this claim, e.g., skipping rate is influenced by the visual contrast of the parafoveal information. In addition, well-supported models of eye movement control give a role to parafoveal identification of words. Our favored model, the E-Z Reader model (Pollatsek, Reichle, & Rayner, 2006) claims that a saccade to word $n+1$ is programmed at the end of an initial stage of processing word n (L1), and that when word n is identified (L2), attention moves to word $n+1$. Because of the substantial latency of planning and executing a saccade, the shift of attention can take place before the eye has moved to word $n+1$. Once attention is directed to

word $n+1$, the job of recognizing it can begin. A short, very frequent word may very well be identified before the eyes have moved to it, and if this happens quickly enough, the saccade may be reprogrammed to skip word $n+1$ and go directly to word $n+2$.

Our data suggest a refinement of this process, related to the “integration check” proposed in E-Z Reader 10 (Reichle, Warren, & McConnell, 2009). Reichle et al. propose that the saccade programmed to word $n+1$ can be cancelled if the reader encounters a problem integrating word n into a coherent sentence representation. We propose that moving the eyes forward from word n in a normal fashion requires not only a grammatically coherent representation of the sentence to that point, but a lexically and metrical coherent representation of word n as well. If the reader has difficulty forming a coherent representation of word n , the forward saccade from that word can be cancelled, resulting in longer fixation duration on word n . Taking our critical alternating ambiguous word as word n , we follow E-Z Reader in proposing that attention can shift to word $n+1$ allowing it to be identified before the eye moves to it. If this happens quickly enough (as is possible with a short function word), and the integration check proposed in E-Z-Reader 10 fails, then the saccade from word n can be cancelled. The syntactic revision triggered by integration failure requires the critical word to be identified as a verb, not a noun, which in our Alternating ambiguous items triggers a revision of the metrical representation of word n . Since the required revision from a SW form to a WS form appears to be difficult (cf. Experiment 1), this revision results in prolonging the fixation on word n .

We acknowledge that this interpretation goes beyond existing data, and is somewhat speculative. Other hypotheses should be developed, and tested. One possibility is that we have the causal sequence of events backwards. It is possible, for instance, that our readers were aware of the lexical and metrical ambiguity of our alternating ambiguous words. This could have happened if they became aware of the ambiguity as the experiment progressed, or if they were sensitive to the ambiguity of the words at the outset. This recognition of ambiguity could increase the time readers fixated on these words on some trials, and, during this lengthened fixation, attention could have shifted to the following word ($n+1$). This shift of attention could have permitted word $n+1$ to be recognized parafoveally, and thus skipped. We are inclined to reject this suggestion. First, it is unlikely that these results were due to subjects developing an awareness of the lexical stress ambiguity during the experiment. Specifically, adding trial number to the mixed models we used to analyze our data did not improve the model fits. Order effects would have been unlikely, however, given that the alternative ambiguous items accounted for only 10 out of 132 (7.5%) of all items in the experimental session. Second, and crucially, this suggestion predicts that the effect of skipping should appear for both noun and verb contexts: the ambiguity is present in both cases, and should have resulted in longer fixations and more frequent skips in both. It did not.

We had originally expected to observe an effect of metrical reanalysis in Region 3, the disambiguating region itself. No such effect was present. In attempting to understand this failure, we note that the requirement to disambiguate the ambiguous word as a verb increased the frequency of regressions out of the disambiguating region and the time to re-read previous regions. It did not significantly increase the time to read the disambiguating region itself. We note that the situations under which syntactic reanalysis increases fixation duration vs. the probability of regressive eye movements are far from well understood (cf. the discussion in Altmann, Garnham, & Dennis, 1992, and Rayner & Sereno, 1994, see also Clifton et al., 2003, for an experiment where effects of syntactic disambiguation appeared in first-pass fixation durations vs. in regressive eye movements in different experimental conditions). It does seem to be the case that the identification of a lexical item that can be successfully integrated into the context largely controls the first-pass durations (Reichle et

al., 1998, 2009). We claim that accessing the lexical entry with the proper stress pattern is part of lexical identification, therefore, the requirement to revise lexical stress in our Alternating Ambiguous items could well control forward movement of the eyes from the critical word itself. However, when syntactic reanalysis takes place only after fixating the disambiguating Region 3, it may trigger a regression into and reanalysis of an earlier region. In this reanalysis, lexical items may have to be re-accessed regardless of their metrical structure, so that metrical ambiguity need not affect the time to complete the reanalysis.

General Discussion

This paper presents findings from two eye-tracking studies designed to investigate the role of metrical prosody in silent reading. Both experiments demonstrated that readers are slower to read words when their stress pattern does not conform to expectations. Experiment 1 demonstrated a reading cost when readers encountered a mismatch between the predicted and actual stress pattern of a word in the context of a limerick. Experiment 2 demonstrated a similar cost of a mismatch in stress patterns in a context where the metrical constraint was mediated by lexical category rather than by explicit meter.

