

© 2021, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/xhp0000929

Lexical Stress Representation in Spoken Word Recognition

Angeliki Andrikopoulou¹, Athanassios Protopapas², and Amalia Arvaniti³


¹Department of History and Philosophy of Science, National and Kapodistrian University of Athens


²Department of Special Needs Education, University of Oslo


³Department of English Language and Linguistics, University of Radboud

In Press, Journal of Experimental Psychology: Human Perception and Performance

Author Note

Angeliki Andrikopoulou  <https://orcid.org/0000-0002-9531-7144>

Athanassios Protopapas  <https://orcid.org/0000-0002-7285-8845>

Amalia Arvaniti  <https://orcid.org/0000-0002-1689-1931>

We thank SR Research support specialists Sam Hutton and Kurt Debono for assistance with the eye gaze experiments. This work was supported by the European Union (European Social Fund) and Greek national funds through the operational program “education and lifelong learning” of the National Strategic Reference Framework, research program THALIS-UOA-COGMEK (project 892, MIS 375737). Parts of these data have been presented in the 19th meeting of the European Society for Cognitive Psychology and the 5th Panhellenic Conference of Cognitive Science. The number of counted words is 13,434. Materials, data, and analysis scripts are available at <https://osf.io/nspgd/>

Correspondence should be addressed to Angeliki Andrikopoulou, IFE, Ano Ilissia University Campus, GR-157 71 Zografos, Greece. Email: ang.andrikopoulou@gmail.com

Abstract

According to a popular model of speech production, stress is underspecified in the lexicon, that is, it is specified only for words with stress patterns other than the default, termed the “default metrics” assumption. Alternatively, stress may be fully specified in the lexicon as part of every lexical representation. In the current study the two accounts are tested in the perceptual domain using behavioral and eye-tracking data in Greek. In a first experiment, cross-modal fragment priming was used in a lexical decision task. According to default metrics, priming should occur for targets with antepenultimate- or final-syllable stress but not for targets with the default penultimate-syllable stress. The same word pairs were used in two subsequent visual world experiments. Default metrics predicts an asymmetric pattern of results, namely that incoming spoken words with the default stress pattern should inhibit the activation of lexical representations with nondefault stress, whereas the converse should not be observed; that is, spoken words with nondefault stress should not inhibit representations of words with the default stress. None of the results provided support for the idea of default metrics, leading to alternative conceptualizations regarding the representation of stress.

Keywords: fragment priming; lexical stress; Greek; underspecification; eye tracking

Public significance statement

According to a popular model of speech production, our knowledge of the words in our language does not include information about how every word is stressed. Instead, stress information is only stored for words that diverge from whatever is the most common pattern in the language. In this study we conducted three experiments on recognition of spoken words in Modern Greek to test predictions based on this hypothesis and we found no evidence for it. Rather, our results are consistent with the idea that all words are represented along with their stress information. This highlights the importance of testing theoretical assumptions across perception and production and across languages.

Lexical Stress Representation in Spoken Word Recognition

In this study we are concerned with the representation of stress in the mental lexicon. The two major theoretical approaches to lexical representation suggest that words are represented in the lexicon either in the form of multiple episodic traces (Goldinger, 1998) or in phonologically abstract forms (McQueen et al., 2006). However, neither approach has considered prosodic¹ information, such as that associated with lexical stress, and how it might be represented in the lexicon along with segmental information.

Lexical stress refers to an abstract prosodic property consisting in the prominence of one syllable within a word. Its phonetic correlates typically include increased duration and amplitude (Beckman, 1986; Laver, 1994; but are not limited to these, also involving segmental quality in many languages; de Jong, 1995). In languages in which stress position can vary it can be contrastive, that is, used to differentiate lexical items (Revithiadou, 1999). In such languages, the distribution of different stress patterns may be far from uniform, with some stress patterns attested much more frequently than others. In the lexical stress literature, the term “dominant” is used to refer to the statistical fact that one stress pattern is observed more frequently than others. In contrast, terms such as “default” or “regular” are of a more theoretical nature, referring to a case that is considered special in some principled way (e.g., “regular” depends on positing a “rule”).² For simplicity purposes, in the current study we adopt

¹ The term “suprasegmental” is often used in the literature to refer to such features. Here we use the term “prosodic” instead, which has now largely replaced “suprasegmental” in the linguistic literature and avoids the “layering” metaphor implicit in “suprasegmental”.

² From a theoretical point of view, a “default” need not apply to a majority of cases. For example, only a minority of nouns conforms to the presumed default in German plural

the term “dominant”, in accordance with most of the relevant literature, and only refer to “default” when the distinction is of theoretical significance.

Stress Effects in Word Recognition

Evidence from a variety of languages indicates the importance of stress in both early and late stages of word recognition. For example, cross-modal fragment priming studies in Spanish, Italian, Dutch, and English, using lexical decision tasks, have consistently revealed effects of stress on word recognition (Cooper et al., 2002; Soto-Faraco et al., 2001; Tagliapietra and Tabossi, 2005; Van Donselaar et al., 2005). Across languages, auditory fragment primes identical to ensuing visual targets both segmentally and prosodically (e.g., PRINci- for the target PRINcipe) facilitated responses to the targets, compared to a neutral condition, whereas primes segmentally identical but differing prosodically from the targets (e.g., prinCI- from the word prinCIpio, for the target PRINcipe) either inhibited or failed to affect response latencies (these Spanish examples are from Soto-Faraco et al.).

Studies employing word spotting or lexical decision tasks without fragment priming have also indicated a role of stress in word recognition. In Dutch, Cutler and Van Donselaar (2001) found that participants were less likely to reject a spoken pseudoword when the onset matched a real word both segmentally and in terms of stress than when it only matched segmentally. In Italian, Colombo and Sulpizio (2015; Sulpizio & Colombo, 2017) found that visually presented target words with the dominant stress pattern were responded to faster and more accurately than words with nondominant stress.

Support for the role of stress in word recognition also comes from eye tracking.

inflection (Marcus, Brinkmann, Clahsen, Wiese, & Pinker, 1995). Thus it is theoretically possible for “dominant” and “default” to refer to different stress patterns.

Ashby and Clifton (2005) found that readers took longer to read words with two stressed syllables (e.g., “fundamental”) than frequency-matched words with one stressed syllable (e.g., “significant”). Shatzman and McQueen (2006), while not directly studying the effects of lexical stress, found that prior prosodic knowledge (i.e., first syllable duration) affected listeners’ eye movements to newly acquired words and proposed that abstract prosodic properties of known words affect the recognition process. Subsequently, Reinisch et al. (2009) and Sulpizio and McQueen (2012) found that the stress pattern of a spoken word affected recognition of a visual target as soon as it was available, earlier than distinguishing segmental information. Finally, Breen and Clifton (2011, 2013) found more fixations to visual target words when both syntactical and metrical reanalysis was required than when syntactic reanalysis alone sufficed.

Taken together, these findings indicate that stress information is used online during lexical access and constrains word recognition as soon as it becomes available. For this to be possible, lexical representations must contain information that can be matched against the incoming acoustic signal in terms of prosodic properties associated with stress.

The Representation of Stress

Although the effects of stress on word recognition seem well established, little is known regarding the representation of stress in the lexicon. Despite the fact that alternative experimental paradigms have been employed, from cross-modal priming to the visual world paradigm, central questions remain unanswered concerning the representation of stress: Is stress an integral part of each lexical representation or represented separately from segmental information? Is the dominant stress pattern lexically represented or assigned extra-lexically by rule?

Considering integrated representations of segmental and stress information in

the lexicon, Schild et al. (2014a) suggested that “stress might be tightly linked to phonemes both at the pre-lexical and lexical level” (p. 32). For example, they argued that an incoming spoken word with a long /u/ in the first syllable would match a pre-lexical representation with a long /u/, which would in turn match lexical representations with long /u/ in the first syllable. This would explain the facilitation observed in priming studies. Alternatively, segmental and stress information might be represented separately. In this scenario, abstract (“phoneme-free”) prosodic representations could be mapped onto lexical representations with a long vowel in their first syllable, regardless of the precise identity of the vowel. In a series of priming studies in German using Event Related Potentials (ERPs), Schild et al. (2014a, b) found no interaction between prosodic and phoneme priming, concluding in favor of phoneme-free prosodic representations.

The conceptualization of abstract prosodic representations is not new. It has also been suggested in the theory of lexical access in speech production by Levelt et al. (1999), as implemented in their speech production model WEAVER ++. Levelt et al. related lexical stress to the application of metrical frames onto word representations. They proposed that each word includes stored metrical information consisting of the number of syllables and the position of main stress. In this model, a “default metrics” assumption is made, stating that the “regular default” metrical pattern of a given language is not stored in the lexicon (p. 22), whereas other patterns are. According to Levelt et al., “regular default” stress is the most frequent stress pattern of words and is assigned by a rule during the later stages of production. In contrast, stress patterns other than the most frequent one need to be fully specified in the lexicon along with the corresponding segmental representations. In other words, one might say that in WEAVER++ the representation of stress is *underspecified*, in the sense that only

nondominant patterns are stored in the lexicon, whereas the dominant stress pattern is left unspecified (i.e., not represented in the lexicon). Underspecification contrasts with *full specification*, according to which all stress patterns are stored in the lexicon rather than only the dominant one.

There is some support for the independence of segmental and lexical stress representations from studies of speech production and reading. Specifically, studies of aphasic patients in Italian have demonstrated selective impairments in stress assignment mainly for words with less frequent stress patterns, indicating (a) separate representations of segmental and stress information and (b) representation of stress patterns only for words in which stress cannot be assigned by default (Cappa et al., 1997; Laganaro et al., 2002). Additionally, significant priming of stress patterns has been reported in production studies in Italian (Colombo & Zevin, 2009; Sulpizio et al., 2012; Sulpizio et al., 2016). In contrast, no stress priming was found in perception using lexical decision tasks in Greek (Protopapas et al., 2016) and English (Slowiaczek et al., 2006). Indeed, only fragment priming has consistently produced stress effects in perception; we are not aware of any reports of significant stress priming effects in spoken word recognition using whole words.

Beyond speech production and perception, cognitive models of reading have addressed the representation of stress (Perry et al., 2010, 2014), emphasizing the role of a sublexical mechanism underlying effects of stress regularity and stress consistency (observed in Italian: Burani and Arduino, 2004; Colombo, 1992; Sulpizio, Boureux, et al., 2012; Sulpizio & Colombo, 2013). Specifically, according to the connectionist dual process model of reading aloud (CDP++; Perry et al., 2014), stress patterns can be assigned by combining activation from both a lexical and a sublexical mechanism working in parallel. In CDP++ the most frequent stress pattern is not assigned by a rule.

Instead, all stress patterns are fully specified in the lexicon. A sublexical associative network (i.e., not involving the lexicon) supports reading of unknown words by mapping stress patterns to sequences of graphemes.

Behavioral studies of the interplay between lexical and sublexical mechanisms have tended to focus on one or the other. An almost completely lexical route of processing has often been posited for languages with varying stress position (as opposed to languages with fixed stress), based on the general assumption that the involvement of the lexicon is crucial for highly unpredictable dimensions of a language (e.g., in Russian: Lukyanenko et al., 2011). On the other hand, stress consistency effects have also been observed in languages with varying stress position, pointing toward a sublexical route of processing (in Italian and Russian: Colombo, 1992; Burani and Arduino, 2004; Jouravlev and Lupker, 2014).

In sum, the representation of stress in the mental lexicon and its relation to the segmental characteristics of words remains an open issue. Stress could be regarded as an integral part of the lexical representation (if lexically represented and mediated) or assigned by a default rule (if underspecified) or even assigned via sublexical processes not limited to a dominant pattern (as in the dual-route approach). In the current study we aim to address this issue directly by seeking evidence that can help distinguish among these alternatives by studying the effects of lexical stress in Greek.

Unfortunately, no model of spoken word recognition has directly addressed the issue of lexical stress. We have identified two partial and indirect approaches: Shortlist (Norris, 1994) incorporated the effects of stress in word activation by including a distinction between strong and weak syllables, which is only applicable to languages in which stress is associated with significant changes in segmental quality. Fine-Tracker (Scharenborg & Boves, 2010) was developed to account for prosodic effects in word

recognition by incorporating the effects of duration on word access. Thus, the effects of lexical stress per se remain largely unexplored. Inevitably, the lack of a model of spoken word recognition positing lexical stress representations leads us to extract hypotheses based on WEAVER++ (Levelt et al., 1999) and CDP++ (Perry et al., 2014), two models focusing on production and reading, respectively. Although production and perception certainly involve different processes, the assumption of a common lexicon with shared representations for both (Pickering & Garrod, 2004) seems to us sufficiently established to serve as a reasonable starting point.

The Current Study

Greek is a language that lends itself to the investigation of stress representation. Similar to Russian, Spanish and Italian, stress position in Greek can vary, constrained to the last three syllables of a word (Malikouti-Drachman & Drachman, 1989). Stress in Greek has a relatively small effect on segmental quality (Arvaniti, 2007; Fourakis et al., 1999), thus allowing us to study stress largely independently of segmental effects (in contrast to English, where prosodic cues are interdependent with segmental characteristics; Cooper et al., 2002). The Greek orthography is relatively transparent at the grapheme-phoneme level (Protopapas & Vlahou, 2009), similar to Italian. In addition—and unlike Italian—stress in Greek is obligatorily marked with a special diacritic (accent) on the vowel of the stressed syllable of every word with two or more syllables; e.g. *τραπέζι* /tra.'pe.zi/ “table”.

The acoustic correlates of stress in Greek involve longer duration and higher amplitude of the stressed syllable (as compared to the unstressed syllables of a word) and small quality changes (Arvaniti, 2000; Botinis, 1989; Fourakis et al., 1999). Linguistic studies have identified the syllabic trochee as the default metrical frame (Malikouti-Drachman, 2002; Malikouti-Drachman & Drachman, 1989). Thus the default

stress pattern corresponds to a two-syllable word-final foot stressed on its first syllable, that is, the penultimate syllable of the word. A penultimate-stress default is consistent with developmental data in Greek language acquisition (Kappa, 2002) and with experimental data from studies of reading aloud across skill levels (Protopapas & Gerakaki, 2009; Protopapas et al., 2006, 2007). Penultimate-syllable stress is also the dominant pattern. Specifically, stress on the penult is found on approximately 28% of all word tokens (a relative majority). Considering only words with two or more syllables, approximately 45% of tokens have penultimate stress, compared to 25% having stress on the antepenult and 30% having stress on the final syllable (Protopapas, 2006).

