Continued Access to Syntactic Information in Reading

Reinold Funke, Jasmin Sieger

Pädagogische Hochschule Heidelberg

Institut für Deutsche Sprache und Literatur und ihre Didaktik

Funke, Reinold, & Sieger, Jasmin (2012). Continued access to syntactic information in reading. *Reading and Writing: An Interdisciplinary Journal, 25*, 1769-1794, DOI: 10.1007/s11145-011-9341-8.

The final publication is available at www.springerlink.com. (http://www.springerlink.com/content/1474342g0607h288/)

Reliable access to syntactic information beyond its use in first-pass reading was probed using a task labeled *syntactic reading task*. This task draws on an orthographic marker for syntactic structure, the capitalization of nouns in German. Subjects read written text strings containing a unit that could be taken as a noun or as a verb. The strings were disambiguated by the unit being capitalized or not. Thus subjects could interpret the text only when taking the capitalization marker into account. Data were collected from 231 German-speaking secondary school students. The results show that even subjects who were capable of reliably using capitalization markers in writing varied considerably in their use of these markers to identify syntactic structure in reading. High-achieving subjects profited more than others in situations where they could tie syntactic information to a prominent capitalization marker. This suggests that focus on syntactic form alone does not guarantee reliable access to syntactic information.

Keywords: Reading, Syntactic awareness, Syntactic sensitivity, Working memory, Capitalization

Syntactic processing in reading is understood to be necessary because it enables readers "to formulate hypotheses about the meaning of the sentence" (Verhoeven & Perfetti, 2008, p. 295). Readers process text by using syntactic cues to construct a semantic description of the message conveyed. They may consider word order or morphological case markers when selecting a propositional frame that accurately represents the meaning of the given sentence. This process is sometimes referred to as *syntactic decoding* (Millis, Magliano, & Todaro, 2006; Pugh et al., 2001). Syntactic decoding begins as soon as a sentence is read. It is, however, commonly assumed that the information gained gets lost after a short time. As Just and Carpenter (1987) put it, "syntactic information … tends to be forgotten" (p. 130) after it has fulfilled its function in constructing meaning.

Linguistic analyses have suggested that syntactic information in reading may serve purposes that go beyond syntactic decoding. One of these purposes is to indicate textual features. Word order and case marking signal not only the thematic roles of propositional arguments but also the perspective of a message. The sentence *Paul meets Jane* may convey the same idea as *Jane meets Paul*, but presents it from a different perspective (Dik, 1997). The first sentence also indicates that Paul continues to be the textual theme (Halliday, 2004). In addition, syntactic information serves to adjust word meaning to context. Verbs' actionality type and selectional restrictions can serve as examples of this function (Grimshaw, 2005). Therefore, readers must consider syntactic features when interpreting word meanings. For these reasons, they may need to access syntactic information beyond their first-pass reading.

When readers interpret text, pondering it one more time, textual features like perspective and thematicity play important roles. Considering alternative contextual meanings of a word in the text requires the ability to return to a passage and read it in greater analytical detail. Thus, to interpret both textual and word-level features, readers need to have continued access to syntactic information beyond its use in syntactic decoding.

Various tasks have been employed to assess whether readers dispose of this type of information. Some of them, such as maze or cloze tasks (Guthrie, 1973; Nation & Snowling, 2000) and tasks that measure the use of context to identify words (Isakson & Miller, 1976; Bentin, Deutsch, & Liberman, 1990), probe achievements that are close to spontaneous syntactic processing. In other tasks, an attempt is made to eliminate the influence of spontaneous syntactic processing by forcing subjects to consciously attend to linguistic form instead. Examples of this include the correction of grammatical errors on demand (Bowey, 1986) or, conversely, their deliberate reproduction (Gaux & Gombert, 1999).

Two main theoretical questions have emerged as to the secondary use of syntactic information in reading. Firstly, there are studies which suggest that subjects' ability to engage with it no longer has any predictive power for reading comprehension when the effects of other variables, such as phonological sensitivity (Gottardo, Stanovich, & Siegel, 1996) or verbal working memory (Shankweiler et al., 1995), are controlled for. Secondly, it remains unclear what distinguishes the secondary access to syntactic information from syntactic processing in reading itself. Some authors suggest that this access is characterized by the reader taking a metacognitive stance (Gombert, 1990) or attending to linguistic form instead of message (Bowey, 1993). Others have been less explicit on this question.

We propose that the secondary use of syntactic information in reading is distinguished from spontaneous syntactic processing by the fact that readers repeatedly address the same syntactic information without distraction or confusion. This proposal underlies the empirical tool employed in the present investigation.

The Syntactic Reading Task

In various writing systems, syntactic features leave traces in the orthographic form of the written text. The capitalization of syntactic nouns in German orthography is one example of this phenomenon. Studies by Bock (1986) and Müsseler, Nißlein, and Koriat (2005) have demonstrated that advanced readers of German use the capitalization marker to identify syntactic structure. Novice and disabled readers seem to be less able to profit from capitalization (Bock, 1990). In these studies, subjects read continuous unambiguous text while using capitalization as a cue for syntactic decoding. A different situation emerges when readers are confronted with text in which the resolution of a syntactic ambiguity depends exclusively on capitalization. An example illustrates this case: In the passage *I often avoid discussions, because most harm* […], the unit *harm* may be read as a noun or a verb. The written expression can be disambiguated by adding a continuation. If the continuation reads *comes with them*, it becomes clear that *harm* is meant to be a noun (noun reading), and if it reads *rather than help*, *harm* is meant to be a verb (verb reading). In written German, the noun and the verb readings of the whole expression would be distinguished independently of the continuation simply by using an upper-case or lower-case letter. Applying the rules of German orthography to the example, one would have to write *I often avoid discussions, because most Harm* […] for the noun reading and *I often avoid discussions, because most harm* […] for the verb reading. Whether a reader is attentive to and profits from the information conveyed by upper- and lower-case letters can be examined by having him or her choose the appropriate continuation from the two possibilities. To solve this task, readers must take into account whether or not the initial letter is capitalized. We call this task *syntactic reading task* (SRT) and the process it assesses *syntactic reading*.

The process the SRT taps into is different from the use of capitalization in reading "normal", unambiguous texts as assessed by Bock (1990) and Müsseler, Nißlein, and Koriat (2005). In "normal" texts, the information conveyed by the capitalization of nouns is usually redundant because it converges with information provided by sentential context (Mentrup, 1993). When readers nevertheless draw on capitalization to detect syntactic structures, they are supported by additional substantiating information. The SRT, however, requires readers to activate or select an appropriate syntactic template on the sole basis of capitalization. They must do this independently, without contextual support, and perhaps even in opposition to contextual cues. To activate a syntactic template without contextual support, readers must be able to focus on it repeatedly. This is the rationale for assuming that the SRT is suited to measuring continued access to syntactic information, as defined above.

The SRT is a specifically syntactic task. Firstly, syntactic reading, as measured by the SRT, is different from the recognition of lexical categories using orthographic cues which has been found to occur during reading (Arciuli & Monaghan, 2009; Kemp, Nilsson, & Arciuli, 2009). Lexical categories are features that may

be ascribed to words without regard to context. In the SRT, an expression must be assigned a syntactic category determining its function in the given context. Secondly, syntactic reading, as recorded by the SRT, draws on orthographic knowledge, but this knowledge must be supported by syntactic experience in order to serve its purpose in reading. The German capitalization rule laconically states that "nouns are capitalized" (Rat für deutsche Rechtschreibung, 2006, rule #55). The rule is commonly explained to students by telling them that nouns are words linked to an article. This criterion is tailored for use in writing, where students interpret a word as a noun and are consequently expected to capitalize it. In the SRT, the inverse situation occurs: A word is capitalized and has to be interpreted as a noun. Knowing how to correctly capitalize nouns in writing does not necessarily imply an ability to interpret capitalization in reading.

In summary, the SRT is assumed to measure continued access to syntactic information during the reading process. Although the task is tied to German orthography, it should be noted that the achievement it is intended to measure is by no means unique to this language. Reading situations in which continued access to syntactic information is crucial appear across orthographies (for the signaling of textual features by syntactic means, see Givón, 2001; Schleppegrell, 2004; for the use of context to determine word meanings, see Nagy & Scott, 2000; Nagy, 2007). The case of German capitalization simply offers an opportunity to scrutinize how syntactic structure is grasped in reading. This is done using a procedure which captures the ongoing reading process (though it is not completely unobtrusive). Incidentally, SRT-like tasks may be devised in different writing systems that signal syntactic categories by orthographic means, such as the apostrophe in English (Bryant, Devine, Ledword, & Nunes, 1997) and markers of verb morphology in French (Bautier & Branca-Rosoff, 2002).

Hypotheses

The written string of text *I often avoid discussions, because most harm* […] has two possible readings, one in which *harm* is interpreted as a noun and one in which it is regarded as a verb. A case like this is called a category ambiguity because the interpretation of the expression depends on the syntactic category assigned to a single written unit (*harm* in the example cited above). In the following, this unit is labeled *critical unit*.

How readers handle category ambiguities has been studied extensively. Research has been predominantly driven by "how the reader … determines the contextually appropriate meaning" (MacDonald, Pearlmutter, & Seidenberg, 1994, p. 676), with frequency of use (Duffy, Morris, & Rayner, 1988) and contextual constraints (Boland, 1997) identified as the most influential factors. If, as is the case in the SRT, readers must deal with critical units written according to German capitalization norms, there is in principle no ambiguity. However, one may assume that, even in this case, two syntactic templates for interpreting the text get activated and compete with each other. Readers' behavior will thus probably be influenced by the factors described in the research. Nevertheless, the SRT is not intended to reveal which influence will prevail in this situation, but to assess the extent to which readers are able to access syntactic information without being supported or misled by contextual cues. This purpose, which is different from that of most category ambiguities research, must be kept in mind when formulating

hypotheses about SRT achievement. Accordingly, we advance three hypotheses concerning SRT performance.