Note the similarity in the direction of stress shift across the experiments. In both cases, readers had difficulty going from a SW to a WS. In Experiment 1, we observed a cost only when the metrical context created an expectation of a SW word like *PREsent*, but the lexical item was *preSENT*. In other words, shifting stress to the right was costly. Similarly, in Experiment 2, readers encountered difficulty when they had to reanalyze a SW noun as a WS verb. The metrical reanalysis always necessitated a shift in stress to the right, as whenever a word in English has variable stress depending on whether it is a noun or verb, the noun is always SW, and the verb is always WS. Future work will be required to determine if leftward shifts are ever costly.

There are, to be sure, dissimilarities between the results of the two experiments. Apart from the decreased frequency of skipping the last word of the second line in Experiment 1 when it required a shift from of a metrically-expected SW pattern to a lexically-required WS implicit pronunciation compared to a shift from SW to WS, the effects observed in Experiment 1 involved increased rereading of earlier material and increased re-reading of the critical word itself. As stressed earlier, this effect occurred only in the inconsistent prosody, WS word condition. In Experiment 2, the time to read the WS word itself was increased, without any clear sign of increased re-reading of earlier material. We believe that there are at least two reasons for the difference in how disruption appeared. The first was mentioned earlier: In Experiment 1, the critical word ended a clause and a line, which is a common place from which to launch a regression. In Experiment 2, the critical word was in the middle of a line and did not occur at a clause boundary. The second reason is that the evidence that led to conflicting metrical and syntactic requirements in Experiment 1 came before the critical word. The material earlier on the line induced a rhythmic pattern, leading to the expectation of a line-final word with a particular stress pattern. The syntactic category of this critical word was unambiguously clear from context, but fixating on the word itself (since it was a content word, generally not identified parafoveally) provided the information that its stress was inconsistent with the anticipated SW pattern in the condition that proved to be disruptive. In Experiment 2, the context preceding the critical word gave no clear reason to anticipate a particular metrical structure. Rather, initially identifying the critical word as a noun was the event that determined its metrical structure (SW in the alternating ambiguous condition). Only when the following short function word was read (frequently parafoveally, we have argued) was the inconsistency between the initial and the required metrical structure apparent. The difference between expectation and final resolution involved a clash between the critical word and material preceding it in Experiment 1, but between the critical

word and the following word in Experiment 2. We suggest that this difference in the locus of the inconsistency would encourage regressive eye movements in Experiment 1 but not in Experiment 2.

Our data suggest that forming a metrical representation of a word plays an important role in eye movement control that may require extensions to existing models of eye movement control in reading (e.g., Reichle et al., 1998, 2009). It may be that, before the eyes move on from a word, the reader must form a metrical representation of the word that is consistent with its lexical identity (e.g., a WS representation for a WS verb). This process may be slowed when an initially-anticipated SW pattern must be changed to a WS representation, but perhaps not when the change has to go in the opposite direction (because of the frequency of such changes in normal production). Following suggestions made by Ashby and Clifton (2005), we do not claim that the WS item must actually be subvocalized. Rather, we propose that what controls eye movement is the creation of an implicit program for subvocalizing the item. We view this proposal as similar (at least in spirit) to the addition of a syntactic and semantic integration stage with an ‘integration check’ as proposed by Reichle et al. (2009). In both cases, the claim is that a grammatically coherent representation must be formed, or at least, effective preparations for forming such a representation must be made, before the eyes move on.

Conclusions

At least since Huey (1908/1968), experimental psychologists have been curious about the nature and function of the ‘inner voice’ people experience when they read. They have debated about whether it plays a causal role in language comprehension and memory, or whether it is purely an epiphenomenal experience. They have attempted to discern details of its nature, e.g., does it contain phonetic detail (e.g., Abramson & Goldinger, 1997) or only more abstract phonemic information (Oppenheim & Dell, 2008), and does it include suprasegmental information such as prosodic phrasing and accent (Fodor, 1998).

The work reported here has led us to conclude that the inner voice does contain some suprasegmental information, information about the metrical structure of words, and that this information does play a causal role in reading comprehension and the control of eye movements. Data from our two eye-tracking experiments provide evidence for these claims during silent reading of two very different types of material. It appears that a reader generally prepares or creates a lexical representation that is veridical in that it honors the metrical stress pattern of the identified word, and that if prior context misleads the reader into creating the wrong stress pattern, reading can be disrupted while the reader corrects the error. The contexts used in the experiments varied substantially, being either a metrical context (Experiment 1) or a syntactic context (Experiment 2), but disruption was observed in both. The exact nature of the disruption did differ between the two contexts, leaving interesting questions about the nature of eye movement control to be explored, but we submit that the data strongly supports the conclusion that metrical structure is created during silent reading and plays a causal role in the reading process.

References

- Abramson M, Goldinger SD. What the reader’s eye tells the mind’s ear: Silent reading activates inner speech. *Perception & Psychophysics*. 1997; 59:1059–1068. [PubMed: 9360478]
- Altmann GTM, Garnham A, Dennis Y. Avoiding the garden path: Eye movements in context. *Journal of Memory and Language*. 1992; 31:685–712.

