

Lexical stress assignment as a problem of probabilistic inference

Olessia Jouravlev · Stephen J. Lupker

Published online: 31 January 2015
© Psychonomic Society, Inc. 2015

Abstract A new conceptualization of the process of stress assignment, couched in the principles of (Bayesian) probabilistic inference, is introduced in this paper. According to this approach, in deciding where to place stress in a polysyllabic word, a reader estimates the posterior probabilities of alternative stress patterns. This estimation is accomplished by adjusting a prior belief about the likelihoods of alternative stress patterns (derived from experience with the distribution of stress patterns in the language) by using lexical and non-lexical sources of evidence for stress derived from the orthographic input. The proposed theoretical framework was used to compute probabilities of stress patterns for Russian disyllabic words and nonwords which were then compared with the performance of readers. The results showed that the estimated probabilities of stress patterns were reflective of actual stress assignment performance and of naming latencies, suggesting that the mechanisms that are involved in the process of stress assignment might indeed be inferentially-based.

Keywords Lexical stress assignment · Reading · Bayesian probabilities · Russian · Word naming · Nonword naming

Introduction

The processes by which readers determine a word's stress, a process that has been shown to play an important role in overt and silent reading of polysyllabic words (Ashby & Clifton, 2005; Breen & Clifton, 2011), has become an object of

increased scientific interest in the field of reading research. Both of the main computational approaches to reading (i.e., the dual-route view and the connectionist view) have acknowledged that any truly successful model of word reading needs to provide an explanation of the principles of lexical stress assignment. As a result, attempts to model this process have been made within both frameworks (Arciuli, Monaghan, & Ševa, 2010; Perry, Ziegler, & Zorzi, 2010; Rastle & Coltheart, 2000; Ševa, Monaghan, & Arciuli, 2009).

Within the dual-route approach, it has been suggested that, in English, stress can be assigned lexically via retrieval of stress information from memory or as a result of computations by a non-lexical, rule-based system using an algorithm that is mainly based on the knowledge of associations between morphemes and stress patterns in that language, and, to some extent, additional orthographic (e.g., bigram legality) and phonological (e.g., schwa vowel) information (Rastle & Coltheart, 2000). Within the connectionist framework, the models of Ševa et al. (2009) and Arciuli et al. (2010) present three-layer supervised feed-forward networks that map the orthography of English disyllables onto stress patterns. The orthographic input layer is connected to the stress pattern output layer via a layer of hidden units. In the process of training, the connectionist models of stress assignment learn weights on connections among units that reflect the statistical relations that exist between the orthographies of words and their stress patterns. Finally, the CDP++ (Perry et al., 2010) is a model of word reading built on the strengths of the dual-route and the connectionist models. Similar to the dual-route model, within the CDP++, orthography is mapped onto stress representations both lexically and sub-lexically. However, the sub-lexical route is represented by a connectionist two-layer network of phonological assembly (the TLA network) that encodes statistical regularities rather than by a set of rules.

As seen in Fig. 1, these models demonstrate good performance in assigning stress to words and reasonably high

O. Jouravlev (✉) · S. J. Lupker
Department of Psychology, The University of Western Ontario,
London, Ontario N6A 5C2, Canada
e-mail: ozhuravl@uwo.ca

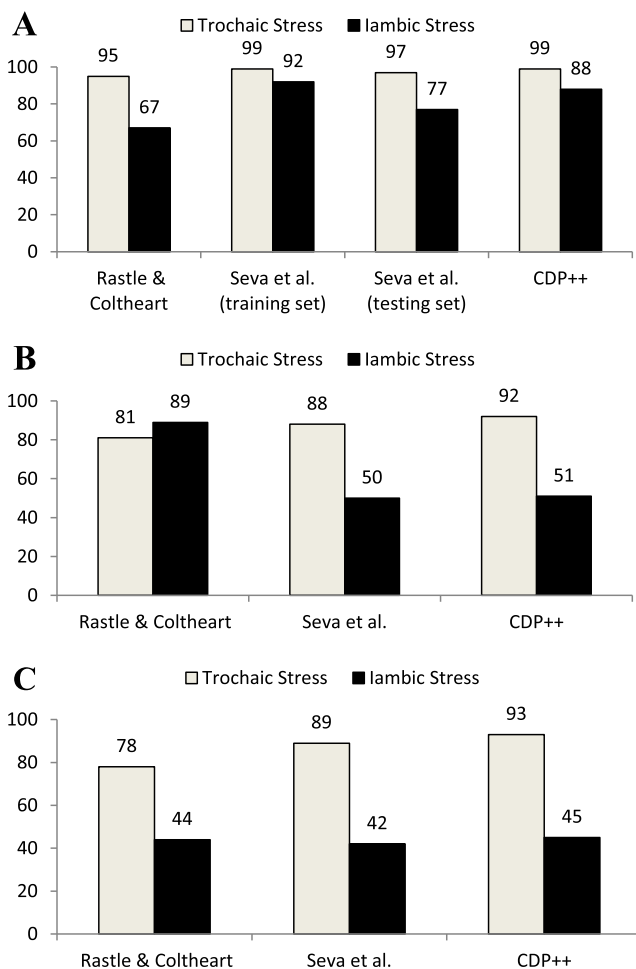


Fig. 1 Correct stress agreement (percentage) for the model of Rastle and Coltheart (2000), the model of Ševa et al. (2009), and the CDP++ on (A) English disyllabic words from the CELEX database (Baayen et al., 1995), (B) Rastle and Coltheart's (2000) English nonwords, and (C) Kelly's (2004) English nonwords

agreement on stress assignment for nonwords that are named with first syllable stress by native speakers. Their performance on the naming of nonwords that are assigned second syllable stress by speakers is considerably less impressive. Relative to other models, the CDP++ performs better, but it is still far from perfect, suggesting that alternative approaches to explaining the mechanisms of stress assignment should also be considered.

In particular, one may conceptualize the process of stress assignment in reading within the framework of probabilistic approaches to human cognition, approaches that are becoming increasingly popular in cognitive psychology and neuroscience (Clark, 2013; Pothos & Busemeyer, 2013). In this paper, we assess whether the process of stress assignment in reading disyllabic words and nonwords in Russian, a language with an extremely opaque lexical stress system, can be conceived of as a problem of probabilistic inference making (hereafter referred to as the *probabilistic inferential approach*). We start the paper with an overview of the general principles of probabilistic

inference as related to human cognition, then we describe a possible implementation of the process of stress assignment in reading based on probabilistic inference making. We finish by examining whether the patterns of stress assignment performance predicted by this probabilistically based process are, in fact, observed in the behavior of Russian readers naming disyllabic words and nonwords.

The probabilistic nature of human cognition

People constantly face the challenge of interpreting uncertain signals coming from a noisy environment and acting in the face of incomplete knowledge. One of the ways of dealing with this uncertainty is to process information using a probabilistic framework. In the presence of uncertainty, a person can make rational decisions by considering estimates of the probabilities of events. Thus, the human mind can potentially be perceived as an evaluator of the likelihoods of events with the aim being to make near optimal decisions (Anderson, 1991). The view of the human mind as a probability estimator, which is associated with Bayesian theory, has been widely adopted to explain various cognitive processes (for a review see Griffiths, Kemp, & Tenenbaum, 2008), although this approach also finds its opponents (Bowers & Davis, 2012; Jones & Love, 2011).

Probability can be viewed as a numerical measure of the relative frequency of an event or the strength of a belief in a certain proposition. In a probabilistic system, one considers the probability of various possible hypotheses about the state of the environment by considering the sensory input received from this environment and prior knowledge about the state of the world. Such probability calculations may very well then be based on some form of Bayesian inference which is based upon a simple formula known as Bayes' rule (Bayes, 1763/1958):

$$P(h|d) = \frac{P(d|h)P(h)}{P(d)} \quad (1)$$

where h refers to a hypothesis, and d stands for some data used as evidence in the process of inference. In computing the probability of the hypothesis given the data, also known as the *posterior probability*, one uses knowledge of the probability of the observed data given the hypothesis, or the *likelihood of evidence*, $P(d|h)$, the probability of the hypothesis before any data were observed, or the *prior probability*, $P(h)$, and the overall probability of observing those data regardless of the hypothesis, $P(d)$.

The complexity of the world that our mind has to grasp makes such metaphors as “the Bayesian brain” or “the probabilistic mind” very popular. In fact, the idea of human cognition being probabilistic in nature has been described as “the

most exciting and revolutionary paradigm to hit cognitive science since connectionism” (Movellan & Nelson, 2001, p.691). According to this idea, people learn probabilities of the occurrence of various events that they observe in the world as they experience the world and use this knowledge to estimate likelihoods of those events in particular situations (Peterson & Beach, 1967; Zacks & Hasher, 2002). In this way, the human mind operates like a statistician, although people are often vulnerable to incorrect assumptions about the relevance of any observed sample to the population, which can give rise to incorrect assessments of probabilities and, therefore, various errors and biases in predictions (for a review, see Hansson, Juslin, & Winman, 2008).

Originally, the Bayesian view of cognition was thought to be able to explain the computational level of processing only. Currently, however, there is growing evidence that probabilistic analysis is relevant to human cognition at the neuronal level as well (for a review, see Doya, Ishii, Pouget, & Rao, 2007), and, therefore, probabilistic models of cognition appear to also have biological plausibility.

Probabilistic approaches have been widely applied to explain many areas of human cognition including visual perception (Feldman, 2001), object recognition (Kersten, Mamassian, & Yuille, 2004), eye movements (Najemnik & Geisler, 2009), and theory of the mind (Baker, Saxe, & Tenenbaum, 2009). Most relevant to the present paper, another aspect of cognition that many researchers have evaluated using a probabilistic inferential approach is language (for a review, see Jurafsky, 2003). Traditionally, language has been viewed as involving a set of abstract units that are generated and used according to some formal, deterministic rules. However, in reality, language is characterized by the presence of significant noise and ambiguity that speakers must successfully deal with, possibly by following the principles of probabilistic constraint satisfaction (McRae, Spivey-Knowlton, & Tanenhaus, 1998; Seidenberg & MacDonald, 1999). In fact, the principles of probabilistic inference have already been adopted by researchers in some areas of language research (Chater & Manning, 2006; Norris, 2006; Norris & Kinoshita, 2008; Norris & McQueen, 2008; Xu & Tenenbaum, 2007).