Since mastering the orthographic capitalization rule does not, in itself, involve the knowledge needed to cope with the SRT, subjects' achievement in the SRT will constitute a dimension of interindividual variation beyond orthographic capitalization skills. Therefore, we expect SRT scores to vary even among subjects who have mastered capitalization in writing (Hypothesis A).

When working on a SRT item, a subject will solve the task either by relying on the capitalization marker or by drawing on other characteristics of the two competing alternatives. If he or she does not attend to capitalization, his or her chance of solving the item will depend on factors such as frequency and contextual preponderance. If she or he uses capitalization, however, the chance of solving the item will not be determined by these factors. In a group of individuals who differ systematically in their ability and propensity to access syntactic information independently, subjects will consistently tend to behave either one way or the other. Therefore, the distribution of scores in the SRT should scatter around a lower value, near chance level, and an upper value which is near ceiling (Hypothesis B).

A subject's continued access to syntactic information manifests itself in the ability to reaccess this information when forced by a reading task to do so. In the case of an SRT, a subject may become aware of the necessity to focus on syntactic information if the critical unit attracts attention. The capitalization marker is more likely to be noticed if the critical unit is made prominent. In this case, the thought process involved in solving the task will increasingly depend on the subject's ability to resume access to syntactic information. For this reason, we hypothesize that the level of critical unit prominence will have a different effect on subjects who are able to reliably refer to syntactic information compared to those who are not (Hypothesis C).

Method

Participants

The study's participants were students from grades five to seven attending schools in a provincial region of southern Germany. Students from schools with strict admission criteria (Gymnasien) were not included. Participation was conditional on the informed consent of both students and parents. Subjects first took an orthographic writing test designed to check their ability to correctly capitalize written words. Afterwards, they were given a syntactic reading task. Details on both tasks are given below.

Participants were also asked to indicate whether they had learned German prior to or after having started school. Data from the latter subjects were excluded from the analysis. Ultimately, only data from students who obtained full credit in the orthographic writing task were analyzed. This was done to make sure that subjects had been successfully taught the orthographic capitalization rule, thus giving them the chance to solve the syntactic reading task. Of 794 students who had

been tested, data from 231 native speaker subjects with perfect orthographic writing scores were retained for analysis (44 fifth-graders, 46 sixth-graders, and 141 seventh-graders).

Tasks

Syntactic reading task. The syntactic reading task was composed of 20 short written texts, each featuring a unit that could be regarded as either a noun or a verb. Whether or not the unit was capitalized resolved the ambiguity. Each text was followed by a task to check whether the unit had been properly read (see Appendix A). Such a task is illustrated in the following two examples, freely translated into English with the critical unit written according to German orthography.

Example 1:

> *Derek says, "Nowadays, so many people are divorced after only a few years of marriage. Most love*
> ☐ *… ... someone else after a while."*
> ☐ *… ... ends sadly."*

Example 2:

> *Christina states, "In my eyes people should be more careful who they believe because most Trust in others*
> ☐ *… ... although they don't know them well enough."*
> ☐ *… ... is given too hastily."*

The item consists of two parts—a short text that roughly outlines a fictitious situation (*starter text*), and two possible continuations of it (*alternatives*). The starter text sometimes consists of several sentences, but it always ends with a sentence fragment that contains a critical unit (in the examples above, these are *love* and *Trust,* respectively). This unit may be presented with the initial letter capitalized—therefore to be read as a noun, as in example 2—or it may be presented with a lower-case initial letter, as in example 1, urging the reader to interpret it as an inflected verb. One of the options matches the noun reading and the other the verb reading. Subjects are instructed to mark the correct option, but nothing is done to direct their attention to the capitalization. Their choice indicates whether they have interpreted the critical unit correctly. An item is considered to be solved if the noun option for a capitalized critical unit and the verb option for a critical unit with a lower-case letter are marked. Each subject is accorded the sum of solutions as his or her raw score; this may be any value between 0 and 20. A subject's SRT achievement is referred to as *syntactic reading*.

The items of the SRT differ with regard to two position parameters. One of them is *position of critical unit*: The critical unit can be found either at the end or within the starter text. In the first case, position of critical unit is *final*; in the second, it is *embedded*. It is assumed that critical units in the final position are more prominent than critical units in the embedded position. In example 1, the critical unit is final; in example 2, it is embedded. The other position parameter is *position of solution*: In half the items, the solution appears first; in the other half, it is presented after the incorrect option. Thus, position of solution can be either *first* or *second*. It is first in example 1; in example 2, it is second. As each position parameter takes two values, there are four potential item conditions (*final*/*first*, *fi-*

nal/*second*, *embedded*/*first*, and *embedded*/*second*). We ensured that the SRT-items were equally distributed among these conditions.

Two versions of the SRT (A and B) were used. Both versions contained the same starter texts and alternatives but differed in the initial letter of the critical units. When the critical unit was capitalized in one version, it was written with a lower-case letter in the other version. Half the critical units were capitalized; the others were not. Moreover, the items were ordered differently in each version.

When constructing items of this type, one is faced with the problem that the noun and verb readings may differ in the degree to which they "sound natural". In the context *In my eyes people should be more careful who they believe because most trust in others* [...], for example, most people would presumably tend to take *trust* to be a verb, and some might miss the noun reading completely. This feature, which influences the difficulty of syntactic reading items, is labeled *polarity* (referring to the distinction between polarized and balanced items in Duffy, Morris, & Rayner, 1988). Thus, a syntactic reading item may be polarized into a verb or into a noun reading.

A post hoc estimation of item polarity is possible if one restricts oneself to subjects who obviously did not consider capitalization systematically when working on the SRT (that is, subjects whose raw score is on chance level). Half of the subjects had to solve an item with the critical unit capitalized and thus needing to be read as a noun (one version). The other half had the item with the critical unit uncapitalized and thus needing to be read as a verb (other version). If responses are exclusively determined by item polarity, the probability p of solving an item correctly in one version and the probability p' of solving it correctly in the other version must fulfill the condition $p = 1 - p'$ for subjects whose raw score is on chance level. As long as this condition is met, item polarity may be defined by any expression depending exclusively on p. In the case given, tables were created for the noun items in each version by sorting subjects' responses with a raw score of 10 according to the scheme displayed in Table 1.

[Please insert Table 1 about here]

Given a table of this type, $G = ((a + c) - (b + d))/N$ may be calculated where $N = a + b + c + d$ (Lautsch & Lienert, 1993). If the condition of definability mentioned above is met, for an item with difficulty p, G equals $2p - 1 = p - (1 - p)$ and may be used as a polarity index which ranges between -1 and 1 and takes value 0 if the item is not polarized. This polarity index may be tested for significance because under null hypothesis the numerator of G approximately follows a normal distribution with expectation 0 and variance N. Whether the definability condition is fulfilled may be checked by subjecting the table to a χ^2-test of independence.

Estimated polarity indices for the 20 items of the SRT are provided in Appendix B. They were calculated based on the responses of the subset of 144 subjects with raw score 10 out of the whole sample (N = 794). Four items did not meet the condition of definability. In two cases, subjects solved the item significantly more often than should be expected under the condition of random choice. In the other cases, subjects took the critical unit significantly more often for a verb when it was capitalized than when it was uncapitalized (and vice versa). Thus, when considering subjects whose responses were, on the whole, on chance level, one should be cautious before concluding that they did not attend to capitalization at all. Without getting further into this matter, we will use G as a measure to gain descriptive information on item polarities.

According to their polarity index, most items were significantly polarized (see Appendix B). It is important to note that this is not a flaw. If one uses the SRT to measure access to syntactic information in reading, polarized items are simply difficult ones, not defective ones. Evidence for this is given in the Results section. If, however, one is to use the SRT to check for differences in subjects' responses to the four item conditions described above, item polarities should be balanced across those conditions.

In the present study, an ex ante estimation of item polarity was undertaken to make sure that item polarities are equally distributed between versions A and B of the SRT. Six advanced graduate students rated items according to degree (*neutral, moderate,* or *strong*) and direction (*verbal* or *nominal*) of polarity. When assigning a capital or lower-case letter to the critical unit of items in either version A or B, an attempt was made to balance the degree of difficulty between the two versions. A post hoc analysis on the basis of the data confirmed that the item difficulties were indeed balanced between the two versions.

Cronbach's α for the SRT was .65 (version A) and .69 (version B). Both values are low. Given the item intercorrelations, this may be due to chance variation in low-achieving subjects. The item intercorrelations computed for them, as opposed to those computed for the high-achieving group, in many cases approached zero. This is to be expected on condition that low-achieving subjects responded by chance, and leads to a reduction of reliability. It is therefore important to observe that raw scores from the SRT as it is presented in Appendix A are not suited for measuring the syntactic achievement of individual subjects.

