Delineating H* and L+H* in Southern British English

Jiseung Kim, Na Hu, Stella Gryllia, Riccardo Orrico, Amalia Arvaniti

Radboud University, The Netherlands

jiseung.kim@ru.nl, na.hu@ru.nl, stella.gryllia@ru.nl, riccardo.orrico@ru.nl, amalia.arvaniti@ru.nl

Abstract

In English, H* is said to encode new information and be realized as high pitch, while L+H* encodes degrees of contrastivity and realized as rising pitch. However, empirical evidence for this distinction is sparse, especially in Southern British English (SBE). To gain a better understanding, we examined 2,126 words with high and rising accents in an SBE corpus of unscripted speech. The accents were separately annotated for (i) f0 shape (high or rising) in PRAAT and (ii) pragmatic function (corrective, contrastive, and non-contrastive) based on written transcripts only. The data were modelled using Functional Principal Component Analysis and GAMMs. Phonetically H* and L+H* were distinct: H* was realized as a fall and L+H* as a rise-fall. However, these shapes did not neatly map onto pragmatic functions: corrective accents were likely to be L+H*s and were, thus, distinct from non-contrastive accents, which were likely to be H*s, but there was no phonetic difference between contrastive accents and non-contrastive accents and corrective accents, indicating that the mapping between phonetic form and pragmatic function was not one-to-one. By separating the shape- from the meaning-based annotation, the relation between the f0 shapes and the pragmatic functions of these accents is thus better understood.

Index Terms: f0, pragmatics, autosegmental-metrical theory, pitch accents, British English

1. Introduction

The distinction between H* and L+H* in English is a widely known source of disagreement within the autosegmental-metrical theory of intonational phonology (henceforth AM). In [1], [2] and [3], these accents were argued to be phonologically distinct, at least in American English. The substantial rising f0 leading up to the H* target in L+H* accents is considered as evidence for a L target ([4]), in contrast to H* which has a shallower rise. Further, the accents are said to differ in terms of their information-structure (IS) related functions: by and large, [1–3] argue that H* encodes new information, while L+H* is used to mark contrastive information. This dichotomy in terms of form and function does not completely hold up to empirical evidence, however. Perception studies [5], [6], [7] provide evidence that both H* and L+H* can be used to signal both new and contrastive referents. As a result of these empirical findings, the Mainstream American English (MAE) ToBI authors [4] conclude that both accents can be used in a variety of contexts and note that different accent shapes may arise from a specific context. These ambivalent findings and conclusions in studies on American English cast doubt on the H* ~ L+H* contrast and provide some support for [8] (inter alia) who argues that L+H* is an emphatic rendition of H*. According to this view, H* and L+H* form a continuum of realizations of a single accent category, rather than being separate phonological categories. This question has not been empirically tested in Southern British English as extensively as in American English, a gap the present study intends to fill.

In addition to the relative paucity of empirical data, accounts on British English intonation differ in their descriptions of the shapes and pragmatic intent of these accents. Regarding shape, [9] and [10] describe low and high falls but not rising accents (comparable to the typical description of L+H*). [11] examined nuclear accents in declarative utterances produced by London and Cambridge speakers, as part of the IViE (Intonational Variation in English) corpus and concluded that they were mostly falls. Examining the same corpus, [13] observed that many accented syllables were produced with high pitch but also noted that large rising pitch excursions were rare. Still others have argued that accents described as either H* or L+H* include a low tonal target, and are, thus, rises, [8], [13]. In short, there is little consensus on accent phonetics.

There is similarly limited consensus with respect to pragmatic function. For instance, within the British School of intonation, which is largely based on impressionistic data, high and low falls are sometimes seen as a continuum indicating speaker involvement [10], or as two distinct categories, with high falls described with labels such as “lively” [9]. Empirical research on the pragmatic functions of pitch accents in SBE is relatively limited and has largely focused on examining the role of H* and L+H* in marking new information, e.g., [14].