Stress assignment as a problem of probabilistic inference

In the present paper, the principles of probabilistic inference are extended to the area of lexical stress assignment which can be viewed as the process of posterior probability estimation for alternative hypotheses concerning the position of stress. There would, presumably, be as many hypotheses considered for a word as there are syllables in the word. This theorizing, of course, implies that a reader is aware of the syllabic length of a word before the probability of each hypothesis is computed. Although there is currently no strong evidence showing a

time-period when the discrimination of words according to their syllabic length occurs, it seems likely that it happens at an early stage of processing, and there is reasonable empirical support for this claim (Ashby & Rayner, 2004; Ashby & Martin, 2008). Further, the assumption that readers consider only those stress patterns that are possible in a word has been made in all previous models of polysyllabic word reading and models of stress assignment (Perry et al., 2010; Ševa et al., 2009). As in the CDP++ model (Perry et al., 2010), a reasonable assumption is that the decision concerning the number of syllables that a word has is most likely based on information about the number of vowel graphemes the word contains.¹ A similar assumption, that readers infer the number of syllables based on the number of identified vowel categories, was made in Perry, Ziegler, and Zorzi’s (2013) most recent extension of the CDP++ model.

Within our proposed framework, the computation of the posterior probability of each stress pattern in a word given some evidence, $P(\text{stress}|\text{evidence})$, further assumes that readers have some knowledge of the prior probability that any word in a given language has a hypothesized stress pattern, $P(\text{stress})$, as well as knowledge of the probability that the evidence made available from processing the word is associated with that hypothesized stress pattern, $P(\text{evidence}|\text{stress})$. The product of these two terms is divided by the sum of the products of prior probabilities and probabilities of evidence for all alternatives (stress') in the hypothesis space ($STRESS$) to calculate $P(\text{stress}|\text{evidence})$. Thus, Bayes’ formula in the present circumstances can be written as follows:

$$P(\text{stress}|\text{evidence}) = \frac{P(\text{evidence}|\text{stress})P(\text{stress})}{\sum_{\text{stress}' \in STRESS} P(\text{evidence}|\text{stress}')P(\text{stress}')} \quad (2)$$

A reader’s estimates of the prior probabilities of stress patterns in a language, $P(\text{stress})$ would, presumably, be based on that individual’s experience with the language. The assumption made here is that the probability estimates used by readers are essentially equivalent to the probabilities that actually exist in their language. Equating the prior probability of a stress pattern with the relative frequency of that stress pattern in the language is motivated by the fact that in many realms of language processing, readers have been shown to be sensitive

¹ This assumption does raise the question of how readers cope with situations in which the number of syllables and the number of vowel graphemes differ (e.g., when a silent vowel –e occurs at the end of some monosyllabic English words or for the so-called “hiatus” words). Those types of words do appear to be more difficult for readers to deal with (Chetail & Content, 2012, 2013).

to frequency (e.g., word frequency (Balota & Chumbley, 1985), frequency of meanings of polysemous words (Dell, 1990), frequency of co-occurrence of words (Saffran, Newport, & Aslin, 1996), etc.). Therefore, it is quite likely that readers would have a reasonably good impression of the frequency with which various stress patterns occur in their language. Indeed, there have been a number of empirical studies showing exactly that (Colombo, 1992; Monsell, Doyle, & Haggard, 1989). Some researchers even have claimed that readers not only know the most frequent stress pattern but also assign it by default (Black & Byng, 1986; Colombo, 1992) in many circumstances. Note that, within the framework being proposed here, the claim is not being made that there is a default assignment of the more frequent stress pattern. Rather, information about the frequency of stress patterns only establishes a bias towards the more frequent pattern that can be diminished or reversed whenever evidence strongly associated with an alternative stress pattern exists.

In order to calculate the posterior probability of a hypothesized stress pattern, it is, of course, also necessary to identify the sources of evidence for stress. What evidence do readers use in trying to decide what syllable of a word should be stressed? In processing a polysyllabic written word, readers may assign lexical stress by retrieving it from memory following lexical access, or by computing it sub-lexically based on some stress cues present in the language's orthography. Thus, there are two sources of evidence for stress: lexical and non-lexical.

Lexical evidence (LE) for stress is not probabilistic, but rather deterministic. If a reader accesses the lexical representation of a word and retrieves its corresponding stress pattern from memory, then s/he knows what the correct stress pattern for this word is. In other words, $P(LE|stress) = 1$, if a hypothesized stress pattern is a pattern that is stored with the lexical representation of the word. Due to the deterministic nature of lexical evidence, the posterior probability of a stress pattern computed based on the lexical evidence is also deterministic:

$$P(stress|LE) = \frac{P(LE|stress)P(stress)}{\sum_{stress' \in STRESS} P(LE|stress')P(stress')} \\ = \frac{1 * P(stress)}{1 * P(stress) + 0 * P(stress')} = 1. \quad (3)$$

Lexical evidence for stress, therefore, has perfect reliability; however, the likelihood of reliance on this evidence will vary across words due to differences in the speed of lexical access and, hence, the speed of retrieval of stress information from lexical memory. Thus, lexical evidence for stress would be more likely to be used when the representation of the word in the mental lexicon gets activated rapidly (e.g., when the

word is a high frequency word). On the other hand, when lexical access is relatively slower (e.g., the word is a low frequency word), the chance that perfectly reliable, but not readily available, lexical evidence for stress would be used by readers declines, meaning that the probability of using less reliable, but more readily available, non-lexical evidence for stress, evidence that the orthography provides, increases. Indeed, it has been empirically demonstrated that lexical stress can be assigned based on the non-lexical evidence provided in the word's orthography (Arciuli et al. 2010; Burani & Arduino, 2004; Colombo, 1992; Jouravlev & Lupker, 2014), and also that the reliance on non-lexical evidence in assigning stress, in particular, on the spelling-to-stress consistency of word endings, is greater for low frequency than for high frequency words (Colombo, 1992).

The second source of evidence for stress, therefore, is non-lexical evidence (non-LE) provided by the orthography. In prior research, a number of non-lexical cues to stress, including diacritics (Gutiérrez-Palma & Palma-Reyes, 2008; Protopapas, 2006), orthographic complexity of word onsets and codas (Kelly, Morris, & Verrekia, 1998; Kelly, 2004; Jouravlev & Lupker, submitted), and the orthography of various word segments (Arciuli & Cupples, 2006; Colombo, 1992; Jouravlev & Lupker, 2014, submitted) among others, have been identified. In any given language, there are likely a number of sources of non-lexical evidence for stress. Each source of evidence, or stress cue, used in the calculation of stress must, above all, be reliable, meaning that it should be in a significant probabilistic relation with a specific stress pattern (i.e., the cue should have high "validity") and it also should be known to be used by readers (i.e., the cue should have high "utility").

One assumption made here is that readers consider multiple sources of evidence in making stress assignment decisions in a sequential fashion, starting with the most informative cue to stress in a language and going to the least informative one (although, in the mathematical formulation, the order in which the cues are considered is usually irrelevant).² Thus, if there are two non-lexical stress cues in the language A and B, and A is a more informative one, first, the posterior probabilities of the stress patterns given evidence A, $P(stress|A)$, are

² In those cases when valid sources of evidence for stress overlap (e.g., if the orthography of a syllable and, at the same time, the orthography of some sub-component of that syllable, are used as stress cues), the order in which the cues are considered is important due to the fact that some of the information that they are based on is the same. In this case, the evidence from the source that is larger in size is entered into the equation first. Then, the evidence from the other source, corrected for conditional dependence with the larger source of evidence, is entered. The process of correction for conditional dependence of two overlapping components is described in more detail later in the paper.

calculated by using the following equation:

$$P(\textit{stress}|\textit{non-LE}_A) = \frac{P(\textit{non-LE}_A|\textit{stress})P(\textit{stress})}{\sum_{\textit{stress}' \in \textit{STRESS}} P(\textit{non-LE}_A|\textit{stress}')P(\textit{stress}')} \quad (4)$$

In this equation, the likelihood of evidence A given a hypothesized stress pattern, $P(\textit{non-LE}_A|\textit{stress})$, is computed by finding the proportion of words containing this particular evidence across all words of the corpus that have the hypothesized stress pattern. $P(\textit{non-LE}_A|\textit{stress}')$ is estimated by finding the proportion of words containing this particular evidence across all words of the corpus that do not have the hypothesized stress pattern. For example, in establishing the likelihood of evidence for words ending in “-et” in English (assuming that that word ending is a valid source of evidence for stress in this language) given a trochaic stress pattern, one would determine the number of trochaically stressed words that have that word ending and, then, one would find what proportion in all English words with trochaic stress that this particular group of words (i.e., the -et word group) constitutes.

Next, the posterior probabilities of the stress patterns given both evidence A and evidence B is calculated. At this stage, the prior probabilities are not those that reflect the frequency of various stress patterns in the language, but rather the posterior probabilities calculated based on the existence of evidence A (i.e., $P(\textit{stress}|\textit{non-LE}_A)$, which will be referred to as $P(\textit{stress})^*$). Thus, at the second stage of the computation, the formula is:

$$P(\textit{stress}|\textit{non-LE}_A, B) = \frac{P(\textit{non-LE}_B|\textit{stress})P(\textit{stress})^*}{\sum_{\textit{stress}' \in \textit{STRESS}} P(\textit{non-LE}_B|\textit{stress}') (1 - P(\textit{stress})^*)} \quad (5)$$

This process continues until all the relevant stress cues used by readers have been considered. Note that the stepwise approach in updating the posterior probabilities using several sources of evidence assumes that these sources are conditionally independent of each other. Two sources can be considered independent if the existence of one of them does not change the impact (i.e., the probability associated with the various possible stress assignments) of the other source. In the case of cues to stress this is not always the case, in that some sources might be correlated. Thus, the posterior probabilities computed following the above equation may be slightly

inaccurate. However, it is possible to essentially remove the statistical dependency, as will be discussed below.

In order to model the stress assignment process in any language, one must, of course, have a good idea of what the non-lexical sources of evidence for stress are that are used/acquired by readers of that language. There are likely to be a number of sources of evidence for stress that are in probabilistic relations with stress patterns in the language. However, some of these sources might be providing the same cue to the word’s stress pattern, and, thus, given this redundancy, one or more of these stress cues may be ignored by readers. Instead, readers might base their stress assignment decisions only on the most reliable subset of the sources of evidence for stress present in the language in most situations. We assume that during the process of language learning, readers attempt to find the simplest possible set of orthographic stress cues that would still allow them to assign stress to most of the words successfully. This process of converging on a certain set of sources of evidence for stress during the process of learning, however, is not a topic addressed in the present research. For present purposes, we will rely on the work of Jouravlev (2014) and Jouravlev and Lupker (submitted) who have identified three important sources of non-lexical evidence for stress assignment in Russian.