Orthographic writing task. The orthographic writing task was a completion test composed of ten sentences with sentence-final gaps, each to be filled on dictation by three consecutive, syntactically coherent words. When compiling the sentences, we selected words most frequently miscapitalized by German students according to the list published by Menzel (1985). Eight of the words dictated were nouns, five were verbs, five were adjectives, one was a pronoun and one was an adverb. Care was taken not to include concrete nouns that might be capitalized correctly without regard to syntactic function. Furthermore, the sentences were designed to contain syntactic constructions that are known to cause capitalization problems to students, such as noun phrases with adjunct adjectives (Heckel, 1980) and non-finite verb forms (Funke, 1995). A post hoc analysis confirmed that the orthographic writing task was considerably more challenging with respect to capitalization than common dictations in schools seem to be. The proportion of capitalization errors ranged from 24.1% for grade 5 students from basic schools to 5.2% for grade 7 students from more advanced schools. Pomm, Mewes, and Schüttler (1974), using a common school dictation, found this proportion to range from 5.3% (grade 5) to 3.7% (grade 7) for students with comparable school affiliations. It is thus unlikely that subjects could reach high scores in the task without having been successfully taught the orthographic capitalization rule.

When evaluating the orthographic writing task, we scored each word written with the proper capitalization as correct. However, only the second and the last of the three words dictated for each gap were considered. This was done because experience shows that German-speaking students tend to capitalize words extraordinarily often when they fill gaps in completion tests (Eichler, 2002), possibly believing their task is to write a context-free unit. Thus, scores on the writing task run between 0 and 20.

Cronbach's α for the orthographic writing task was .78.

Results

Scaling Considerations

In each item of the syntactic reading assignment, subjects are presented two options. By simply making a guess on each item without considering the others, they will receive scores that are binomially distributed with parameters n = 20 (number of trials) and π = ½ (probability of a hit). From this, it follows that a subject needs to have at least 15 correct solutions to exclude the possibility of chance level behavior (criterion for the significance level of .05). If one is to assess the achievement of individual subjects, the raw scores of the syntactic reading assignment must be transformed into binary values (*below criterion/above or equal to criterion*). This study compares groups of subjects and thus analyzes raw scores or scores preserving raw score order . However, only nonparametric statistical procedures are employed on these scores.

Distribution of Raw Scores

Versions A and B of the SRT did not differ in their overall difficulty. In version A, mean and standard deviation were $M = 12.82$, $SD = 3.08$; in version B they were $M = 12.84$, $SD = 3.28$. The median was 12 in both versions. As evaluated by a Mann-Whitney U-test, score distributions did not differ significantly ($z = 0.06$, $p = .97$). Thus, the raw scores of both versions may be aggregated when one considers SRT scores as a whole.

Although it can be assumed that orthographic capitalization skill is homogenous in the sample, SRT raw scores vary considerably, as displayed in Figure 1. Evaluation of the syntactic reading results revealed a variability in literacy-related abilities among students from grades five to seven that is not evident in their overt writing behavior.

[Please insert Figure 1 about here]

Only 30.7% of the subjects reached the criterion of at least 15 correct solutions in the SRT. One cannot exclude the possibility that the thought process of the others was guided by considerations not systematically related to the issue in question, which is to say, the information conveyed by capitalized and lower-case letters. While this seems to suggest that most students in the sample either did not attend to capitalization or were not able to interpret it, another picture arises when one considers the complete sample instead of individual subjects. The distribution of syntactic reading scores significantly deviates from what would result if subjects had randomly marked options ($\chi^2(2) = 111.83$, $p < .001$). This is true even when one considers the scores of versions A and B separately ($\chi^2(2) = 53.66$ for version A; $\chi^2(2) = 58.52$ for version B; $p < .001$ in each case). In other words, the deviation of the raw score distribution from a random one does not result from subjects being led to correct solutions by the polarity of items rather

than through the use of information conveyed by capitalization. If a reliance on polarity had indeed caused the deviation of the raw score distribution from a random one, raw score gains in one version would have cancelled out the losses in the other. Consequently, the assumption can be made that even subjects with raw scores below the criterion often considered capitalization. However, they do not seem to have proceeded systematically while doing so.

The distribution of syntactic reading scores shows three peaks, one in the upper range. However, the latter is not very marked.

Raw score distributions were essentially the same in each grade. The average score was M = 12.64 in fifth-graders, M = 12.15 in sixth-graders and M = 13.11 in seventh-graders, the median being 12 in all grades. No significant deviations were found to exist according to a Kruskal-Wallis test on the SRT scores with grade as a between-subjects factor ($H(2) = 4.56$, $p = .24$). In considering this, one must take into account that the sample was homogenized by the exclusive selection of subjects with perfect orthographic writing scores. The effect of this selection can also be seen in the fact that, while 30.7% of subjects in the sample reached the criterion of 15 solutions in the SRT, only 6.2% of the subjects with orthographic writing scores below 20 reached the criterion.

Effects of Item Polarity on Item Difficulty

Figures 2 to 4 show how item polarity influenced item difficulty in different achievement groups. The figures were created by calculating the ratio of the odds of an item being solved correctly in a given achievement group to the odds of it being solved correctly in a baseline condition where subjects do not evaluate capitalization systematically (that is, the group of subjects with an SRT score of 10). The odds ratios were computed separately for version A and B and transformed into logits by taking the logarithm. Item difficulties for the baseline condition were calculated based on the same data taken from the whole sample (N = 794) as used for the estimation of item polarities. Item difficulties for the achievement groups were calculated based on the selected sample (N = 231).

[Please insert Figures 2, 3 and 4 about here]

In the figures, each item is represented by a cross. The inserted line serves illustrative purposes, indicating which values are to be expected if item polarity is the only determinant of item difficulty and uniformly influences the odds of an item being solved correctly across all achievement groups. This hypothetical line has slope -1 because, if item polarity is the only determinant of item difficulty, an item with difficulty p in one version must have difficulty $1 - p$ in the other version under the baseline condition (see above). In this case, the odds for one version must be the inverse of the odds in the other version; consequently, the logits must have the same absolute value but differ by sign. Furthermore, the line intercept will be 0 in the baseline condition. For each achievement group considered, the line was adjusted to the data by shifting it to achieve a median split on the logits of the group.

As Figure 2 shows, logits for the lowest achievement group did not differ much from the baseline condition. In this condition, items with low polarity appear near the origin of the coordinates. The higher the polarity of an item, the more top-left

or bottom-right the cross representing it will be when one follows the inserted line. In Figure 3, logits for medium-achievers are displayed; the inserted line has been shifted uniformly to the top-right to conform to the raised attainment level. The line fits acceptably with the logits observed, suggesting that, on the whole, item polarity affected item difficulty in this achievement group no differently than for subjects of the lowest achieving group. This seems to be true even for highly polarized items. The picture is less clear in the highest achievement group (Figure 4). In this group, noticeable deviations from the hypothetical line occur. It is difficult, though, to attribute them specifically to differences between item polarities. In a multivariate test based on a repeated measures analysis of variance on the item logits with whole-plot factor degree of polarity (*below median*, *above median*, referring to the absolute value of the polarity index) and sub-plot factors SRT achievement group (*scores 0-11*, *scores 12-14*, *scores 15-20*) and version (*A*, *B*), there was no significant interaction of polarity and achievement group ($F(2, 17) = 1.60$, $p = .23$). It seems probable that the deviations of the logits from the expected values found in the highest achievement group are due to factors other than item polarity.

Effects of Position

On the basis of the number of items a subject has responded to correctly, she or he is assigned a subscore ranging from 0 to 5 for each of the four item conditions *final*/*first*, *final*/*second*, *embedded*/*first*, *embedded*/*second*. To test the hypothesis that the item conditions influence high- and low-achieving subjects differently, the sample was divided into three achievement groups of approximately equal sizes according to SRT achievement: scores 0-11 ($n = 87$), 12-14 ($n = 73$), and 15-20 ($n = 71$). Consequently, each subject's subscores may be analyzed for the effects of the factors achievement group (3 levels), position of critical unit (2 levels), and position of solution (2 levels). Position of critical unit and position of solution are repeated measures factors because each subject is assigned two scores for each.

To test the effects and interactions of factors, we used a nonparametric statistical model for repeated measures data (Brunner, Domhoff, & Langer, 2002). The procedure applied is labeled F1-LD-F2-model, meaning that there is one whole-plot (that is, between-subjects) factor and two sub-plot (that is, within-subjects) factors. Models of this type test the null hypothesis that score ranks are equally distributed across factors and factor combinations. The main statistic is the rank statistic Q^*, which is distributed approximately according to $F(f, \infty)$, where the degree of freedom f must be estimated from the data. Factor effects are measured using relative treatment effects (RTE) ranging from 0 to 1, which, under null hypothesis, take the value ½.

When one applies this model to the data in order to detect effects of the four item conditions, item polarities should be balanced across those conditions because they influence item difficulty. This turned out not to be the case in the post hoc estimation of item polarities as documented in Appendix B. Mean item polarities varied between item conditions and, in some cases, between versions A and B. Across all item conditions, they ranged from .011 to .397. Consequently, the distributions of raw scores for each item condition and, partially, each version were shifted with respect to the others, their median values ranging from 2 to 4. In this case, an analysis based on raw scores is likely to lead to erroneous conclusions

concerning the influences of item conditions because effects and interactions may occur that are caused by the raw score distributions differing across conditions. Additionally, scores from version A and B may be pooled only in the analysis of item condition effects if no interactions are observed between those effects and version. However, a check based on an F1-LD-F2-model with version as whole-plot factor and position of critical unit and position of solution as sub-plot factors revealed that there was a significant interaction of version and position of solution ($Q^*(1) = 37.85$, $p < .001$).