In our work we have employed experimental procedures previously used to study the effects of stress on word recognition. Specifically, Experiment 1 involved a lexical decision task in a fragment priming paradigm, aiming to replicate findings in other languages consistent with the contribution of stress to lexical activation. However, lexical decision tasks cannot inform about the online processing of stress or the time course of competition between critical words. Therefore, Experiments 2 and 3 employed eye-tracking in a visual world paradigm to address these issues. All three experiments were designed to reveal the asymmetric effects predicted by underspecification (i.e., default metrics, in the WEAVER++ terminology).

Experiment 1

In the first experiment we tested whether disyllabic word fragments would prime segmentally and prosodically identical targets (match condition; e.g., κέρα/'ce.ra/ – κέρασε/'ce.ra.se/) and/or inhibit targets that are segmentally identical but differing in stress (mismatch condition; e.g., κέρα/'ce.ra/ – κέράσι /ce.'ra.si/). We expected to replicate fragment priming experiments in Spanish (Soto-Faraco et al., 2001) and Dutch (Van Donselaar et al., 2005). Specifically, we expected facilitation for matching word

targets and inhibition for mismatching word targets. Under the hypothesis of full specification of lexical stress, these priming patterns should not depend on the stress pattern of the target, because stressed and unstressed syllables would match or mismatch regardless of the metrical pattern to which they belong. However, according to underspecification, no stress priming should be observed for word targets with a dominant stress pattern because there is no stress-related representation in the lexicon to match or mismatch the prime. In contrast, priming should be evident for word targets with stress patterns other than the dominant, because they are fully represented in the lexicon and can be compared against prime stress patterns. Thus, Experiment 1, we used a lexical decision task to test whether priming effects can be observed for word targets with stress on the antepenultimate or final syllable, which are fully specified in the lexicon (i.e., including stress pattern information), but not for targets with penultimate-syllable stress, because their (dominant) stress pattern is not part of the words' lexical representation.

Furthermore, pseudoword prime and target pairs were also included. Obviously, pseudowords have no pre-existing representation in the lexicon to match or mismatch the input, either segmentally or prosodically. Thus they afford testing of theoretical hypotheses regarding sublexical mechanisms of stress processing. Specifically, pseudowords can help disambiguate any priming effects found with words, which could conceivably have a lexical or sublexical processing origin. Beyond theoretical proposals (such as the aforementioned CDP++), there is also empirical precedent: Specifically, in a series of reading aloud experiments, Colombo and Zevin (2009) reported stress priming in pseudowords with both penultimate and antepenultimate stress, even when stimuli shared minimal segmental similarity. If a sublexical route of processing is involved in assigning stress to specific grapheme-phoneme correspondences in reading (aloud), it

could also be involved in the online processing of stress patterns in spoken word recognition. In that case, stress priming should also be evident in pseudowords.

Method

Participants. Eighty-eight adults (18–35 years old) participated in this experiment. Most were undergraduate students at the University of Athens or Panteion University and received class credit for participation. All were native Greek speakers with normal or corrected-to-normal vision and no reported hearing or reading difficulties. This sample size is comparable to that of earlier studies (43/51/96 participants in the three experiments of Soto-Faraco et al., 2001; 62/60/80 participants in the three experiments of Van Donselaar et al., 2005).

When it comes to the interaction of stress pattern by stress congruence, the effect size we can reasonably expect to be able to detect can be estimated from the only previous fragment-priming experiment in Greek (Experiment 5 in Protopapas, Panagaki, et al., 2016), which produced a 50-ms congruence effect. Retaining the observed variance-covariance matrix of that 2×2 design (stress congruence by stress position) and varying the mean difference of differences in steps of 2.5 ms in a simulation with 10,000 random samples of size 80 at each step, we found 90% power to detect a reduction of the congruence effect by 20 ms or an increase by 27.5 ms, i.e., about half of the original full effect. This compares favorably with the prediction of a complete elimination of the effect in the case of penultimate-stress targets.

Materials. The stimuli were derived from the C corpus of the ILSP Psycholinguistic Resource (IPLR; speech.ilsp.gr/iplr; Protopapas et al., 2012). Seventy-two critical pairs of words were constructed, sixty-nine of them matching segmentally in their first two syllables and up to the onset of the third syllable but not in stress, and three pairs matching segmentally in all three syllables and differing only in

stress. The two words in each pair were semantically unrelated and were used as both targets and primes. Primes consisted of the first two syllables of each word (disyllabic fragments). None of the primes constituted a real Greek word. Target stress position was equally distributed among the final, penultimate, and antepenultimate syllable.

The 72 critical pairs formed 3 stress contrast sets: 24 of the pairs had one member stressed on the final syllable and the other on the penultimate (e.g., γενναίων /je.'ne.on/ “brave_{M.GEN.PL}” – γενεών /je.ne.'on/ “generation_{GEN.PL}”); 24 had words with penultimate and antepenultimate stress (e.g., κέρασε /'ce.ra.se/ “treat_{3RD.SG.PAST.FIN}” – κέρασι /ce.'ra.si/ “cherry_{NOM.SG}”), and 24 with final and antepenultimate stress (e.g., άπορους /'a.po.rus/ “poor_{M.ACC.PL}” – απορώ /a.po.'ro/ “wonder_{1ST.SG.PRES}”). Word groups stressed on each syllable were matched on mean log frequency and number of letters (based on IPLR). In addition, 144 words were chosen to serve as neutral primes, matched to the critical pairs on mean log frequency, number of letters, and number of syllables but sharing no phonological, orthographic, or stress pattern overlap, and bearing no semantic relationship with them.

For each target (e.g., κέρασι /ce'rasi/ “cherry_{NOM.SG}”), two-syllable fragment primes fell into one of three stress congruence conditions: (a) match, in which the fragment was segmentally and prosodically identical to the first two syllables of the target (e.g., /ce'ra/ – κέρασι), (b) mismatch, in which the fragment was segmentally identical but different in stress pattern (e.g., /'cera/ – κέρασι), and (c) a neutral condition, in which the fragment differed from the target, both segmentally and in stress pattern (e.g., /sa'la/ – κέρασι). Three experimental lists were created, in each of which every target appeared once. The three stress congruence conditions were equally distributed over the three lists.

Seventy-two pairs of pseudowords were created, patterned after the critical

words, with the same number of letters and syllables. Every pair of pseudowords were segmentally identical in their first two syllables (and in some cases in the third syllable as well) and differed in stress pattern. Pseudoword fragments served as primes in three conditions, as for the words, divided over the three experimental lists. Finally, there were 10 practice prime-target pairs (five including word targets and five pseudoword targets) and an extra set of 24 filler pairs (12 word–pseudoword pairs and 12 word–word pairs) with no segmental or prosodic overlap and sharing no semantic relation.

All words and pseudowords were recorded by a female native speaker of Greek in the context of neutral sentences (not biasing toward the meaning of the target word) and ended with the same irrelevant sentence-final word (i.e., “now you will hear... clearly”). This frame puts all the test words in focal position, so they will be clearly pronounced and carry a rising pitch accent on the stressed syllable (Arvaniti & Baltazani, 2005). The sentences were digitized and cutoff points at the end of the second syllable of the prime item were established using Praat (Boersma & Weenink, 2007). Table 1 presents acoustic measures of the word stimuli used in the experiment.

Complete lists of materials and experimental scripts can be found at <https://osf.io/nspgd/>.

Procedure. Participants were tested individually in a quiet, dimly illuminated room. Auditory stimuli were delivered over headphones and visual stimuli were displayed on a 15.5” LCD computer screen using DMDX (Forster & Forster, 2003).

Each trial began with a fixation cross at the center of the screen for 250 ms, followed by a blank screen for 250 ms. Immediately thereafter, the carrier sentence ending with the auditory fragment prime was heard while dashes (“-----”) were displayed on the screen, followed immediately by the visual target (word or pseudoword) presented centered in the middle of the screen in lowercase white 20-pt

Times New Roman on black background for 2000 ms or until response. The intertrial interval was 1000 ms.

Participants were instructed to listen to the auditory sentences and respond to the words on the screen as fast and as accurately as possible, using two keys on the keyboard for words and pseudowords. The key for words was always assigned to the participant's dominant hand. The entire experimental procedure lasted about 20 minutes, with a break halfway.

Results

Based on the reported exclusion criteria in Soto-Faraco et al. (2001), items and/or participants with error rates higher than 15% were excluded from further analysis. Fourteen (out of 88) participants were thus excluded,³ leaving data from 74 participants for further analysis. Response times (RTs) lower than 250 ms (0.01%) or greater than 2000 ms (none) were also excluded. The total proportion of incorrect responses (falsely responding for word or pseudoword) was 7.53%. Table 2 shows the mean RTs and error rates as a function of stress congruence (three levels: match, mismatch, and neutral) for word and pseudoword targets.

RTs, for correct responses only, were logarithmically transformed and analyzed with general linear mixed-effects models with crossed random effects for participants and items (Baayen, Davidson, and Bates, 2008) using function `lmer` of the `lme4` package (version 1.1-21; Bates, Maechler, Bolker, & Walker, 2015) in R 3.6.2 (R Core Team, 2019). The models included fixed effects of stress congruence, trial order (centered) and RTs in the immediately preceding trial (centered), as well as random slopes for

³ Alternative analyses following a more lax exclusion criterion (<30% errors, only 4 participants excluded) are available on OSF (<https://osf.io/nspgd/>) along with the analyses reported here. Analysis of this more inclusive dataset did not deviate from the reported pattern of results in any substantial way.

stress congruence, per item, and for order, per participant. Follow-up models included the stress pattern of the target interacting with stress congruence. In our analyses we started off with maximal random structures but ended up with simplified random structures after more complex models failed to converge or resulted in singular fits, consistent with overparameterization.

Accuracy was analyzed using generalized linear mixed-effects models, with function `glmer` of the `lme4` package for binomial data. The model formula included fixed effects of stress congruence. Random slopes were excluded due to convergence problems. Follow-up models included an interaction with stress pattern.

For words, stress congruence conditions differed significantly in both accuracy and latency. More specifically, there was a significant effect of stress congruence on RTs, with the match condition ($\beta = -0.059, t = -7.528, p < .001$) as well as the mismatch condition ($\beta = 0.051, t = 7.404, p < .001$) differing significantly from the neutral one, with fastest RTs observed in the match condition and slowest in the mismatch condition. In accuracy, the neutral condition also differed significantly from the match ($\beta = -0.300, z = -3.480, p < .001$) and from the mismatch condition ($\beta = 0.301, z = 3.771, p < .001$), in which most errors were observed.

The interaction between stress congruence and stress pattern was (barely) significant in latencies only for final-stressed targets ($\beta = 0.039, t = 2.040, p = .043$), which differed from penultimate-stressed targets in match vs. neutral condition, but not for antepenultimate-stressed targets ($\beta = 0.023, t = 1.206, p = .230$). No interaction was significant for mismatch vs. neutral (final-stressed targets vs. penultimate: $\beta = 0.012, t = 0.723, p = .471$; antepenultimate-stressed targets vs. penultimate: $\beta = 0.029, t = 1.716, p = .089$). In accuracy, the interaction between stress congruence and stress pattern was not significant in match vs. neutral (final-stressed targets vs. penultimate: $\beta = -0.225, z$

= -0.976, $p = .329$; antepenultimate-stressed targets vs. penultimate: $\beta = 0.003$, $z = -0.014$, $p = .989$), whereas in mismatch vs. neutral the interaction was significant for antepenultimate-stressed targets vs. penultimate ($\beta = 0.384$, $z = 2.098$, $p = .036$) but not for the final-stressed targets vs. penultimate ($\beta = 0.097$, $z = 0.463$, $p = .643$).

A different pattern was observed for pseudowords. RTs differed significantly between the neutral condition and the match ($\beta = -0.049$, $t = -7.946$, $p < .001$) and mismatch ($\beta = -0.028$, $t = -5.035$, $p < .001$) conditions. Here, the fastest RTs were observed in the match condition whereas the slowest were observed in the neutral condition. In accuracy, there was no significant difference in errors between the neutral condition and the match ($\beta = -0.157$, $z = -1.082$, $p = 0.279$) or the mismatch ($\beta = -0.092$, $z = -0.643$, $p = 0.520$) condition.

The interaction between stress congruence and stress pattern was not significant either in the match condition vs. neutral (final-stressed targets vs. penultimate: $\beta = -0.007$, $t = -0.484$, $p = .629$; antepenultimate-stressed targets vs. penultimate: $\beta = -0.012$, $t = -0.800$, $p = .425$) or in the mismatch vs. neutral (final-stressed targets: $\beta = -0.022$, $t = -1.642$, $p = .103$; antepenultimate-stressed targets: $\beta = -0.013$, $t = -0.951$, $p = .343$). In accuracy, the interaction was not significant for the match condition vs. neutral (final-stressed targets vs. penultimate: $\beta = 0.379$, $z = 1.052$, $p = .293$; antepenultimate-stressed targets vs. penultimate: $\beta = 0.716$, $z = 1.014$, $p = .056$). In the mismatch condition vs. neutral the interaction was significant for final-stressed vs. penultimate ($\beta = -0.884$, $z = -2.516$, $p = .012$) but not for antepenultimate-stressed targets vs. penultimate ($\beta = -0.103$, $z = -0.300$, $p = .764$).

Figure 1 shows the model estimates (and associated 95% CI) for latency (RT) and accuracy (errors) for word and pseudoword targets with each stress pattern.

The aforementioned analyses concern differences from the neutral condition,

which include both segmental and stress effects. To home in specifically on the stress effects, separate analyses were conducted focusing on differences between match and mismatch conditions, separately for each target stress pattern. Table 3 presents RTs per stress position for word and pseudoword targets in the match and mismatch conditions, which concern segmentally identical fragments. The models included fixed effects of stress position of prime, trial order (centered) and RTs in the immediately preceding trial (centered), as well as random slopes for stress position, per item, and for order, per participant. Further simplifications of the random structure proved necessary for some of the models to converge.