In summary, in the process of stress assignment, a reader might utilize deterministic, lexical evidence (i.e., the stress pattern is retrieved directly from memory) or probabilistic, non-lexical evidence (i.e., the stress pattern is computed based on the orthographic stress cues contained in a word). How likely is one or the other source of evidence for stress to be used for a particular word? As has been already mentioned, these probabilities will likely depend on the ease of lexical access. If the lexical representation of a word is easily accessed by readers, then we should see greater reliance on stress pattern information retrieved from lexical memory than for words for which lexical access is relatively slow. Prior research provides us with an extensive list of various item characteristics that are known to have an impact on the ease of lexical access (for an overview of major item effects see Adelman, 2012), with word frequency derived from the occurrence of words in a language corpus being the most widely used index of the speed of lexical access. Therefore, a word frequency measure, in particular a relatively new measure of word frequency, the Zipf-scale (van Heuven, Walter, Mandera, Keuleers, and Brysbaert, 2014), was used as an estimate of the likelihood that readers would base their stress assignment decisions on lexical evidence.

The Zipf score is a transformation of a more traditional measure of frequency, frequency per million words (fpmw), using the following formula:

$$\textit{Zipf}_{score} = \log_{10}(\textit{fpmw}) + 3. \quad (6)$$

The Zipf score ranges from 1 to 7, where 1 corresponds to words of very low frequency ($f_{pmw} = .01$) and 7 corresponds to words of very high frequency ($f_{pmw} = 10,000$). The Zipf score is assumed to be related to the ease of lexical access and, therefore, it should be related to the likelihood of utilization of lexical evidence for stress ($P(LE)$) in the process of stress assignment. We assumed that when readers process very frequent words, that is, those with the maximum Zipf score of 7, they would always assign stress based on the lexical evidence (i.e., $P(LE) = 1$). On the other hand, when extremely low frequency words with the minimum Zipf score of 1 as well as nonwords are processed, lexical evidence would not be used at all in the process of stress assignment ($P(LE) = 0$). For intermediate Zipf scores, the assumption was made that there is a linear relationship between the Zipf score and the likelihood of utilization of lexical evidence for stress. For instance, for a word with the Zipf score of 4, a score that is located right in the middle of the scale from 1 to 7, $P(LE) = .50$.

If stress is not assigned based on the lexical evidence for stress, then it is assigned based on the non-lexical evidence, hence, the likelihood of utilization of non-lexical evidence ($P(non-LE)$) can be computed through the following formula:

$$P(non-LE) = 1 - P(LE) \quad (7)$$

Finally, knowing the likelihoods of utilization of the two sources of evidence for stress (i.e., lexical versus non-lexical) and the posterior probabilities of a hypothesized stress pattern given the lexical and non-lexical evidence for a given word, we can estimate the combined posterior probability of a given stress pattern for that word through the following formula:

$$P(stress | LE, non-LE) = P(LE) * P(stress | LE) + P(non-LE) * P(stress | non-LE) \quad (8)$$

To conclude, the present approach is based on two assumptions about the sources of evidence for stress used by readers in the process of stress assignment. The first one is that there are two sources of evidence for stress: lexical and non-lexical. These two sources differ in reliability (with lexical evidence being more reliable) and accessibility (with lexical evidence being more accessible for high frequency than low frequency words). At the same time, the two sources are used in the process of stress assignment in essentially identical ways: i.e., readers rely on them in adjusting their prior beliefs in order to establish their posterior beliefs about likelihoods of stress patterns for items. The second assumption is that the increase of the likelihood of utilization of lexical

evidence for some words (related to the frequency of the word in a language) is associated with a decrease of the impact of the non-lexical evidence for stress on stress assignment. There is some empirical support for these assumptions coming from prior research. It has been reported in English and Italian (Colombo, 1992; Rastle & Coltheart, 2000), for example, that readers' performance is impacted significantly more by non-lexical sources of evidence for stress when readers name words of low rather than high frequency. The proposed probabilistic inferential view of stress assignment should be able to simulate this interaction of lexical (frequency) and non-lexical (e.g., consistency of orthography-to-stress mappings of word endings) factors.

The computed estimates of posterior probabilities for stress patterns are assumed to be reflective of the ease of stress assignment both in terms of speed and accuracy. Readers should be more accurate in assigning stress to words with high compared to low estimated posterior probabilities of the stress patterns that these words actually have. Further, we expect that for nonwords with high estimated posterior probabilities of a stress pattern, there should be greater agreement among readers in regards to the stress patterns assigned to those items compared to nonwords with estimated posterior probabilities of stress patterns that are near 50:50. Finally, the estimated posterior probabilities of stress patterns should correlate with the naming response latencies. In particular, a word with a high estimated posterior probability of the stress pattern that the word actually has and nonwords with a high posterior probability of a given stress pattern should show a latency advantage.

Calculating probabilities of stress patterns

For purposes of illustration, the computations of the posterior probabilities of trochaic and iambic stress patterns for the trochaically stressed word *beltik* based on a corpus of 30 disyllabic words of a fictitious language (see Table 1) are described. Let's assume that this word has a relatively low word frequency of 5 per million words. Hence, its Zipf score equals 3.7 ($=\log_{10}(5) + 3$). Based on the assumption of a linear relationship between the Zipf scores in the range of 1 to 7 and corresponding likelihoods of utilization of lexical evidence for stress in the range of 0 to 1, the likelihood of utilization of the lexical evidence for stress for this word with a Zipf score of 3.7 can be computed as being equal to .45. The likelihood of utilization of non-lexical evidence for stress for this word then equals .55.

Based on the lexical evidence, the probability that the correct trochaic stress pattern would be assigned to this word is 1.0. The probabilities of trochaic and iambic stress patterns given non-lexical evidence for stress are computed based on the specific non-lexical stress cues that are present in this

Table 1 The corpus of disyllabic words of a fictitious language used to illustrate the computation of the posterior probabilities within our proposed Bayesian framework

Words with trochaic stress	Words with iambic stress
BELTIK	BELTOP
BELKOP	BENRET
BETNIK	BELTET
BELSIK	DOLMAT
BENSET	DOLTIK
BELRAT	DOLNOP
BELMOT	FAPLOP
BERMAT	KILPIK
DOMRET	LIPSOP
FAPRET	MERLON
FAMLIK	
KOLTIK	
LIPSET	
MERLIK	
MOLTET	
MONPIK	
NERMET	
NELTIT	
POMLOP	
TERLIP	

language. Let's say that there are three sources of non-lexical evidence for stress in the language: the orthography of the ending of the second syllable (i.e., the second vowel of a word and all following consonants), the orthography of the first syllable, and, finally, the orthography of the beginning of the first syllable (i.e., all graphemes up to and including the vowel of the first syllable).³ These sources of evidence are integrated into the computation of posterior probabilities of stress patterns given non-lexical evidence in the following way. First, the prior distributions of stress patterns in this language are considered. As 67 % of words have a trochaic stress pattern, the prior probability that a word has trochaic stress, $P(\text{Stress1})$, is .67, while the prior probability that a word has iambic stress, $P(\text{Stress2})$, is .33. Next, the evidence provided in the orthography of the word is accounted for. As the ending *-ik* of the word *beltik* is present in 35 % of words with trochaic and 20 % of words with iambic stress, the posterior probability of a word having a trochaic stress given the presence of the ending *-ik* is

$$P(\text{Stress1}|-ik) = \frac{P(-ik|\text{Stress1})P(\text{Stress1})}{P(-ik|\text{Stress1})P(\text{Stress1}) + P(-ik|\text{Stress2})P(\text{Stress2})} = \frac{(.35)(.67)}{(.35)(.67) + (.20)(.33)} = \frac{.24}{.31} = .77 \quad (9)$$

³ The three sources of evidence for stress that were used to illustrate the computation of posterior probabilities of stress patterns in this fictitious language were selected merely for illustration purposes and are not the same sources that were used in the computation of posterior probabilities of stress patterns in Russian.

In the case of disyllabic words, the two stress hypotheses are in a trade-off relationship with one another such that increasing the belief in one decreases the belief for the other. Therefore, the posterior probability of the other hypothesis can be directly calculated:

$$P(\text{Stress2}|-ik) = 1 - P(\text{Stress1}|-ik) = 1 - .77 = .23 \quad (10)$$

The evidence provided by the first syllable *bel-* is present in 25 % of words with trochaic stress and 20 % of words with iambic stress. This stress cue is used to update earlier beliefs about stress patterns that were based on the presence of the evidence *-ik* in the word. Thus, $P(\text{Stress1}|-ik)$, referred to as $P(\text{Stress1})^*$, serves as the new prior probability of a trochaic stress pattern, while $P(\text{Stress2}|-ik)$, referred to as $P(\text{Stress2})^*$, is the new prior probability of an iambic stress pattern. The posterior probability that the word *belpik* has trochaic stress given the presence of *bel-* and *-ik* is calculated as follows:

$$P(\text{Stress1}|bel-, -ik) = \frac{P(bel-|\text{Stress1})P(\text{Stress1})^*}{P(bel-|\text{Stress1})P(\text{Stress1})^* + P(bel-|\text{Stress2})P(\text{Stress2})^*} = \frac{(.25)(.77)}{(.25)(.77) + (.20)(.23)} = \frac{.19}{.24} = .79 \quad (11)$$

As the final step in the calculation, the evidence provided by the beginning *be-* is considered. In assessing the likelihood of evidence provided by this orthographic component, the calculations cannot be based simply on the scope of representation of this component in words with trochaic versus iambic stress due to the fact that the beginning *be-* is a part of the first syllable *bel-* that has already been accounted for. This issue is addressed by considering the distribution of this component in all words except the words that have a first syllable of *bel-* ($P(\text{be-}, \neg \text{bel})|\text{Stress1}$ and $P(\text{be-}, \neg \text{bel})|\text{Stress2}$). Out of the words that meet that criterion, 15 % have trochaic stress and 10 % have iambic stress. These probabilities are used to update previous calculations. Thus, $P(\text{Stress1}|bel-, -ik)$, further referred to as $P(\text{Stress1})^{**}$, serves as the new prior probability of trochaic stress and $P(\text{Stress2}|bel-, -ik)$, further referred to as $P(\text{Stress2})^{**}$, is the new prior probability of iambic stress. Based on that, the posterior probability that the word *belpik* has a

trochaic stress pattern given the evidence *be-*, *bel-*, and *-ik*, is calculated as follows:

$$\begin{aligned}
 & P(\text{Stress1} | \text{be-}, \text{bel-}, \text{-ik}) \\
 &= \frac{P(\text{be-} | \text{Stress1})P(\text{Stress1})^{**}}{P(\text{be-} | \text{Stress1})P(\text{Stress1})^{**} + P(\text{be-} | \text{Stress2})P(\text{Stress2})^{**}} \\
 &= \frac{(.15)(.79)}{(.15)(.79) + (.10)(.21)} = \frac{.12}{.14} = .86
 \end{aligned}
 \tag{12}$$

Thus, based on the prior knowledge of the distribution of stress patterns in the language and the relationships that exist between the three types of orthographic cues and stress patterns in the words of this language, the posterior probabilities of trochaic and iambic stress patterns given non-lexical evidence for stress can be computed for the word *beltik*.