As noted, the problem stems from the raw score distributions shifting across item conditions and, partially, versions. One way of coping with this is to consider the relative locations of subjects within the raw score distribution of each item condition and version instead of the raw scores. This can be achieved by determining, for each combination of item conditions and versions, the percentile ranks assigned to raw scores in the respective raw score distribution. Percentile ranks may subsequently be converted to standardized values (Borg & Staufenbiel, 2007). The conversion is different from the common calculation of standardized values by linear transformation because it is based on relative locations instead of sample parameters. Its objective is to eliminate the effects on the analysis of item polarities differing across item conditions and versions. As item polarity was shown to affect achievement groups in the sample fairly homogeneously, one may assume that the standardizing transformation will serve this purpose. Indeed, when working with standardized values instead of raw scores, data from both versions may be aggregated. This was shown by an F1-LD-F2-model employed on the standardized values with version as whole-plot factor and position of critical unit and position of solution as sub-plot factors. The results are displayed in Table 2.

[Please insert Table 2 about here]

No significant interactions of version and item conditions were left in the standardized values. Thus, we pooled data from both versions and tested hypothesis C by subjecting the standardized values to an F1-LD-F2-analysis with achievement group (*low*, *medium*, *high*) as whole-plot factor and position of critical unit (*final*, *embedded*) and position of solution (*first*, *second*) as sub-plot factors. The test statistics are compiled in Table 3.

[Please insert Table 3 about here]

When considering Table 3, one should be aware that it is not possible to detect main effects of the sub-plot factors in the analysis of standardized values because values are constrained by the standardizing transformation and must yield equal rank sums in all item conditions. It is possible, though, to detect interactions of achievement level and item conditions. Such interactions are predicted in hypothesis C.

Apart from the trivial effect of achievement group, the statistics displayed in Table 3 indicate that there might be an association of achievement group and position of solution. The interaction of both factors just failed to reach significance ($p = .05$). The advantage of the high-achieving subjects compared to the other subjects was more marked in items with final critical unit than it was in items with embedded critical unit. One should refrain from a closer consideration of this effect not only because its reliability was not conclusively established, but also because it was superimposed by a threefold interaction of achievement group, position of critical unit, and position of solution that was significant ($p = 0.01$). This

interaction means that the relative success on different item conditions varied across the three achievement groups, as illustrated in Figures 5 to 7.

[Please insert Figures 5, 6 and 7 about here]

When working on items with final critical unit, high-achieving subjects' success was enhanced when the solution was presented as the first of the two alternatives. This effect was absent in low- and medium-achievers (see left-hand side of Figures 5 to 7). When working on items with embedded critical unit, high-achievers' success might at first glance seem to have been enhanced by solutions positioned in second position instead (see the right-hand side of Figure 7). One must keep in mind, however, that the analysis refers to relative locations of subjects, not to absolute success. Thus, it should not be concluded from the relative treatment effects displayed in Figure 7 that high-achieving subjects really did better in the item condition *embedded/second* than in the item condition *embedded/first*. The F1-LD-F2-procedure, when calculating effects and interactions, first computes ranks for all values and then pools the ranks across item conditions. High-achievers' second-position advantage in items with embedded critical unit mirrors a first-position advantage in the low-achieving group (see Figure 5). Both deviations contribute to the interaction effect that was found. Whether one of the effects—the seemingly second-position advantage in high-achievers and the seemingly first-position advantage in low-achievers—may be considered reliable independently of the other cannot be disentangled on the basis of the present analysis.

In summary, high-achieving subjects' success seems to have been enhanced in items with final critical unit and the solution being presented first. Moreover, their success seems to have been relatively unaffected compared to other subjects in items with embedded critical unit and the solution being presented second. This may be called a position-related response shift which is observed in high-achievers but not in low- and medium-achievers.

Discussion

According to hypothesis A, there will be considerable interindividual variation in syntactic reading even in subjects who seem to have mastered capitalization in writing. This is confirmed by the data. The SRT scores are distributed over a wide range. This is not simply the result of subjects solving the tasks by guessing because the score distribution differs significantly from that which would be obtained were that the case.

Hypothesis B predicts that the distribution of SRT scores will have two peaks, one in the lower range and one near ceiling. This was not confirmed by the data. There are two peaks in the lower range, both well above chance level. One peak is near ceiling, but this might be due to random variation. The failure of Hypothesis B concerning the ceiling level peak might be explained by the fact that its chance of being corroborated depends on the distribution of the linguistic abilities in the sample. In the given case, the sample did not include students of the most academically-rigorous secondary school type in Germany (Gymnasium) who can be expected to have higher proficiencies than the subjects tested. Alternatively,

one could assume that Hypothesis B is definitively wrong because it does not account for the conditions under which a bimodal score distribution emerges. One of these conditions might be that syntactic structure is more clearly put into focus by the given task than in the SRT. In data on the detection of the syntactic difference between noun and verb use with homograph units among fifth-grade to seventh-grade students, a bimodal score distribution was obtained (Funke, 2005). In this study, subjects were given assignments that focused strongly (though not explicitly) on syntactic structure. Perfect achievement was observed in 17.4% of the subjects, compared to 2.6% in the SRT. In contrast, in a syntactic reading study in which all critical units were positioned in complete paragraphs rather than in sentence fragments and where subjects had to paraphrase the paragraphs without being alerted to their potential ambiguity by presenting two possible continuations to them, no more than 0.8% of the participants obtained a perfect score. This was true even though the subjects were older than those in the current study (Funke & Sieger, 2009).

In Hypothesis C, it is proposed that position effects are different for low- and high-achieving subjects. This claim is substantiated by the position-related response shift found in the high-achieving subjects.

Readers' ability to quickly access both meanings of a potentially ambiguous string of text and to immediately determine which meaning to maintain and which to discard is a precondition for the effective functioning of orthographic markers of syntactic structure in writing systems. In the SRT, capitalization serves as an orthographic marker of syntactic structure. If the condition of quick access is fulfilled, the reader may envisage what the marker signals in the case given. This condition is likely to be met more frequently by readers with sufficient interpretive abilities than by others because the former may be assumed to activate both potential meanings of the written string without relevant delay even if one of them is less common or less contextually appropriate than the other. If the condition of immediate determination is fulfilled, readers will probably consider capitalization during the first-pass reading because they give it a meaning in the context given. Both conditions taken together describe what it means to evaluate capitalization in reading. In the SRT, subjects incidentally have to decide which alternative to mark. This process must be distinguished from the evaluation of capitalization in the reading process itself, for it is a postponed deliberation that occurs after reading. However, if a reader did evaluate capitalization, his or her choice of which alternative to mark will be based on the access to syntactic information gained during reading. In this case, the decision process may be described as information-driven.

If the condition of quick access or the condition of immediate determination is not fulfilled, the reader is likely, when deciding later on the solution of an SRT item, to either consider only one of both potential meanings of the text or to need to deal with the unsettled affair of the two concurring meanings. In the first case, she or he may readily respond to the item, but the chance of solving it will generally be random, even if, in any single item given, it may be raised by favorable item polarity. In the second case, the subject must decide upon an alternative because the task has to be solved, but lacks a hint about what to attend to. As the evaluation of capitalization in the reading process itself has not ended, the syntactic information needed for a deliberate reexamination of the text will not be reliably accessible. The kind of decision resulting from this may be described as task-driven.

A task-driven decision process is likely to be laborious and to give rise to frequent confusion because it puts a heavy load on working memory. This might lead to a preference for marking alternatives presented in a specific position in order to short-cut the process, resulting in effects of position of solution. Such preferences were not demonstrated in the present study, but they have been observed in multiple choice assignments with lengthy or difficult alternatives (Porst, 1985; Wellenreuther, 2000) and also in a syntactic task based on a multiple choice format (Funke, 2005).

In an information-driven decision process, in contrast, one may expect intervening effects of position of critical unit. A critical unit in a final position is likely to be prominent. In the event that a reader interprets the text based on the evaluation of capitalization, he or she will be able, by this prominence, to tie his or her interpretation to the spelling of the critical unit by taking an extra glance at it after the first-pass reading. If the solution is presented as the first alternative in this case, it will not be necessary to seriously consider the second before deciding which one to mark. Processing load will be reduced as well as the resultant confusion. If, on the other hand, the critical unit is embedded, the interpretation of the text is less likely to be corroborated by a second glance at the critical unit. Therefore, it is more probable that the choice of an alternative results from a postponed consideration taking place after reading both alternatives. However, contrary to the situation given in a task-driven process, subjects may have an idea (albeit vague) of what they need to look for—specifically, that they must look for syntactic features. This might sustain their inclination to consider both alternatives equally closely and eliminate positional preferences.

Based on this consideration, one may question whether the secondary use of syntactic information in reading is a consequence of metacognitive control or attention to linguistic form. When a reader exerts metacognitive control, she or he deliberately monitors reading. By doing so, she or he may, when considering syntactic structure, be led to focus on formal features of the text regardless of whether the appropriate syntactic information is reliably disposable. According to the data presented, that will not favor access to syntactic information because this access results from the reader being information-driven instead of task-driven.

Open questions concern proposals seeking to explain differences in the continued access to syntactic information by working memory variation. In the case of the SRT, working memory load may be assumed to be potentially high because two syntactic patterns must be kept in mind while checking the two alternatives. The importance of working memory functioning for the processing of ambiguous text has been demonstrated repeatedly (Fiebach, Vos, & Friederici, 2004; Friederici, Steinhauer, Mecklinger, & Meyer, 1998; Just & Carpenter, 1992; MacDonald, Just, & Carpenter, 1992; Miyake, Just, & Carpenter, 1994; Vos, Gunter, Schriefers, & Friederici, 2001).

