Interspeech 2024
1-5 September 2024, Kos, Greece

Quantity-sensitivity affects recall performance of word stress

Constantijn Kaland[1] *and Maria Lialiou*[2]

[1]Institute of Linguistics-Phonetics, University of Cologne, Germany
[2]Department of German Language and Literature I, University of Cologne, Germany
ckaland, mlialiou {@uni-koeln.de}

Abstract

Previous studies showed that word stress patterns need to be lexically stored when they are highly variable. Hence, listeners are better at memorizing stress patterns in recall tasks when their native language has variable stress patterns than when their native language has fixed stress patterns. The current study hypothesizes that in quantity-sensitive stress languages, segmental information facilitates stress perception and therefore makes smaller demands on lexical storage than in languages with quantity-insensitive stress patterns. This prediction is tested in two stress recall tasks with German (var. stress, q-sensitive) and Greek (var. stress, q-insensitive) listeners. Results show that German listeners indeed perform worse than Greek listeners. The outcomes provide an important novel perspective on the interaction between segmental and suprasegmental information in speech perception and nuance lexical statistics accounts of word stress.

Index Terms: word stress, lexical storage, prosody, stress deafness, German, Greek

1. Introduction

Word stress, the occurrence of a single acoustically most prominent syllable in a word [1], serves a number of communicative functions: listeners use their expectations of the position of the stressed syllable to detect word boundaries [2]. Furthermore, listeners may reduce the number of possible word candidates on the basis of stress information, which helps them with word identification. The extent to which listeners actually make use of these functions, depends largely on how they store stress information in their mental lexicon, which is in turn determined by the stress pattern variation in their language. Languages that make use of word stress vary largely in the way the position(s) of the stressed syllable are determined. The position of the stressed syllable may be rather fixed, in the sense that it predominantly occurs on the same syllable position (e.g., the penult, as in Polish; [3]) or always on the same syllable (e.g., Finnish; [4]). In other languages, the position of the stressed syllable varies among two, three or even more positions (e.g., ante-penult, penult or ult, as in Spanish; [5]).

Stress recall tasks have been carried out with languages that differ largely in the variability of their stress patterns (e.g., [6],[7],[8],[9]). It has been shown that speakers of languages with more variable stress position perform better on recalling sequences of words with different stress patterns than those of languages with fixed stress patterns. The (in)sensitivity to varying locations of word stress in tasks which involve processes like memorizing and recalling these contrasts has been referred to as the stress 'deafness' effect (e.g., [6]). The degree to which the stress position is predictable was found to be the crucial underlying factor determining listeners' recall performance [8], corroborating lexical statistics accounts in phonology [10]. That is, for languages with a large number of exceptions to the regular stress position(s), such as Spanish, it was found that listeners were better in recalling the patterns [8]. Other studies have furthermore shown that the type of cues specified in the lexicon does not matter, as long as there is variability of word prosodic patterns in the lexicon. That is, Japanese listeners, used to varying positions of lexical tone, were shown to outperform listeners from languages in which no stress or tone was lexically specified (i.e. Indonesian, Persian and French) [11].

The study of word stress heavily relies on keeping segments separate from suprasegments, i.e., distinguishing between vowels and consonants as syllable structure and the acoustic cues to word stress (i.e. duration, spectral tilt) that may affect them. The fact that there are languages in which the segmental makeup of syllables determines the variability in the position of stress, i.e. phonological *weight* in languages with quantity-sensitive stress, suggests that there is a limit to which this theoretical distinction can be maintained. That is, in quantity-sensitive stress languages segmental and suprasegmental information of lexical items are intertwined. Many stress languages are quantity-sensitive to some degree [12]. In other languages the variability in stress positions is minimally or not determined by weight, but rather by other factors, such as morphological processes (Greek; [13]). Thus, in quantity-sensitive languages, storing the segmental information of a lexical item entails storing information needed to locate the stressed syllable. Additional phonological rules might still be needed to account for some variability in stress patterns, if the language has them. We predict, therefore, an inverse relationship between quantity-sensitivity and recall performance. Thus, listeners are expected to have worse recall performance when the syllable structure of their languages helps them to predict the stress position. We investigate this by replicating the stress recall task from [11] with German and Greek listeners.