Finally, the posterior probability of a trochaic stress pattern given lexical and non-lexical evidence for stress for the word *beltik* is computed:

$$\begin{aligned}
 & P(\text{stress1} | \text{LE}, \text{non-LE}) = P(\text{LE}) * P(\text{stress1} | \text{LE}) \\
 & \quad + P(\text{non-LE}) * P(\text{stress1} | \text{non-LE}) \\
 & = (.45 * 1.0) + (.55 * .86) = .45 + .47 = .92
 \end{aligned}
 \tag{13}$$

Based on these calculations, one would expect that proper stress assignment should be rapid for the word *beltik* and the error rate should be quite low, potentially around 8 % in a speeded naming situation.

The calculation of probabilities of stress patterns does, of course, reflect the behavior of an ideal observer, who computes the most probable stress pattern given the whole corpus of the language. The real patterns of behavior are expected to be correlated with the patterns predicted based on the computations we have described, but are unlikely to be identical. First of all, although humans appear to be good statisticians, they are not perfect, while the presented computations are error-free. As various errors, biases, and heuristics are common features of human cognition (Tversky & Kahneman, 1974), departures from optimal behavior are expected in human performance.

Another point to note here relates to the process of estimation of the likelihoods. In the computation of posterior probabilities of stress patterns described above, these likelihoods were calculated online from the full lexicon. It is unlikely that during actual reading these values would be computed anew every time. It is more likely that being exposed to the corpus of words has caused readers to learn likelihoods for various recurring units in their language which can simply be used in

language processing. That is, readers have knowledge about the likelihoods with which stress patterns are associated with particular non-lexical orthographic cues and use this information when deciding what syllable in a word to stress. It should be noted, of course, that the reader's estimate of the likelihood of an orthographic component being associated with a given stress pattern would be a dynamic representation that would be constantly updated in the light of new experiences.

Stress assignment in Russian disyllables as a problem of probabilistic inference

As noted, the language used to evaluate the idea that stress assignment can be perceived as a problem of probabilistic inference was Russian. The process of stress assignment appears to be complicated in this language because stress is not explicitly marked in the orthography and it does not conform to any clear implicit rules (see Coats, 1976; Gouskova & Roon, 2013; Halle, 1973; Jouravlev & Lupker, 2014; Zaliznjak, 1985 for more information about the system of lexical stress in Russian). Prior investigations of the lexical stress system in Russian provided the information required for the implementation of our framework to stress assignment. Specifically, from the analysis of the corpus of Russian disyllabic words (Jouravlev & Lupker, 2014), we know the distributions of stress patterns in this language. These distributions are used as the priors in the computations of the posterior probabilities of stress patterns. In particular, the prior probability of a trochaic stress pattern in Russian ($P(\text{Stress1})$) is .55, while the prior probability of an iambic stress pattern ($P(\text{Stress2})$) is .45. Note that these prior probabilities reflect stress pattern distributions in the Russian language only. The priors used in the computation of posterior probabilities of stress patterns in other languages will be different (e.g., in English, $P(\text{Stress1}) = .8$ and $P(\text{Stress2}) = .2$).

As is the case for the priors of stress patterns in a language, it is also the case that non-lexical sources of evidence for stress would differ between languages. To conclude that a certain set of non-lexical stress cues is used in the process of stress assignment in a language, one needs to demonstrate that these cues are probabilistically associated with stress patterns in that language and that readers are sensitive to these cues. An investigation of this type has been recently conducted in Russian (Jouravlev, 2014; Jouravlev & Lupker, submitted). The 11 potential sources of evidence for stress that were examined were Grammatical Category, Log Frequency, Length, Word Onset Complexity, Word Coda Complexity, and a set of six orthographic segments. The orthographic segments were the First Syllable (referred to as CVC1), the Beginning of the First Syllable (CV1), the Ending of the First Syllable (VC1), the Second Syllable (CVC2), the Beginning of the Second Syllable (CV2), and the Ending of the Second Syllable

(VC2). It should be noted that it is not a particular identity of some component (e.g., “-et” in “market”), but an orthographic component per se (e.g., VC2 in all words of the corpus) that was considered as the source of evidence for stress in this research (although the calculations in the present research are based on the probabilities associated with the identities of the particular components themselves in each word being named). If some orthographic component is a valid source of evidence for stress, we should observe improved stress-assignment performance when the particular identity of that component cues a stress pattern that is consistent with the actual stress that a word has (e.g., if the VC2 component is a valid source, the VC2 component “-et”, the identity of which is associated with trochaic stress, correctly provides evidence for trochaic stress in the word “market”) and disrupted performance if the particular identity of this component cues a stress pattern that is inconsistent with the actual stress that a word has (e.g., the VC2 component “-et” incorrectly provides evidence for trochaic stress in the word “cadet”). Essentially, one would expect to observe spelling-to-stress consistency effects for the components that act as strong sources of evidence for stress.⁴

In the first regression study (Study 1) reported in Jouravlev and Lupker (submitted), the predictability of stress patterns in a corpus of disyllabic words based on a number of potential cues was assessed. Onset complexity, coda complexity, the orthography of the first syllable (CVC1), of the second syllable (CVC2), and of the ending of the second syllable (VC2) were found to be probabilistically associated with stress in the corpus of Russian disyllabic words. In addition to examining relationships existing in this normative corpus, real language use was investigated as well due to the fact that it might deviate significantly from the stress information reflected in the dictionaries. In their second study, Jouravlev and Lupker ran a set of regression analyses assessing the effects of the same eleven potential sources of evidence for stress on the naming latencies and stress-assignment accuracy data for a large number of disyllabic words named by native speakers of Russian. The results

showed that only three of the eleven cues (CVC1, CVC2, and VC2) were predictive of the behavioral data. Finally, in order to validate these three sources of evidence for stress identified in the regression studies, these sources were examined again in a study employing a factorial design. Those results corroborated the finding that Russian speakers do use the orthography of CVC1, CVC2, and VC2 when assigning stress. To conclude, there appear to be three predominant non-lexical sources of evidence that are probabilistically associated with stress patterns in Russian and are used by speakers of that language: the orthography of the first syllable (CVC1), of the second syllable (CVC2), and of the ending of the second syllable (VC2). These three sources of evidence for stress were used in the computation of posterior probabilities of stress given non-lexical evidence in the present analyses.⁵

The orthographic segments CVC2 and VC2 are not completely independent from each other as one is a part of the other. Thus, when we account for evidence provided by CVC2, we also account for some, but not all, evidence that VC2 provides. This issue was addressed in the way that was already described earlier in the example from the fictitious language. In particular, we considered the distribution of specific instances of the VC2 cue (e.g., VC2 “-оHT” present in the word “рeмoHT” for which posterior probabilities of stress patterns are computed) in all words except the words that have the same CVC2 as the word of interest (e.g., all words that have VC2 “оHT”, but do not have CVC2 “мoHT”).

The computations of the likelihood of evidence based on these three orthographic components were derived from the lexicon of 13,942 Russian disyllabic words. It should be noted that only the words with a frequency of more than one per million were included in this corpus. Further, only nouns that describe a class of entities (i.e., common nouns), but not unique entities (i.e., proper nouns), are in this corpus. Thus, the corpus that was used for these calculations did not include all disyllabic words in the Russian language. The fact that the corpus used was not exhaustive might lead to a slight distortion in the computation of the likelihoods of evidence. Although this limitation should not change the predictions to any noticeable degree in the majority of cases, it might matter when certain orthographic patterns are very low in frequency in the language overall and, therefore, those patterns are not well represented in the selected corpus. For example, in the corpus, there is just one word *мольберт* that has the

⁴ Both type and token spelling-to-stress consistency measures were analyzed by Jouravlev and Lupker (submitted). Words sharing an orthographic component under consideration were defined as the words in the target’s neighborhood. In calculating the type consistency measure, the proportion of words with trochaic stress in the neighborhood was calculated by dividing the number of words in the neighborhood having trochaic stress by the total number of words in the neighborhood. High values meant that in the specified orthographic neighborhood the number of words with trochaic stress is higher than the number of words with iambic stress, while low values meant that there are more words with iambic than trochaic stress in the neighborhood. Values near .50 indicate that the neighborhood has approximately equal percentages of the two stress types (i.e., it is an inconsistent neighborhood). Token consistency measures, corresponding to the summed frequency of words with trochaic stress in an orthographic neighborhood divided by the summed frequency of all words in that orthographic neighborhood, were calculated in a similar fashion.

⁵ Although the orthography of CVC1, CVC2, and VC2 were identified as sources of evidence for stress with high utility and validity in Russian (and are used in the computation of the posterior probabilities of assigning stress in Russian disyllables in the present paper), there is no reason to believe that the set of non-lexical sources of evidence for stress will be the same in other languages nor is there reason to believe that no other non-lexical sources of evidence for stress are ever used by Russian readers.

orthographic component *-бepm* (CVC2) and it has an iambic stress. Thus, in calculating the posterior probability of a trochaic stress for this word based on the information provided by CVC2, the prediction is that there is no chance that this word would be assigned a trochaic stress pattern:

$$P(\text{Stress1}|-бepm) = \frac{(0/7668) \times .55}{((0/7668) \times .55) + ((1/6274) \times .45)} = .00 \quad (14)$$

This situation likely will not create a problem when the proposed computational approach is assessed on its ability to predict a stress pattern for the single word *мольбepm* that is a part of the selected lexicon; however, it can become a problem if the approach is assessed on its ability to predict a stress pattern for another word with the orthographic component *-бepm* (CVC2) that is not a part of the selected corpus and that has a trochaic stress pattern (e.g., the proper name *uУбepm*). The issue of a distortion in the representation of evidence in the corpus due to having a frequency cutoff is less problematic when that evidence is represented widely in the language. In that situation, even if a few words having a particular orthographic component did not make it into the corpus due to their very low frequency, the relative strength of the evidence based on that component for the two alternative hypotheses should not depart greatly from the distribution present in the language.

The issue of incorrect calculations of posterior probabilities due to the underrepresentation of certain orthographic patterns in the corpus was addressed in the following way. To reflect the possibility that there might be a word present in the language that has a certain orthographic pattern, but that simply did not make it into the corpus, one instance was added in the calculations of likelihoods of evidence of both trochaic stress and iambic stress.⁶ That is, in calculating the posterior probability of a trochaic stress given *-бepm* (CVC2), the counts for trochaic versus iambic stress were not 0 and 1 (meaning that in the lexicon, 0 words with that orthographic pattern have trochaic stress and 1 word with that pattern has iambic stress), but rather 1 and 2 (meaning that there is potentially 1 word with that pattern that has trochaic stress

and 2 words with that pattern that have iambic stress). Following this way of estimating the likelihood of evidence, the posterior probability of a trochaic stress given the evidence *-бepm* (CVC2) is:

$$P(\text{Stress1}|-бepm) = \frac{(1/7668) \times .55}{((1/7668) \times .55) + ((2/6274) \times .45)} = .33 \quad (15)$$

The inclusion of a parameter reflecting the possibility that there might be other words with the same stress patterns that were simply not included in the corpus allows for more reasonable estimates of probabilities of stress patterns not just for words like *мольбepm*, but also for words like *uУбepm* as well as nonwords containing the orthographic pattern in question.