The data presented are not suited for evaluating explanations of the observed position effects based on the assumption of working memory limitations. Nevertheless, they highlight an interesting question concerning it. Some accounts of working memory assume that the information needed to cope with syntactic tasks is preserved in working memory by phonological coding (Crain & Shankweiler, 1988; Shankweiler, 1999). According to Crain and Shankweiler, poor comprehenders experience difficulties establishing or maintaining high-quality phonological representations of reading input. Consequently, they lack a solid base on which to sustain syntactic information in working memory. As for the

SRT, one might ask how phonological representations may be used to preserve the syntactic information required. In this task, subjects must deal with two possible readings of a sentential fragment which differ phonologically, if at all, only in their prosodic features. More specifically, if prosodic differences exist, they would concern sentential prosody, not lexical prosody. However, relying on sentential prosody in order to preserve syntactic information seems to be risky and difficult. Informal experience suggests that the conscious interpretation of sentential prosody is no easier than preserving syntactic information (Redder, 1982). In the case of the SRT, maintaining one reading of an item (either the noun or verb reading) based on a phonological representation carries the risk of maintaining the other reading as well, thereby giving rise to confusion. For this reason, one may question whether high-achievers in the SRT really rely on phonological encoding to solve the items.

Several methodological limitations should be noted that might reduce the force of the arguments drawn from the results.

First, data on participants' reading abilities were not collected. It would be important to make sure that subjects were not hampered by insufficient reading skills when working on the SRT. It seems unlikely, though, that the results of the study were extensively biased by subjects' deficient reading abilities. The sample was selected to reach full score in the orthographic writing task. In a study using the same task with sixth-grade students (Funke, Wieland, Schönenberg, & Melzer, 2011), subjects who obtained perfect scores significantly outperformed the other subjects on a self-conceived test of reading comprehension ($t(254) = 3.27$; $p = .001$). Their average reading score was about one-half standard deviation above that of the rest of the group.

Second, although an attempt was made to eliminate the biasing effects of different item polarity means in the four item conditions considered, the results may nevertheless have been influenced by different item polarity distributions across the conditions. This possibility can be ruled out only by further investigation. Another point that must be considered is the fact that some items of the SRT did not fulfill the definability condition of polarity. Even so, the possible effects of this feature in one version of the SRT should cancel out the adverse effects in the other.

Third, one has to bear in mind the nature of the syntactic reading task. In order to cope with this task, a subject needs not only to activate syntactic information but also to attend to capitalization. Therefore, one might conclude from their SRT achievement that some subjects do not seem to reliably dispose of syntactic information even though they actually do. The benefits of the SRT—its ability to assess a reader's disposal of syntactic information in the reading process itself—must be weighed against its shortcomings, which include a possible underestimation of individuals' abilities.

References

Arciuli, J., & Monaghan, P. (2009). Probabilistic cues to grammatical category in English orthography and their influence during reading. *Scientific Studies of Reading, 13*, 73-92.

Bautier, É., & Branca-Rosoff, S. (2002). Pratiques linguistiques des élèves en échec scolaire et enseignement [Linguistic behavior of students failing in school and instruction]. *Ville-Ecole-Intégration Enjeux, 130*, 196-213.

Bentin, S., Deutsch, A., & Liberman, I. Y. (1990). Syntactic competence and reading ability in children. *Journal of Experimental Child Psychology, 48*, 147-172.

Bock, M. (1986). Cognitive aspects of upper and lower case for initial letters in German. In G. Augst (Ed.), *New trends in graphemics and orthography* (pp. 287-299). Berlin, Germany: de Gruyter.

Bock, M. (1990). Zur Funktion der deutschen Groß- und Kleinschreibung: Einflüsse von Wortform, Muttersprache, Lesealter, Legasthenie und lautem versus leisem Lesen [The function of capitalization in German: Effects of word shape, readers' first language, reading age, reading disability, and reading aloud vs. reading silently]. In C. Stetter (Ed.), *Zu einer Theorie der Orthographie* [*Contributions to a theory of orthography*] (pp. 1-32). Tübingen, Germany: Niemeyer.

Boland, J. A. (1997). Resolving syntactic category ambiguities in discourse context. *Journal of Memory and Language, 36*, 588-615.

Borg, I., & Staufenbiel, T. (2007). Theorien und Methoden der Skalierung [Theories and methods of scaling]. 4th ed. Bern, Switzerland: Huber.

Bowey, J. A. (1986). Syntactic awareness in relation to reading skill and ongoing comprehension monitoring. *Journal of Experimental Child Psychology, 41*, 282-299.

Bowey, J. A. (1993). Grammatical awareness and learning to read: A critique. In E. M. H. Assink (Ed.), *Literacy acquisition and social context* (pp. 122-149). New York, NY: Harvester Wheatsheaf.

Brunner, E., Domhof, S., & Langer, F. (2002). *Nonparametric analysis of longitudinal data in factorial experiments*. New York, NY: Wiley.

Bryant, P., Devine, M., Ledword, A., & Nunes, T. (1997). Spelling with apostrophes and understanding possession. *British Journal of Psychology, 67*, 93-112.

Crain, S., & Shankweiler, D. (1988). Syntactic complexity and reading acquisition. In A. Davison & G. M. Green (Eds.), *Linguistic complexity and text comprehension* (pp. 167-192). Hillsdale, NJ.: Erlbaum.

Dik, S. C. (1997). *The theory of functional grammar. Part 1*. 2nd ed. Berlin, Germany: de Gruyter.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language, 27*, 429-446.

Eichler, W. (2002). *Rechtschreiben lernen* [Learning to spell]. Oldenburg, Germany: Universität Oldenburg.

Fiebach, C., Vos, S. H., & Friederici, A. D. (2004). Neural correlates of syntactic ambiguity in sentence comprehension for low and high span readers. *Journal of Cognitive Neuroscience, 16*, 1562-1575.

Friederici, A. D., Steinhauer, K., Mecklinger, A., & Meyer, M. (1998). Working memory constraints on syntactic ambiguity resolution as revealed by electrical brain responses. *Biological Psychology, 47*, 193-221.

Funke, R. (1995). Grammatik in Funktion [Grammar at work]. *Der Deutschunterricht, 48*, 430-434.

Funke, R. (2005). *Sprachliches im Blickfeld des Wissens* [The expliciteness of linguistic knowledge]. Tübingen, Germany: Niemeyer.

Funke, R., & Sieger, J. (2009). Die Nutzung von orthographischen Hinweisen auf syntaktische Strukturen und ihre Bedeutung für das Leseverstehen [The use of orthographic clues to gain access to syntactic structure and its role in reading comprehension]. *Didaktik Deutsch, 26*, 31-53.

Funke, R., Wieland, R., Schönenberg, S., & Melzer, F. (2011). *Die Lernwirksamkeit sprachreflexiven Unterrichts. Abschlussbericht* [*The effects of reflecting on language in language education. Final report*]. Heidelberg: Pädagogische Hochschule Heidelberg.

Gaux, C., & Gombert, J. E. (1999). Implicit and explicit syntactic knowledge and reading in pre-adolescents. *British Journal of Developmental Psychology, 17*, 169-188.

Givón, T. (2001). *Syntax: An introduction*. Vol. 1. Rev. ed. Amsterdam, The Netherlands: Benjamins.

Gombert, J. É. (1990). *Le développement métalinguistique* [*Metalinguistic development*]. Paris, France: Presses Universitaires de France.

Gottardo, A., Stanovich, K. E., & Siegel, L. S. (1996). The relationships between phonological sensitivity, syntactic processing, and verbal working memory in the reading performance of third-grade children. *Jounal of Experimental Child Psychology, 63*, 563-582.

Grimshaw, J. (2005). *Words and structure*. Stanford, CA: Center for the Study of Language and Information.

Guthrie, J. T. (1973). Reading comprehension and syntactic responses in good and poor readers. *Journal of Educational Psychology, 65*, 294-799.

Halliday, M. A. K. (2004). *An introduction to functional grammar*. 3rd ed., revised by C. M. I. M. Matthiessen. London, United Kingdom: Hodder Arnold.

Heckel, B. (1980). Nach wie vor ein Fehlerschwerpunkt: die Groß- und Kleinschreibung [Persistent troubles with capitalization]. *Deutschunterricht, 33*, 411-415.

Isakson, R. L., & Miller, J. W. (1976). Sensitivity to syntactic and semantic cues in good and poor comprehenders. *Journal of Educational Psychology, 68*, 787-792.

Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension*. Newton, MA: Allyn and Bacon.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122-149.

Kemp, N., Nilsson, J., & Arciuli, J. (2009). Noun or verb? Adult readers' sensitivity to spelling cues to grammatical category in word endings. *Reading and Writing: An Interdisciplinary Journal, 22*, 661-686.

Lautsch, E., & Lienert, G. A. (1993). *Binärdatenanalyse* [Analysis of binary data]. Weinheim, Germany: Psychologie Verlags-Union.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology, 24*, 56-98.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676-703.

Mentrup, W. (1993). *Wo liegt eigentlich der Fehler? Zur Rechtschreibreform und ihren Hintergründen* [*So what's wrong? The reform of orthography and its backgrounds*]. Stuttgart, Germany: Klett.

Menzel, W. (1985). *Rechtschreibunterricht* [*Spelling instruction*]. Seelze, Germany: Friedrich.

Millis, K., Magliano, J., & Todaro, S. (2006). Measuring discourse-level processes with verbal protocols and latent semantic analysis. *Scientific Studies of Reading, 10*, 225-240.

Miyake, A., Just, M. A., & Carpenter, P. A. (1994). Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple representations in neutral contexts. *Journal of Memory and Language, 33*, 175-202.