Starting with words, RTs to targets with penultimate-syllable stress differed significantly between the match condition (i.e. preceded by fragments stressed on their penultimate syllable) and the mismatch conditions (fragments stressed on the antepenultimate: $\beta = 0.130, t = 7.459, p < .001$; fragments with two unstressed syllables: $\beta = 0.103, t = 4.671, p < .001$). Similarly, RTs to targets stressed on their final syllable were significantly faster after matching than mismatching fragments (stressed on the antepenultimate syllable: $\beta = 0.087, t = 4.799, p < .001$; stressed on the penultimate syllable: $\beta = 0.100, t = 6.399, p < .001$). The same pattern also appeared for targets stressed on the antepenultimate syllable, where RTs were faster in the match condition than in the mismatch condition (second-syllable-stress fragments: $\beta = 0.110, t = 5.091, p < .001$; fragments with two unstressed syllables: $\beta = 0.134, t = 9.963, p < .001$).

Turning to pseudowords, RTs to penultimate-syllable-stress targets differed significantly between the match and the mismatch conditions (antepenultimate-stress fragments: $\beta = 0.031, t = 2.862, p = .005$; unstressed fragments: $\beta = 0.023, t = 2.140, p = .044$). In contrast, RTs to final-syllable-stress targets did not differ significantly between the match and mismatch conditions (penultimate-syllable-stress fragments: $\beta = -0.003,$

$t = -0.233, p = 0.816$; antepenultimate-syllable-stress fragments: $\beta = 0.020, t = 1.615, p = .118$). Finally, RTs to the antepenultimate-syllable-stress targets differed significantly between the match and mismatch condition for fragments with two unstressed syllables ($\beta = 0.041, t = 3.229, p = .003$) but not for fragments stressed on their second syllable ($\beta = 0.009, t = 0.673, p = .505$).

As a final step toward evaluating potential differences in stress priming across stress patterns, we calculated the Bayes factor for a model with the interaction in comparison to a model with no interaction using package BayesFactor (Morey & Rouder, 2018). Models of RT to word targets in the match and mismatch conditions were fit by function `lmBF`, including fixed effects of target stress pattern, stress congruence, trial order, and RTs in the immediately preceding trial, as well as random intercepts for participants and items (random slopes are not supported). The result was 0.1, indicating moderate to strong support for the null hypothesis of no interaction.

Discussion

The results of Experiment 1 replicated previous findings in Spanish (Soto-Faraco et al., 2001) and Dutch (Van Donselaar et al., 2005). Specifically, for word targets, facilitation was observed in the match condition and inhibition in the mismatch condition, compared to the neutral condition. Focusing on the stress differences alone, significant priming effects were evident in every combination of stress patterns in targets and primes.

A different pattern was observed for pseudoword pairs, where facilitation was evident in both the match and mismatch conditions, relative to the neutral condition, a finding consistent with segmental priming. However, stress-specific priming (facilitation) was also evident in penultimate-syllable-stress targets, and partly in antepenultimate- (but not in final-) syllable-stress targets. This pattern suggests the

involvement of a sublexical mechanism for nonword stimuli, whereby transient representations based on the incoming signal carry stress information that can affect the processing of accent-bearing visual stimuli to some extent. This effect was much smaller than the sizeable segmental priming observed with the pseudowords, and dwarfed by the—ostensibly lexically based—stress-specific effect observed in words, which evidently overcame the segmental match in the stress-mismatch condition to result in significant inhibition compared to the neutral condition.

Most importantly, the results provide no support for the default metrics assumption (i.e., underspecification). Specifically, there was no evidence consistent with the predicted asymmetries in priming effects as a function of the target's stress pattern. On the contrary, substantial stress priming was observed both in targets with nondominant stress patterns (i.e. antepenultimate, and final stress), as well as in targets bearing the dominant pattern (i.e. penultimate stress targets).

Experiment 2

Although the results from the lexical decision task failed to confirm our prediction based on default metrics, a more sensitive paradigm tracking the gradual accumulation of activation in support of specific lexical candidates might allow us to observe fine differences among stress patterns and their forms of representation in the lexicon. In our second experiment, we first sought to replicate findings from Reinisch et al. (2009) and Sulpizio & McQueen (2012), that is, that lexical stress affects online processing of words as soon as it becomes available in the acoustic signal, even before the words differ segmentally from each other.

The specific predictions for this (and the following) experiment were derived from a rationale modeled on the approach of Mitterer (2011). Mitterer compared an optimal-perception account with the theory of a Featurally Underspecified Lexicon

(FUL; Lahiri & Reetz, 2010). According to FUL, (a) some features of the acoustic signal may not be part of the lexical representation (but can be generated by general or morpheme-specific production rules), and (b) the most predictable features are not specified in the lexicon. As the incoming signal is parsed into features, a mapping process matches these features to those stored in the lexicon, with three possible outcomes: (a) match (when a feature from the signal is identical to the one in the lexicon), (b) mismatch (when features from the signal and from the lexicon are mutually incompatible), and (c) a no-mismatch (when a feature from the signal cannot match but also fails to mismatch features in the lexicon, because the feature is not specified there).

These three alternative outcomes of the mapping process lead to important asymmetries, since a mismatching input can activate lexical representations not fully specified in the lexicon (by forming a “no-mismatch” relation), whereas fully represented characteristics would fail to be activated by a mismatching input (because they would form a clear “mismatch”). Mitterer (2011) argued that these asymmetries invoked by the “match”, “mismatch”, and “no -mismatch” relations of the acoustic input to the lexical entries should be reflected in the competition observed in a visual world paradigm, with more looks to the competitor when the target is fully specified and the competitor is unspecified and fewer looks to the competitor in the reverse condition.

It should be made clear that the FUL model concerns phonological distinctive features and says nothing about stress. Indeed, there can be no phonological theory of stress underspecification, because from a linguistic point of view stress is not a feature associated with any particular segment but, rather, a relational property concerning relative prominence, as a linguistic manifestation of rhythmic structure (Hayes, 1995). Thus, by adopting Mitterer’s (2011) rationale we do not in any way purport to test a

linguistic theory of stress or the FUL model. Instead, our concerns are strictly of a psycholinguistic nature: as noted above, prosodic aspects of speech that are associated with lexical stress seem to be involved in the online uptake of information during speech perception and to bias the activation and selection of lexical candidates as soon as they become available. Therefore, it seems reasonable to probe the nature of lexical representations to determine what enables the relevant stress-related computations.

In other words, from a psycholinguistic point of view we need a theory of what—and how—features associated with lexical stress are represented in the lexicon so that their online uptake and utilization can be supported. Given that WEAVER++, a major model of speech production, clearly posits what essentially amounts to an underspecified representation of lexical stress, we have adopted and extended the rationale of Mitterer (2011) by applying it to word stress. To achieve this, we hypothesize that stress is lexically represented by reference to prosodic features for processing purposes. With this linking assumption we can obtain clear predictions on the basis of a default metrics (underspecification) hypothesis for stress. Hence, we predicted asymmetries in the competition between targets with dominant stress and competitors with nondominant stress patterns, on the one hand, and between targets with nondominant stress and competitors with dominant stress, on the other.

In particular, we predicted that incoming spoken words with nondominant stress patterns (i.e. antepenultimate and final stress) would constitute a clear “match” with the corresponding lexical representations and the fully specified stress patterns therein. These incoming words would stand in a “no-mismatch” relation to lexical entries with the dominant stress pattern, which is not specified in the lexicon, leading to more competition among target and competitor, hence smaller differences in the corresponding number of fixations. In contrast, spoken words with the dominant stress

pattern would constitute a clear “mismatch” to the lexical entries of competitors with antepenultimate or final stress, leading to overall less competition, hence greater difference in looks to targets and competitors compared to the aforementioned case.

In comparison, the full specification hypothesis leads to the prediction of no differences in competition among targets and distractors on the basis of their stress patterns, as the prosodic features of stressed and unstressed syllables are always present in the signal and specified in the lexicon, producing matches and mismatches accordingly.

Method

Participants. Thirty-seven adults (18–35 years old) participated in this experiment. Most were undergraduate students at the University of Athens or Panteion University and received class credit for participation. All were native Greek speakers with normal or corrected-to-normal vision and no reported hearing or reading difficulties. The sample size is larger than that of earlier studies (24 participants in Reinisch et al., 2009; 32/22 in the two experiments of Sulpizio & McQueen, 2012, and 24/24 in Mitterer, 2011).

The size of effects we can reasonably expect to detect based on the literature is difficult to determine as there are no previous comparable studies in Greek. A reanalysis of the data from Experiment 2 (in Italian) of Sulpizio and McQueen (2012; data kindly provided by Simone Sulpizio) produced a 95% confidence interval for the difference between target and distractor looks of about .45 when modeled in a way similar to our analyses (see below). Adjusted for 37 participants (rather than the original 22), this amounts to a detectable difference of about 7% in proportion of looks to target around the point of 25% target looks (before curves begin to diverge), or about 9% around the point of 50% target looks (when the target becomes dominant). This turned out to be a

gross underestimate, further underscoring the difficulty of obtaining reliable expectations regarding power across languages, samples, items, and data screening and clean-up protocols.

Material. For maximum comparability a subset of the word pairs used in the previous experiment was used here as well. Specifically, 30 critical pairs of trisyllabic words were selected. Each consisted of words matching segmentally in their first two syllables but differing in their third syllable and in their stress pattern, with the exception of three pairs where words matched segmentally in all three syllables and differed only in stress. Of the 30 critical pairs, ten consisted of words stressed on the antepenultimate and the final syllable; ten consisted of words stressed on the antepenultimate and the penultimate syllable; and ten consisted of words stressed on the penultimate and on the final syllable. As all words were trisyllabic, the antepenultimate was always the first syllable, the penultimate the second, and the final the third syllable.

Every word in each pair appeared in the experiment both as a competitor and a target, leading to six conditions with respect to the contrast between the stress patterns of target and competitor: (a) antepenultimate stress target–penultimate stress competitor (antT_penC), (b) antepenultimate target–final competitor (antT_finC), (c) penultimate target–antepenultimate competitor (penT_antC), (d) penultimate target–final competitor (penT_finC), (e) final target–antepenultimate competitor (finT_antC), and (f) final target–penultimate competitor (finT_penC). As a consequence, we could examine potential differences for every possible stress combination between target and competitor. Table 4 presents the conditions with word examples and matching relations to illustrate the rationale and theoretical predictions under the underspecification hypothesis.

Thirty filler pairs were also included, consisting of trisyllabic pairs of words overlapping phonologically in their first two syllables but differing orthographically in every syllable as well as in their stress patterns. Twelve additional pairs were provided for practice.

The auditory stimuli were based on the same recordings as in the previous experiment. For the purposes of this experiment digitized sentences were further processed and cutoff points were established at the end of the target word using Praat (Boersma & Weenink, 2007). Thus, the entire word was included in this experiment (rather than only the first two syllables, as in Experiment 1). Additional marks were placed at the beginning of the target word and at the beginning of its second and third syllables, to be used for temporal alignment with eye movement metrics.

Procedure. Participants were seated at a distance of approximately 90 cm in front of a 21.5" LCD computer screen. Eye movements were recorded monocularly (right eye) using a desktop-mount EyeLink 1000 Plus eye-tracker (SR Research Ltd) at a sampling rate of 1000 Hz.

The experimental procedure consisted of 3 phases: familiarization, practice, and the main experimental procedure. In familiarization, all words appeared on screen in lower case, one by one, and participants had to read each word aloud. No feedback was given. The eye-tracking experiment followed immediately thereafter. Practice consisted of 12 trials. The main experimental procedure consisted of 4 blocks of 30 trials each. The same one of the four blocks always appeared first, while the three remaining blocks were counterbalanced across participants. The first block contained targets and competitors from the critical and filler pairs with antepenultimate, penultimate, or final stress. Targets in the remaining three blocks could be words that also appeared as targets in the first block, their competitors, fillers, or words appearing as targets for the

first time. As a consequence, participants could not predict which word was going to be the target in any given trial. Trials within blocks appeared in a random order.

In every trial, a fixation dot appeared for 500 ms, followed by a blank screen for 200 ms. Then, four words appeared on screen (including the target, the competitor and two filler words) and remained for 5000 ms or until a mouse click. The 4 words were displayed in 30-pt light grey Verdana normal font, on black background. Each word was centered in one of one of the four quadrants of the screen. The auditory stimuli (including the carrier sentence ending with the target word, e.g., “Now you will hear ...”) were presented via headphones, timed such that the onset of the target word occurred 1200 ms after the appearance of the words on screen. Following the mouse click or timeout period, a blank screen was displayed for 480 ms, serving as an inter-trial interval. A drift check was performed at every fifth trial.

Participants were instructed to listen to the auditory sentences and click on the last word they heard. The entire experimental procedure, including eye tracking calibration and validation, lasted approximately thirty minutes.

Data Analysis. In the past, studies using the visual world paradigm have typically employed methods such as ANOVA or linear mixed-effects models to analyze fixation proportions to targets vs. competitors summed over large time windows such as first syllable, first and a half syllable, and second syllable, after correcting for saccade planning latency by an offset of 200 ms. Fixation proportions have been independently compared within each time window (Reinisch et al., 2009; Sulpizio & McQueen, 2012), despite the obvious nonindependence of the partially overlapping data. Thus, techniques typically used earlier have failed to capture the essential nature of the data in a visual world paradigm, that is, their continuous change over time.

Alternative statistical approaches have been proposed that are more suitable for

time series data and their nonlinear dynamics (i.e., Barr, 2008; Mirman et al., 2008). However, it is only recently that more appropriate methods for this type of data have been made available, namely generalized additive mixed models (Baayen et al., 2017; Wood, 2006) and corresponding analysis packages and workflows for eye gaze data (Poretta et al., 2017). Generalized additive mixed models (GAMMs) are regression models that, in addition to the standard linear predictors, can also include “smooth” terms to model arbitrary curvilinear relationships between the dependent variable and one or more predictors. Smooth terms are built using piecewise polynomials (splines) and are fit by controlling the degree of smoothness. GAMMs have the advantage of not requiring aggregation of continuous data or pre-specification of time windows for analysis, thus enabling observation of patterns in dynamic time series data (Wieling, 2018), while at the same time accounting for random effects of subjects and items.

Besides the obvious difference from linear models in accounting for nonlinear effects, GAMMs present novel challenges in significance testing for the smooth terms, because what is significant in the model (i.e., whether a term departs from a straight line) is not necessarily what the researcher is interested in (i.e., whether two curves differ at some point in time). Thus a combination of approaches is used, including summary statistics for smooth terms, visualization of the smooth terms and their differences (i.e., difference plots), as well as multiple model comparisons. There are no explicit corrections for multiple comparisons; however, testing alternative combinations of predictors in each model along with the specification of parameters, defines the final choice of model (for a more detailed overview, also see van Rij, 2015 ; Wieling, 2018).