This slight transformation of the likelihoods of evidence for stress by adding one “potentially possible” item with trochaic stress and one with iambic stress also allows for the computation of posterior probabilities of stress patterns for nonwords with some non-extant components. For example, the Russian nonword *блинбиж* has two components that are present in Russian (CVC1 *блин-* and VC2 *-иж*) and one component that does not appear in words of the Russian language (CVC2 *-биж*). We assume that there is potentially one trochaically stressed word and one iambically stressed word in Russian that have the non-extant component *-биж*. Based on this assumption the posterior probabilities of stress patterns given this non-extant component can be computed:

$$P(\text{Stress1}|-биж) = \frac{(1/7668) \times .55}{((1/7668) \times .55) + ((1/6274) \times .45)} = .50 \quad (16)$$

As can be seen here, the impact of non-extant orthographic components on the posterior probabilities of stress patterns is minimal, which is probably a correct reflection of the absence of formed associations between any non-extant orthographic components and stress patterns in the language. Therefore, the estimated posterior probabilities of stress patterns would mainly rely on the likelihoods of evidence based on those orthographic components that are present in the language (e.g., CVC1 *блин-* and VC2 *-иж* for the nonword *блинбиж*). Based on the set of assumptions described above, the estimates of stress pattern probabilities can now be compared with actual stress pattern assignment of readers of Russian naming disyllabic words (Experiment 1) and nonwords (Experiment 2).

⁶ The strategy of adjusting the likelihoods of evidence given stress patterns by adding one potential exemplar to the computation was applied in all cases. For instance, if a certain component was present in five words with trochaic and two words with iambic stress patterns, the count used for the trochaic stress pattern was 6 (=5 + 1), while the count used for the iambic stress pattern was 3 (=2 + 1).

Experiment 1: Word naming task

The goal of Experiment 1 was to assess whether the computed posterior probabilities of stress patterns are predictive of stress assignment performance and of response latencies of native speakers of Russian naming disyllabic words. In most of the prior computational studies of the process of stress assignment (Rastle & Coltheart, 2000; Ševa et al., 2009; Perry et al., 2010), the authors compared the stress pattern predictions of their models with the stress patterns that disyllabic words have in the language rather than actual stress assignment performance of readers on those words. The reason that this approach was used was either that these models are unable to simulate stress assignment probabilities across individuals due to the fact that their output is deterministic (i.e., trochaic or iambic stress patterns) rather than probabilistic (Rastle & Coltheart, 2000), or because the modelers (Perry et al., 2010; Ševa et al. 2009) preferred to transform their continuous probability values (e.g., activation levels for the various stress patterns in the model) into binary stress outputs (with the stress pattern node achieving the highest level of activation being considered as determining the stress pattern that the model would assign to a word). In either situation, it was, therefore, these binary outputs that were then compared against stress patterns in the relevant corpus.

Our main analysis is somewhat different than in the previous literature. Our aim was to compare the predicted probabilities of stress assignment against the actual probabilities of speakers producing the various stress patterns. This process of comparing the predicted probabilities of stress patterns with actual stress assignment performance is a novel and, presumably, more diagnostic way of evaluating the computational principles of stress assignment as well as the theoretical ideas underlying those principles. Further, we also assessed whether the speed of naming is related to the predicted probabilities of the most frequently produced stress pattern as our framework also speaks to that issue. Nonetheless, for purposes of comparison to the previous literature, we also analyzed stress pattern predictions by creating binary outputs (i.e., we report a binary classification analysis). However, the binary output predictions were compared not against the stress patterns listed in a dictionary but rather against the more frequent (binary) stress assignments produced by our participants (for about 2 % of our words participants produced a different stress pattern than that listed in the dictionary).

Method

Participants

Thirty-four undergraduate students (age 17–23 years; $M = 19$) from Altay State University (Barnaul, Russia) took part in this experiment for a small (approximately \$5 US) monetary

remuneration. All were native speakers of Russian. None of the participants reported high proficiency in any second language. The same group of participants was used in prior studies aimed at identifying the sources of evidence for stress (Jouravlev & Lupker, *submitted*).

Materials

A set of 500 disyllabic words was randomly selected from the corpus of Russian disyllabic words. None of these items were used in the experiments of Jouravlev and Lupker (*submitted*) in which the three non-lexical sources of evidence for stress were identified. The words were selected so that the distribution of stress patterns and grammatical categories in this set of experimental items was similar to that in the Russian language. There were thirty-four words that had ambiguous stress (also known as minimal stress pairs) because they corresponded to two lexical items that differed in stress pattern only (e.g., *напом* – Instrumental case for “steam” vs. *напом* – Nominative case for “ferry”). For each of these words, the stress pattern for the more frequent word of the pair was chosen as the stress pattern that should be assigned when stress assignment is based on lexical evidence.

Procedure

The behavioral data against which the simulation results were compared was collected in the following way. Participants were instructed to read aloud words presented on the screen as quickly and as accurately as possible. Instructions and stimuli were presented using the DMDX display system (Forster & Forster, 2003). The list of 500 items was presented in two blocks of trials. Every participant named all 500 items. The order of blocks and of items within blocks was randomized for each participant. Each trial started with the presentation of a fixation point for 500 ms. The target word in upper-case appeared in white on a black background (Courier New, 12-point font) for 2,000 ms or until the participant responded. The intertrial interval was 1,000 ms. Responses were marked using CheckVocal (Protopapas, 2007) by the first author and by two other native speakers of Russian. The predicted posterior probabilities of trochaic and iambic stress patterns were computed for each word using the method described above.

Results

Analysis of stress pattern assignment

In our main analysis, we evaluated the actual posterior probability predictions, specifically, whether there is the predicted relationship between the posterior probabilities of stress patterns generated by our calculations and the stress assignment probabilities demonstrated by the participants (see Fig. 2A for

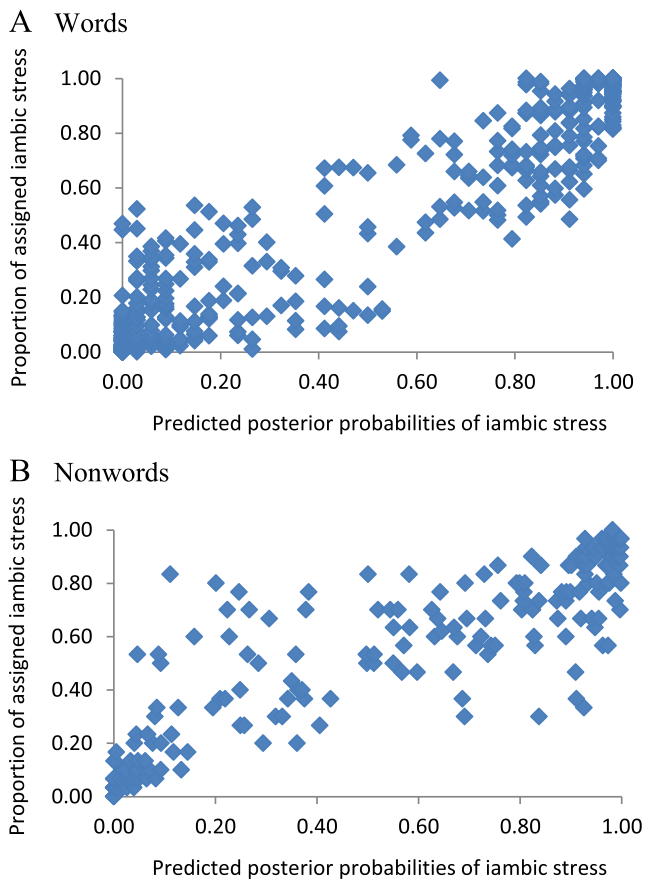


Fig. 2 Scatterplots of predicted posterior probabilities of iambic stress patterns in relation to the proportion of iambic stress patterns being assigned by participants naming (A) the words of Experiment 1 and (B) the nonwords of Experiment 2

the scatterplot of predicted and assigned stress patterns). For this analysis, the computed posterior probability of an iambic stress pattern was correlated with the proportion of responses with iambic stress that ranged from 0 (meaning that all participants named a word with a trochaic stress pattern) to 1 (meaning that all participants named a word with an iambic stress pattern). This analysis showed that the posterior probabilities of iambic stress patterns were predictive of the likelihood that the readers would pronounce those words with iambic stress, $r(498) = .94$, $F(1,498) = 3960.17$, $p < .001$.⁷

For purposes of the more traditional, binary classification analysis, a posterior probability of a stress pattern that exceeded .50 was interpreted as the prediction that that word has that stress pattern. These predictions were compared against the stress patterns produced by the participants. If a stress pattern produced by the participants for a given word was produced more than 50 % of the time, the word was assumed to have that

⁷ The predicted posterior probabilities of iambic stress patterns and likelihoods that that stress pattern was assigned by readers were also strongly correlated in the analysis that did not contain the minimal stress pair words, $r(464) = .88$, $F(1,464) = 1688.81$, $p < .001$.

stress pattern. Interestingly, of the 500 words, five words were assigned the two stress patterns by participants equally often, while eight were assigned the incorrect stress pattern more frequently than the correct pattern.⁸ Therefore the present classification analysis is slightly different than the analysis typically used in the literature in which the model's predictions are compared against the stress patterns listed in a dictionary. In the classification analysis, the stress patterns in 96 % of the analyzed words were predicted correctly, with no difference in the accuracy of predictions for the trochaically versus iambically stressed words (see Table 2A for the distribution of human stress pattern classification and computed stress pattern predictions). This level of predictive success was highly significant: $(493) = .93$, $\chi^2(493) = 424.90$, $p < .001$. Thus, on the basis of the criterion commonly used to evaluate models of stress assignment in the literature, our predictions were quite good.⁹

Analysis of response latencies

Response latencies for words that were assigned the expected stress pattern were analyzed using a linear mixed effects logistic regression model. In this model, *Subjects* and *Items* were entered as crossed random factors, while estimated *Posterior Probability* of trochaic (for trochaically stressed words) or iambic (for iambically stressed words) stress patterns, *Log Frequency*, *Length*, *Onset Complexity*, and *Characteristics of the First Phoneme* were entered as fixed factors. The effect of the estimated *Posterior Probability* on naming response latencies is of primary theoretical interest in this analysis. The other predictor variables, all of which have been previously shown to impact the speed of word naming in Russian (Jouravlev, 2014; Jouravlev & Lupker, submitted), were included to remove any variance associated with them. The analysis was conducted using the R package *lme4* (Bates & Maechler, 2010). The results showed that the following variables were significant predictors of response latencies: *Posterior Probability* ($t(16378) = -4.67$, $p < .001$), *Log Frequency* ($t(16378) = -8.30$, $p < .001$), *Length* ($t(1286) = 2.79$, $p = .05$), and *Onset Complexity* ($t(1286) = 3.31$, $p = .01$). Most importantly, the *Posterior Probability* factor was strongly and negatively related to readers' speed of naming. Higher predicted posterior probabilities of the correct stress pattern were associated with shorter latencies.¹⁰

⁸ The five words that were assigned the two stress patterns equally often were not included in the classification analysis.