1.1. German

Several studies have described stress assignment rules in German, for an overview see [14]. All studies agree that German stress is quantity-sensitive and is assigned within a three syllable window counted from the right word edge; i.e. the antepenult (APU), penult (PU) or ultimate (U) syllable [15]. In a nonce word production task [14] it was indeed found that the rhyme structure of the syllable determines the stress position to a large extent. That is, syllables that are closed or have a diphthong tend to attract stress (APU or U), whereas open syllables tend to default to PU stress. In addition, when the onset

10.21437/Interspeech.2024-182

structure of the PU syllable is more complex, a trend in attracting PU stress was observed. In a corpus study ([15] based on CELEX2 data [16]), German words were shown to have the following frequencies of the respective stress positions: 23.4% (APU), 48.8% (PU) and 27.7% (U), largely comparable to an earlier study [17]. Acoustic studies have further indicated that duration is the primary correlate of German word stress, mainly as the result of vowel reduction [18].

1.2. Greek

The Greek stress system is determined by morphology, meaning that stress is lexically-encoded. Whereas stress is lexically specified for many individual morphemes, the language also has a phonologically default stress on the APU, emerging in the absence of the lexically specified stress (also known as non-lexically inflicted stress) [19] [20] [21] [22] [23]. Thus, it has been suggested that Greek stress is mainly assigned by a "complex interaction between morphology and phonology" [24, p.11]. All previous work agrees on stress assignment in a three syllable window (APU, PU or U). The predictability of the word stress position depends largely on word class. Adjectives generally show APU stress, nouns vary in stress position depending on declension and inflection, and verbs show complex stress assignment based on an interaction of conjugation class, distance from the stem and tense [13]. For example, an additional syllable *e-* is inserted such that the obligatory APU stress can be realised in past tense forms that have two syllables otherwise. This strong effect of morphology has lead to the proposal to treat Greek stress as a morpheme [25]. Lexical database frequencies show the following division of stress positions: 32.6% (APU), 38.4% (PU) and 28.2% (U) [26]. Acoustic analyses showed that the combination of duration and amplitude (as integral) is the strongest stress correlate in Greek [24].

1.3. Research question and hypotheses

In sum, German and Greek word stress seem similar when observing their surface distributions; both languages have variable stress in a three-syllable window with comparable distributional frequencies. Both languages have the highest relative amount of PU stress, although in German this tendency is stronger than in Greek (cf. [15] and [26]). Importantly, for both languages, (considerable) individual differences among speakers were reported on the production of word stress and its cues. The crucial difference between German and Greek word stress pertains to quantity-sensitivity, which is mainly found in the former. The question underlying the research in this paper is therefore whether differences in quantity-sensitivity affect listeners' memorization and recall of stress patterns.

We hypothesize that in quantity-sensitive stress systems such as the German one, the segmental material facilitates the detection of the position of the stressed syllable. This can be seen as a form of redundancy in the speech production system; in case the prosodic cues would be missing (e.g., in noisy conditions) listeners might be able to rely on the segments in quantity-sensitive stress languages. Accordingly, there is less pressure on knowing the stress pattern 'by heart' in these languages. If stress is entirely unpredictable from the segments, as in Greek, listeners can only rely on their separately stored lexical stress information, and are consequently fluent in doing so. Note that the writing system in the latter type of language reflects this; Greek spelling requires an accent mark on the vowel to indicate the stressed syllable. Spanish, also having highly variable stress, is similar in this respect as it also marks stressed syllables in writing [27]. To investigate the above mentioned issue, two sequence recall tasks were carried out with German and Greek listeners respectively.

2. Methodology

If not explicitly mentioned below, all aspects of the setup, stimuli and procedure are identical to the task in [11]. Note that the speakers of the stimulus materials therefore have Dutch and Persian as native languages (explanation in [11]). A web-based and on-site version of the experiment and stimuli used in this study is available for testing with other languages [28].

2.1. Participants

In total 26 native speakers of German completed the experiment; 14 female, 12 male, mean age: 31, age range: 20-64. In addition, 24 native speakers of Greek completed the experiment; 14 male, 10 female, mean age: 32, age range: 18-57. All participants carried out the experiment in a quiet room using headphones. None of them reported having hearing problems.