Taking advantage of these developments, the sample report produced by DataViewer (SR Research) was preprocessed with the VWPre package (Porretta et al.,

2016) and then submitted to analysis via GAMM and associated visualization tools as implemented in the R packages *mgcv* (Wood, 2006, 2011) and *itsadug* (van Rij et al., 2015) to track and compare relative fixation proportions between targets and competitors over time. Fixations were assigned to a word when falling within a 9.1-cm-side square centered on the middle of the word.

Results

Only trials in which participants had responded correctly (by clicking on the target word) were selected for further analysis. The total proportion of correct responses was 87.7%.⁴ Five items were excluded due to experimenter error causing display of the wrong target. Table 5 presents response time and accuracy data per condition. Collected response times were not further analyzed because (a) they included mouse movement to reach the response area prior to clicking, adding temporal noise in the process; and (b) occasions were reported in which the mouse would temporarily freeze on screen, making any recorded response susceptible to error. Figure 2 illustrates the visual world dynamics of lexical competition, displaying raw fixation proportions for targets, competitors, and fillers in each stress contrast condition averaged over 20-ms bins.

Our analysis focused on three main aspects. First, we aimed to identify the point in time when fixations to targets differed significantly from fixations to competitors for each stress contrast condition. Second, we tested whether competitor stress pattern affected fixations to a given target. Last, and most important for the evaluation of underspecification, we searched for asymmetries in the competition between stress

⁴ This is a higher proportion of errors than we expected, however we are unable to provide further insight on the matter as participants did not report facing particular difficulties upon debriefing.

conditions involving the dominant vs. other stress patterns. For the purposes of the current analysis, competition was defined as difference in looks to targets relative to competitors in each stress condition. The comparison of competition between two stress type conditions was defined as the difference between the curves of relative looks proportions over time for differently stressed targets.

Multiple model comparisons, along with their corresponding visualizations, were conducted for the selection of a model with an adequate number of parameters that would best fit the data. The selected model included fixed effects of stress pattern (coding the six contrasting conditions), a smooth term of trial order (to capture long-term trends), a smooth term of within-trial time (to capture the overall dynamic of relative looks to target vs. competitor), as well as an interaction of time and stress pattern, modeled by six condition-specific time smooths to capture any differences in target looks dynamics associated with stress pattern contrasts. Random effects terms included random smooths of within-trial time per participant and per item, to capture individual differences in lexical competition dynamics. Complete model results, along with all data and analysis scripts and additional graphs, can be found at <https://osf.io/nspgd/>.

Figure 3 displays the relative proportions of looks to target vs. distractor for each condition. In every condition, the difference in looks proportions between targets and competitors became significant within the second syllable of the target words (adjusted by 200 ms to account for oculomotor programming), that is, before segmental disambiguation of target and competitor. Direct comparison of target-competitor relative proportions curves between conditions (i.e., target-competitor stress pattern combination) revealed no significant differences for any stress pattern. Indeed, there was no evidence to suggest that the stress pattern of the competitor influenced the time

course of looks to the target, consistent with disambiguation based solely on the acoustic input matching the target—and mismatching the distractor.

The main goal of this experiment was to examine differences in competition between target and competitor with respect to predictions stemming from the underspecification hypothesis. This involves comparing the relative looks proportions to target vs. competitor for pairs of stress conditions contrasting target and competitor stress; that is, antT_penC vs. penT_antC, penT_finC vs. finT_penC, and antT_finC vs. finT_penC. Underspecification considerations (see Table 4) led to the prediction of differences for the first two of these, which involve a dominant-nondominant asymmetry, but no difference in the third one, which involves only nondominant patterns.

Figure 4 displays the difference plots of the summed effects corresponding to these pairs of stress conditions. There were no significant differences between the curves in any critical pair, as all differences remained very close to zero throughout the trial duration. Beyond the lack of statistical significance, there does not appear to be any trend distinguishing the pairs involving dominant stress from the pair not involving it in any way. This pattern of results is consistent with no difference in competition dynamics between conditions that were predicted to differ crucially in matching status (“mismatch” vs. “no mismatch”) according to underspecification.

Notably, the 95% confidence interval for comparisons between pairs of conditions ranged between .74 and .78 logits, suggesting that we can detect a difference of $\pm 18\%$ over a 50% reference. This was much larger than the .45 computed above for Experiment 2 in Sulpizio and McQueen (2012) (leading us to expect to be able to detect a difference of $\pm 9\%$). This deviation may be due to methodological differences between the studies, for example the fact that Sulpizio and McQueen used novel words in their

Experiment 2, as opposed to the familiar words we used in the present study.

Discussion

These results replicate and extend previous findings using the visual world paradigm in Dutch and Italian (Reinisch et al., 2009; Sulpizio and McQueen, 2012). Specifically, significantly more fixations were recorded for targets before target and competitor were segmentally disambiguated, thus supporting the view that the dynamic uptake of acoustic cues related to lexical stress affects word recognition at very early stages. There was no evidence to suggest that alternative stress patterns of competitors differentially affect recognition of the target, for any combination of target and competitor stress pattern.

Crucially, there was no evidence in favor of nonspecified stress in this experiment. Levels of competition, as defined by the difference in the proportions of looks to the target relative to the competitor over time, did not differ significantly in the critical conditions. According to underspecification, targets with dominant (i.e., penultimate-syllable) stress should not have been mismatched by acoustic signals indicating other stress positions, because dominant stress is not lexically specified, leading to asymmetric competition curves for these target-competitor pairs. There is no hint of such an effect in our data. Indeed, the curves for Pen targets were not even consistently below the curves for nondominant targets.

However, there is a possibility that the obligatory diacritic mark of stress in the orthography of the visually presented words may have influenced the results. As was previously mentioned, the Greek orthography requires a special diacritic on the vowel of the stressed syllable of each word. If the orthographic cue of stress in each word affects the online processing and activation of critical words and corresponding representations, then it remains possible that its presence on visual stimuli may have

led to increased fixations to the target as opposed to the competitor from an early point in time. Specifically, the presence of the diacritic may have caused activation of a fully specified stress pattern for each target, including for penultimate-stress targets, thus obscuring the difference between lexically specified and unspecified patterns.

Therefore, before drawing a final conclusion, an additional experiment was conducted, in which the diacritic, and therefore its potential to affect the results, was eliminated.

Experiment 3

Experiment 3 was designed to alleviate concerns over possible influence stemming from orthographic cues to stress. We specifically aimed to clarify and exclude effects of the stress diacritic in the visually presented words. Therefore, we presented all visual stimuli without a stress diacritic. If the diacritic affected the processing of experimental stimuli by causing stress patterns to be activated regardless of lexical specification, this could have effectively neutralized differences arising due to underspecification. In that case, elimination of the diacritic should lead to a different pattern of results in Experiment 3. If, however, the same pattern of results is obtained in Experiment 3, then we can conclude that the orthographic cue to stress did not influence the outcome.

Method

Participants. Participants were 42 undergraduate students from the Panteion University (aged 18–35), native speakers of Greek, with no reported hearing or reading difficulties and with normal or corrected-to-normal vision. All participants received class credit for their participation.

Material. The exact same material was used as in the previous experiment, except that all words appeared on screen without a stress diacritic on the vowel of the stressed syllable.

Procedure. The procedure was the same as in Experiment 2.

Results

Preprocessing and analysis of the data was identical to Experiment 2 (indeed the exact same script was used). The same five items were excluded. One participant was excluded due to very low accuracy. The total proportion of correct responses was 94.6%. Descriptive statistics for accuracy and response time per condition are listed in Table 6. Figure 5 presents raw fixation proportions for targets, competitors, and fillers in each stress condition averaged over 20-ms bins.

Figure 6 displays the relative proportions of looks to target vs. distractor for each condition, relative to the average position of syllable boundaries. There is again evidence for early disambiguation in most conditions; however, in the conditions with a penultimate stress competitor, the curves diverge significantly from zero in the beginning of the third syllable (after accounting for oculomotor delay). Figure 7 displays the difference plots of the summed effects corresponding to the critical pairs of stress conditions. As in Experiment 2, there were no significant differences between the curves, and no discernible trend distinguishing the two pairs involving dominant stress from the one pair not involving it. The results thus fail to support the predictions that were based on underspecification.

As a final step, we performed a direct comparison between Experiments 2 and 3, by modeling the data from the two experiments in a series of combined models that included experiment as a fixed-effects factor as well as smooths of within-trial time per experiment, to model any overall differences in the temporal dynamics (regardless of stress patterns). We specifically contrasted a model with smooths for the six stress conditions not distinguishing between experiments, and a more complex model with different smooths for each experiment, thereby modeling an interaction between experiment and stress condition. The two models accounted for precisely the same

proportion of deviance and the simpler one achieved a lower (i.e., better) smoothing parameter selection score, consistent with no difference between experiments in the within-trial temporal dynamics of any stress contrast conditions.

Discussion

The results of Experiment 3 replicated the results of Experiment 2. Fixations to targets differed significantly from looks to competitors before segmental disambiguation of the words in most cases, suggesting that the presence (or, in this case, the absence) of the stress diacritic in the displayed words had little impact on word recognition processes.

It may be worth noting that the two conditions in which the divergence between target and competitor curves achieved significance latest (in the beginning of the third syllable) were those with penultimate-syllable-stress competitors. That is, they were the two conditions that were predicted to have the highest levels of competition based on the default metrics hypothesis. However, this observation must be tempered by the facts that (a) this does not imply that the interval of divergence starts significantly later than in other conditions, as there is no way to test differences along the temporal dimension; (b) significant divergence necessarily lags the actual divergence point, which is located within the second syllable; and (c) the curve pairs for the relevant contrasting conditions show no supporting trend in one of the two cases, namely final vs. penultimate (Figure 7, middle); only the antepenultimate vs. penultimate curves (Figure 7, left) show the trend (which is far from statistical significance).

Overall, and based on the statistical comparisons, we can conclude that the stress pattern of the competitors did not significantly affect processing and activation of the targets. In particular, the results of Experiment 3 did not provide reliable evidence in favor of underspecification. The predicted asymmetries in the competition between

targets and competitors with the dominant and less frequent stress patterns were not confirmed, consistent with similar processing of all stress patterns.

General Discussion

One fragment priming and two visual world experiments were conducted to explore the nature of stress representations in the lexicon. In particular, we focused on the “default metrics” assumption of WEAVER++ (Levelt et al., 1999; here termed “stress underspecification”), according to which stress is not represented in the lexicon for words with the dominant (i.e., most frequent) stress pattern of the language. In contrast, stress is supposed to be represented in the lexicon—independently of the corresponding segmental specification—for words with nondominant stress patterns. To test this assumption we capitalized on predicted asymmetries in the priming effects for differently stressed targets (Experiment 1) and in the competition between targets and competitors with different stress patterns (Experiments 2 and 3).

The results of Experiment 1 revealed priming effects across stress patterns, that is, for target words with antepenultimate-, penultimate-, and final-syllable stress. The positive finding of significant—and substantial—stress priming for targets with penultimate stress runs directly against the prediction based on stress underspecification, namely that priming should only be evident for words with nondominant stress patterns, that is, for targets stressed on the antepenultimate or the final syllable. Beyond the lack of a significant interaction, for which an argument concerning power can always be made, there is no hint (Table 3 and Figure 1) that stress priming (i.e., the difference between the match and mismatch conditions) might be consistently less for penultimate-syllable-stress words than for targets with nondominant stress patterns (see also supplementary analyses at <https://osf.io/nspgd/>). In all, we interpret the results of Experiment 1, and in particular

the significant stress priming for dominant-stress targets, as supporting the notion that all stress patterns are represented in the lexicon.

The results of Experiments 2 and 3 corroborated those of Experiment 1 by failing to reveal any asymmetries in the competition between targets and stress competitors with different stress patterns. In particular, the prediction based on stress underspecification was that there would be more competition between targets with nondominant stress and competitors with dominant stress, compared to the other way around. However, the difference in looks to targets versus looks to competitors did not appear to be modulated by target stress pattern, whether words were spelled with the obligatory stress diacritic (Experiment 2) or when this orthographic cue to stress was removed (Experiment 3; thus precluding sublexical activation of dominant stress patterns). Figures 4 and 6 do not reveal a consistent association of one stress pattern with higher competition that might have been too weak for our sample size to detect as statistically significant. Although it is always possible that sampling noise may have obscured a very weak signal somewhere, our interpretation of the results of Experiment 2 and 3 is that they are consistent with full representation of all stress patterns in the lexicon.

Sublexical processing

As our focus here is on the lexical representation of stress, it was important to be able to distinguish between lexical and potential sublexical processes. Processing the incoming speech certainly results in a transient representation that includes stress (otherwise pseudoword repetition would be impossible). At the same time, it is plausible that general knowledge of the language can result in stress effects originating in lexical statistics rather than individual representations (as proposed by CDP++, for example). The use of pseudowords in Experiment 1 was therefore a key element for

making the necessary distinctions.

Pseudowords do not have a representation in the lexicon to match or mismatch a given target segmentally, and obviously they lack a stress representation as well. Thus, if stress priming is mediated by the lexicon, then it should only be evident in word stimuli. In contrast, if a sublexical mechanism can produce stress effects through the temporary formation of an online representation, its effects would be observable with both words and pseudowords. In this case, priming effects on pseudowords would constitute evidence for such sublexical stress representations, while effects on words would be interpretable as lexically mediated only to the extent they exceed effects on pseudowords. Experiment 1 provides evidence consistent with this pattern.

Recall that neutral primes differed both segmentally and prosodically from the target, whereas match and mismatch primes were both segmentally identical to it. Thus differences from the neutral condition reflect both segmental and prosodic effects: in synergy, in the match condition (which is both segmentally and prosodically identical to the target), and in opposition, in the mismatch condition (which matches segmentally but mismatches prosodically). The contrast between the match and mismatch condition must then reflect pure effects of stress. In this light, the difference between words and pseudowords can be interpreted as reflecting different processing mechanisms, some of which may be specific to the processing of stress.

For words, the effects of stress on response times were robust, clearly showing the significant role of stress in word recognition. It seems that, in the matching condition, the activation of the correct candidate in the lexicon is facilitated by both segmental and prosodic information, leading to faster response times. In the mismatching condition, the fragment prime inhibits the activation of the same candidate because a part of the available information (i.e., the prosodic features associated with

the stress pattern of the disyllable fragment in the acoustic input) is inconsistent with the lexical representation of the target word. The inhibition due to prosodic mismatch is greater than any facilitation due to segmental match, resulting in net inhibition.