- Ashby J, Martin AE. Prosodic phonological representations early in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*. 2008; 34:224–236. [PubMed: 18248150]
- Ashby J, Clifton C Jr. The prosodic property of lexical stress affects eye movements during silent reading. *Cognition*. 2005; 96:B89–100. [PubMed: 15913592]
- Ashby J, Treiman R, Kessler B, Rayner K. Vowel processing during silent reading: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2006; 32:416–424.
- Augurzky, P. Prosodic balance constrains argument structure interpretation in German. Poster presented at the 14th Conference on Architectures and Mechanisms for Language Processing; Cambridge, England. September, 2008; 2008.
- Baayen, RH. Analyzing linguistic data: a practical introduction to statistics using R. Cambridge: Cambridge University Press; 2008.
- Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*. 2008; 59:390–412.
- Baayen, RH.; Piepenbrock, R.; van Rijn, H. The CELEX lexical database (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania; 1993.
- Bader, M. Prosodic influences on reading syntactically ambiguous sentences. In: Fodor, J.; Ferreira, J., editors. *Reanalysis in sentence processing*. Dordrecht: Kluwer; 1998. p. 1–46.
- Bolinger, DL. A theory of pitch accent in English. In: Abe, I.; Kanekiyo, T., editors. *Forms of English: Accent, Morpheme, Order*. Cambridge, MA: Harvard University Press; 1965. p. 17–55.
- Clifton C Jr, Traxler M, Mohamed MT, Williams RS, Morris RK, Rayner K. The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*. 2003; 49:317–334.
- Cutler A, Dahan D, Van Donselaar W. Prosody in the comprehension of spoken language: A literature review. *Language and Speech*. 1997; 40:141–201. [PubMed: 9509577]
- Cutler, A.; Clifton, CE. The use of prosodic information in word recognition. In: Bouma, H.; Bouwhuis, DG., editors. *Attention and performance X: Control of language processes*. Hillsdale, NJ: Erlbaum; 1984. p. 183–196.
- Cutler A, Norris D. The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*. 1988; 14:113–121.
- Dell, GS.; Repka, RJ. Errors in inner speech. In: Baars, BJ., editor. *Experimental slips and human error: Exploring the architecture of volition*. New York: Plenum; 1992. p. 237–262.
- Drieghe D. Foveal processing and word skipping during reading. *Psychonomic Bulletin & Review*. 2008; 15:856–860. [PubMed: 18792516]
- Drieghe D, Rayner K, Pollatsek A. Eye movements and word skipping during reading revisited. *Journal of Experimental Psychology: Human Perception and Performance*. 2005; 31:954–969. [PubMed: 16262491]
- Fodor JD. Learning to parse? *Journal of Psycholinguistic Research*. 1998; 27:285–319.
- Francis, WN.; Ku era, H. *Frequency analysis of English usage*. Boston: Houghton–Mifflin; 1982.
- Frazier, L. *On comprehending sentences: Syntactic parsing strategies*. Bloomington, IN: Indiana University Linguistics Club; 1979.
- Fry DB. Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*. 1955; 27:765–768.
- Grabe, E.; Warren, P. Stress shift: do speakers do it or do listeners hear it?. In: Connell, B.; Arvanti, A., editors. *Phonology and phonetic evidence: Papers in Laboratory Phonology IV*. Cambridge: Cambridge University Press; 1995.
- Henderson JM, Dixon P, Petersen A, Twilley L, Ferreira F. Evidence for the use of phonological representations during trans saccadic word recognition. *Journal of Experimental Psychology: Human Perception and Performance*. 1995; 21:82–97.
- Hirose Y. Recycling prosodic boundaries. *Journal of Psycholinguistic Research*. 2003; 32:167–195. [PubMed: 12690830]