2.2. Stimuli

Stimuli consisted of 48 audio files with spoken versions of nonce words, balanced for *contrast* (phoneme, stress), *speaker gender* (male, female), and *speaker language* (Dutch, Persian). Each contrast was generated by two nonce words (A and B). That is, the phonemic contrast occurred between [ˈmu.ku]^{A} and [ˈmu.nu]^{B}, and the stress contrast occurred between [ˈnu.mi]^{A} / [nu.ˈmi]^{B}. Thus, the phonemic contrast was based on a consonantal difference in the second unstressed syllable: [ku] or [nu]. The stress contrast was based on a combination of cues: duration, intensity, f0, and formant frequencies (with higher values obtained for the stressed syllable than for the unstressed syllable). Detailed acoustic characteristics are described in [11, Table 2] and f0 contours are analysed in detail in [29, p. 161]. The nonce words were indeed meaningless in German and Greek. There were three versions of each contrast/gender/language combination, resulting in 4 * 2 * 2 * 3 = 48 words.

2.3. Design

The task tested the recall of sequences consisting of three, four or five nonce words (five sequences per length). The sequences were designed in such a way that there was either a phoneme or a stress contrast present. No sequence was a repetition of a single version of an item. The sequences used for each contrast are listed in Table 1 and are identical to the ones in [11].

Table 1: *Nonce word sequences per length (N words) as used in the test block with A and B referring to each item member of the contrast (either phoneme or stress).*

Length	Sequences				
3	AAB	ABA	ABB	BAA	BAB
4	ABAA	ABBA	BAAB	BABB	BBAB
5	ABAAB	ABABB	ABBAB	BABAA	BABBA

2.4. Procedure

Before the start of the experiment, participants were asked which language(s) they spoke to assess their level of native language proficiency and multilingualism. Thereafter, they received instructions about the experimental task. The experiment was run using PsyToolkit ([30]; [31]). The procedure described in [11, p. 6] was implemented in the PsyToolkit online environment (html-based), such that the experiment could be run remotely via a web-browser ([28]). Instructions were translated into the respective languages. The contrast were tested in two block orders, either phoneme contrasts first or stress contrasts first, which were fully randomized.

Each part consisted of a training session, a warm-up round and a test block. In the training session, participants learned to associate member A of the contrast with key '1' and member B with key '2'. This was done by instructing them to press any of the keys and listen to all items associated to that key (12 items for each key). Then, participants pressed any of the keys and listened to a single item chosen randomly from the set associated to that key. They could press either key as often as they wanted until they indicated that they had learned the item-key associations. After that, participants heard a single randomly chosen item and pressed the key that they just learned was associated to that item. Feedback was given on whether participants correctly or incorrectly identified the item. The training session ended after correct identification of eight items.

In the warm-up round participants listened to sequences of two nonce words (AA, AB, BA, BB - for each language set). Their task was to identify the sequence by pressing the correct key combination. They needed to complete all eight sequences flawlessly before they could continue. Each sequence was presented (repeatedly) until participants had correctly identified it.

In the test block, participants identified sequences of three, four or five nonce words (Table 1), which were presented in random order, different for each participant. Within each sequence, items were chosen randomly from the same speaker gender and language and therefore only varied in contrast and version (no items were repeated within sequences). Participants received no feedback on their performance during the test block. A test block consisted of 30 trials (5 sequences * 3 lengths * 2 languages). The data of participants who gave a correct answer in less than half of all trials or who did not learn the language before the age of 7 was discarded (German: four participants, Greek: five participants; excluded from the counts in Sec. 2.1).