In pseudowords, however, the situation must be different, as there is no lexical representation to be matched or mismatched. A matching fragment (segmentally and prosodically) evidently facilitates the correct response by activating a fully consistent transient sublexical representation. However, in the mismatching condition there are no lexical candidates to be inhibited by the mismatch. Rather, the matching segments have a facilitating effect by supporting the construction of a sublexical segmental and/or orthographic representation. In this transient representation stress seems to play a rather small role, as facilitation (with respect to neutral) was observed in both the match and the mismatch condition. Yet this facilitation was not equal in the two conditions: The difference between match and mismatch was significant for pseudowords, consistent with a weak stress effect. This effect must be due to the transient prosodic representation of the incoming spoken pseudoword, which supports (or hinders) the prosodic decoding of the visual stimulus. The small size of this stress effect in pseudowords is what permits us to confidently interpret the large effect observed in words as lexical in nature.

On the nature of the observed stress effects

As noted, we interpret the evidence from all three experiments taken together as consistent with the dominant stress pattern being represented in lexical entries in the same way as nondominant patterns. A possible alternative explanation might hinge on lexical access processes, without the involvement of the lexicon itself,⁵ if we allow such

⁵ We are grateful to James McQueen for suggesting this explanation.

processes to be affected by “pan-lexical” knowledge. For example, lexical access processes may be inherently biased toward the dominant stress pattern to the extent licensed by the observed stress pattern distributions in the lexicon. In this scenario, the representation of the spoken input can compete either with the content of specific lexical entries or with the lexical access process itself expressing an overall bias amounting to an implicit “default”. Words with nondominant patterns would have to overcome the process bias whereas words with dominant patterns would have to overcome competition from lexical items with mismatching stress patterns. Thus dominant patterns may be unspecified within the lexical entries but can still effectively match and mismatch the input like nondominant patterns.

This kind of approach has the inherent epistemological disadvantage that it can be very difficult to test empirically, absent a method for distinguishing the contents of the lexicon from the effects of the processes accessing it. We acknowledge that theoretical alternatives of this sort may be possible to disentangle on the basis of computational models, at a future time when sufficiently detailed models of speech perception and lexical representation become available. However, there are further reasons to remain skeptical towards this alternative explanation. From a purely theoretical point of view, it is unclear how such knowledge can emerge if not on the basis of one’s language processing history. And it is unclear how and why an emerging language processing system would choose to systematically discard a certain stress pattern from its—otherwise gradually refined and enriched—stored representations. Further, if lexical access processes are systematically biased in favor of a more frequent pattern, this would necessarily be the case regardless of the current input or any active lexical candidates. Therefore the effects of pan-lexical knowledge would be in addition to any lexical effects, so that an asymmetry in representation would still surface in

differential priming patterns, as it would without the biased process.

From an empirical standpoint, priming effects have typically been associated in the literature with stored forms of representations in the lexicon, whereas the absence of priming effects has been associated with lack of lexical mediation. Specifically concerning stress representations, in a series of four speech production experiments in Dutch, Schiller et al. (2004) found no stress priming and suggested that perhaps neither stress pattern is stored in the lexicon. Conversely, in a study exploring stress priming effects in a naming task in Italian, Sulpizio et al. (2012) found equal stress priming effects in both dominant and nondominant word stress patterns, suggesting lexical representation of both penultimate and antepenultimate stress patterns. Following up along these lines, our Experiment 1 extends the findings of priming to a lexical decision task and to more than two stress patterns, and concludes in favor of the explicit representation of all stress patterns.

On default stress patterns and their effects

Turning to the visual world paradigm, Sulpizio and McQueen (2012) reported similar competition among penultimate and antepenultimate real word targets in Italian, in agreement with our findings in Experiments 2 and 3. However, in a second experiment involving production of an artificial lexicon, they found higher levels of competition for antepenultimate stress targets relative to the dominant penultimate. Penultimate stress targets were responded to faster, leading the authors to conclude that (a) Italian speakers have abstract prosodic knowledge that can be generalized to an artificial lexicon, and (b) the significantly faster recognition of penultimate stress targets is better explained via the use of a default mechanism. However, differences found in the treatment of dominant vs nondominant stress patterns could also reflect the interplay between the lexicon and a sublexical mechanism enhancing the process of

stress identification due to statistical preponderance of the dominant stress, rather than the operation of a default mechanism per se. The involvement of a sublexical mechanism in stress effects in Italian is in fact supported by several studies (Burani & Arduino, 2004; Colombo & Zevin, 2009; Sulpizio, Boureux, et al., 2012; Sulpizio, Job, et al., 2012).

The idea that “default” stress effects are statistical in origin, that is, they are effects of stress dominance, is also consistent with a number of findings from reading studies in Greek (Grimani & Protopapas, 2016; Protopapas & Gerakaki, 2009; Protopapas et al., 2006, 2007). In these studies, the dominant stress pattern seemed to apply only when orthographic, morphological, or lexical information was unavailable (i.e., no resemblance to a word, no recognizable morpheme, and no diacritic). In all of these studies, findings are consistent with the existence of a penultimate “default” that expresses the statistical dominance of this stress pattern, rather than with the operation of an abstract psycholinguistic operation. This is further consistent with a probabilistic, rather than rule-based, approach to stress assignment more generally (and not only in Greek; cf. Arciuli et al., 2010; Mousikou et al., 2017). The statistical nature of the default—and the preference for using the term “dominant” instead—indicates that all stress patterns must be fully specified in the lexicon so that their cumulative effects can emerge on the basis of all the individually stored instances.

One question that might arise in this context is, can we be certain that the penultimate-syllable stress pattern is the default one in Greek? From a theoretical point of view, statistical preponderance is not necessarily associated with unmarkedness. Therefore the superficial lexical statistics that unambiguously indicate penultimate-syllable stress as dominant may not be entirely convincing in this regard. As noted in the introduction, theoretical accounts from phonology are consistent with

the bisyllabic trochee being the default, despite further theory-internal considerations that may introduce ambiguity. More importantly, the penultimate default is also what the empirical evidence indicates, across diverse fields ranging from language acquisition (Kappa, 2002) to reading aloud (Protopapas & Gerakaki, 2009; Protopapas et al., 2006). And finally, one may note that the interpretation of our findings is not affected by which stress pattern is taken to be the default. This is because all stress patterns exhibited priming in Experiment 1 (Figure 1) and all contrasting stress pairs were examined in Experiments 2 and 3 (Figures 4 and 7, respectively). None of the three stress patterns emerged as distinct, to support claims of special status. We can thus be confident in concluding that, taken together with the results of previous studies, the current evidence supports the notion that prosodic features associated with stress patterns are fully represented in the lexicon, for all stress patterns.

On the relationship between segmental and prosodic representations

How do our findings bear on the issue of exactly how prosodic and segmental information is represented in the lexicon? The results of Experiments 2 and 3 revealed in most cases significantly more fixations to targets, relative to competitors, before segmental disambiguation of the two, in some cases already significant during the first syllable (after taking oculomotor planning into account). This could be explained as the effect of intonation (in our stimuli high F_0) on the stressed syllable acting as an acoustic cue for the classification of the current stretch of the incoming signal as being part of a stressed (vs. unstressed) syllable. From this point of view, phonetic and (stress-related) prosodic information is extracted from the speech signal to be separately matched against the corresponding aspects of lexical representations. That is, prosodic features are matched against lexical stress representations and phonetic features are matched against lexical segmental representations, in two potentially independent information

streams. This would presumably be akin to the “phoneme-free” view of prosodic representations espoused by Schild et al. (2014a).

From another angle, stress-based lexical disambiguation might suggest that each syllable is individually represented in sufficient prosodic detail in the lexicon, regardless of whether it bears lexical stress or not, so as to be differentially activated by acoustic input consistent with stressed vs. unstressed syllables. In other words, prosodic representations could be viewed as intimately connected to segmental representations rather than abstracted away from them as standalone metrical information streams. In this approach there is no separate matching for segmental and prosodic information but, rather, a single domain of representation that includes both prosodic and phonetic features. Thus, in the spirit of Schild et al. (2014a), such combined phoneme-prosody representations would not constitute good matches to phonemically unrelated sequences on the basis of prosodic match alone.

Both of these approaches concern relatively low levels of abstraction, not far removed from specific features (but certainly well above specific acoustic episodes; cf. McQueen et al., 2006). In our opinion the notion of fully abstract metrical templates cannot easily be applied to speech perception. If an abstract prosodic template is represented separately from segmental characteristics of a word to be retrieved upon lexical access, then one would expect that in speech perception lexical stress could only play a minimal—if any—role in word identification, because stress is a relational property defined over two or more syllables, and therefore the proper metrical template could only be activated following lexical access. That is, an abstract metrical template is difficult to reconcile with a processing view that emphasizes continuous uptake of all relevant information, including prosodic information; a view that is both popular and consistent with the results of Experiments 2 and 3.

Beyond the current experiments, empirical support for the involvement of abstract metrical templates in spoken word recognition has been difficult to obtain, insofar as there is no published demonstration of pure stress priming, that is, priming among words with the same stress pattern in the absence of segmental match (in English: Slowiaczek et al., 2006; in Greek: Protopapas et al., 2016). This led Protopapas et al. (2016) to argue for an integrated segmental and prosodic representation rather than abstract stress pattern information stored in the lexicon.

Future implementations of psycholinguistic models of spoken word recognition can help clarify the relation of segmental and prosodic information in lexical representations. Shortlist (Norris, 1994; Norris & McQueen, 2008) and Fine-Tracker (Scharenborg & Boves, 2010) have already incorporated effects of certain correlates of stress such as vowel segmental quality and duration, respectively, and may be extended toward more encompassing prosodic representations. It may also be possible to add prosodic features to the TRACE model (McClelland & Elman, 1986) across the three levels of processing (from input to lexical representations). However, TRACE seems to be better tailored to separate rather than integrated representation of segmental and prosodic features because of the abstract phoneme level between input and the lexicon (Weber & Scharenborg, 2012). It seems fair to say that models of spoken word recognition have tended to focus on selected aspects of representation and processing and that more inclusive implementations will be needed to account for a wider range of effects, including those of lexical stress.

On the role of the diacritic

With respect to the impact of the stress diacritic in the Greek orthography, Experiment 3 provided a somewhat mixed picture. In most stress conditions targets disambiguated early from their competitors, similar to Experiment 2. In the two

conditions in which the difference between targets and competitors did not attain significance before segmental disambiguation, initial curve divergence appeared well before. However, comparison of Figures 3 and 5 suggests that disambiguation may have been generally somewhat delayed in Experiment 3 relative to Experiment 2. A direct comparison of the experiment-specific smooths (see supplementary material at <https://osf.io/nspgd/>) showed that the slope of the target-competitor curve was less steep in Experiment 3 and remained lower (the difference attaining statistical significance after 818 ms), consistent with the longer RTs in Experiment 3 (cf. Tables 5 and 6). Barring uncontrolled differences in the two samples, this suggests overall higher competition in Experiment 3 as a result of the missing stress diacritic.

A possible source of this effect is that the lack of the diacritic may have delayed full lexical activation of the target due to the incomplete orthographic form. Effects of the lack of the stress diacritic have not been previously established (cf. Protopapas & Gerakaki, 2009), but perhaps the higher sensitivity of the visual world experimental procedure may have allowed them to be observed. If this explanation is on the right track then it supports the idea that the diacritic does not operate independently at a sublexical level but forms part of a single coherent orthographic word form. This means that the diacritic could not have worked against the underspecification hypothesis in Experiment 2 by supporting the construction of a metrical frame sublexically. This interpretation also bolsters previous suggestions that the typical origin of stress pattern activation in reading Greek is lexical rather than assembled on the basis of the diacritic (Protopapas, 2016). The present findings thus leave the door open for a disambiguating role of the stress diacritic in the case of stress minimal pairs. Further research is required to get a better understanding of the role of the stress diacritic in the recognition of written words.

At any rate, and notwithstanding the elevated rate of stress diacritic omissions by poor spellers (Protopapas et al., 2013), it must be kept in mind that the stress diacritic is obligatory in Greek spelling and its omission is a frank spelling error. Therefore, any findings from experiments with stimuli presented without diacritics must be interpreted with caution due to the potential for diminished ecological validity.

Limitations of the current study

In the current study we have derived our predictions based on theoretical assumptions made in WEAVER++ (Levelt, et al., 1999), a model of speech production that makes no reference to how metrical representations in the lexicon may impact receptive word processing and activation of corresponding representations. Thus our findings do not contradict this model in its original domain, but only an ad-hoc extension to perception. The same applies for CDP++ (Perry et al., 2014), a model of reading aloud that makes no explicit claims regarding spoken word recognition, which is the topic of the current study.

One may question whether a model of language production is an optimal source of hypotheses regarding the representation of stress in spoken word recognition. However, given that no model of spoken word recognition makes any specific reference to the representation of lexical stress, empirical work in this domain would either have to wait for further theoretical developments or seek alternative sources for relevant testable assumptions. In this light, models of word production can be seen as highly relevant for work on word recognition to the extent that production and perception processes are viewed as tightly linked, interacting, and sharing common representations (Meyer et al., 2016; Pickering, & Garrod, 2013).

An assumption of shared representations among production and comprehension is not uncommon in the literature, even if the precise degree of sharing remains under

debate. Levelt et al. (1999) posited that production and perception share common representations from the lemma level onwards (that is, mostly conceptual and syntactic representations). Pickering and Garrod (2004) suggested complete sharing of representations at all linguistic levels across production and comprehension. Despite a lack of unanimity among researchers, there seems to be evidence for shared representations for the phonological level as well (see Gambi & Pickering, 2017, for a review). In sum, and bearing in mind that production and perception must involve different processes and may operate under different constraints, it remains likely that both operate via accessing the same lexicon. Our study is thus based on the (not uncommon) assumption of a single lexicon with lexical representations underlying both production and recognition, while acknowledging that perception and production may use different components or aspects of these representations, or use them in different ways.

An additional limitation concerns linguistic specificity. Our experiments were carried out in Greek and therefore the conclusions hold for Greek and may be generalized—with caution—only to languages with similar properties in lexical stress. The conclusions may not apply for languages with different stress systems or other overriding characteristics. Greek is a free stress language with a word-final trisyllabic stress domain, very small segmental quality changes associated with stress, a dominant penultimate stress pattern, and a relatively transparent orthography that marks stress with a diacritic. Comparable work in other languages can reveal which of these properties may be critical for constraining or generalizing the conclusions.