Müsseler, J., Nißlein, M., & Koriat, A. (2005). German capitalization of nouns and the detection of letters in continuous text. *Canadian Journal of Experimental Psychology, 59*, 143-158.

Nagy, W. (2007). Metalinguistic awareness and the vocabulary-comprehension connection. In R. K. Wagner, A. E. Muse, & K. R. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 52-77). New York, NY: Guilford.

Nagy, W., & Scott, J. A. (2000). Vocabulary processes. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. III, pp. 269-284). Mahwah, NJ: Erlbaum.

Nation, K., & Snowling, M. (2000). Factors influencing syntactic awareness skills in normal readers and poor comprehenders. *Applied Psycholinguistics, 21*, 229-241.

Pomm, H. P., Mewes, U., & Schüttler, H. (1974). Die Entwicklung der Rechtschreibleistungen von Schulkindern unter besonderer Berücksichtigung von Reformvorschlägen [The development of spelling achievement in students considered with respect to proposals for orthography reform]. In G. Augst (Ed.), *Deutsche Rechtschreibung mangelhaft?* [*Failing grades in German spelling?*]. (pp. 59-78). Paderborn, Germany: Schöningh.

Porst, R. (1985). *Praxis der Umfrageforschung* [Doing inquiries in social research]. Stuttgart, Germany: Metzler.

Pugh, K. R., Mencl, W. E, Jenner, A. R., Katz, L., Frost, S. J., Lee, J. R. … Shaywitz, B. A. (2001). Neurobiological studies of reading and reading disability. *Journal of Communication Disorders, 34*, 479-492.

R Development Core Team (2007). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Rat für deutsche Rechtschreibung (2006). *Deutsche Rechtschreibung. Regeln und Wörterverzeichnis. Amtliche Regelung* [*German orthography. Rules and word list. Official definition*]. Tübingen, Germany: Narr.

Redder, A. (1982). *Schulstunden 1. Transkripte* [School lessons 1. Transcripts]. Tübingen, Germany: Narr.

Ruoff, A. (1990). *Häufigkeitswörterbuch gesprochener Sprache* [Frequency dictionary of spoken language]. 2nd ed. Tübingen, Germany: Niemeyer.

Schleppegrell, M. J. (2004). *The language of schooling. A functional linguistics perspective*. Mahwah, NJ: Erlbaum.

Shankweiler, D. (1999). Words to meanings. *Scientific Studies of Reading, 3*, 113-127.

Shankweiler, D., Crain, S., Katz, L., Fowler, A. E., Liberman, A. M., Brady, S. A. … Shaywitz, B. A. (1995). Cognitive profiles of reading-disabled children: Comparison of language skills in phonology, morphology, and syntax. *Psychological Science, 6*, 149-155.

Verhoeven, L., & Perfetti, C. A. (2008). Advances in text comprehension: Model, process and development. *Applied Cognitive Psychology, 22*, 293-301.

Vos, S. H., Gunter, T. C., Schriefers, H., & Friederici, A. D. (2001). Syntactic parsing and working memory: The effects of syntactic complexity, reading span, and concurrent load. *Language and Cognitive Processes, 16*, 65-103.

Wellenreuther, M. (2000). *Quantitative Forschungsmethoden in der Erziehungswissenschaft* [*Methods of quantitative research in educational science*]. Weinheim, Germany: Juventa.

Table 1 Classification of responses to SRT items with critical unit presented as noun

Critical unit is read by the subject as …	Critical unit is presented to the subject as …	
	… noun	… verb
… noun	a	b
… verb	c	d

Note. Frequency counts for the left column are taken from one version and frequency counts for the right column from the other version. For items in which the critical unit was presented as a verb, rows and columns have to be interchanged.

Table 2 Effects and interactions of version and item conditions for standardized values

Factor	*Q**	*f*	*p*
Version	0.17	1	.68
Position of critical unit	0.21	1	.65
Position of solution	0.00	1	.99
Version × Position of critical unit	2.48	1	.12
Position of solution × Position of critical unit	0.21	1	.64
Version × Position of solution	1.73	1	.19
Version × Position of critical unit × Position of solution	0.03	1	.86

Note. *Q** - rank statistic (Box-type statistic), *f* – estimated degree of freedom, *p* – probability under null hypothesis.

Computations were run in the R statistical environment 2.7.3 (cf. R Development Core Team, 2007). Functions written in R for the nonparametric models of the Brunner et al. (2002) group are supplied on their homepage (http://www.ams.med.uni-goettingen.de/de/sof/index.html). These functions were applied with some corrections.

Table 3 Effects and interactions of achievement levels and item conditions for standardized values

Factor	*Q**	*f*	*p*
Achievement group	498.80	1.79	.00
Position of critical unit	0.07	1	.80
Position of solution	0.01	1	.91
Achievement group × Position of critical unit	3.00	1.99	.05
Position of solution × Position of critical unit	0.43	1	.51
Achievement group × Position of solution	0.78	1.84	.45
Achievement group × Position of critical unit × Position of solution	5.31	1.94	.01

*Note. Q** - rank statistic (Box-type statistic), *f* – estimated degree of freedom, *p* – probability under null hypothesis.

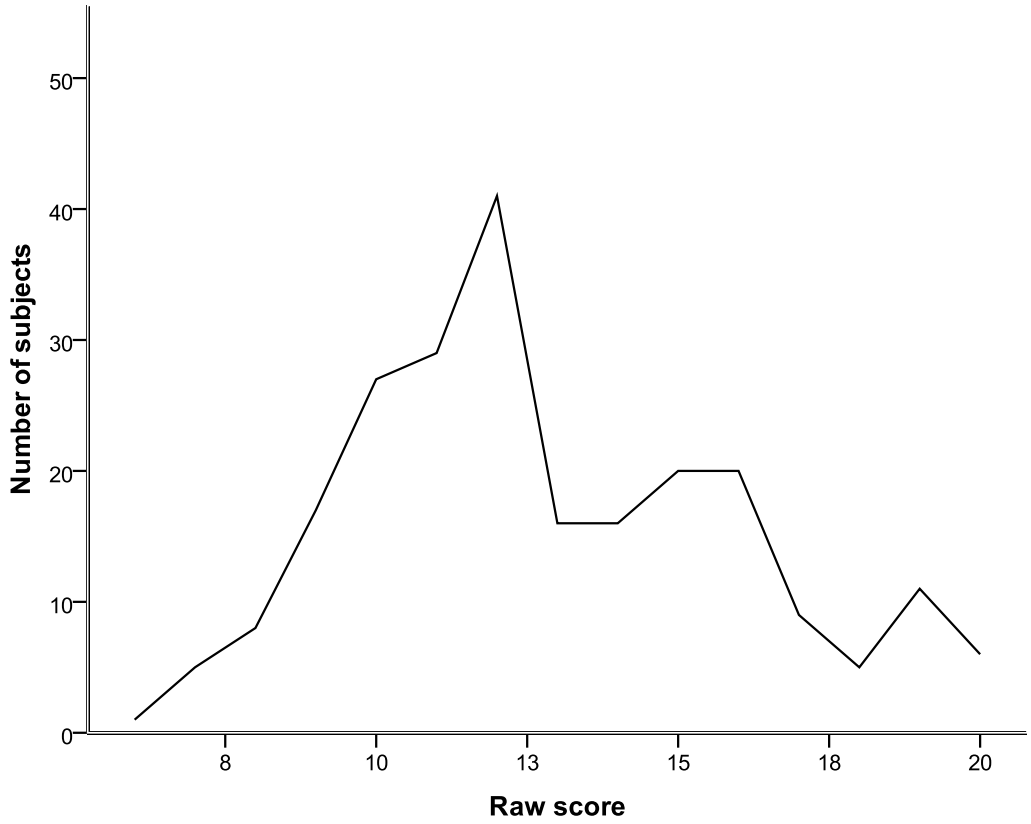

Fig. 1 Distribution of syntactic reading scores.

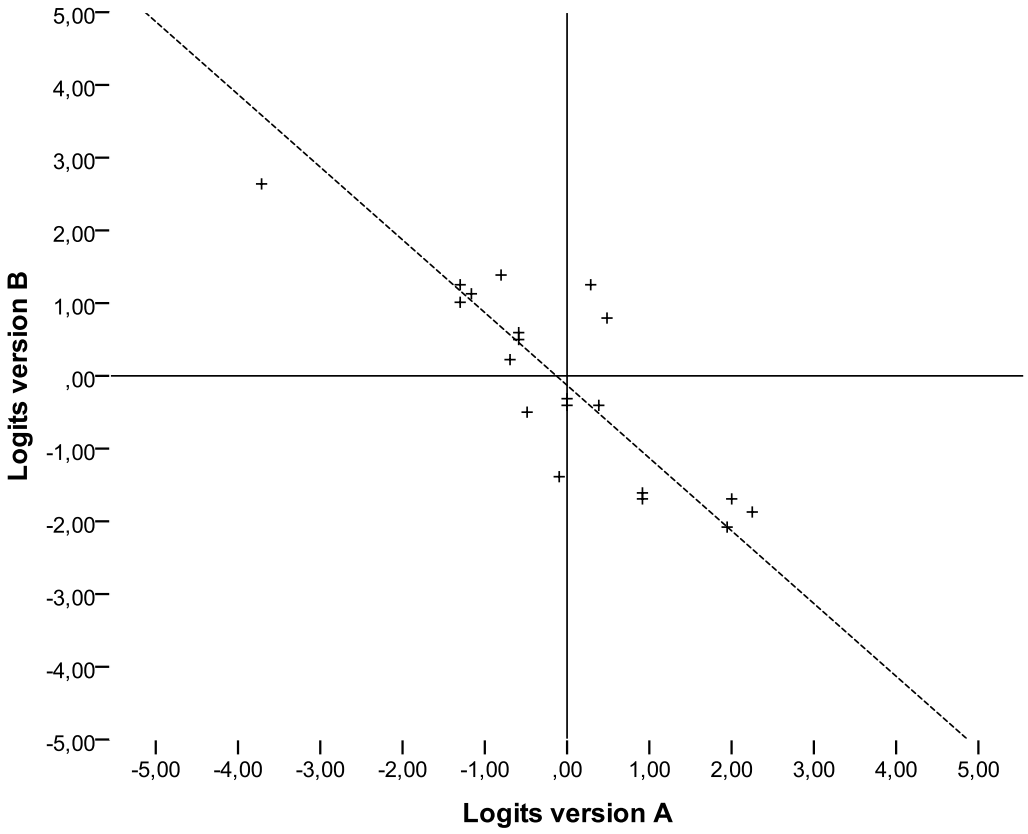

Fig. 2 Relative odds of an item being solved correctly by subjects with raw scores between 0 and 11 (n = 87) compared to the baseline condition. Odds are displayed as logits. Each item is represented by a cross. The inserted line indicates which values are to be expected if item polarity is the only determinant of item difficulty and uniformly influences the odds of an item being solved correctly in all achievement groups.