2.5. Statistical analysis

A generalized linear mixed model (GLMM) analysis was run on the correctness scores (0 for incorrect, 1 for correct), with the interaction of *experiment language* (two levels: German, Greek) and *contrast* (two levels: phoneme, stress), and *sequence length* (successive differences contrast coded, three levels: 3, 4, 5) as fixed factors. The model also included random intercepts per participant and item. Another model including the additional fixed factors *speaker language* (two levels: Dutch, Persian), *speaker gender* (two levels: male, female), *block order* (two levels: phoneme first, stress first) was also run to check for any effects originating from the stimuli or procedure. None of these were found to be significant and therefore not included in the final model. Thus, for the final model the following formula was used: `glmer(score ~ lg * contrast + seq.length + (1|item) + (1|pp), data, contrasts = list(seq.length = contr.sdif), family = binomial)`

3. Results

The results are given in Table 2, Figure 1 and Table 3. They show that German listeners performed significantly worse than Greek listeners. An effect of contrast was found in that phoneme recall was better than stress recall. The effects of sequence length showed a significant decrease in recall performance with each longer sequence. The interaction effect of language and contrast indicated that the difference between phoneme and stress recall was larger in German than in Greek. It can furthermore be observed that phoneme recall showed a steady decline over the sequence lengths in both languages, whereas the stress recall showed a comparable steady decline only between length 3 and length 4, and a much steeper decline between length 4 and length 5 (Figure 1).

Table 2: *Mean recall scores (and standard deviations) for language, sequence length and contrast.*

language	seq.length	stress	phonemic
German	3	0.77 (0.42)	0.85 (0.36)
	4	0.74 (0.44)	0.79 (0.41)
	5	0.53 (0.50)	0.72 (0.45)
Greek	3	0.92 (0.28)	0.93 (0.26)
	4	0.88 (0.33)	0.90 (0.31)
	5	0.73 (0.45)	0.80 (0.40)

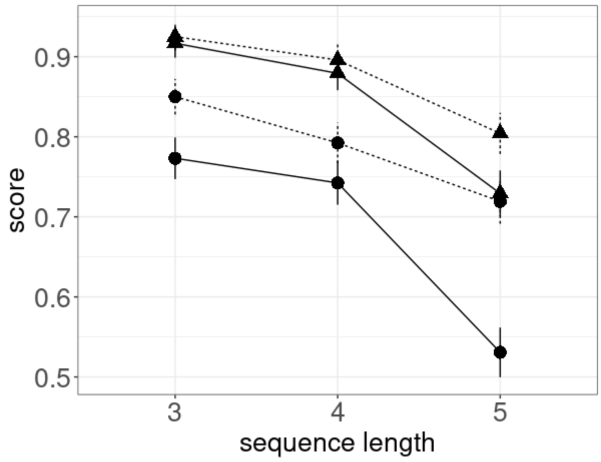

Figure 1: *Mean recall scores (std errors) per sequence length.* ● *German* ▲ *Greek* - - - *phonemic* — *stress.*

Table 3: *Overview of the factors and effects in the final GLMM.*

	β	SE	z	$p <$
(Intercept)	0.96	0.25	3.79	0.001
lg:Greek	1.16	0.37	3.18	0.01
contrast:ph	0.75	0.14	5.48	0.001
seq.length: 4-3	-0.39	0.19	-2.00	0.05
seq.length: 5-4	-0.96	0.18	-5.23	0.001
lg:Gr*contr:ph	-0.42	0.21	-2.00	0.05

4. Conclusions

The current study shows that German and Greek listeners differ in their ability to recall stress patterns at the word level. German listeners were found to perform overall worse than Greek listeners, even though both languages have variable stress patterns that show highly similar distributions in a three-syllable window. It was hypothesized that German and Greek listeners would differ because their languages differ in the extent to which the segments facilitate prediction of the stress position.

This hypothesis was confirmed in that German listeners, who are used to quantity-sensitive stress, perform worse than Greek listeners, who cannot predict the stress position from the segments and consequently are better at storing stress information entirely separate from the segmental information.