Finally, a statistical limitation concerns the interpretation of the nonsignificant differences between different stress pattern conditions in Experiments 2 and 3. These findings are consistent with no effect of stress pattern (i.e., a true null effect), but they

may also be consistent with a small nonnull effect. Unfortunately, it is not possible to conduct Bayesian Factor analysis within the GMM framework to evaluate the strength of support for the null, as was done in Experiment 1. We are thus limited to stating that there was no evidence for an effect of stress pattern in these experiments.

Conclusion

The results of the current study provide evidence in favor of fully specified lexically represented stress patterns and an interplay among the lexicon and a sublexical mechanism accountable for the processing of words and pseudowords. While no direct evidence is provided regarding exactly how stress is represented and related to the segmental characteristics of words, the results taken together with those of previous studies are consistent with prosodic representations being integrated within the lexical representations of words rather than abstract.

References

- Arciuli, J., Monaghan, P. & Seva, N. (2010). Learning to assign lexical stress during reading aloud: Corpus, behavioral, and computational investigations. *Journal of Memory and Language*, *63*, 180–196.
- Arvaniti, A., (2000). The Phonetics of stress in Greek. *Journal of Greek Linguistics*, *1*, 9-38.
- Arvaniti, A., (2007). Greek phonetics: The state of the art. *Journal of Greek Linguistics*, *8*, 97–208.
- Arvaniti, A., & Baltazani, M. (2005). Intonational analysis and prosodic annotation of Greek spoken corpora. In J. Sun-Ah (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 84–117). Oxford, UK: Oxford University Press.
- Ashby, J., & Clifton, C. (2005). The prosodic property of lexical stress affects eye movements during silent reading. *Cognition*, *96*, B89–B100.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.
- Baayen, H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language*, *94*, 206–234.
- Barr, D. J. (2008). Analyzing ‘visual world’ eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*, 457–474.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme. *Journal of Statistical Software*, *67*, 1–48.
- Beckman, M. E. (1986). *Stress and mon-stress accent*. Dordrecht: Foris.
- Boersma, P., & Weenink, D. (2007). PRAAT, a system for doing phonetics by computer (version 5.1.32) [computer program]. Retrieved from <http://www.praat.org>

- Botinis, A., (1989). *Stress and prosodic structure in Greek: A phonological, acoustic, physiological and perceptual study*. Lund, Sweden: Lund University Press.
- Breen, M., & Clifton, C. (2011). Stress matters: Effects of anticipated lexical stress on silent reading. *Journal of Memory and Language*, *64*, 53–170.
- Breen, M., & Clifton Jr, C. (2013). Stress matters revisited: A boundary change experiment. *The Quarterly Journal of Experimental Psychology*, *66*, 1896–1909.
- Burani, C., & Arduino, L. S. (2004). Stress regularity or consistency? Reading aloud Italian polysyllables with different stress patterns. *Brain and Language*, *90*, 318–325.
- Cappa, S. F., Nespore, M., Ielasi, W., & Miozzo, A. (1997). The representation of stress: Evidence from an aphasic patient. *Cognition*, *65*, 1-13.
- Colombo, L. (1992). Lexical stress effect and its interaction with frequency in word pronunciation. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 987–1003.
- Colombo L., & Zevin, J. (2009) Stress priming in reading and the selective modulation of lexical and sub-lexical pathways. *PLoS ONE*, *4*, e7219.
- Colombo, L., & Sulpizio, S. (2015). When orthography is not enough: The effect of lexical stress in lexical decision. *Memory and Cognition*, *43*, 811–24.
- Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech*, *45*, 207–228.
- Cutler, A., & van Donselaar, W. (2001). Voornaam is not (really) a homophone: Lexical prosody and lexical access. *Language and Speech*, *44*, 171–195.
- De Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *The Journal of the Acoustical Society of*

America, 97, 491–504.

Fourakis, M., Botinis, A., & Katsaiti, M. (1999). Acoustic characteristics of Greek vowels. *Phonetica*, 56, 28–43.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers*, 35, 116–124.

Gambi, C., & Pickering, M. J. (2017). Models linking production and comprehension. In E. Fernández, and H. Cairns (Eds.), *Handbook of Psycholinguistics*. (pp 157-182).

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251–271.

Grimani, A., & Protopapas, A. (2016). Derivational suffixes as cues to stress position in reading Greek. *Journal of Research in Reading*, 40, 23–41.

Hayes, B., (1995). *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.

Jouravlev, O., & Lupker, S. J. (2014). Stress consistency and stress regularity effects in Russian. *Language, Cognition and Neuroscience*, 29, 605–619.

Kappa, I., (2002). On the acquisition of syllabic structure in Greek. *Journal of Greek Linguistics*, 3, 1–52.

Laganaro, M., Vacheresse, F., & Frauenfelder, U. H. (2002). Selective Impairment of Lexical Stress Assignment in an Italian-Speaking Aphasic Patient. *Brain and Language*, 81, 601–609.

Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*, 38, 44–59.

Laver, J. (1994). *Principles of phonetics*. Cambridge University Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech

- production. *Behavioral and Brain Sciences*, 22, 1–75.
- Lukyanenko, A., Idsardi, W. J., & Jiang, N. (2011). Opening Your Ears: The Role of L1 in Processing of Nonnative Prosodic Contrasts. In G. Granena, J. Koeth, S. Lee-Ellis, A. Lukyanenko, G. Prieto Botana, & E. Rhoades (Eds.), *Selected proceedings of the Second Language Research Forum 2010*. Somerville, MA: Cascadilla Press.
- Malikouti-Drachman, A. (2002). Review Article: Greek Phonology: A contemporary perspective. *Journal of Greek Linguistics*, vol. 2 (1), 187–243.
- Malikouti-Drachman, A., & Drachman, G. (1989). Τονισμός στα ελληνικά [Stress in Greek]. *Studies in Greek Linguistics*, 9, 127-143.
- Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology*, 29, 189–256.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
- McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological Abstraction in the Mental Lexicon. *Cognitive Science*, 30, 1113–1126.
- Meyer, A. S., Huettig, F., & Levelt, W. J. M. (Eds.).(2016). Same, different, or closely related: What is the relationship between language production and comprehension? *Journal of Memory and Language*, 89, pp. 1–7.
- Mirman, D., Dixon, J. A., & Magnuson, J. S.(2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59, 475–494.
- Mitterer, H. (2011). The mental lexicon is fully specified: Evidence from eye-tracking. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 496–513.
- Morey, R. D., & Rouder, J. N. (2018). BayesFactor: Computation of Bayes factors for

- common designs. R package version 0.9.12-4.2.
- Mousikou, P., Sadat, J., Lucas, R., & Rastle, K. (2017). Moving beyond the monosyllable in models of skilled reading: Mega-study of disyllabic nonword reading. *Journal of Memory and Language, 93*, 169–192.
- Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition, 52*, 189–234.
- Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115*, 357–395.
- 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*, 106–151.
- Perry, C., Ziegler, J. C., & Zorzi, M. (2014). CDP++ Italian: Modelling sublexical and supralexicalexical inconsistency in a shallow orthography. *Plos ONE, 9*(4):e94291.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*(4), 329–392.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(02), 169–190.
- Porretta, V., Kyröläinen, A., van Rij, J., & Järvikivi, J. (2016). VWPre: Tools for preprocessing visual world data. R package version 0.9.5.
- Porretta, V., Kyröläinen, A. J., van Rij, J., & Järvikivi, J. (2017, June). Visual world paradigm data: From preprocessing to nonlinear time-course analysis. In I. Czarnowski, R. Howlett, & L. Jain (Eds.), *International Conference on Intelligent Decision Technologies* (pp. 268–277). Springer, Cham.
- Protopapas, A. (2006). On the use and usefulness of stress diacritics in reading Greek. *Reading & Writing: An Interdisciplinary Journal, 19*, 171–198.

- Protopapas, A. (2016). From diacritics to the mental lexicon: Where is the stress? In J. Thomson & L. Jarmulowics (Eds.), *Linguistic rhythm and literacy* (pp. 237–264). Amsterdam: John Benjamins.
- Protopapas, A., Fakou, A., Drakopoulou, S., Skaloumbakas, C., & Mouzaki, A. (2013). What do spelling errors tell us? Classification and analysis of spelling errors of Greek schoolchildren with and without dyslexia. *Reading & Writing, 26*, 615–646.
- Protopapas, A., & Gerakaki, S. (2009). Development of processing stress diacritics in reading Greek. *Scientific Studies of Reading, 13*, 453–483.
- Protopapas, A., Gerakaki, S., & Alexandri, S. (2006). Lexical and default stress assignment in reading Greek. *Journal of Research in Reading, 29*, 418–432.
- Protopapas, A., Gerakaki, S., & Alexandri, S. (2007). Sources of information for stress assignment in reading Greek. *Applied Psycholinguistics, 28*, 695–720.
- Protopapas, A., Panagaki, E., Andrikopoulou, A., Gutiérrez Palma, N., & Arvaniti, A. (2016). Priming stress patterns in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 42*, 1739–1760.
- Protopapas, A., Tzakosta, M., Chalamandaris, A., & Tsiakoulis, P. (2012). IPLR: An online resource for Greek word-level and sublexical information. *Language Resources & Evaluation, 46*, 449–459.
- Protopapas, A., Vlahou, E. (2009). A comparative quantitative analysis of Greek orthographic transparency. *Behavior Research Methods, 41*(4), 991–1008.
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>
- Reinisch, E., Jesse, A., & McQueen, J. M. (2009). Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately. *The*

Quarterly Journal of Experimental Psychology, 63, 772–783.

Revithiadou, A. (1999). *Headmost Accent Wins: Head Dominance and Ideal Prosodic Form in Lexical Accent Systems*. LOT Dissertation Series 15 (HIL/Leiden Universiteit).

Holland Academic Graphics, The Hague.

Scharenborg, O., & Boves, L. (2010). Computational modelling of spoken-word recognition processes: Design choices and evaluation. *Pragmatics & Cognition*, 18, 136–164.

Schild, U., Becker, A. B. C., & Friedrich, C. K. (2014a). Phoneme-free prosodic representations are involved in pre-lexical and lexical neurobiological mechanisms underlying spoken word processing. *Brain & Language*, 136, 31–43.

Schild, U., Becker, A. B. C., & Friedrich, C. K. (2014b). Processing of syllable stress is functionally different from phoneme processing and does not profit from literacy acquisition. *Frontiers in Psychology*, 5, 530.

Schiller, N. O., Fikkert, P., Levelt, C.C. (2004). Stress priming in picture naming: An SOA study. *Brain & Language*, 90, 231–240.

Shatzman, K. B., & McQueen, J. M. (2006). Prosodic knowledge affects the recognition of newly acquired words. *Psychological Science*, 17, 372–377.

Slowiaczek, L. M., Soltano, E. G., & Bernstein, H. L. (2006). Lexical and metrical stress in word recognition: Lexical or pre-lexical influences? *Journal of Psycholinguistic Research*, 35, 491–512.

Soto-Faraco, S., Sebastian-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, 45, 412–432.

Sulpizio, S., Arduino, L. S., Paizi, D., & Burani, C. (2013). Stress assignment in reading Italian polysyllabic pseudowords. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 51–68.

- Sulpizio, S., Boureux, M., Burani, C., Deguchi, C., & Colombo, L. (2012). Stress assignment in the development of reading aloud: Nonword priming effects on Italian children. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.) *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp.2369–2374). Austin: Cognitive Science Society.
- Sulpizio, S., & Colombo, L. (2017). Early markers of lexical stress in visual word recognition. *Journal of Memory and Language*, *45*, 1398–1410.
- Sulpizio, S., & Colombo, L. (2013). Lexical stress, frequency, and stress neighbourhood effects in the early stages of Italian reading development. *The Quarterly Journal of Experimental Psychology*, *66:10*, 2073–2084.
- Sulpizio, S., Job, R., & Burani, C. (2012). Priming lexical stress in reading Italian aloud. *Language and Cognitive Processes*, *27*, 808–820.
- Sulpizio, S., & McQueen, J. M. (2012). Italians use abstract knowledge about lexical stress during spoken-word recognition. *Journal of Memory and Language*, *66*, 177–193.
- Sulpizio, S., Vespignani, F., & Job, R. (2016). On the time course of lexical stress priming in speech production: Behavioral and ERPs evidence from a free-stress language. *Brain Research*, *vol.1648 (A)*, 202–213.
- Tagliapietra, L., & Tabossi, P. (2005). Lexical stress effects in Italian spoken word recognition. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the XXVII Annual Conference of the Cognitive Science Society* (pp. 2140–2144). Mahwah, NJ: Erlbaum.
- van Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *Quarterly Journal of Experimental Psychology*, *58A*, 251–273.
- van Rij, J. (2015). Overview of GAMM analysis of time series data.
<https://jacolienvanrij.com/Tutorials/GAMM.html>

- van Rij, J., Wieling, M., Baayen, R. J., & van Rijn, H. (2015). *itsadug: Interpreting time series and autocorrelated data using GAMMs*. R package version 2.2.
- Weber, A., & Scharenborg, O. (2012). Models of spoken-word recognition. Wiley Interdisciplinary Reviews: *Cognitive Science*, 3, 387–401.
- Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, vol. 70, 86–116.
- Wood S. N. (2006). *Generalized additive models: An introduction with R*. Chapman and Hall/CRC Press.
- Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73, 3–36.

Table 1*Mean acoustic measures of word stimuli used as targets and /or primes in Experiment 1*

Targets	Duration (ms)		F_0 (Hz)		Intensity (dB)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>Antepenultimate-syllable stress</i>						
1st syllable	175.2	30.4	208.2	12.2	74.5	2.1
2nd syllable	124.2	29.5	163.4	14.1	75.3	11.5
3rd syllable	164.9	25.8	143.9	6.1	71.7	2.8
<i>Penultimate-syllable stress</i>						
1st syllable	119.5	41.9	174.6	15.8	72.4	10.8
2nd syllable	206.3	39.4	197.6	14.4	75.4	1.9
3rd syllable	144.8	35.8	169.1	12.7	71.6	5.5
<i>Final-syllable stress</i>						
1st syllable	116.0	42.5	183.9	12.7	73.3	2.8
2nd syllable	145.5	37.6	160.9	13.1	74.5	2.3
3rd syllable	215.6	41.8	179.1	9.6	73.1	2.2

Note. F_0 = Fundamental Frequency. For each syllable in each item, F_0 and intensity were obtained using the “Get mean...” Praat function of the Pitch and intensity contour, respectively, over the marked duration of the syllable.

