

From Sublexical to Lexical Processing: Developmental Effects of Word Characteristics on Word Recognition

Participant-visible method context. Results, discussion, captions, and the target figure are excluded.

Participants The sample consisted of 334 elementary schoolchildren recruited from two primary schools (22 classes) in Bavaria, Germany (Grade 2: $n = 127$, 54% female; Grade 3: $n = 124$, 56% female; Grade 4: $n = 83$, 55% female). On average, 15 children per class participated ($SD = 2.72$). The data collection was fully anonymized because of data protection regulations imposed by the school authorities, which means that the migration background and age of participants could not be obtained. However, in German primary schools, the age range of children in Grade 2 is 7–8 years, in Grade 3, 8–9 years, and in Grade 4, 9–10 years, with typically little variations. Data collection took place over a 3-week period in December 2018. Only children whose parents provided written informed consent and who gave their own assent prior to participation were included in the study. The study was conducted in line with the Declaration of Helsinki. Word recognition task A computerized lexical decision task was created using the childLex corpus that contains linguistic data from 500 German children's books with an intended reading age from 6 to 12 years (Schroeder et al., 2015). This task format was chosen because it captures accuracy and response latency in visual word recognition and is well suited for investigating sublexical and lexical processing (Coltheart et al., 2001). Lexical decision tasks have been shown to rely more heavily on lexical processing than naming tasks (Burani et al., 2002), making them particularly important for studying the extent that the importance of lexical processing changes during elementary school (van den Boer & de Jong, 2015). The task consisted of 96 written German words (e.g., Jahr [year]) and 96 written pseudowords (e.g., Klatz) in randomized order, and the children were required to decide whether the presented letter string was a real word.

SCIENTIFIC STUDIES OF READING

Pseudowords were orthographically and phonologically legal strings of letters and varied in their similarity to actual German words. Pseudowords similar to actual words were constructed by changing the first character of an existing word (e.g., Vogel → Rogel). Pseudowords dissimilar to actual words were constructed by combining the first syllable of an existing word with other syllables of existing words. For example, the pseudoword Sonnfön was constructed by combining the first syllable of the word Sonntag and the last syllable of the word Telefon. The existing German words selected for this study consisted of nouns that were representative of the vocabulary level and content of the target group. All words varied systematically in regularity, length, frequency, and number of orthographic neighbors. Regularity was classified according to the German grapheme–phoneme correspondence rules (Ziegler et al., 2000) and supplemented by the orthographic complexity proposed by Schmalz et al. (2015). This classification relied on the principles of the nonlexical route of the German DRC model, assuming that the lexical route was switched off. This deactivation allowed us to assess the extent that each word could be accurately decoded using the German grapheme–phoneme correspondence rules. We applied the three main types of rules from the original DRC: single-letter, multiletter, and context-sensitive rules (Ziegler et al., 2000). By comparing the orthographic input and phonological output lexicons, each grapheme–phoneme mapping was assigned to either single-letter rules, multi-letter rules or context-sensitive rules. Single-letter rules were considered the most regular because they involve consistent one-to-one correspondence between individual graphemes and phonemes. Multiletter rules were considered slightly less regular, as they require the combination of multiple graphemes to form a predictable phoneme sequence. Contextsensitive rules were classified as the least regular, as they depend on specific graphemic environments and thus show greater variability across lexical units. Words that could be decoded using only singleletter rules (e.g., Papa) were classified as regular with low complexity (Level 1). If decoding required at least one multiletter rule (e.g., Chaos), the word was classified as regular with moderate complexity (Level 2), whereas words that contained at least one context-sensitive rule (e.g., Spaziergang) were classified as regular with high complexity (Level 3). Any word that was named incorrectly by the nonlexical route was, by definition, irregular, as it violated the grapheme–phoneme correspondence rules (Level 4: e.g., Jeep). All other psycholinguistic word characteristics, including word length, frequency, and orthographic neighborhood size, were derived from the childLex database (Schroeder et al., 2015). Word length was determined by the number of letters in the word. For frequency, we applied the normed lemma frequencies from the corpus childLex (Schroeder et al., 2015). Finally, the number of orthographic neighbors was operationalized using Coltheart's N (ON; Coltheart et al., 1977), defined as the number of words that can be produced by changing a letter in a word of the same length. For example, the letter string H-A-U-S (which translates to “house” in English) is

likely to activate not only the orthographic representation for Haus, but also the representations of other words with similar letter combinations, such as Hans or Hals [throat]. The words had an average regularity score of 2.97 (SD = 0.64; min = 1; max = 4). They had an average (logarithmic) frequency of 3.68 (SD = 1.36; min = 1.55; max = 6.51), retrieved from the online database childLex, an average length of 5.60 characters (SD = 1.62; min = 3; max = 11) and on average 4.62 orthographic neighbors (SD = 4.99; min = 0; max = 21). The distributions of the word characteristics are illustrated in Figure 1. An overview of the words and detailed information on their characteristics can be found in Table S1 in the online supplementary materials on OSF.