The results give a new insight into the interplay of segmental and prosodic information at the word level. Previous work showed that the number of exceptions to the default stress pattern determines listeners' recall ability [8]. This was shown by comparing Spanish to (South-Eastern) French, Finnish, Hungarian and Polish. Among these languages, Spanish is the one with the largest number of exceptions (17%) to the default PU stress. The other languages in the study mainly had a single stress position, with Polish allowing a minimal amount of variability (0.1%) [8]. In the current study, we have compared two languages with highly variable stress patterns and it can be questioned to what extent they have a 'default' stress position when considering their surface distributions (Section 1.1 and 1.2). This has been claimed for German to be PU [32], but see [14] for a nuanced view. Even though both languages show a majority of PU stress patterns, they make up less than half of all possible patterns, making the number of exceptions to that pattern larger than the number of the alleged 'default' ones. While it might be that the phonological (i.e. underlying) default is indeed PU in German, this claim cannot be maintained for Greek due to its interaction with morphology. That is, in Greek, some word classes have their own default pattern. Lexical statistics accounts (e.g., [10]; [8]) focus heavily on the number of exceptional (non-default) stress patterns as the main determiner of stress recall performance. As the current study shows, this lexical statistic is not the only factor determining the way stress patterns are stored. It also matters to what extent word stress variability is determined by phonological weight.

Thus, we can reconcile the current results with the previous literature (e.g [8]; [11]) by dividing the free-fixed stress continuum of languages into three stages depending on the role of segmental information in determining the stress position (Figure 2): 1) variable quantity-*in*sensitive stress languages in which the segments are not useful to predict the stress position and hence listeners need to store the patterns independently (e.g., Greek); 2) variable quantity-sensitive stress language in which the segments are useful to predict the stress position and listeners rely on them accordingly (e.g., German); 3) fixed stress languages, regardless of whether they are quantity-sensitive, this feature does not play a role in storage and recall, as the stress position does not vary (e.g., Finnish, see [8]). From the perspective of recall performance, the difference between the respective variable stress languages (e.g., German and Greek in the current study) is likely smaller than the difference between fixed and variable stress languages (e.g., Finnish and Spanish in [8]), hence the effect of weight-sensitivity applies only to 'free' stress end of the scale (Figure 2). Inspection of the Greek results (Table 2) reveals a (near-)ceiling score for the 3- and 4-word sequences with minimal difference between segmental and prosodic recall. It is therefore likely that listeners of languages such as Greek are among the ones that perform best in this task, mainly showing a decline in performance due to the number of items that are needed to recall rather than due to the type of contrast.

It is furthermore interesting to observe that phoneme recall is better in Greek than in German. This observation provides another indication that the storage and recall of phoneme and stress information are heavily correlated, posing limits on the traditional theoretical divide between segments and suprasegments. A priori, there is no difference to expect in the extent to which listeners of different languages are able to recall segments, as all languages make use of them. This expectation was also not met in [11], in which Dutch and Japanese listeners outperformed Persian, French and Indonesian ones on stress *and* phoneme recall. There are more trends in the data of the current study (Figure 1) and in [11] that suggest strong correlations in the types of recall. That is, in languages for which listeners show overall worse stress recall, the decline in recall performance for longer sequences, the decline of stress recall performance compared to the decline of phoneme recall performance, as well as the overall difference between stress and phoneme recall performance are all steeper/larger than in languages for which listeners show better stress recall performance.

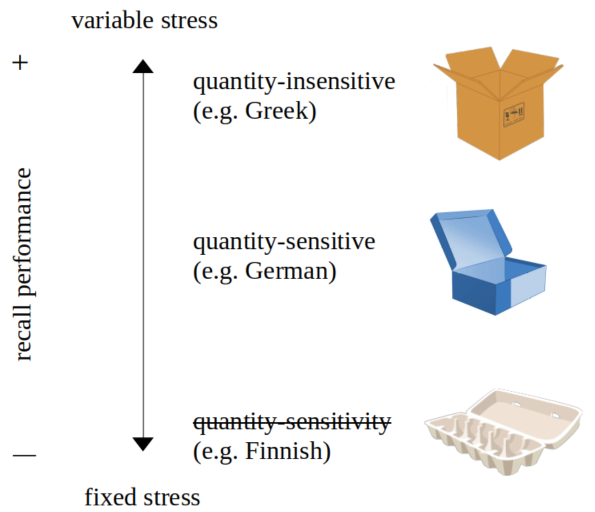

Figure 2: *Schematic illustration of the interplay between (left-to-right) stress recall performance, pattern variability, and quantity-sensitivity with their analogous carriers as cardboard box types (illustrations from [33]).*