In this study, we investigated how putative H* and L+H* accents were produced by native speakers of Southern British English (henceforth, SBE), by testing (1) whether they differ phonetically and (2) how their phonetic realization and pragmatic function are related. These research questions derive from AM work on the phonology of H* and L+H* [1], and their pragmatics [2]. (We note that although [1] and [2] refer specifically to American English, their analysis is often seen as appropriate for British English as well; cf. [8].) Specifically, following [1] we would expect to see accents with distinct shapes, viz. high (H*) and rise (L+H*) in the corpus. Further, if [2] are correct in their analysis of the pragmatics of these accents, we would anticipate that, by and large, H* accents would be associated with items conveying new information, while L+H* accents would be associated with items conveying contrastive and corrective information. In other words, based on [1] and [2], we predicted that we would observe reasonable correspondence between the phonetic realization and pragmatic function of the accents.

2. Methods

2.1. Participants

Eight native speakers of SBE (5 F) participated in the study. They were from the Greater London area and between 18 and

54 years of age (median: 25) at the time of participation. They reported no history of speech or hearing disorders.

2.2. Recording procedure

The recordings were made in quiet rooms in private houses in Canterbury, UK, using an H4N recorder with the built-in stereo microphones at the sample rate of 44.1 kHz at 24 bits. Each recording session consisted of several tasks, including the three tasks the data from which are examined in the present study: *story cube task*, *map task*, and *unusual objects task*. In the story cube task, each participant threw six Story Cube dice (www.storycubes.com) three times and told three short stories based on the icons on the dice. The aim of the task was to generate discourse that involved largely new information in the form of narratives that included elements of surprise (due to the unusual combinations of the dice icons). In the map task (groups.inf.ed.ac.uk/maptask), an instruction giver described the route on their map using the map landmarks, while a follower recreated the route on their map which had a slightly different set of landmarks. All participants took up the role of instruction giver and follower with different sets of maps. The aim of the task was to generate conversation that included several instances of corrective and contrastive information (due to the discrepancies between the instruction giver and follower maps). Finally, in the unusual objects task, participants in pairs were given a bag containing seven unusual objects (e.g., an egg separator) and discussed their potential functions. The aim of the task was to provide a balanced mix of accented items conveying new, contrastive, and corrective information.

2.3. Data annotation

To prepare the data for annotation, the audio recordings were transcribed verbatim before being split into utterance-sized files. Crucially, the phonetic and pragmatic categorization of the accents were determined independently of one another, to ensure that the evaluation of the pragmatic functions of accented items was not influenced by their auditory impression, and the description of the accents' phonetic categorization was not guided by the pragmatic context in which they appeared.

The phonetic annotation was done by inspecting visually and auditorily the f0 curves of individual utterances. All words deemed accented with high or rising accents were labeled as carrying a H* or a L+H* using the following criteria: If there was a deliberate f0 dip at the onset of the accented syllable, the word was labeled as L+H*, otherwise as H*; in addition, the rise should take place largely during the accented syllable, so as to exclude L*+Hs. The annotation was carried out by a trained assistant who was a near-native English speaker. Her annotations were inspected by the first author to ensure the criteria were used consistently.

The unit of annotation, as well as the window for the FPCA analysis (see 2.4), was the *prosodic* word, i.e., the accented word and any applicable function words before or after it: for example, '*look at me*', where the accented word is underlined, formed one analysis window. On occasion, additional unaccented material was added to the analysis window to ensure the selected stretch of f0 curve was adequate for FCPA.

The pragmatic annotation was based on the fully punctuated orthographic transcripts. It was undertaken by two of the authors working together, based on a classification schema originally devised by a pragmaticist who is a native SBE speaker. The annotators consciously avoided using their own implicit (or read-aloud) prosody as a means of pragmatic classification. Following [2], constituents that were overt corrections of some previously mentioned item were marked as *corrective*; e.g., in Table 1, speaker B's "lavender" is an overt correction of speaker A's "aubergine"; constituents that generated alternatives or closed sets from which an alternative was chosen and contrasted with the rest were labeled *contrastive*; e.g., in Table 1, "in the kitchen" and "in the lab" were contrasting in a parallel construction. That is, the primary criterion for contrastivity was the discourse context rather than the speaker's intention.

Table 1. *Examples of pragmatic annotation*

Type	Example
Corrective	A: Have you got a purple aubergine right there? B: I've got purple *lavender*.
Contrastive	So they've got to be used for something *in the kitchen*, or something *in a lab*, or something like that.