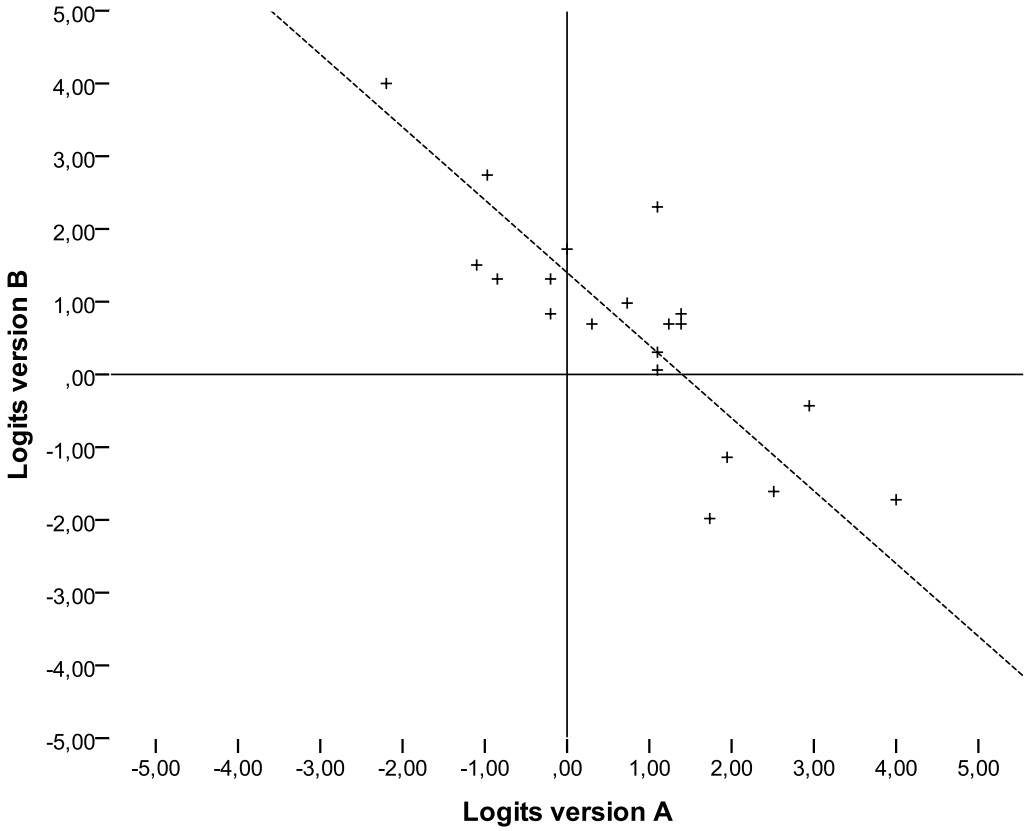

Fig. 3 Relative odds of an item being solved correctly by subjects with raw scores between 12 and 14 (*n* = 73) compared to the baseline condition. For further comments, see Fig. 2.

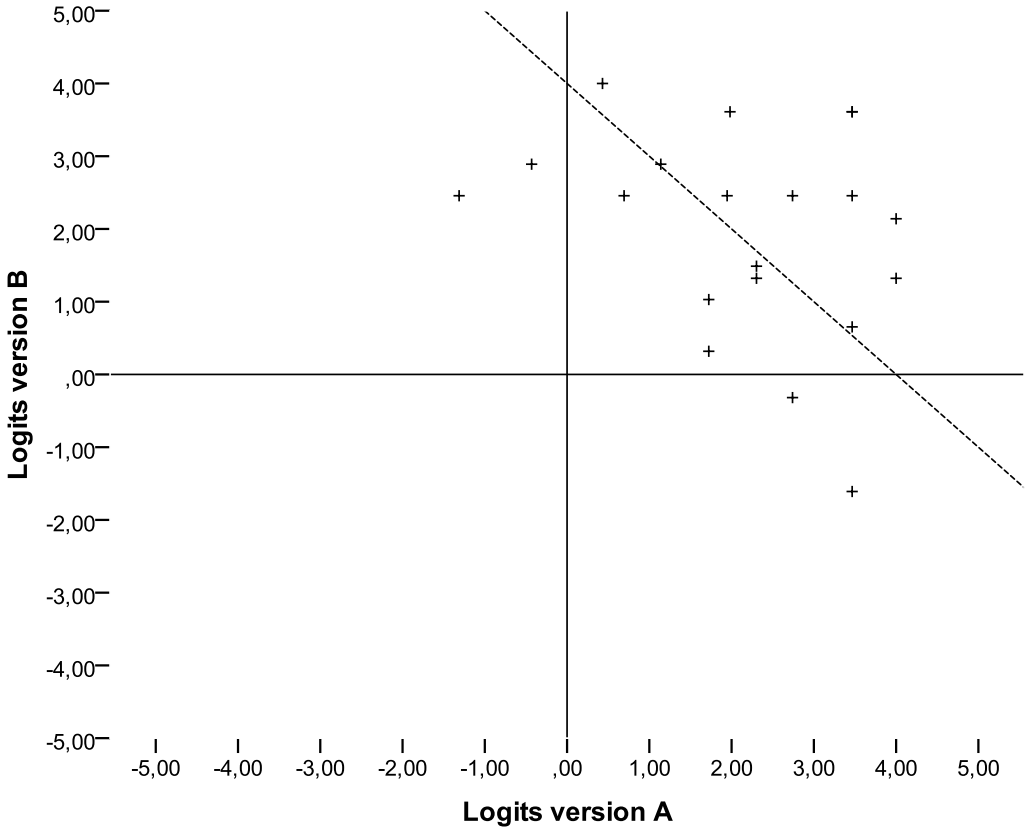

Fig. 4 Relative odds of an item being solved correctly by subjects with raw scores between 15 and 20 (*n* = 71) compared to the baseline condition. For further comments, see Fig. 2.

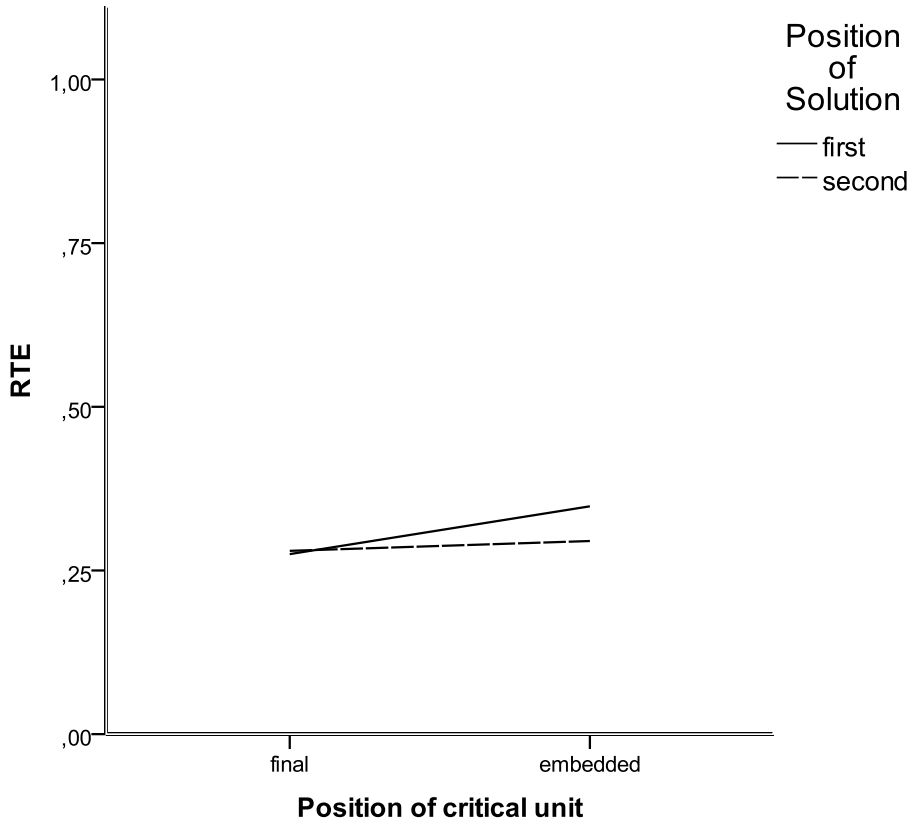

Fig. 5 Effect of position of critical unit on the choice of solutions presented first or second in the lowest-achieving group of the SRT (with scores between 0 and 11). Effects are measured by relative treatment effects (RTE).