- Huey, EB. The psychology and pedagogy of reading. Cambridge, MA: M.I.T. Press; 1968. (Original work published 1908)
- Hwang H, Schafer AJ. Constituent Length Affects Prosody and Processing for a Dative NP Ambiguity in Korean. *Journal of Psycholinguistic Research*. 2009; 38:151–175. [PubMed: 19067169]
- Jaeger TF. Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language*. 2008; 59:434–446. [PubMed: 19884961]
- Kennison SM, Sieck JP, Briesch KA. Evidence for a late-occurring effect of phoneme repetition in silent reading. *Journal of Psycholinguistic Research*. 2003; 32:297–312. [PubMed: 12845941]
- Kennison SM. The effect of phonemic repetition on syntactic ambiguity resolution: Implications for models of working memory. *Journal of Psycholinguistic Research*. 2004; 33:493–516. [PubMed: 15614991]
- Kjelgaard M, Speer S. Prosodic facilitation and interference in the resolution of temporary syntactic ambiguity. *Journal of Memory and Language*. 1999; 40:153–194.
- Lieberman P. Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America*. 1960; 32(4):451–454.
- Lieberman M, Prince A. On stress and linguistic rhythm. *Linguistic Inquiry*. 1977; 2:249–336.
- Mattys SL, Samuel AG. How lexical stress affects speech segmentation and interactivity: Evidence from the migration paradigm. *Journal of Memory and Language*. 1997; 36:87–116.
- McCutchen D, Perfetti CA. The visual tongue-twister effect: Phonological activation in silent reading. *Journal of Verbal Learning and Verbal Behavior*. 1982; 21:672–687.
- Oppenheim GM, Dell GS. Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*. 2008; 106:528–537. [PubMed: 17407776]
- Pitt MA, Samuel AG. The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human Perception and Performance*. 1990; 16:564–573. [PubMed: 2144571]
- Pollatsek A, Reichle ED, Rayner K. Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology*. 2006; 52:1–56. [PubMed: 16289074]
- Pollatsek A, Lesch M, Morris RK, Rayner K. Phonological codes are used in integrating information across saccades in word identification and reading. *Journal of Experimental Psychology: Human Perception and Performance*. 1992; 18:148–162. [PubMed: 1532185]
- Price PJ, Ostendorf M, Shattuck-Hufnagel S, Fong C. The use of prosody in syntactic disambiguation. *Journal of Acoustical Society of America*. 1991; 90:2956–2970.
- R Development Core Team. R Foundation for Statistical Computing; Vienna, Austria: 2007. R: A language and environment for statistical computing. URL <http://www.R-project.org>
- Rayner K. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*. 1998; 124:372–422. [PubMed: 9849112]
- Rayner, K.; Duffy, SA. On-line comprehension processes and eye movements in reading. In: Daneman, M.; MacKinnon, GE.; Waller, TG., editors. *Reading research: Advances in theory and practice*. New York: Academic press; 1988. p. 13–66.
- Rayner, K.; Pollatsek, A. The psychology of reading. Englewood Cliffs, NJ: Prentice-Hall; 1989.
- Rayner K, Sereno SC. Regressive eye movements and sentence parsing: On the use of regression-contingent analyses. *Memory & Cognition*. 1994; 22:281–285.
- Rayner K, Sereno S, Morris R, Schmauder R, Clifton CJ. Eye movements and on-line language comprehension processes. *Language and Cognitive Processes*. 1989; 4:SI 21–50.
- Reichle ED, Pollatsek A, Fisher DF, Rayner K. Toward a model of eye movement control in reading. *Psychological Review*. 1998; 105(1):125–156. [PubMed: 9450374]
- Reichle E, Warren T, McConnell K. Using E-Z Reader to model the effects of higher-level language processing on eye movements during reading. *Psychonomic Bulletin & Review*. 2009; 16:1–21. [PubMed: 19145006]
- Rohde, D. Linger. 2008. Retrieved 9/15/2008 from <http://tedlab.mit.edu/~dr/Linger/>
- Schafer AJ, Speer SR, Warren P, White SD. Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*. 2000; 29:169–182. [PubMed: 10709182]
- Shattuck-Hufnagel S, Ostendorf M, Ross K. Stress shift and early pitch accent placement in lexical items in American English. *Journal of Phonetics*. 1994; 22:357–388.

- Shields JL, McHugh A, Martin JG. Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology*. 1974; 102:250–255.
- Slowiaczek ML, Clifton C. Subvocalization and reading for meaning. *Journal of Verbal Learning & Verbal Behavior*. 1980; 19:573–582.
- Sternberg, S.; Monsell, S.; Knoll, RL.; Wright, CE. The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In: Stelmach, GE., editor. *Information processing in motor control and learning*. SanDiego, CA: Academic Press; 1978. p. 117-152.
- Swets B, Desmet T, Hambrick D, Ferreira F. The role of working memory in syntactic ambiguity resolution: A working memory approach. *Journal of Experimental Psychology: General*. 2007; 136:64–81. [PubMed: 17324085]
- Taft, L. Unpublished doctoral dissertation. University of Massachusetts; Massachusetts: 1984. Prosodic constraints and lexical parsing strategies.
- Treiman R, Freyd JJ, Baron J. Phonological mediation and use of spelling–sound rules in reading of sentences. *Journal of Verbal Learning & Verbal Behavior*. 1983; 22:682–700.
- Van Orden GC. A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition*. 1987; 15:181–198.
- Warren, S.; Morris, R. Phonological similarity effects in reading. Talk presented at the 15th European Conference on Eye Movements; Southampton, England. August, 2009; 2009.
- Zwicky, AM.; Zwicky, A. Patterns first, exceptions later. In: Channon, R.; Shockey, L., editors. *In honor of Ilse Lehiste*. Dordrecht: Foris; 1987. p. 525-37.

Appendix A: Noun-verb (or noun-adjective) homographs that served as the critical word in Experiment 1

Non-reduced

complex, decrease, defect, digest, discount, extract, import, insult, object, permit, pervert, proceed, protest, rebound, recall, refund, survey, suspect, transplant, transport

Reduced

address, ally, combat, combine, commune, compact, compound, conflict, contest, convict, desert, detail, entrance, present, produce, progress, project, rebel, record, relay

Appendix B: Materials used in Experiment 2. Full paradigm illustrated for Item 1; alternative words are indicated for the remaining items

Alternating Ambiguous: Noun The brilliant abstract was accepted at the prestigious conference.