For each annotation type, an additional author (who was not involved in the original annotation) annotated 8.2% of the data using the same criteria. The unweighted Cohen's Kappa score was calculated based on whether a word was labeled as H*, L+H*, or not labeled as accented, which resulted in very high agreement (0.84, *C.I.* = 0.78 – 0.89). The unweighted Cohen's Kappa score for the inter-labeler reliability of the pragmatic annotation, calculated based on contrastive (inclusive of correctives) vs. non-contrastive, was 0.72 (*C.I.* = 0.62 – 0.82).

The pragmatic labels in the transcripts were imported to the Praat textgrids after these had been annotated with the phonetic labels. Thus, only the items that had already been labelled as H* or L+H* also received a contrastive or corrective label if applicable; any remaining items labelled H* or L+H* were marked as *non-contrastive*. A total of 2,450 prosodic words were annotated for both phonetics and pragmatics. Due to an issue with the time-warping procedure (see 2.4), we excluded from further analysis 323 items in which the temporal landmarks used for time normalization overlapped with the onset or offset of the window of analysis. The summary of the phonetic and pragmatic labels of the remaining 2,127 annotated prosodic words is given in Table 2. (We note that, as FPCA operates on the data prior to linguistic categorization, its outcome is not influenced by the distribution of the data into linguistic categories, thus the uneven samples are unlikely to have had a strong influence on the FPCA output.)

Table 2. *Summary of the annotated items*

	Non-contrastive	Contrastive	Corrective
H*	1600	119	89
L+H*	212	73	33

2.4. Statistical analysis

The f0 curves of the annotated items were smoothed and normalized for speaker and time. Time normalization was done by warping the intervals before, between, and after two temporal landmarks, the onset and offset of the accented vowel. These curves were then submitted to Functional Principal Component Analysis (FPCA; [15]). FPCA returns as the output of the model the dominant modes of variation within the data in the form of Principal Component (PC) curves and the corresponding coefficients (or *scores*) that quantify the independent contribution of the aspects of the curves captured by each PC. The equation given in (1) expresses how each f0 curve in the data is modelled:

$$f(t) \approx \mu(t) + s1 \times PC1(t) + s2 \times PC2(t) + \ \dots \qquad (1)$$

FPCA was chosen as the analysis method because it breaks down complex f0 curves by identifying independent components that can be further examined. Importantly, FPCA is a data-driven approach as it does not require the input data to have a priori categorization motivated by a theoretical account. Therefore, if the coefficients of a principal component are used as the dependent variable in a statistical model and the model shows they are linked to the – independently undertaken – categorization of the f0 curves into accents, such a finding tells us which component(s) of the f0 curves differentiate accents in terms of their phonetic realization. Additionally, by using independent annotations for phonetics and pragmatics, the PC scores can be used to compare how the phonetically-labeled accents relate to pragmatically-labeled accents. , To this effect, the PC scores were analyzed in a series of Linear Mixed-Effects Models (LMEMs; [16]) in R ([17]), with PHONETICS (H*, L+H*), PRAGMATICS (contrastive, corrective, non-contrastive), and their interaction as fixed effects, and SPEAKER and ITEM as random effects.

In addition to FPCA, the smoothed and speaker- and time-normalized f0 curves were also analyzed in a Generalized Additive Mixed-effects Model (GAMM), using the *mgcv* ([18]) and *itsadug* ([19]) packages, with PHONETICS (H*, L+H*) and PRAGMATICS (contrastive, corrective, non-contrastive) as fixed intercepts and smooth terms, as well as SPEAKER as a factor smooth (random slope and intercept). The reason for analyzing the same set of data using two different methods is that FPCA and GAMM offer different insights to the research question. FPCA identifies independent components that contribute to distinguishing the f0 curves, which we used to test the association between these components and the theory-driven labels. On the other hand, a GAMM directly tests whether the f0 curves categorized in AM terms differ in their shapes and identifies where the differences lie along the curve, while not making an assumption about how the f0 curvatures are shaped for the different phonetic/pragmatic categories.

3. Results

Table 3 summarizes the conditional and marginal R^2 values ([20]) calculated from the LMEMs on the PC scores of the first five PCs. Below we report the results for only the first three PCs for two reasons: first, their R^2 values were much higher than those of the other PCs, indicating that the aspects of the curves explained by these PCs were more closely related to the linguistic categories of interest to this study; second, the changes in the f0 curve shapes identified by the first three PCs collectively explained 96.5% of the variance in the input f0 curves. The first three PC curves are shown in Figure 1.