Procedure The lexical decision task was administered to children collectively in classrooms as part of a cross-sectional study investigating a new tablet-based error-detection test (Endlich et al., 2020). Children were required to complete the lexical decision task in a quiet, separate room while they were monitored by student research assistants (i.e., psychology undergraduates). These student research assistants received prior training and were supervised by the authors.

J. HEß ET AL.

Figure 1. Distributions of the word characteristics.

The words and pseudowords were presented in a fixed order in the center of the screen, one at a time. The children's task was to decide as rapidly as possible whether a presented letter string was a real German word or a pseudoword by using two color-coded response keys (yes in green/no in red). These response keys were displayed on the tablet screen to the left and right of the pseudo(word) to be processed. Prior to the presentation of the test items, two practice items were presented, which were repeated until the children provided the correct answer. The total of 192 items were divided into four blocks and were randomly presented within each block. Between the blocks was a screen with positive feedback, and children could proceed to the next block when they felt ready to do so. Response accuracy and log-transformed response latencies (measured from stimulus onset to the press of a response button) were recorded as dependent variables.

Plan of analysis The first objective of our study was to determine a model that reliably predicts word difficulty by word characteristics. To this end, we used explanatory item response models, following a standard threestep procedure (e.g., see Hartig et al., 2012; Schindler et al., 2018; Schulz et al., 2023). In the first step,

SCIENTIFIC STUDIES OF READING

we determined the empirical difficulties of each word. In a second step, we estimated models to predict word difficulties based on theoretically derived word characteristics. In a third step, we correlated the empirical and estimated word difficulties to determine the proportion of variance in the word difficulties that can be explained by word characteristics. This three-step procedure allows for examining the extent to which the models that predict word difficulties based on theoretically derived word characteristics capture individual differences in the processing of written words. For the estimation of empirical word difficulties in Step 1, we used a one-parameter logistic model (1PL or Rasch model) for the accuracy data, implemented as a Generalized Linear Mixed Model (GLMM; Dixon, 2008) with a logit-link. Similarly, the word difficulties (time intensities) for the response times were estimated using a Rasch-analogous Linear Mixed Model (LMM; Baayen et al., 2008) for the (log-)transformed response latency (van Breukelen, 2005). The words were dummy-coded with the first word serving as the reference category and included as predictors (fixed effects) for response accuracy and log-transformed response latency as dependent variables. The intercept was allowed to vary randomly between participants (random intercept), representing the person parameter and between classes (random intercept) to account for potential clustering effects on the class level. Formula 1 shows the model equation for the GLMM for logit-transformed response accuracy (i.e., logtransformed odds ratios for correct responses) as dependent variable:

The LMM for logarithmically transformed response latency as outcome variable was set up in the same way as a random-intercept model with fixed item effects. In Step 2, response accuracy was modeled in a Logistic Linear Test Model (LLTM, Fischer, 1974) and response latency in an analogous linear mixed model (LLT-analogous model) with the theoretically derived word characteristics regularity of the grapheme–phoneme correspondence (z-standardized), length (z-standardized), frequency (z-standardized), and number of orthographic neighbors (z-standardized) as fixed item-level predictors. Both models included random effects of items, participants, and classes. Again, a GLMM with a logit-link for response accuracy was estimated:

The LMM for the logarithmically transformed response latency as outcome variable was set up in the same way. The intercepts and slopes from the models of Step 2 were used to derive estimated word difficulties for each item. In Step 3, empirical word difficulties obtained in Step 1 were correlated with the estimated word difficulties of Step 2 to determine how closely word characteristics predict the word difficulties. To examine whether the effects of word characteristics in the lexical decision task varied systematically across grade levels, we tested for overall differences between grades and for interactions between word characteristics and grade levels using a Type-II Wald χ^2 test as implemented in the R package car (Fox & Weisberg, 2019). To investigate how the effects of word characteristics on word difficulties change in a lexical decision task from Grade 2 to Grade 4, grade level was included as a contrast coded predictor variable (Grade 2 vs. Grade 3 and Grade 3 vs. Grade 4), in both the LLTM and the LLT-

J. HEß ET AL.

analogous models of Step 2. Moreover, grade level interaction terms with the word characteristics were added to the models. These interaction terms accounted for possible developmental changes in the extent to which different word characteristics predict word difficulties. For interpreting significant interactions, we estimated simple slopes (Aiken & West, 1991; Cohen et al., 2003, Chapter 9).