⁹ The exclusion of minimal stress pair words from the analysis did not change the results with stress patterns being predicted correctly for 97 % of the remaining words: $(460) = .94$, $\chi^2(460) = 410.22$, $p < .001$.

¹⁰ The interaction of the Frequency and Posterior Probability factors was not entered into this analysis as the computation of posterior probabilities, in particular of the likelihood of utilization of lexical evidence for stress, already incorporates word frequency information.

Table 2 Contingency tables showing the distribution of human stress pattern classification and computed stress pattern predictions in (A) the words of Experiment 1 (N=495) and (B) the nonwords of Experiment 2 (N=195)

Assigned Stress	Predicted Stress		Total
	Trochaic	Iambic	
A. Words			
Trochaic	268	9	277
Iambic	10	208	218
Total	278	218	
B. Nonwords			
Trochaic	64	9	73
Iambic	15	107	122
Total	79	116	

Experiment 2: Nonword naming task

Simulating performance of readers on nonwords, in particular their decisions in assigning stress patterns, is generally considered the gold standard in validating theories and computational models of reading and stress assignment (Perry et al., 2010). One of the challenges in this task is that stress assignment in nonword naming is characterized by great inter-subject variance (Zevin & Joanisse, 2000). All existing models of stress assignment have been previously tested by their modelers on their ability to predict stress pattern placement in nonword naming; however, as noted above, because models produce (or the researchers reported only) binary, deterministic output (trochaic or iambic stress patterns), they would be unable to account for any of this variability. That is, the three most well-known models of stress assignment only predict the most frequent stress pattern that participants assign to a nonword, rather than the ratio of responses with trochaic versus iambic stress patterns assigned to that nonword. In contrast, the computational implementation of the stress assignment process that is being presented here can provide predictions of the distribution of trochaic and iambic responses that speakers should produce in naming nonwords as well as the most frequent response to each nonword that should be given by participants and its latency. The aim of Experiment 2 was to assess whether estimated posterior probabilities of stress patterns were correlated with the actual stress assignment performance and response latencies of readers naming nonwords.

Method

Participants

Thirty undergraduate students (age 17–23 years; $M = 19$) from Altay State University (Barnaul, Russia) took part in this experiment for a small (approximately \$5 US) monetary

remuneration. All were native speakers of Russian. None of the participants reported high proficiency in any second language.

Materials

A set of 200 disyllabic nonwords was created by randomly combining first syllables and second syllables of Russian disyllabic words. Syllables were combined in such a way that no real words or pseudohomophones were created. All nonwords were pronounceable and did not violate any ortho-phonological constraints present in Russian. To minimize the possibility that stress assignment is completed by analogy to a real word, no nonword that is an orthographic neighbor of a real word (Coltheart, Davelaar, Jonasson, & Besner, 1977) was included as a stimulus in this experiment.

Procedure

The behavioral data against which the simulation results were compared were collected in the following way. Participants were instructed to read aloud novel words that would be presented on the screen. Instructions and stimuli were presented using the DMDX display system (Forster & Forster, 2003). The list of 200 items was presented in two blocks of trials. Every participant named all 200 items. The order of blocks and of items within blocks was randomized for each participant. Each trial started with the presentation of a fixation point for 500 ms. The target nonword in upper-case appeared in white on a black background (Courier New, 12 font) for 2000 ms or until the participant responded. The intertrial interval was 1000 ms. Responses were marked using CheckVocal (Protopapas, 2007) by the first author and by two other native speakers of Russian. The predicted posterior probabilities of trochaic and iambic stress patterns were computed for each word using the technique described above.

Results

Analysis of stress pattern assignment

As before, the correlational analysis is taken to be a better means of evaluating the predictions derived from our framework. Those predictions, in the form of posterior probabilities of an iambic stress pattern, were correlated with the ratio of iambic stress responses made by participants to the nonwords (see Fig. 2B for the scatterplot of predicted and assigned stress patterns). The results showed that the estimations of posterior probabilities of stress patterns computed were quite reflective of actual performance, $r(198) = .87$, $F(1, 198) = 600.35$, $p < .001$.

As with the words, a conventional classification analysis of stress assignment was undertaken. Again, for purposes of the

classification analysis, a posterior probability of a stress pattern that exceeded .50 was interpreted as meaning that that pattern was the “correct” pattern for that nonword. In the classification analysis, correct predictions about the stress pattern that is more likely to be realized by participants were made for 176 out of the 200 nonwords (88 % correct): $\phi(198) = .75$, $\chi^2(198) = 112.69$, $p < .001$. For both trochaically and iambically stressed nonwords, the accuracy of prediction reached 88 % (see Table 2B for the distribution of human stress pattern classification and computed stress pattern predictions).

Analysis of response latencies

Response latencies were analyzed using a linear mixed effects logistic regression model. In this model, *Subjects* and *Items* were entered as crossed random factors. *Length*, *Onset Complexity*, and *Characteristics of the First Phoneme* were fixed factors that were of no theoretical interest in this analysis and were included to remove any variance associated with them. The variable of interest was the estimated *Posterior Probability* of that stress pattern (trochaic or iambic) for which this estimate was higher than that for the other stress pattern. For example, if the estimated posterior probability of a trochaic stress pattern was greater than that of an iambic stress pattern for a given nonword, the posterior probability of trochaic stress pattern was used in the analysis of the naming latencies of that nonword when it was named with a trochaic stress. Higher values for the *Posterior Probability* variable meant that stronger predictions about the likelihood of either a trochaic or an iambic stress pattern are made which should translate into more rapid naming. The analysis was conducted using the R package *lme4* (Bates & Maechler, 2010). The results showed that only the *Posterior Probability* variable was significantly (and negatively) related to response latencies: $t(5964) = -2.16$, $p = .03$. Thus, participants were faster to name those nonwords for which the posterior probabilities of trochaic or iambic stress patterns were high than those nonwords for which the computed posterior probabilities of trochaic or iambic stress patterns were nearly equal.

General discussion

In this paper, we proposed and evaluated the idea that the human mind might be approaching the task of stress assignment as a problem of probabilistic inference, i.e., in assigning stress, readers evaluate the likelihood of potential stress patterns when making a stress assignment decision. The process is assumed to be accomplished by considering prior

probabilities of stress patterns in a language and the evidence for each stress pattern provided in the stimulus being named. The prior probabilities of stress patterns in a language used by a speaker are assumed to reasonably accurately reflect the frequency of those stress patterns in the language. The evidence for stress involves any type of information present in stimulus that is probabilistically associated with stress patterns in the language (i.e., has high validity) and that speakers use in making their stress assignment decisions (i.e., has high utility). We posit that stress assignment may be completed based on two sources of evidence for stress: lexical and non-lexical. Lexical evidence corresponds to stress pattern information retrieved by a reader from the mental lexical representation of a word. This evidence has perfect validity and utility (i.e., each word is generally associated just with one stress pattern and readers would be expected to know what that stress pattern is). Non-lexical evidence for stress corresponds to stress cues present in the orthography of items. The likelihood of utilization of lexical versus non-lexical sources of evidence for stress in the process of stress assignment will depend on the ease of lexical access for a particular item. For items that do have lexical representations (i.e., words), but for which lexical access is relatively effortful and slow (low frequency words) stress assignment will be more likely to be based on non-lexical rather than lexical evidence. In contrast, readers are expected to use mainly lexical, rather than non-lexical, evidence for stress in processing items for which lexical access is relatively effortless and fast (high frequency words). Readers will assign stress to items for which there is no lexical representation (i.e., nonwords) based on non-lexical evidence only.

The role of the overall frequency of a particular stress pattern in a language in the process of stress assignment is an object of significant debate. One view is that readers assign the most frequent stress pattern by default and, thus, that there is an essential difference in the processing of words with more frequent versus less frequent stress patterns (i.e., stress patterns are computed by readers based on some stress cues only for words with less frequent stress patterns; Black & Byng, 1986; Colombo, 1992). The polar opposite view is that the frequency of stress patterns plays no role in processing at all and, hence, the mechanisms of processing words with a more frequent stress pattern are exactly the same as those for words with a less frequent stress pattern (Burani & Arduino, 2004; Sulpizio, Arduino, Paizi, & Burani, 2013). We take an intermediate position: Knowledge of the overall distribution of stress patterns in a language is used to create a prior belief about the likelihood with which words will have a particular stress pattern or, in other words, as a baseline for further computations of probabilities of stress patterns. Thus, in the proposed probabilistic inferential framework, there are no differences in the mechanisms of processing of words with the more frequent versus the less frequent stress pattern, although words with a more frequent stress pattern do enjoy somewhat of a

head start. However, this initial situation can be easily changed if the word contains a number of orthographic cues that are probabilistically associated with the less frequent stress pattern.

The analysis being offered here is, therefore, based on the idea that non-lexical sources of evidence for stress (or “stress cues”) play a vital role in assigning stress. Previously, a wide range of potential non-lexical stress cues has been suggested, including graphemic complexity of onsets and codas (Kelly, 2004; Kelly, Morris, & Verrekia, 1998), orthography of word beginnings and endings (Arciuli et al., 2010), grammatical status of a word (Arciuli & Cupples, 2006), etc. In making stress assignment decisions, readers may evaluate many sources of evidence for stress present in their language. However, due to time constraints and due to the excessive amount of evidence for stress patterns, some of which is redundant, readers, in general, are likely to rely on a limited set of highly informative stress cues. Doing so would allow readers to assign stress with high accuracy and speed in the majority of cases.

The idea proposed here, that stress assignment can be thought of as a problem of probabilistic inference, can be applied to any language that utilizes lexical stress. The exact prior probabilities and sources of evidence for stress would, of course, be language-specific. In this paper, the idea that readers consider prior probabilities of stress patterns and various sources of evidence for stress when assigning stress to an item and the probable computational implementation of this idea have been tested against stress assignment performance of native speakers of Russian who were naming disyllabic words and nonwords. The choice of Russian for the present research is explained by the fact that despite the importance of lexical stress for word recognition in Russian, its assignment is very complex and is often a source of speech errors. Further, prior research on the lexical stress system in Russian (Jouravlev & Lupker, 2014; Jouravlev & Lupker, *submitted*) provides us with information about the distribution of stress patterns in that language (i.e., prior probabilities of stress patterns) and the nature of the cues that are probabilistically associated with stress patterns (i.e., sources of evidence for stress) that are needed for the computational implementation of the probabilistic inferential approach to stress assignment examined in the present paper. Thus, we know that in Russian the prior probability of a trochaic stress pattern is .55, while the prior probability of an iambic stress pattern is .45. Jouravlev and Lupker’s (*submitted*) empirical investigations also demonstrated that there are three reliable non-lexical stress cues in Russian: the orthography of the first syllable (CVC1), of the second syllable (CVC2), and of the ending of the second syllable (VC2).