The output of the LMEM showed that higher PC1 scores were associated with L+H* (red curves in the PC1 panel in Fig.1), while lower PC1 scores were associated with H* (blue curves in the same panel). This indicates that L+H* was produced with higher f0 than H* ($\beta = 13.86$, $p < .001$). However, the interaction between PHONETICS and PRAGMATICS ($\beta = -9.03$, $p < .05$) showed that this f0 scaling difference was only significant for the non-contrastive items.

The output of the LMEM for PC2 showed significant main effects of both PHONETICS ($\beta = -6.26$, $p < .001$) and PRAGMATICS ($\beta = -2.00$, $p < .01$), as well as a significant interaction between them ($\beta = 3.27$, $p < .05$): higher PC2 scores were associated with H* (red curves), meaning that H* were produced with falling f0 shapes, while lower PC2 scores were associated with L+H*, which was produced with rising f0 shapes. However, this shape distinction was only observed in the non-contrastive and contrastive items, not in the corrective items.

Lastly, for PC3, only the main effect of PHONETICS was significant ($\beta = -3.45$, $p < .001$): higher PC3 scores were associated with H* (red curves), meaning that H* was produced with a falling f0 shape throughout the accented vowel, while lower PC3 scores were associated with L+H*, which was produced with a rise-fall shape and a f0 peak approximately in the middle of the accented vowel.

Table 3. *Conditional and marginal R^2 values for the scores of each PC*

	PC1	PC2	PC3	PC4	PC5
R^2_c	0.35	0.32	0.36	0.18	0.18
R^2_m	0.06	0.09	0.07	0.01	0.01

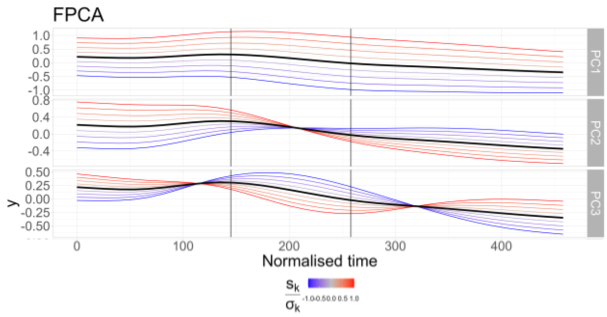

Figure 1. *The first three PCs. The thicker black lines represent the average curve; the red and blue lines indicate curves with higher and lower than mean PC scores, respectively. Vertical black lines denote the onset and offset of the accented vowel.*

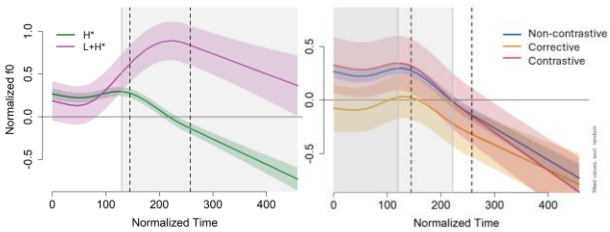

Figure 2. *Smooth curves for* PHONETICS *(left) and* PRAGMATICS *(right). The intervals of significant difference between H* and L+H* (130–458 ms), between contrastive and corrective (0–120 ms), and between non-contrastive and corrective (0–222 ms) are marked as grey shades. The vertical dashed lines indicate the onset and offset of the accented vowel.*

Turning to the results of the GAMM, the left-hand panel of Fig. 2 presents the smooth curves for H* and L+H*, which are the output of fitting the normalized and smoothed f0 input curves using the phonetic labels. The interval of significant difference covered a substantial stretch which included the vowel of the accented word. This suggests that H* accents were realized as falling f0 whereas L+H* were realized as rise-falls with a peak within the accented vowel. On the other hand, the right-hand panel of Fig. 2 shows the smooth curves modeled the same set of data, but this time using the labels from the classification that was based on the pragmatic annotation instead of the phonetic labels (that is, H* and L+H* were pooled). Corrective accents were produced with lower f0 and a lower f0 peak compared to non-contrastive and contrastive accents, whereas no difference was found between the smooth curves for non-contrastive and contrastive accents. The difference in f0 scaling was found in the accented vowel only for the non-contrastive vs. corrective set.