Table 2*Response times (ms) and error rates (proportion) for words and pseudowords per condition in Experiment 1*

Stress	Words				Pseudowords			
	Response time		Error rate		Response time		Error rate	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Match	714.7	134.5	0.09	0.08	741.4	128.9	0.03	0.10
Neutral	756.1	146.9	0.11	0.09	779.2	144.2	0.04	0.10
Mismatch	796.5	163.6	0.14	0.08	759.9	142.2	0.03	0.10

Table 3

Response times (ms) to word and pseudoword targets as a function of prime and target stress positions in Experiment 1

Target stress position	Prime stress position					
	Antepenultimate		Penultimate		Final	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>Words</i>						
Antepenultimate	740.9	143.3	832.0	193.2	843.0	215.4
Penultimate	808.7	193.7	709.6	150.7	787.2	175.7
Final	761.5	164.7	755.1	153.3	695.2	133.9
<i>Pseudowords</i>						
Antepenultimate	715.7	123.2	728.5	146.2	749.8	152.0
Penultimate	758.6	160.7	742.8	128.0	771.6	152.0
Final	780.5	161.1	767.3	155.5	765.2	149.8

Table 4*Conditions and predictions for Experiments 2 and 3*

Condition	Target				Competitor				Predicted Competition
	Stress Position	Word example	Stress Specification	Relation to Signal	Stress position	Word example	Stress Specification	Relation to Signal	
antT_penC	1 (Ant)	κέρασε /'cerase/	Yes	match	2 (Pen)	κεράσι /ce'rasi/	No	no-mismatch	High
penT_antC	2 (Pen)	κεράσι /ce'rasi/	No	no-mismatch	1 (Ant)	κέρασε /'cerase/	Yes	mismatch	Low
finT_antC	3 (Fin)	απορώ /apo'ro/	Yes	match	1 (Ant)	άπορους /'aporus/	Yes	mismatch	Low
antT_finC	1 (Ant)	άπορους /'aporus/	Yes	match	3 (Fin)	απορώ /apo'ro/	Yes	mismatch	Low

finT_penC	3 (Fin)	γενεών	Yes	match	2 (Pen)	γενναίων	No	no-mismatch	High
		/jɛnɛ'on/				/jɛ'neon/			
penT_finC	2 (Pen)	γενναίων	No	no-mismatch	3 (Fin)	γενεών	Yes	mismatch	Low
		/jɛ'neon/				/jɛnɛ'on/			

Note: Ant=antepenultimate (1st syllable), Pen=penultimate (2nd), Fin=final (3rd). Stress specification refers to the presumed lexical representation of the target under the default metrics assumption.

Table 5*Accuracy (proportion) and response times (ms) to targets per condition in Experiment 2*

Condition	Response Time		Accuracy	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
AntT_PenC	1273.0	259.1	1.00	.00
AntT_FinC	1217.9	241.7	1.00	.00
PenT_AntC	1293.2	268.1	1.00	.02
PenT_FinC	1300.2	242.3	0.99	.03
FinT_AntC	1285.1	270.2	1.00	.00
FinT_PenC	1308.0	245.9	0.99	.03

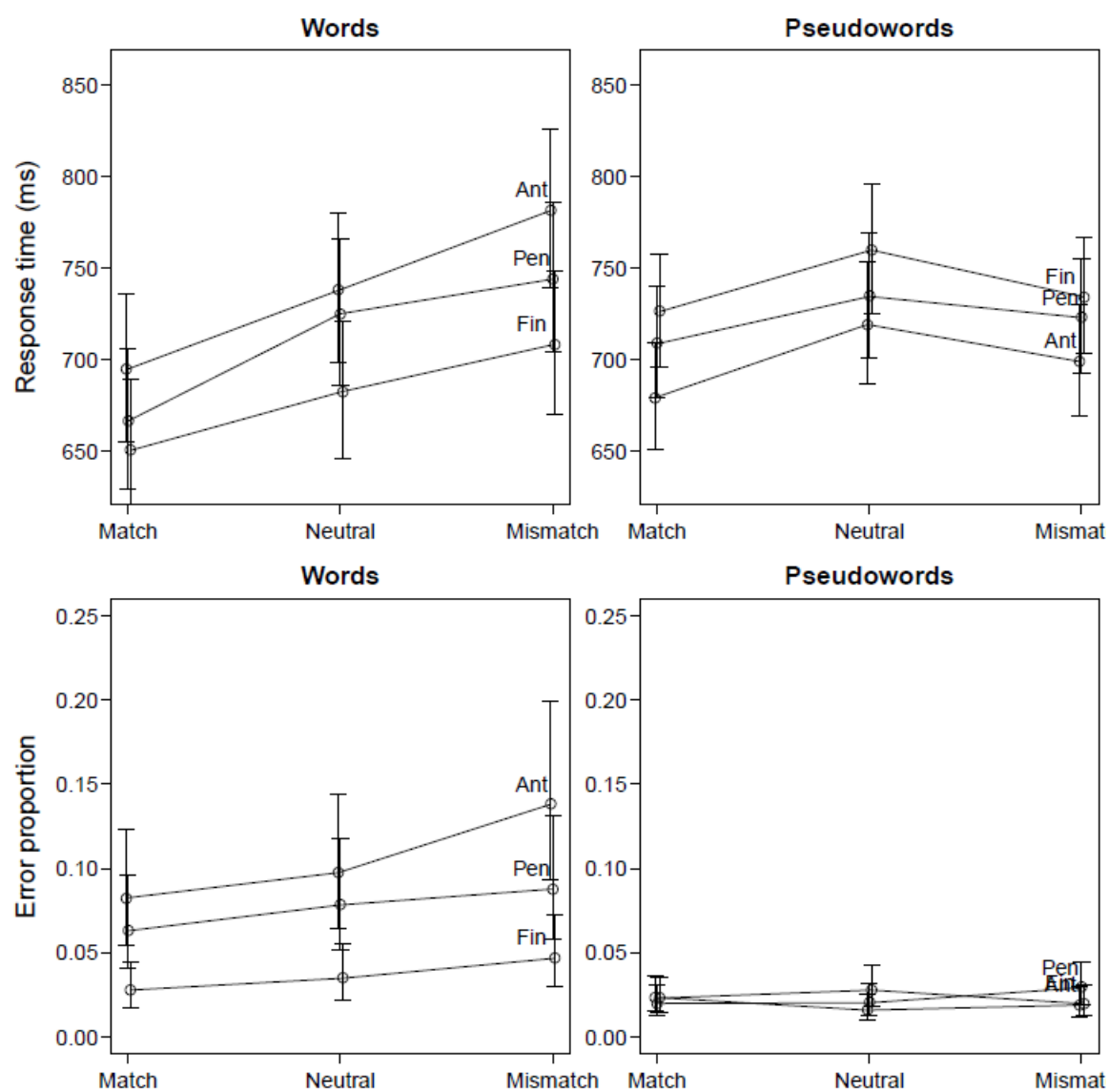
Table 6*Accuracy (proportion) and response times (ms) to targets per condition in Experiment 3*

Condition	Response Time		Accuracy	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
AntT_PenC	1414.8	276.5	1.00	0.02
AntT_FinC	1453.4	304.2	0.99	0.09
PenT_AntC	1560.2	328.8	0.99	0.06
PenT_FinC	1555.6	343.9	0.98	0.10
FinT_AntC	1447.0	287.5	0.97	0.05
FinT_PenC	1470.4	276.1	1.00	0.02

Figure 1

Mean response times and error rates for words and pseudowords per condition in

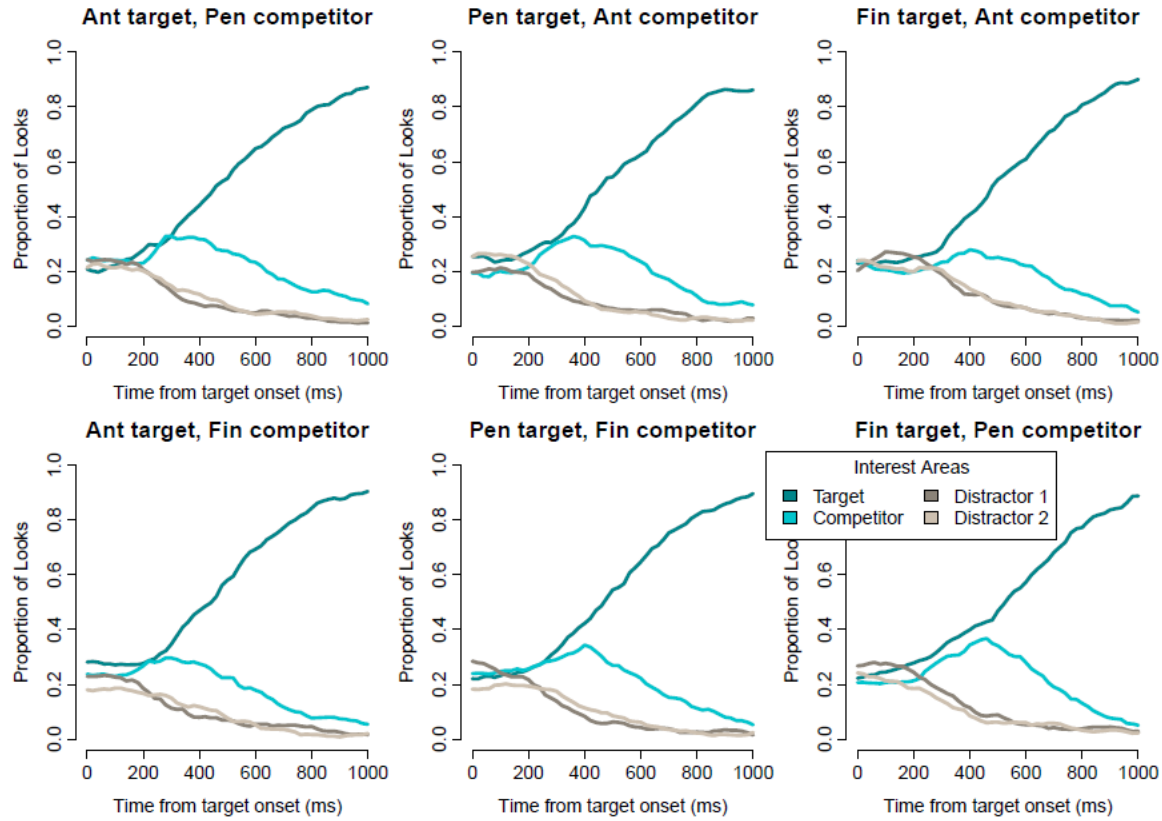
Experiment 1



Note. Model-based estimates and associated 95% CI for response time and errors, back-transformed to ms and proportion, respectively. Ant = antepenultimate-syllable-stress targets; Pen = penultimate-syllable-stress targets; Fin = final-syllable-stress targets.

Figure 2

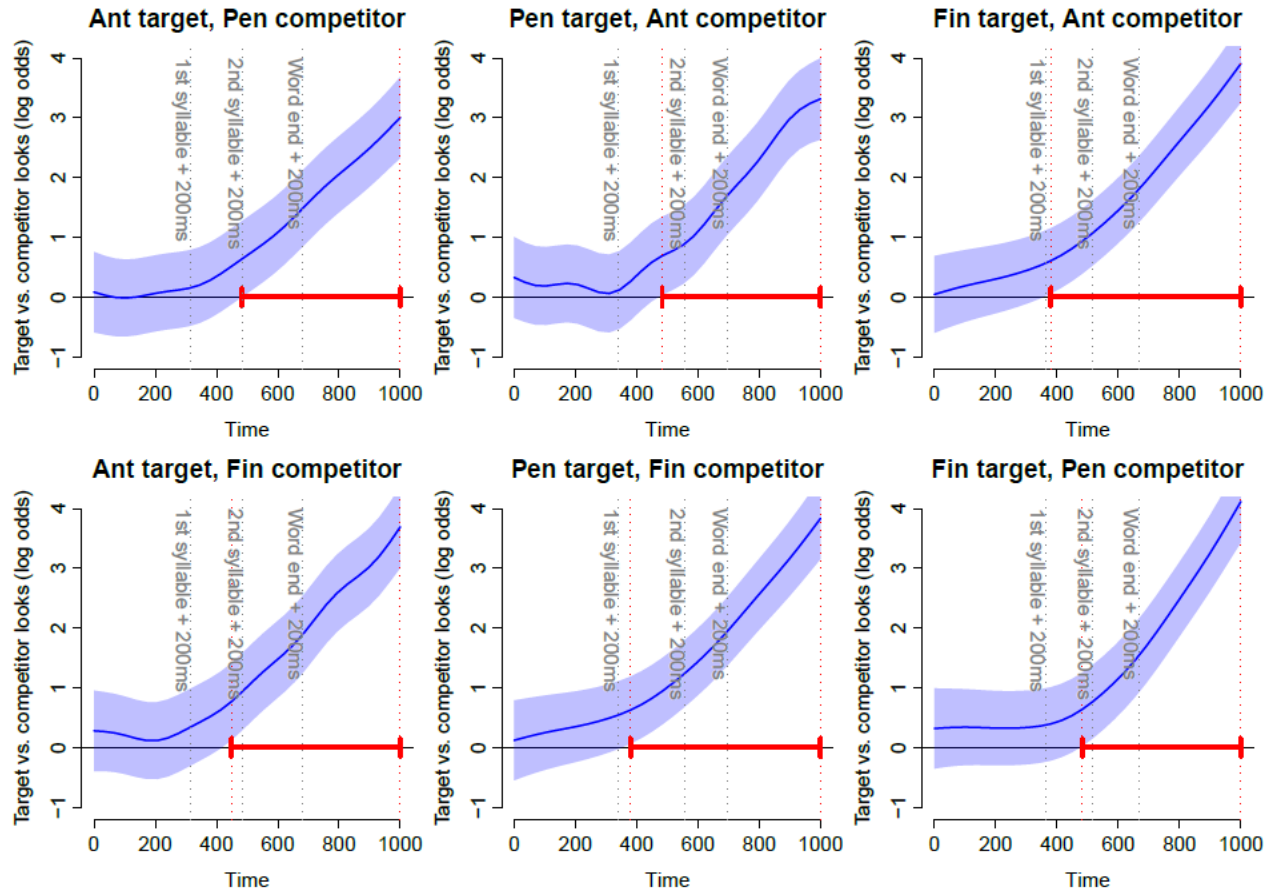
Proportion of looks to interest areas over time in each stress contrast condition in Experiment 2



Note. Ant = antepenultimate-syllable-stress; Pen = penultimate-syllable-stress; Fin = final-syllable-stress.