In Experiment 1, the computed posterior probabilities of stress patterns for a set of Russian disyllabic words were compared with the words’ assigned stress and with stress

assignment probabilities and response latencies of native speakers of Russian naming these words. The results showed that the computed probabilities of stress patterns were predictive of the most common stress patterns for the majority of the words tested and also of the probabilities of stress assignments demonstrated by the readers. In addition, there was a strong relationship between the computed posterior probabilities of stress patterns and response latencies. More specifically, words for which the posterior probability of the correct stress pattern was around 50:50 were not only often assigned the incorrect stress pattern, they also had longer response latencies.

The aim of Experiment 2 was to assess whether the posterior probabilities of stress patterns computed for Russian disyllables according to the technique described in this paper correlate with stress patterns that speakers of Russian assign to nonwords and the speed with which they named those nonwords. Once again, the technique produced good predictions. The posterior probabilities were reflective of the relative ratio of trochaic versus iambic responses given by the participants. Further, readers’ response latencies to nonwords for which the computed posterior probabilities of either a trochaic or an iambic stress patterns were not high were longer in comparison to nonwords for which the posterior probability of one of the two alternative stress patterns was high.

Theoretical implications

According to the view proposed here, the process of stress assignment is a problem of probabilistic inference in which readers consider estimates of the probabilities of alternative stress patterns for each item. These estimates are computed based either on lexical or non-lexical evidence for stress provided by information concerning the item being named. Thus, similar to the dual-route theory of reading (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), to the stress assignment algorithm by Rastle and Coltheart (2000), and to the CDP++ (Perry et al., 2010), this proposal does distinguish between the effects of lexical and non-lexical information on the stress assignment process; however, these two sources of information are not conceived of as two separate routes that reflect essentially different cognitive mechanisms of reading (e.g., retrieved vs. computed phonology), but rather as two different sources of evidence for stress that are used by the same cognitive mechanism.

Another important difference between the present approach and the mechanisms of stress assignment implemented within the CDP++ model is the usage of knowledge of prior probabilities of stress patterns. In the present approach, this distributional knowledge is used as a baseline reflecting a prior belief about the likelihood that a word has a particular stress pattern. In the CDP++ model and in connectionist models in general, no baseline differences are assumed. Is the

knowledge of prior probabilities a required component in our framework? How sensitive are the posterior predicted probabilities of stress patterns to variations in the priors? To get some insight into this issue, we compared the predictions of stress assignment against performance of native readers of Russian on the words and nonwords used in Experiments 1 and 2 as the priors used in the calculations were varied. As can be seen in Fig. 3, the estimated predictions that were based on the correct prior probabilities of stress patterns (.55 and .45) fit the behavioral data best. The goodness of fit of predictions with different priors depended on the degree of difference. Slight differences in prior probabilities (by $\pm .20$) did not have a dramatic impact on the goodness of fit, while significant differences (by more than $\pm .20$) resulted in a substantial drop in the ability of our framework to predict the data. This pattern suggests that the notion of prior probabilities of stress patterns is an important construct and that readers are sensitive to this knowledge. However, very precise estimations of priors do not appear to be essential. The behavioral patterns observed here would follow even if priors were somewhat inaccurately estimated by our participants (i.e., even if their probability learning process was less than perfect).

A more fundamental point of divergence between the approach to stress assignment described in the present paper and any connectionist view is the number of required non-lexical sources of evidence for stress that is assumed to be considered in the process of stress pattern computation. More specifically, in our framework, quite accurate predictions of stress patterns can be achieved using only a limited number of non-lexical sources of evidence for stress. That is, the posterior

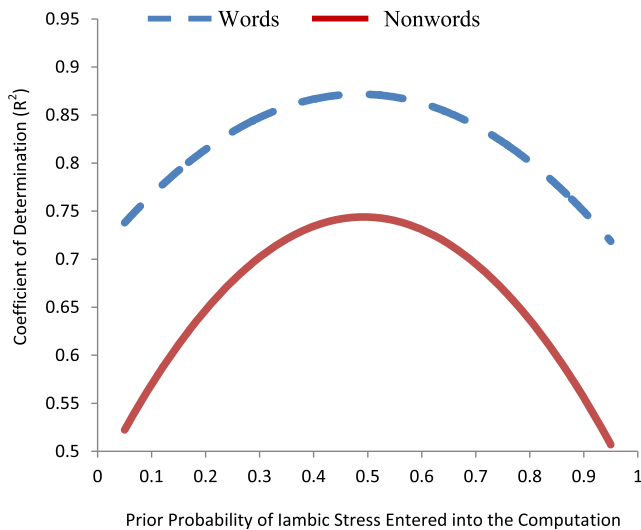


Fig. 3 Goodness of fit of predicted probabilities of stress patterns with different prior probabilities entered into the computations. The predictions are compared to the behavioral performance of readers’ assigning stress to words (striped line) and nonwords (solid line). The predicted probabilities with correct priors (.55 probability of trochaic stress and .45 probability of iambic stress) essentially provide the best fit to the data

probabilities of stress patterns that were highly predictive of actual performance in Experiments 1 and 2 were computed based on three non-lexical sources of evidence for stress (CVC1, CVC2, and VC2). To assess whether all three sources of evidence for stress are, indeed, needed to achieve accurate predictions of stress pattern assignments, posterior probabilities of stress patterns for the words used in Experiment 1 and the nonwords used in Experiment 2 were calculated given (a) only what appears to be the most potent non-lexical cue (CVC1) (Jouravlev & Lupker, *submitted*), (b) what appear to be the two most potent non-lexical cues (CVC1 and CVC2), and, finally, (c) all three non-lexical cues (CVC1, CVC2, and VC2). The comparison of the three sets of posterior probabilities with the actual performance of readers showed that the posterior probabilities of stress patterns with all three sources of evidence for stress provides a better fit to the behavioral data than did posterior probabilities based on the other two sets of cues examined (see Fig. 4).

One additional theoretical point to make is that the framework we have proposed here does not offer a specific processing mechanism for turning calculated probabilities into outputs. What mechanism would allow a word with a predicted probability of trochaic stress of .90 to be assigned trochaic stress more frequently than a word with a predicted probability of .70 with neither word being assigned the same (i.e., trochaic) stress in every instance? In fact, there would seem to be a number of such mechanisms in the literature (e.g., mechanisms that are based on the principles of Luce’s (1950) choice axiom) that might allow this goal to be accomplished. Similarly, one could propose that the mechanism is essentially a random walk process, as in Ratcliff, Gómez and McKoon’s (2004) Diffusion model, with higher posterior probabilities producing higher drift rates. Future research

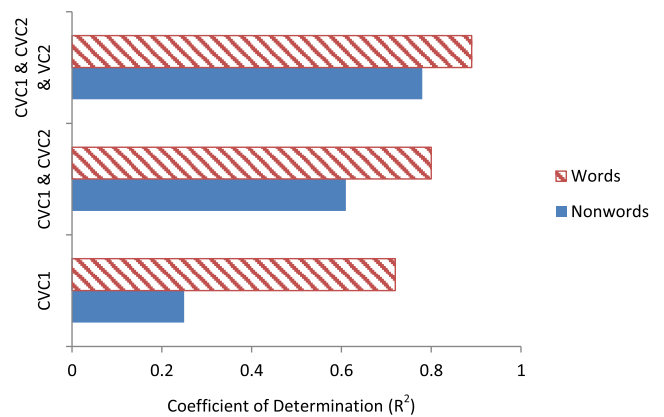


Fig. 4 Goodness of fit of predicted posterior probabilities of stress patterns with different sets of sources of evidence for stress entered into the computations: CVC1 only, CVC1 and CVC2, and CVC1, CVC2, and VC2. The predictions are compared to the behavioral performance of readers’ assigning stress to words (diagonal lines) and nonwords (solid fill). The predicted posterior probabilities of stress patterns based on all three sources of evidence for stress (CVC1, CVC2, and VC2) provide the best fit to the data

should help to clarify which type of mechanism best reflects the process of selecting a stress assignment.

To conclude, although the probabilistic inferential approach to the process of stress assignment does share some theoretical assumptions with a number of existing models of stress assignment and polysyllabic reading, it is a novel computational and theoretical approach that appears to be able to provide new insights into the process of stress assignment in, potentially, a variety of languages.

Limitations and future research

In this paper, we have demonstrated that the process of stress assignment in Russian disyllabic words can be conceived of as a problem of probabilistic inference making. This line of research can be further extended in a number of ways. The first logical step would be to test this theoretical idea and its computational implementation against stress assignment performance in languages other than Russian. It would be of particular interest to examine whether this framework can be applied to stress assignment in English and Italian, languages for which researchers have amassed an extensive amount of experimental evidence regarding the process of stress assignment and for which several models of stress assignment, built on theoretical grounds quite different from the one proposed here, exist.

The framework proposed here should also be tested on its ability to predict stress for polysyllables of various syllabic lengths. It is an open question whether the simulations of stress assignment performance of speakers naming items of more than two syllables would be as successful as that for disyllabic words and nonwords. Although the cornerstone computational principles would remain the same regardless of the syllabic length of an item, there would be some minor differences in how the process was envisioned. First, there would be differences in the number of hypotheses (stress patterns) for which posterior probabilities must be computed. Second, one would need to determine whether the sources of evidence for stress remain the same for items of various syllabic lengths or whether different cues to stress (dependent on the syllabic length of a word) are involved.

Another question concerns the relative time period during which stress pattern information is being processed and the potential interaction of segmental and suprasegmental phonological processing during reading. Although there is some behavioral evidence suggesting that the two phonological processes are distinct and might be selectively impaired in readers (Aichert & Ziegler, 2004), in some languages, segmental and suprasegmental units appear to be perfectly correlated (e.g., in English, the reduced phoneme schwa appears in unstressed syllables only) and, therefore, readers might be using segmental phonological information (like the presence of schwa) to assign stress patterns. Doing so would only be possible if the

output of segmental phonological processing is available to the processing system before readers start assigning stress to words. Indeed, logically, the mapping of orthography onto phonology should precede the process of stress pattern identification. If segmental phonological processing truly precedes suprasegmental phonological processing in reading, then, within the described probabilistic inferential framework of stress assignment, we may need to consider not only orthographic but also segmental phonological sources of evidence for stress. However, for the moment, we have not considered segmental phonology to be one of the sources of evidence for stress (in Russian) as the relative time-frame of activation of segmental versus suprasegmental phonology is still a matter of empirical investigation.