Non-alternating Ambiguous Noun The brilliant report was accepted at the prestigious conference.

Unambiguous Noun The brilliant paper was accepted at the prestigious conference.

Alternating Ambiguous Verb The brilliant abstract the best ideas from the things they read.

Non-alternating Ambiguous Verb The brilliant report the best ideas from the things they read.

Unambiguous Verb The brilliant suggest the best ideas from the things they read.

The dangerous ally/picture/agent was kept a secret from their enemies.

The dangerous ally/picture/enjoy themselves with other shady characters.

The violent combat/defeat/battle resulted in many casualties.
 The violent combat/defeat/resist their enemies with strength and malice.
 The religious commune/worship/vigil was kept secret from all outsiders.
 The religious commune/worship/appear together in holy places.
 The strong contrast/partner/actor was hard to ignore.
 The strong contrast/partner/compete with their weaker friends.
 The handy object/answer/textbook was helpful to the student.
 The handy object/answer/relate to their less able superiors.
 The awkward subject/challenge/topic was discussed by the family.
 The awkward subject/challenge/expose themselves to difficult tasks.
 The phony address/return/memo was discovered by the authorities.
 The phony address/return/compose messages with fake sincerity.
 The epic details/reports/poems of Greek heroes are well known.
 The epic details/reports/depicts the story of Odysseus' journey.
 The local contests/questions/ballgames occupied the entire street.
 The local contests/questions/defies the new zoning laws.
 The subtle conduct/disguise/workings of the secret agent alerted no one to his presence.
 The subtle conduct/disguise/behave themselves in a manner that does not look suspicious.
 The tacky present/outfit/jacket was covered in sequins and feathers.
 The tacky present/outfit/reveal themselves tastelessly.
 The intelligent convict/sentence/statement exhibited surprising coherence.
 The intelligent convict/sentence/condemn criminals after a fair trial.
 The demanding desert/distance/jungle claims the lives of many travellers.
 The demanding desert/distance/ignore the ones who love them.
 The criminal exploits/murders/antics of the mobsters were well-known.
 The criminal exploits/murders/molests anyone he needs to to get ahead.
 The mysterious invite/notice/letter was the topic of much gossip.
 The mysterious invite/notice/dislike the scrutiny of others.
 The speedy relay/process/lecture was finished very quickly.
 The speedy relay/process/obtain important information quickly.
 The foreign permit/practice/custom was not recognized outside the country.
 The foreign permit/practice/allow behaviors that are condemned in America.
 The vulgar insult/surprise/action was a shock to the audience.
 The vulgar insult/surprise/offend people with their inappropriate comments.
 The young produce/blossom/eggplant was expected to mature quickly.

The young produce/blossom/mature and explore new ideas.
The exotic extract/bottle/flavor was sold at the expensive shop.
The exotic extract/bottle/acquire the flavors of herbs and flowers for cooking.
The righteous rebel/struggle/effort was supported by a foreign government.
The righteous rebel/struggle/persist against oppressive regimes.
The famous recall/study/movie was remembered by many parents.
The famous recall/study/ignore their coverage in the press.
The evil suspect/promise/chairman was distressing to the public.
The evil suspect/promise/assume that they will not be caught.
The unfortunate reject/recruit/hippie was run out of town.
The unfortunate reject/recruit/attract offers of help.
The sophisticated survey/design/method was used by many researchers.
The sophisticated survey/design/inspect their surroundings with a critical eye.
The spiritual convert/rally/teacher was known to many of the locals.
The spiritual convert/rally/unite people of many different backgrounds.
The weekly inserts/offers/salesmen were particularly annoying this month.
The weekly inserts/offers/includes useless coupons in every issue.
The sturdy compress/bandage/towel was used to stop the bleeding.
The sturdy compress/bandage/repair the wounds of their fallen comrades.
The suspicious record/shadow.story was discovered to be fake.
The suspicious record/shadow/annoy their friends and colleagues.

Figure 1. Go-past times on Region 3 (the critical word) in Experiment 1. Error bars represent the standard error of the mean.

Figure 2.
Proportion of WS productions of the critical word, Norm 2 for Experiment 2

Figure 3.

First pass reading times on Region 2 (the critical word) in Experiment 2. Error bars represent the standard error of the mean. The word in parentheses is the following word, which served to disambiguate the critical word as a noun or verb.

Figure 4.

First pass reading times on region 2 (the critical word) in Experiment 2 dependent upon whether the reader subsequently fixated (left) or skipped (right) the region 3' (see text for details). Error bars represent the standard error of the mean.

Table 1

Metrical structure of the first two lines of limericks from each of the four conditions in Experiment 1. Light gray shading indicates metrically strong syllables. Dark gray shading indicates syllables which are inconsistent with the metrical structure.