Figure 3

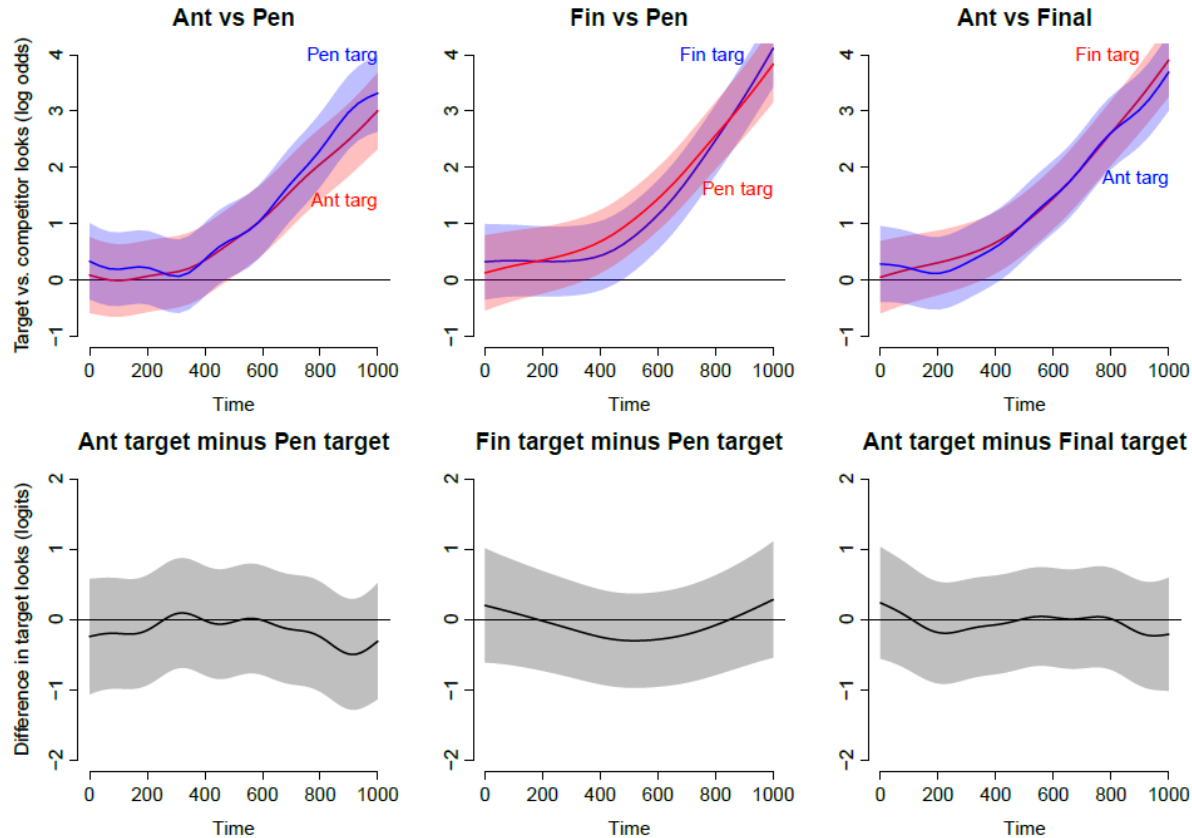
Within-trial dynamics of the difference in looks to target vs. competitor for each stress contrast condition in Experiment 2



Note. Zero time marks the onset of the spoken word. Time intervals of significant difference in looks between targets and competitors are marked in red. The gray dotted vertical lines mark the average location of target syllable boundaries plus a 200-ms offset.

Figure 4

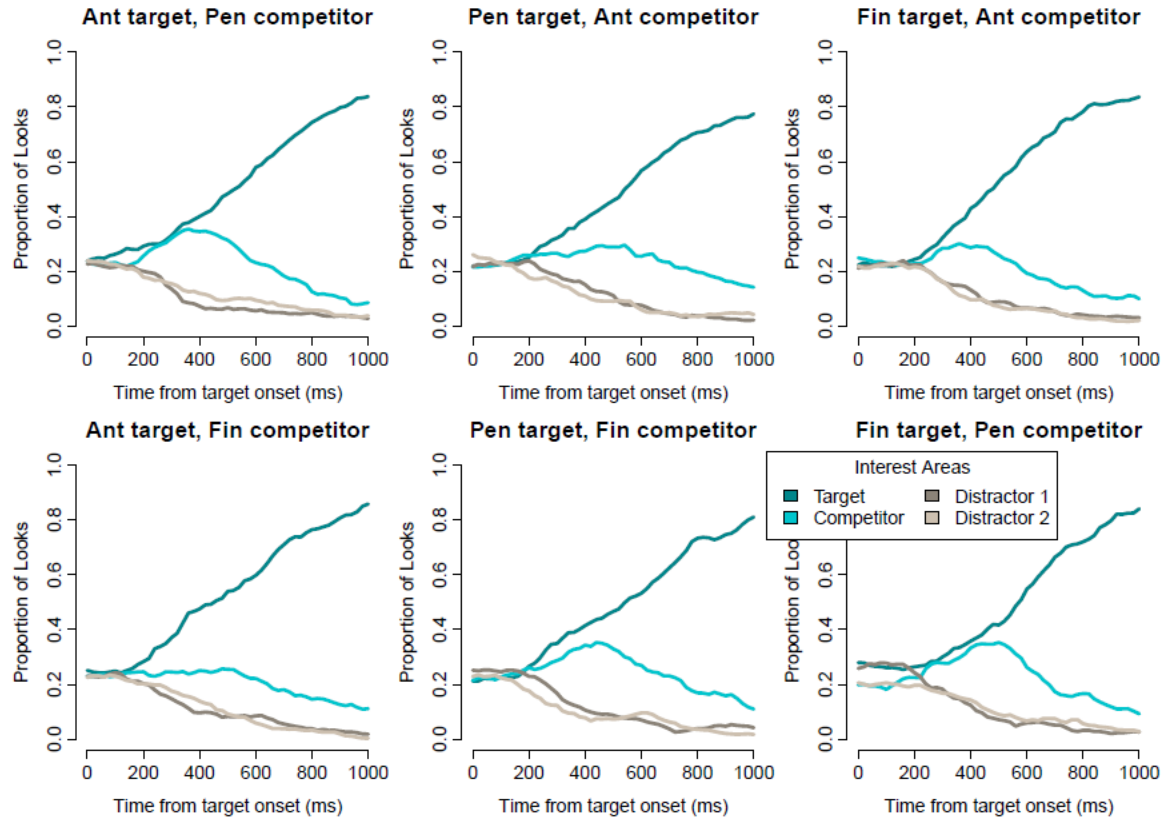
Differences between pairs of conditions predicted to show asymmetrical effects by underspecification in Experiment 2



Note. Each panel contrasts the relative proportions of looks to target and competitor between two conditions. Panels in the top row display the two target-competitor curves together. The bottom row displays the corresponding difference curves and associated 95% CI. Each column corresponds to one pair of stress positions. Antepen = antepenultimate-syllable stress, Pen = penultimate-syllable stress, Fin = final-syllable stress, targ = target

Figure 5

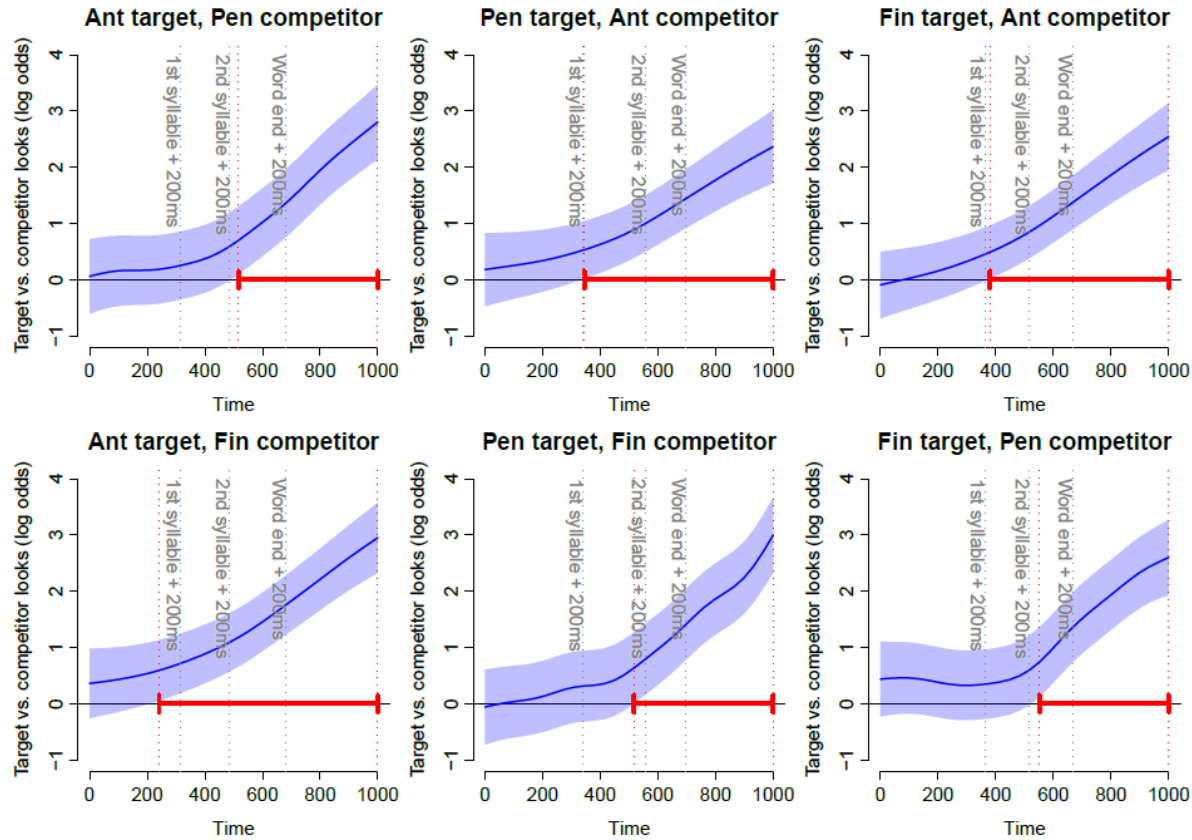
Proportion of looks to interest areas over time in each stress contrast condition in Experiment 3



Note. Ant = antepenultimate-syllable-stress; Pen = penultimate-syllable-stress; Fin = final-syllable-stress.

Figure 6

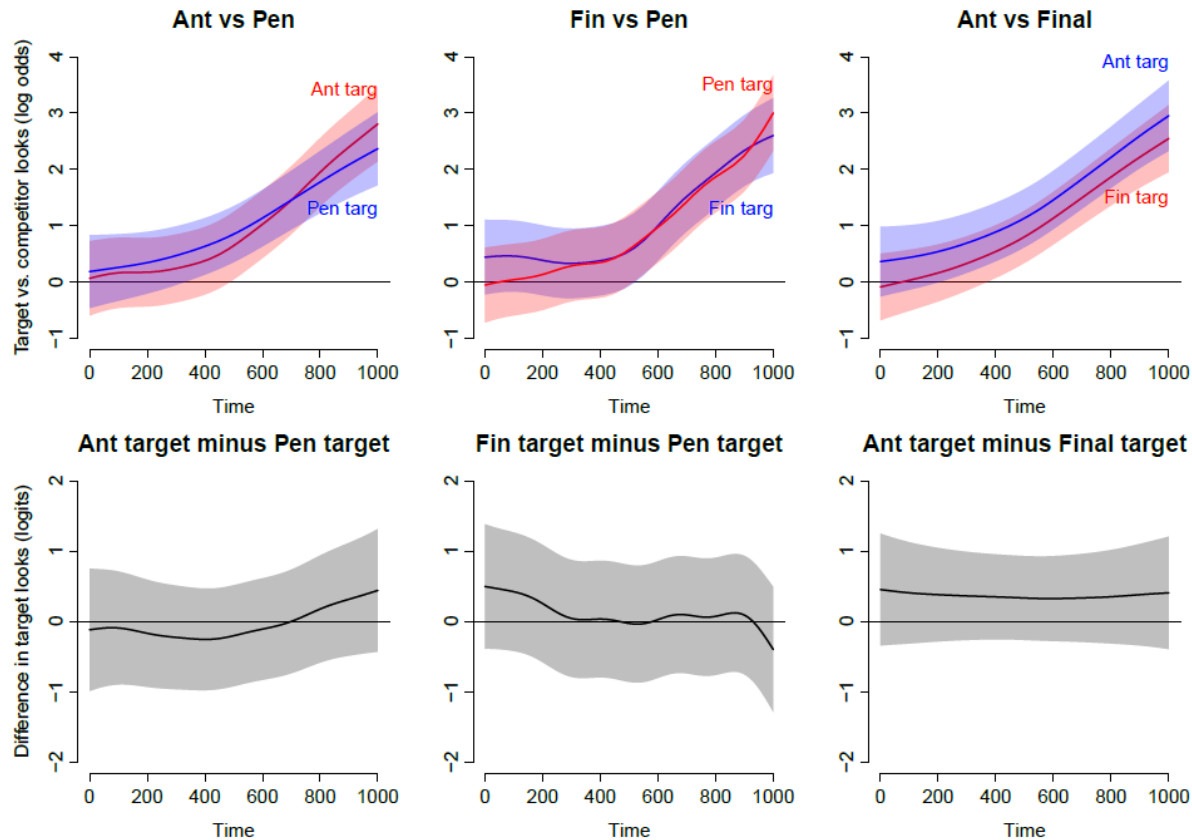
Within-trial dynamics of the difference in looks to target vs. competitor for each stress contrast condition in Experiment 3



Note. Zero time marks the onset of the spoken word. Time intervals of significant difference in looks between targets and competitors are marked in red. The gray dotted vertical lines mark the average location of target syllable boundaries plus a 200-ms offset.

Figure 7

Differences between pairs of conditions predicted to show asymmetrical effects by underspecification in Experiment 3



Note. Each panel contrasts the relative proportions of looks to target and competitor between two conditions. Panels in the top row display the two target-competitor curves together. The bottom row displays the corresponding difference curves and associated 95% CI. Each column corresponds to one pair of stress positions. Antepen = antepenultimate-syllable stress, Pen = penultimate-syllable stress, Fin = final-syllable stress, targ = target