An additional issue that has not been investigated in the present research is whether the proposed framework can simulate individual differences in stress assignment performance. These differences are likely to emerge because individuals might be exposed to different statistical probabilities (both prior probabilities and likelihoods of evidence given stress patterns) even in the same language environment due to variability in the contents and sizes of individuals' lexicons. Based on our analysis of the effect of altering prior probabilities, it would be expected that any individual differences in stress assignment would mainly be driven by variability in likelihoods of evidence given stress patterns rather than by differences in the knowledge of prior probabilities of stress patterns in a language as very precise estimates of priors seem not to be required for good stress assignment behavior. On the other hand, it is possible that seemingly minor differences in individual lexicons might lead to significant differences in estimated likelihoods of evidence given stress patterns.

The idea is simply that due to differences in exposure and corresponding differences in likelihoods of evidence and, possibly to some degree, in prior probabilities, the stress assignment performance of, for example, a highly educated person might be somewhat different from the performance of a person with an impoverished lexicon. Similarly, the statistical probabilities of the lexicon of a 6-year-old child might differ significantly from those of an adult. In fact, Arciuli et al. (2010) analyzed a corpus of children's literature appropriate for various age groups, and demonstrated successive changes in the distributions of stress patterns (prior probabilities) and in reliable stress cues (likelihoods of evidence) that were picked up by children of different ages. By learning the language, children adapted their predictions to more closely reflect the actual structure of the language.

Conclusions

In the present paper, we attempted to conceptualize the process of stress assignment within the framework of probabilistic inference making. In particular, stress assignment was

viewed as the process of estimation of posterior probabilities of stress patterns. In the computation of posterior probabilities of stress patterns, prior probabilities of stress patterns reflecting the frequency of stress patterns in the language are adjusted. This adjustment is done by considering lexical and/or non-lexical sources of evidence for stress. This theoretical idea, couched in the probabilistic inferential context, and, in particular, its computational implementation were successfully tested against stress assignment performance of native speakers of Russian naming disyllabic words and nonwords. One of the greatest advantages of the probabilistic inferential approach to stress assignment is that it can provide predictions concerning differences among stimuli, that is, it can predict not only the most common stress pattern response to a word or nonword, but also the difficulty/likelihood of assigning that stress pattern to that word or nonword across various groups of participants.

Acknowledgments This research was based on a dissertation submitted by Olessia Jouravlev in fulfillment of the requirements for a PhD degree in the Department of Psychology at the University of Western Ontario and was supported by the Natural Sciences and Engineering Research Council of Canada Grant A6333 to Stephen J. Lupker. The authors would like to thank Deb Jared, Marc Joanisse, Kathy Rastle, Serje Robidoux, and an anonymous reviewer for their extremely helpful comments on earlier drafts of this work. Portions of this paper were presented at the Seventh Tucson Lexical Processing Workshop, London, Ontario, November, 2013.

References

- Adelman, J. S. (2012). Methodological issues with words. In J. S. Adelman (Ed.), *Visual word recognition: Models and methods, orthography and phonology* (pp. 116–139). Hove, East Sussex: Psychology Press.
- Aichert, I., & Ziegler, W. (2004). Segmental and metrical encoding in aphasia: Two case reports. *Aphasiology*, *18*(12), 1201–1211.
- Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, *14*(3), 471–485.
- Arciuli, J., & Cupples, L. (2006). The processing of lexical stress during visual word recognition: Typicality effects and orthographic correlates. *Quarterly Journal of Experimental Psychology*, *59*(5), 920–948.
- Arciuli, J., Monaghan, P., & Seva, N. (2010). Learning to assign lexical stress during reading aloud: Corpus, behavioural and computational investigations. *Journal of Memory and Language*, *63*(2), 180–196.
- Ashby, J., & Clifton, C. (2005). The prosodic property of lexical stress affects eye movements during silent reading. *Cognition*, *96*(3), B89–B100.
- Ashby, J., & Martin, A. E. (2008). Prosodic phonological representations early in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(1), 224–236.
- Ashby, J., & Rayner, K. (2004). Representing syllable information during silent reading: Evidence from eye movements. *Language and Cognitive Processes*, *19*(3), 391–426.
- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1995). *The CELEX lexical database. [CD-ROM: Release 2.]*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.
- Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329–349.
- Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language*, *24*(1), 89–106.
- Bates, D., & Maechler, M. (2010). *Package 'lme4'*. Reference manual for the package, available at: <http://cran.r-project.org/web/packages/lme4/lme4.pdf>
- Bayes, T. (1763/1958). Studies in the history of probability and statistics: Thomas Bayes's essay towards solving a problem in the doctrine of chances. *Biometrika*, *45*, 296–315.
- Black, M., & Byng, S. (1986). Prosodic constraints on lexical access in reading. *Cognitive Neuropsychology*, *3*(4), 369–409.
- Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, *138*(3), 389–414.
- Breen, M., & Clifton, C. (2011). Stress matters: Effects of anticipated lexical stress on silent reading. *Journal of Memory and Language*, *64*(2), 153–170.
- Burani, C., & Arduino, L. S. (2004). Stress regularity or consistency? Reading aloud Italian polysyllables with different stress patterns. *Brain and Language*, *90*(1–3), 318–325.
- Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences*, *10*(7), 335–344.
- Chetail, F., & Content, A. (2012). The internal structure of chaos: Letter category determines visual word perceptual units. *Journal of Memory and Language*, *67*(3), 371–388.
- Chetail, F., & Content, A. (2013). Segmentation of written words in French. *Language and Speech*, *56*(1), 125–144.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204.
- Coats, H. B. (1976). *Stress assignment in Russian: Inflection*. Edmonton: Linguistic Research.
- Colombo, L. (1992). Lexical stress and its interaction with frequency in word pronunciation. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 987–1003.
- Coltheart, M., Davelaar, E., Jonasson, J. F., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256.
- Dell, G. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, *5*(4), 313–349.
- Doya, K., Ishii, S., Pouget, A., & Rao, R. (2007). *Bayesian brain: Probabilistic approaches to neural coding*. Cambridge, MA: MIT Press.
- Feldman, J. (2001). Bayesian contour integration. *Perception & Psychophysics*, *63*(7), 1171–1182.
- Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 116–124.
- Gouskova, M., & Roon, K. (2013). Gradient clash, faithfulness, and sonority sequencing effects in Russian compound stress. *Laboratory Phonology*, *4*(2), 383–434.
- Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *Cambridge handbook of computational cognitive modeling* (pp. 59–100). Cambridge, England: Cambridge University Press.
- Gutiérrez-Palma, N., & Palma-Reyes, A. (2008). On the use of lexical stress in reading Spanish. *Reading and Writing*, *21*(6), 645–660.

- Halle, M. (1973). The accentuation of Russian words. *Language*, 49(2), 312–348.
- Hansson, P., Juslin, P., & Winman, A. (2008). The role of short-term memory capacity and task experience for overconfidence in judgment under uncertainty. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(5), 1027–1042.
- Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, 34(4), 169–188.
- Jouravlev, O. (2014). *A Bayesian Model of Stress Assignment in Reading* (Doctoral dissertation, The University of Western Ontario).
- Jouravlev, O., & Lupker, S. J. (submitted). Predicting stress patterns in an unpredictable stress language: The use of non-lexical sources of evidence for stress assignment in Russian.
- Jouravlev, O., & Lupker, S. J. (2014). Stress consistency and stress regularity effects in Russian. *Language, Cognition and Neuroscience*, 29(5), 605–619.
- Jurafsky, D. (2003). Probabilistic Modeling in Psycholinguistics: Linguistic Comprehension and Production. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic Linguistics* (pp. 39–97). London: MIT Press.
- Kelly, M. (2004). Word onset patterns and lexical stress in English. *Journal of Memory and Language*, 50(3), 231–244.
- Kelly, M., Morris, J., & Verrechia, L. (1998). Orthographic cues to lexical stress: Effects on naming and lexical decision. *Memory & Cognition*, 26(4), 822–832.
- Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, 55, 271–304.
- Luce, R. D. (1950). *Individual choice behaviour*. New York: Wiley.
- McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38(3), 283–312.
- Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, 118(1), 43–71.
- Movellan, J. R., & Nelson, J. D. (2001). Probabilistic functionalism: A unifying paradigm for cognitive science. *Behavioral and Brain Sciences*, 24(4), 690–692.
- Najemnik, J., & Geisler, W. S. (2009). Simple summation rule for optimal fixation selection in visual search. *Vision Research*, 49(10), 1286–1294.
- Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, 113(2), 327–357.
- Norris, D., & Kinoshita, S. (2008). Perception as evidence accumulation and Bayesian inference: Insights from masked priming. *Journal of Experimental Psychology: General*, 137(3), 434–455.
- Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115(2), 357.
- Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology*, 61(2), 106–151.
- Perry, C., Ziegler, J. C., & Zorzi, M. (2013). A computational and empirical investigation of graphemes in reading. *Cognitive Science*, 37(5), 800–828.
- Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, 68(1), 29–46.
- Pothos, E. M., & Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences*, 36(3), 255–274.
- Protopapas, A. (2006). On the usefulness of stress diacritics in reading Greek. *Reading and Writing*, 19(2), 171–198.
- Protopapas, A. (2007). CheckVocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, 39(4), 859–862.
- Rastle, K., & Coltheart, M. (2000). Lexical and nonlexical print-to-sound translation of disyllabic words and nonwords. *Journal of Memory and Language*, 42(3), 342–364.
- Ratcliff, R., Gómez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, 111(1), 159–182.
- Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35(4), 606–621.
- Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, 23(4), 569–588.
- Ševa, N., Monaghan, P., & Arciuli, J. (2009). Stressing what is important: Orthographic cues and lexical stress assignment. *Journal of Neurolinguistics*, 22(3), 237–249.
- Sulpizio, S., Arduino, L. S., Paizi, D., & Burani, C. (2013). Stress assignment in reading Italian polysyllabic pseudowords. *Journal of Experimental Psychology: Learning, Memory, Cognition*, 39(1), 51–68.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131.
- van Heuven, Walter, J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, 67(6), 1176–1190.
- Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, 114(2), 245–272.
- Zacks, R. T., & Hasher, L. (2002). Frequency processing: A twenty-five year perspective. (pp. 21–36). In R. T. Zacks & L. Hasher (Eds.), *Frequency processing and cognition*. New York, NY, US: Oxford University Press, New York, NY.
- Zaliznjak, A. A. (1985). *From Proto-Slavic to Russian accentuation*. Moscow: Nauka.
- Zevin, J. D., & Joannisse, M. (2000). Stress assignment in nonword reading. *Journal of Cognitive Neuroscience*, 41B, S5.