	Condition	(W)	(S)	(W)	(W)	(S)	(W)	(W)	(S)	(W)	(W)	(S)	(W)
A.	Strong-Weak/Consistent	There	once	was	a	penn	i	Less	pea	sant			
		Who	could	n't	af	ford	a	nice	PRE	sent			
B.	Strong-Weak/Inconsistent	There	once	was	a	cle	ver	young	gent				
		Who	gave	to	his	girl	a	*PRE	sent				
C.	Weak-Strong/Consistent	There	once	was	a	cle	ver	young	gent				
		Who	had	a	nice	talk	to	pre	SENT				
D.	Weak-Strong/Inconsistent	There	once	was	a	penn	i	less	pea	sant			
		Who	went	to	his	mas	ter	To	*pre	SENT			

Table 2

Mean eyetracking measures, Experiment 1 (with standard errors in parentheses)

Metrical Condition	Region 2		Region 3	
	Lexical Target		Lexical Target	
	Consistent	Inconsistent	Consistent	Inconsistent
Proportion Fixated				
Strong-Weak	.98 (.01)	.98 (.01)	.92 (.02)	.89 (.02)
Weak-Strong	.97 (.01)	.96 (.02)	.90 (.02)	.94 (.01)
Regressions Out (proportion)				
Strong-Weak	.03 (.01)	.01 (.01)	.28 (.03)	.28 (.03)
Weak-Strong	.02 (.01)	.03 (.01)	.24 (.03)	.33 (.03)
First Pass (ms)				
Strong-Weak	867 (21.0)	815 (21.3)	286 (9.6)	266 (9.4)
Weak-Strong	887 (21.8)	956 (23.6)	287 (10.8)	288 (11.1)
Go-Past (ms)				
Strong-Weak	906 (23.0)	857 (25.5)	445 (21.3)	434 (22.8)
Weak-Strong	909 (23.6)	1008 (24.7)	451 (24.4)	535 (31.8)

Table 3

Parameter estimates of models, Analyses of Region 3, Experiment 1

Fixed Effect	Estimate	SE	z or t value	p
Probability of Fixation (logistic analysis)				
Stress (WS vs SW)	−0.37	0.36	1.05	0.29
Consistency with Context	−0.41	0.36	1.15	0.25
Interaction	1.30	0.53	2.46	0.014
Regressions Out (logistic analysis)				
Stress (WS vs SW)	−0.25	0.22	1.16	0.25
Consistency with Context	0.01	0.21	0.03	0.97
Interaction	0.51	0.30	1.73	0.08
First Pass Time				
Stress (WS vs SW)	−0.93	13.19	0.07	0.94
Consistency with Context	−20.17	13.19	1.53	0.13
Interaction	26.23	18.56	1.41	0.16
Go-Past Time				
Stress (WS vs SW)	−6.85	38.66	0.18	n.s.
Consistency with Context	−14.65	34.16	0.43	n.s.
Interaction	102.16	42.10	2.43	0.05
Go-Past Time Residuals				
Stress (WS vs SW)	−8.14	29.12	0.28	n.s.
Consistency with Context	−21.68	27.34	0.79	n.s.
Interaction	95.12	36.78	2.59	0.05

Note: Probability of Fixation and Regressions Out analyses were computed with family binomial, so z values are reported. First pass time analysis was computed with random intercepts, so MCMC estimates of p values are presented. Go-Past time analysis used random slopes, so MCMC was not possible, and at of 2.0 is taken to be significant at the .05 level.

Table 4

Sample set of sentences as used in Experiment 2.

Condition	Disambiguation	Critical Word	Example
A	Noun	Alternating	The brilliant ^{^1} abstract ^{^2} was accepted ^{^3} at the prestigious ^{^4} conference.
B	Noun	Non-alternating	The brilliant ^{^1} report ^{^2} was accepted ^{^3} at the prestigious ^{^4} conference.
C	Noun	Unambiguous	The brilliant ^{^1} paper ^{^2} was accepted ^{^3} at the prestigious ^{^4} conference.
D	Verb	Alternating	The brilliant ^{^1} abstract ^{^2} the best ^{^3} ideas from the ^{^4} things they read.
E	Verb	Non-alternating	The brilliant ^{^1} report ^{^2} the best ^{^3} ideas from the ^{^4} things they read.
F	Verb	Unambiguous	The brilliant ^{^1} suggest ^{^2} the best ^{^3} ideas from the ^{^4} things they read.

Table 5

Means and standard deviations of lengths (in number of letters) and frequencies (in occurrences per million words) of critical words by condition. Standard deviations are in parentheses.