4. Discussion and conclusion

The results from FPCA and GAMM showed that H* and L+H* in SBE differed in f0 shape and scaling. The analyses identified which components in the f0 tracks contributed to this distinction: f0 curves categorized as H* by the LMEM and GAMM were associated with falling f0 shapes with low scaling, whereas f0 curves categorized as L+H* were associated with rise-fall f0 shapes with high scaling. In addition, the accents differed in the location of the f0 peak relative to the temporal landmarks (i.e., the onset and offset of the accented vowel). While H* accents had an early f0 peak close to the onset of the accented vowel, with f0 falling throughout that vowel, L+H*s had a later peak that typically occurred approximately in the middle of the accented vowel. The finding that the accents differed in shape is not perhaps surprising, in that the accents were categorized using criteria that separated falls from rise-falls. However, the GAMM also showed a systematic difference in peak scaling and alignment, which derives from the analysis itself not the annotation criteria. With respect to SBE, what the data show is that both rising and falling accents are attested, but, as discussed in [11] and [12], the rising accents are relatively infrequent compared to falls. Finally, H* is phonetically different from that of American English, in that it is realized as a fall rather than simply high pitch or slightly rising.

The pragmatics-based categorization of the accents did not show as consistent differentiation as the phonetics-based categorization. Pragmatic differences were only associated with systematic changes in PC2 scores: H* accents marked as either contrastive or non-contrastive were realized mostly as falls (higher PC2), while L+H* accents with the same pragmatic labels were rise-falls (lower PC2). This general lack of clear association between pragmatic functions and accent shapes was also observed in the GAMM smooth curves. Taken together, the two analyses indicate that the way the accents were classified by their pragmatic functions did not neatly map onto the f0-based classification. In other words, both H*-shaped and L+H*-shaped accents were used to encode new and contrastive information. Fig. 3 shows, on the other hand, corrective smooth curve looked like L+H* more than non-contrastive and contrastive curves (cf. [21]).

These findings indicate that the information-structure (IS) based distinction between H* and L+H* posited by [2] does not hold for SBE. Further, the results reinforce the notion that contrastivity is not one category when it comes to accentuation: contrastive and corrective items, which are sometimes treated as comparable, were likely to be realized using phonetically different accents (H* and L+H* respectively). Finally, our findings provide insights into the complex relation between the accent shapes and their pragmatic functions in SBE. For example, in our corpus, both H* and L+H* were used to highlight both contrastive non-contrastive information (see Fig. 3). This finding agrees with [5,6,7] and the MAE ToBI guidelines [4]. In addition, our data indicate that L+H* was also used to mark unexpected items, which were frequent in the story cube task in which, as mentioned, participants must create stories based on surprising combinations of items. This is illustrated in Fig. 4, in which *Rapunzel* is new but unexpected information. More examples of the accents encoding meanings beyond information structure are in Table 4. These examples highlight the importance of using various tasks to elicit different pitch accent functions.

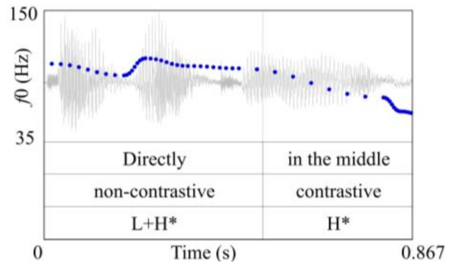

Figure 3. *Sample utterance elicited in the map task.*

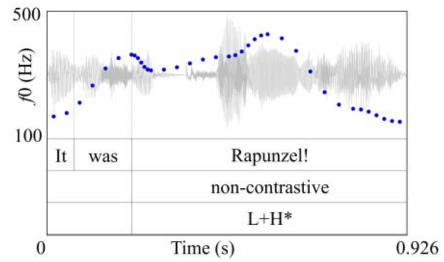

Figure 4. *Sample utterance elicited in the story cube task.*

Table 4. *Examples of accents with non-IS related meaning*

Category	Example
Unexpected (NC, L+H*)	And then I saw *a dinosaur*.
Unexpected (C, L+H*)	But to my horror and surprise, it was just *a wolf*, howling at the moon.