Commonly, segmental structure is seen as the carrier of prosodic patterns (e.g., [34]). Crucially, there are various ways in which the segments may interact with prosody, and depending on the language, some segmental structures have a clearer specification to carry prosodic events than others. Extending the carrier-analogy to cardboard boxes (≈ segmental structure) and how predictable the content (≈ stress) they carry is, the three-way distinction above could be seen as the difference between an egg carton (highly predictable content; eggs), a shoe box (somewhat predictable; i.e. footwear, which could be sandals, sneakers, saddle shoes, etc.), and a moving box (unpredictable content), see Figure 2. As shown by the current study, speech perception mechanisms exploit the information provided by the carrier to the extent the language allows it, and only store suprasegmental information if needed. To further corroborate these findings, more languages with varying degrees of quantity-sensitivity should be investigated.

5. Acknowledgements

Research for this paper was funded by the German Research Foundation (DFG) – Project-ID 281511265 – SFB 1252. The authors thank Sophia Meyer, Alessandra On and Jonathan Reich for the recruitment of participants. The German data are also reported in [35]. Data and script are stored at https://osf.io/usrcd/.

6. References

[1] L. Hyman, "Word-prosodic typology," *Phonology*, vol. 23, no. 02, pp. 225–257, 2006.

[2] A. Cutler, *Native listening: language experience and the recognition of spoken words*. Cambridge, Mass: The MIT Press, 2012.

[3] U. Domahs, J. Knaus, P. Orzechowska, and R. Wiese, "Stress "deafness" in a Language with Fixed Word Stress: An ERP Study on Polish," *Frontiers in Psychology*, vol. 3, 2012.

[4] F. Karlsson, *Finnish: an essential grammar*, 3rd ed., ser. Routledge essential grammars. London New York, NY: Routledge, Taylor & Francis Group, 2015.

[5] A. Domínguez and F. Cuetos, "The contrastive value of lexical stress in visual word recognition: Evidence from Spanish," *Psicothema*, no. 30.3, pp. 276–282, 2018.

[6] E. Dupoux, C. Pallier, N. Sebastian, and J. Mehler, "A Destressing "Deafness" in French?" *Journal of Memory and Language*, vol. 36, no. 3, pp. 406–421, 1997.

[7] S. Peperkamp and E. Dupoux, "A typological study of stress 'deafness'," in *Laboratory Phonology 7*, C. Gussenhoven and N. Warner, Eds. De Gruyter Mouton, 2002, pp. 203–240.

[8] S. Peperkamp, I. Vendelin, and E. Dupoux, "Perception of predictable stress: A cross-linguistic investigation," *Journal of Phonetics*, vol. 38, no. 3, pp. 422–430, 2010.

[9] M. Lialiou, A. Bruggeman, S. Vella, S. Grech, P. Schumacher, and M. Grice, "Word-level prominence and "stress deafness" in Maltese-English bilinguals," in *Proceedings of the 20th International Congress of Phonetic Sciences*, R. Skarnitzl and J. Volín, Eds. Prague (CZ): Guarant International, 2023, pp. 132–136.

[10] J. B. Pierrehumbert, "Probabilistic Phonology: Discrimination and Robustness," in *Probabilistic Linguistics*, R. Bod, J. Hay, and S. Jannedy, Eds. The MIT Press, 2003, pp. 171–222.

[11] H. Rahmani, T. Rietveld, and C. Gussenhoven, "Stress "Deafness" Reveals Absence of Lexical Marking of Stress or Tone in the Adult Grammar," *PLOS ONE*, vol. 10, no. 12, p. e0143968, 2015.

[12] R. Goedemans, J. Heinz, and H. Van der Hulst, "StressTyp2: Home Page," Feb. 2020.

[13] G. Drachman and A. Malikouti-Drachman, "Greek word accent," in *Word Prosodic Systems in the Languages of Europe*, 1999th ed., ser. Eurotyp, G. Bossong, B. Comrie, Y. Matras, and H. V. D. Hulst, Eds. Berlin, New York: Mouton de Gruyter, Mar. 1999, vol. 4, pp. 897–946.