Word type	Length	K-F Frequency
Alternating homophone as Noun	7.6 (0.97)	40 (57)
Non-alternating homophone as Noun	7.9 (1.04)	97 (98)
Unambiguous Noun	7.2 (1.33)	87 (105)
Alternating homophone as Verb	7.6 (0.97)	44 (51)
Non-alternating homophone as Verb	7.9 (1.04)	60 (71)
Unambiguous Verb	7.3 (0.75)	94 (115)

Table 6

Percentages of completions scored as indicating each possible interpretation of the third word in the to-be-completed string, Norm 1 for Experiment 2

Completion Classification	Word Type			
	Alternating Ambiguous	Non-Alternating Ambiguous	Unambiguous Noun	Unambiguous Verb
Noun	79	84	97	13
Verb	17	13	0	74
Adjective	1	0	0	2
Unclassified	4	3	2	11

Table 7
Mean Values (and SEs) of Eyetracking Measures, by Region and Noun vs. Verb Disambiguation, Experiment 2

Condition	Region					
	Region 2		Region 3		Region 4	
	Noun	Verb	Noun	Verb	Noun	Verb
First Fixation						
Nonalt Amb	252 (5.8)	251 (5.3)	250 (5.9)	261 (6.1)	236 (6.0)	238 (6.9)
Unambig	242 (5.1)	267 (6.4)	231 (4.4)	251 (5.9)	249 (6.4)	226 (5.6)
Alt Amb	260 (6.0)	284 (6.9)	257 (6.1)	275 (7.3)	242 (6.5)	242 (6.5)
First Pass						
Nonalt Amb	309 (8.7)	297 (8.0)	422 (12.0)	436 (14.4)	398 (18.1)	439 (22.8)
Unambig	297 (8.1)	332 (10.2)	416 (11.8)	397 (14.8)	414 (19.5)	478 (19.8)
Alt Amb	315 (8.7)	343 (9.0)	444 (13.5)	439 (14.0)	401 (19.6)	435 (18.0)
Go-Past						
Nonalt Amb	411 (13.8)	428 (15.5)	535 (20.8)	645 (28.1)	489 (24.5)	663 (38.1)
Unambig	382 (13.6)	477 (17.2)	503 (16.9)	639 (27.9)	475 (24.1)	558 (23.0)
Alt Amb	402 (14.3)	440 (14.3)	599 (24.2)	634 (26.1)	524 (28.3)	683 (42.8)
Regressions Out						
Nonalt Amb	.19 (.023)	.23 (.025)	.15 (.021)	.25 (.026)	.14 (.024)	.22 (.031)
Unambig	.17 (.022)	.26 (.026)	.11 (.018)	.26 (.025)	.14 (.024)	.13 (.025)
Alt Amb	.16 (.022)	.17 (.022)	.18 (.022)	.24 (.025)	.19 (.028)	.21 (.030)
Regressions in						
Region 1						
Nonalt Amb	.68 (.04)	.75 (.03)	.26 (.026)	.39 (.029)	.24 (.025)	.28 (.026)
Unambig	.68 (.04)	.74 (.03)	.20 (.024)	.35 (.028)	.19 (.022)	.21 (.024)
Alt Amb	.65 (.04)	.76 (.03)	.31 (.027)	.42 (.029)	.20 (.024)	.30 (.027)
Second Pass						
Nonalt Amb	142 (19.1)	255 (23.9)	104 (17.8)	207 (21.1)	94 (19.4)	173 (19.2)
Unambig	132 (20.5)	208 (18.4)	72 (12.1)	155 (20.4)	81 (14.9)	87 (12.7)
Alt Amb	133 (12.7)	239 (27.1)	122 (20.1)	216 (27.7)	87 (13.0)	183 (26.6)

Table 8

Parameter estimates of models, Experiment 2

Fixed Effect	Est	SE	t	pMCMC	Est	SE	t	pMCMC	Est	SE	t	pMCMC
First Fixation (ms)												
Noun vs. Verb (NV)	2.30	8.51	0.27	0.79	12.0	9.09	1.29	0.20	4.45	9.39	0.47	0.64
NonAlt vs. Unambig	-8.64	7.48	1.15	0.25	-18.14	7.76	2.34	0.02	12.43	8.04	1.55	0.12
NonAlt vs. Alternat	10.5	7.51	1.40	0.16	8.66	7.85	1.10	0.27	6.54	8.22	0.92	0.36
NV * NonAlt vs. Unamb	21.97	10.65	2.06	0.04	6.90	11.02	0.63	0.53	-26.25	11.89	2.21	0.03
NV * NonAlt vs. Alt	20.24	10.70	1.89	0.06	4.23	11.15	0.38	0.70	-9.13	11.99	0.76	0.45
First Pass (ms)												
Noun vs. Verb (NV)	-10.29	13.50	0.76	0.45	14.06	27.72	0.51	0.61	34.11	54.51	0.63	0.53
NonAlt vs. Unambig	-11.60	11.36	1.02	0.31	-3.88	16.29	0.24	0.81	6.38	19.33	0.33	0.74
NonAlt vs. Alternat	6.33	11.40	0.55	0.58	20.76	16.50	1.26	0.21	-7.27	19.83	0.37	0.72
NV * NonAlt vs. Unamb	45.17	16.17	2.79	0.01	-37.32	23.17	1.61	0.11	38.67	28.68	1.34	0.18
NV * NonAlt vs. Alt	38.98	16.24	2.40	0.02	-20.28	23.45	0.87	0.39	8.39	28.96	0.29	0.77
Go Past (ms)												
Noun vs. Verb (NV)	18.22	24.05	0.76	0.45	107.05	54.43	1.97	^	184.67	68.56	2.69	*
NonAlt vs. Unambig	-28.13	19.23	1.46	0.15	-33.76	37.99	0.90	n.s.	-16.33	36.40	0.45	n.s.
NonAlt vs. Alternat	-10.06	19.31	0.52	0.60	56.56	34.64	1.63	^	38.75	36.98	1.05	n.s.
NV * NonAlt vs. Unamb	79.63	27.38	2.90	0.01	34.56	51.79	0.67	n.s.	-77.38	51.98	1.49	n.s.
NV * NonAlt vs. Alt	19.35	27.50	0.70	0.48	-64.05	48.52	1.32	n.s.	-21.21	52.93	0.40	n.s.
Regressions Out (prop)												
Noun vs. Verb (NV)	0.26	0.25	1.03	0.31	0.72	0.25	2.86	0.01	0.60	0.34	1.76	0.08
NonAlt vs. Unambig	-0.17	0.23	0.76	0.45	-0.41	0.26	1.60	0.11	0.11	0.30	0.36	0.72
NonAlt vs. Alternat	-0.23	0.23	1.02	0.31	0.21	0.23	0.91	0.36	0.23	0.35	0.66	0.51
NV * NonAlt vs. Unamb	0.37	0.31	1.21	0.23	0.49	0.32	1.51	0.13	-0.90	0.43	2.09	0.04
NV * NonAlt vs. Alternat	-0.25	0.32	0.77	0.44	-0.29	0.31	0.95	0.34	-0.25	0.47	0.53	0.60
Regressions In (prop)												
Noun vs. Verb (NV)	0.45	0.27	1.66	0.10	0.65	0.22	2.96	0.01	0.24	0.23	1.05	0.29
NonAlt vs. Unambig	-0.07	0.24	0.28	0.78	-0.43	0.21	2.05	0.04	-0.33	0.21	1.60	0.11