In conclusion, the current study offers new insight into the phonetic realizations of H* and L+H* accents and how they are related to pragmatic functions in SBE. The results showed that these accents are distinct in terms of f0 shape: H*s were realized as falls with low scaling and an early peak, while L+H*s were realized as rise-falls with high scaling and a late peak. The distinction based on the pragmatic functions of the accents was not as clear-cut, however, in that the different pragmatic functions were not uniquely associated with distinct accents. Rather, both H* and L+H* were used to mark various pragmatic functions. This finding extends the results of the empirical evidence from the perception studies that showed how both H* and L+H* can be interpreted as encoding new, corrective and contrastive information.

5. Acknowledgements

The financial support of the European Research Council through grant ERC-ADG-835263 (SPRINT) to Amalia Arvaniti is hereby gratefully acknowledged. We thank Chris Cummins and Hannah Rohde for their help with the pragmatic annotation schema, our speakers, for participating in the study during the COVID-19 pandemic, and Lisanne Versaevel for data annotation.

6. References

[1] Pierrehumbert, J. (1980). The phonology and phonetics of English intonation [PhD thesis]. MIT.
[2] Pierrehumbert, J. B., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In: P. Cohen, J. Morgan, & M. Pollack (eds.), *Intentions in Communication*, 271–311.
[3] Beckman, M. E., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology*, 3, 255–309.
[4] Brugos, A., Shattuck-Hufnagel, S., & Veilleux, N. (2006). Transcribing Prosodic Structure of Spoken Utterances with ToBI. MIT Open Courseware. Retrieved August 2nd, 2023, from http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-911-transcribing-prosodic-structure-of-spoken-utterances-with-tobi-january-iap-2006/index.htm.
[5] Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, 58(2), 541–573. https://doi.org/10.1016/j.jml.2007.06.013
[6] Metusalem, R., & Ito, K. (2008). The role of L+ H* pitch accent in discourse construction. In *Proceedings of Speech Prosody 2008*, pp. 8776–8780.
[7] Watson, D. G., Tanenhaus, M. K., & Gunlogson, C. A. (2008). Interpreting pitch accents in online comprehension: H* vs. L+ H*. *Cognitive Science*, 32(7), 1232–1244. https://doi.org/10.1080/03640210802138755
[8] Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press.
[9] O'Connor, J. D. & Arnold, G, F. (1973). *Intonation of colloquial English*. London: Longman.
[10] Cruttenden, A. (1997). *Intonation* (2nd ed.). Cambridge: Cambridge University Press.
[11] Grabe, E., Post, B., Nolan, F., & Farrar, K. (2000). Pitch accent realization in four varieties of British English. *Journal of Phonetics*, 28, 161–185.
[12] Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *JASA*, 118(2), 1038–1054.
[13] Ladd, D. R., & Schepman, A. (2003). "Sagging transitions" between high pitch accents in English: Experimental evidence. *Journal of Phonetics*, 31(1), 81–112. https://doi.org/10.1016/S0095-4470(02)00073-6.
[14] Chen, A., den Os, E., & de Ruiter, J. P. (2007). Pitch accent type matters for online processing of information status: Evidence from natural and synthetic speech. *The Linguistic Review*, 24(2–3). https://doi.org/10.1515/TLR.2007.012
[15] Gubian, M., Torreira, F., & Boves, L. (2015). Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts. *Journal of Phonetics*, 49, 16–40.
[16] Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/doi:10.18637/jss.v067.i01.
[17] R Core Team, (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, [Online]. Available: http://www.r-project.org/.
[18] Wood, S. (2017). *Generalized Additive Models: An Introduction with R*, 2nd edition. Chapman and Hall/CRC.
[19] van Rij J., Wieling M, Baayen R, van Rijn H (2022). "itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs." R package version 2.4.1.
[20] Orelien, J. G. & Edwards, L. J. (2007). Fixed-effect variable selection in linear mixed models using R2 statistics. *Computational Statistics & Data Analysis*, 52(4), 1896–1907.
[21] Greif, M., Skopeteas, S. (2021). Correction by focus: Cleft constructions and the cross-linguistic variation in phonological form. *Frontiers in Psychology* 12. https://doi.org/10.3389/fpsyg.2021.648478.