[14] T. B. Röttger, U. Domahs, M. Grande, and F. Domahs, "Structural Factors Affecting the Assignment of Word Stress in German," *Journal of Germanic Linguistics*, vol. 24, no. 1, pp. 53–94, 2012.

[15] U. Domahs, "Untersuchungen zum Wortakzent im Deutschen und Niederländischen," Ph.D. dissertation, Heinrich-Heine-Universität, Düsseldorf, 2003.

[16] H. Baayen, R. Piepenbrock, and L. Gulikers, "CELEX2," 1995.

[17] C. Féry, "German Word Stress in Optimality Theory," *The Journal of Comparative Germanic Linguistics*, vol. 2, no. 2, pp. 101–142, 1998.

[18] G. Dogil, "The phonetic manifestation of word stress in Lithuanian, Polish, German, and Spanish,," in *Word Prosodic Systems in the Languages of Europe*, ser. Eurotyp, H. v. d. Hulst, Ed. Berlin: De Gruyter, 1999, vol. 4, pp. 273–311.

[19] A. Malikouti-Drachman and G. Drachman, "Tonismos sta Elinika (Stress in Greek)," *Studies in Greek Linguistics*, vol. 9, pp. 12–143, 1999.

[20] A. Ralli and L. Touradzidis, "Ipolo?istiki Epekser?asia tu Tonismu tis Elinikis (Computational Processing of Stress in Greek)," *Studies in Greek Linguistics*, vol. 12, pp. 273–289, 1992.

[21] A. Revithiadou, *Headmost accent wins: head dominance and ideal prosodic form in lexical accent systems*, ser. LOT. The Hague: Holland Academic Graphics, 1999, no. 15.

[22] A. Protopapas, S. Gerakaki, and S. Alexandri, "Lexical and default stress assignment in reading Greek," *Journal of Research in Reading*, vol. 29, no. 4, pp. 418–432, 2006.

[23] I. N. Topintzi and E. Kainada, "Acronyms and the placement of default stress in Greek," in *Proceedings of the 10th International Conference on Greek Linguistics*, 2012, pp. 559–570.

[24] Arvaniti, "The phonetics of stress in Greek," *Journal of Greek Linguistics*, vol. 1, no. 1, pp. 9–39, 2000.

[25] M. Van Oostendorp, "Stress as a proclitic in Modern Greek," *Lingua*, vol. 122, no. 11, pp. 1165–1181, 2012.

[26] A. Kyparissiadis, W. J. B. Van Heuven, N. J. Pitchford, and T. Ledgeway, "GreekLex 2: A comprehensive lexical database with part-of-speech, syllabic, phonological, and stress information," *PLOS ONE*, vol. 12, no. 2, p. e0172493, 2017.

[27] T. Mathews, *A complete guide to written accent marks in Spanish with exercises*, third edition ed. San Francisco: Orgle Press, 2013.

[28] C. Kaland, "Stress 'deafness' task PsyToolkit," 2023, https://osf.io/xwmka/.

[29] A. Bruggeman, "Lexical and postlexical prominence in Tashlhiyt Berber and Moroccan Arabic," PhD Thesis, Universität zu Köln, 2020.

[30] G. Stoet, "PsyToolkit: A software package for programming psychological experiments using Linux," *Behavior Research Methods*, vol. 42, no. 4, pp. 1096–1104, 2010.

[31] ——, "PsyToolkit: A Novel Web-Based Method for Running Online Questionnaires and Reaction-Time Experiments," *Teaching of Psychology*, vol. 44, no. 1, pp. 24–31, 2017.

[32] W. J. M. Levelt, A. Roelofs, and A. S. Meyer, "A theory of lexical access in speech production," *Behavioral and Brain Sciences*, vol. 22, no. 01, 1999.

[33] J. Phillips and B. Harrington, "Open Clip Art Library," 2004, https://openclipart.org.

[34] C. Gussenhoven and A. Chen, Eds., *The Oxford Handbook of Language Prosody*, 1st ed. Oxford University Press, 2020.

[35] C. Kaland, "When recall gets stressful: comparing Papuan Malay and German listeners' lexical storage of word stress." *Laboratory Phonology*, accepted.