Fixed Effect	Est	SE	t	pMCMC	Est	SE	t	pMCMC	Est	SE	t	pMCMC
NonAlt vs. Alternat	-0.11	0.24	0.43	0.67	0.22	0.20	1.14	0.25	-0.19	0.21	0.91	0.36
NV * NonAlt vs. Unamb	0.02	0.34	0.07	0.95	0.25	0.28	0.91	0.36	-0.06	0.29	0.21	0.83
NV * NonAlt vs. Alt	0.05	0.34	0.14	0.89	-0.08	0.27	0.29	0.77	0.27	0.28	0.95	0.34
Second Pass (Subjects) (ms)												
Noun vs. Verb (NV)	114.69	20.61	5.56	0.01	102.58	17.53	5.85	*	79.02	19.13	4.13	0.01
NonAlt vs. Unambig	-8.81	20.81	0.42	0.67	-32.19	16.36	1.97	*	-12.54	19.13	0.66	0.51
NonAlt vs. Alternat	-8.49	20.61	0.41	0.68	18.25	17.63	1.04	n.s.	-7.37	19.13	0.39	0.70
NV * NonAlt vs. Unamb	-47.33	29.15	1.62	0.11	-19.38	22.70	0.85	n.s.	-73.49	27.05	2.72	0.01
NV * NonAlt vs. Alt	-13.58	29.02	0.47	0.64	-8.87	22.70	0.39	n.s.	16.41	27.05	0.61	0.55
Second Pass (Items) (ms)												
Noun vs. Verb (NV)	122.33	21.12	5.79	0.01	101.29	21.32	4.75	*	83.01	23.10	3.59	0.01
NonAlt vs. Unambig	4.17	18.09	0.23	0.82	-27.44	21.37	1.28	n.s.	-7.86	23.10	0.34	0.73
NonAlt vs. Alternat	-2.08	18.09	0.12	0.91	18.80	20.11	0.94	n.s.	-3.50	23.29	0.15	0.88
NV * NonAlt vs. Unamb	-59.73	25.46	2.35	0.02	-19.58	30.35	0.65	n.s.	-77.18	32.80	2.35	0.02
NV * NonAlt vs. Alt	-26.44	25.46	1.04	0.30	-15.61	28.45	0.55	n.s.	2.97	32.94	0.09	0.93

* : or ^ or n.s.: random slopes were used in model; pMCMC cannot be calculated; * indicates estimated significance beyond the .05 level; ^ indicated marginal significance

Table 9

First Fixation and First Pass Durations, ms, Region 2, by Noun vs. Verb Disambiguation, Conditional on Fixating/Skipping Region 3'

	Fixate 3'		Skip 3'	
	Noun	Verb	Noun	Verb
	First Fixation			
Non-alternating	254 (7.4)	248 (6.3)	249 (9.2)	259 (9.7)
Unambiguous	241 (6.9)	271 (8.0)	243 (7.6)	255 (9.3)
Alternating	271 (8.2)	280 (8.1)	245 (8.4)	290 (12.8)
	First Pass			
Non-alternating	305 (10.5)	300 (10.1)	315 (15.1)	291 (12.1)
Unambiguous	295 (10.2)	336 (13.0)	301 (13.4)	322 (14.0)
Alternating	329 (11.8)	330 (10.8)	295 (12.6)	366 (15.9)