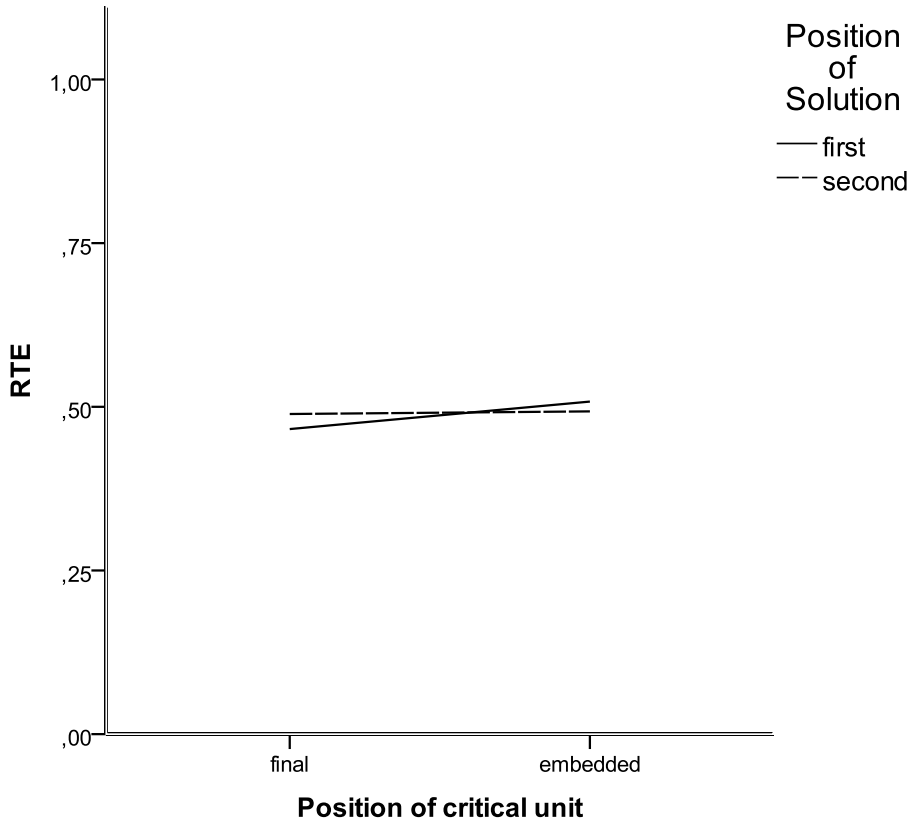

Fig. 6 Effect of position of critical unit on the choice of solutions presented first or second in the medium-achieving group of the SRT (with scores between 12 and 14). Effects are measured by relative treatment effects (RTE).

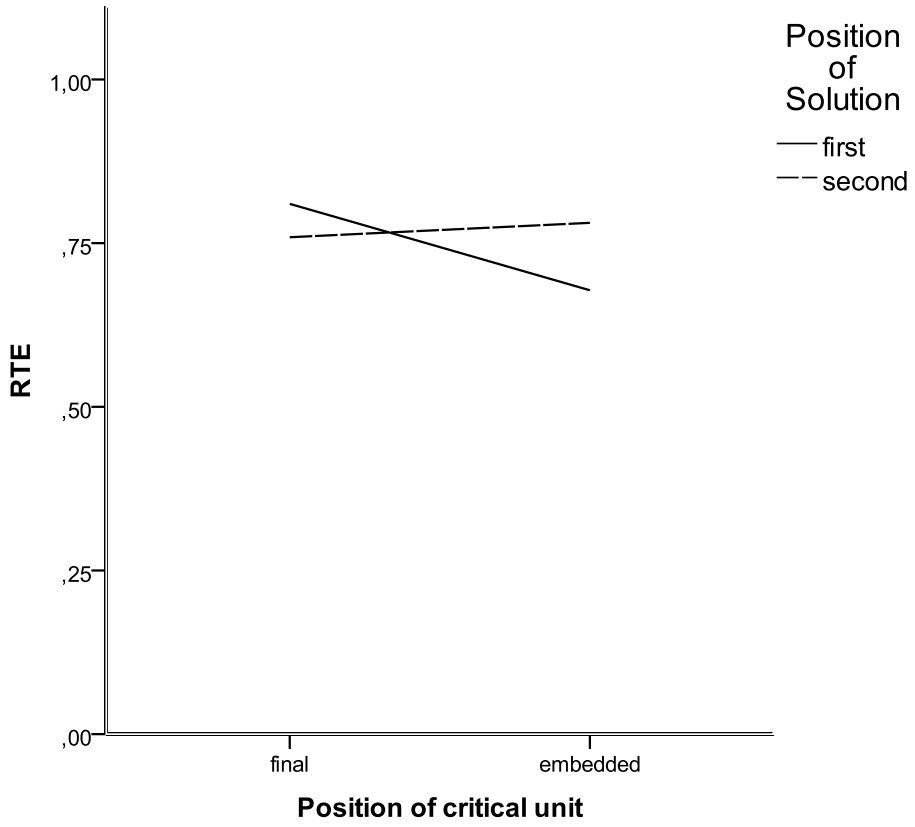

Fig. 7 Effect of position of critical unit on the choice of solutions presented first or second in the highest-achieving group of the SRT (with scores between 15 and 20). Effects are measured by relative treatment effects (RTE).

Appendix A

The Syntactic Reading Task (version A)

Marlies schreibt: „Typische Streber erkennt man daran, dass sie sofort antworten auf jede Frage ...

☐ ... geben müssen.“

☐ ... des Lehrers.“

Gerd freut sich über die neuen Waschräume in der Sporthalle: „Es ist schließlich nicht egal, ob wir nach der Sportstunde duschen ...

☐ ... benutzen können oder nicht.“

☐ ... oder verschwitzt nach Hause gehen.“

Inken meint: „Auf die ganze Briefkasten-Reklame von den Firmen gebe ich gar nichts. Die Versprechen in ihren Prospekten ...

☐ ... stimmen ja doch meistens nicht.“

☐ ... alles Mögliche, halten es aber nicht.“

Uwe schimpft: „Im Winter nerven manche Schüler, wenn sie bei Klassenarbeiten husten...

☐ ... haben und andere stören.“

☐ ... und andere damit ablenken.“

Swantje sagt: „Nach dem Unterricht stehen immer dieselben um die Lehrerin herum. Die Fragen nach den Hausaufgaben, ...

☐ ... bloß weil sie nicht aufgepasst haben.“

☐ ... die die stellen, sind ziemlich dumm.“

Der Kommissar erklärte dem Polizisten: „Daran sehe ich, dass die Gangster vor kurzem noch in der Wohnung waren—weil die Wasserhähne Tropfen ...

☐ ... und noch überall nass sind.“

☐ ... ins Waschbecken fallen lassen.“

Rüdiger meint: „Streitigkeiten gehe ich möglichst aus dem Weg, weil die mehr schaden ...

☐ ... als sie am Ende bringen.“

☐ ... anrichten als Nutzen bringen.“

Bernhard prahlt: „Wir sind richtige Fans, die auf dem Fußballplatz jedes Mal pfeifen ...

☐ ... bei sich haben."

☐ ... so laut sie können."

Jost sagt: „Ich bewundere die 100-Meter-Läufer in einem voll besetzten Stadion. Verfolgt von Geschrei, Blicken ...

☐ ... und Gedanken der Zuschauer, jagen sie auf das Ziel los."

☐ ... sie nur auf eines: Die Ziellinie, die sie vor sich haben."

Wolfgang meint: „Unser Fußballtrainer lobt uns am meisten, wenn wir Kämpfen um den Ball ...

☐ ... statt vor ihm wegzulaufen."

☐ ... nicht aus dem Weg gehen."

Elke erklärt: „Bei Reklame im Radio höre ich gar nicht hin. Meine Mutter behauptet, dass die da lügen wie kein anderer sonst ...

☐ ... auf der Welt."

☐ ... verbreiten."

Leif sagt: „Ich habe öfters Schuppen auf der Jacke. Aber damit werde ich schon fertig. Die Bürste ...

☐ ... hängt ja immer griffbereit neben der Tür."

☐ ... ich mir vor dem Rausgehen schnell ab."

Holger meint: „Ich freue mich, wenn der Tabellenführer zu uns kommt, weil ich gerne Spiele ...

☐ ... gegen starke Gegner mitmache."

☐ ... gegen starke Gegner wie die."

Norman ist begeistert: „Wenn die Brasilianer eine Weltmeisterschaft gewonnen haben, laufen alle echten Fans nach draußen. Die feiern auf den Straßen ...

☐ ... bis spät nach Mitternacht."

☐ ... hören gar nicht mehr auf."

Renate warnt: „Die Hunde werden viel zu fett, wenn sie ständig so viel Fressen wie heute ...

☐ ... und nie damit aufhören."

☐ ... vorgesetzt kriegen."

Ina erzählt: „Manchmal bin ich nachts ganz unruhig. Erst gegen Morgen schlafe ich richtig ein, weil ich dann träume von schönen Sachen ...

☐ ... und mich richtig gut fühle.“

☐ ... in meinem Kopf habe.“

Levke schimpft: „Für unsere Garage müssen wir jeden Monat 100 Euro zahlen. Das kann doch nicht wahr sein! Andere Mieten in dieser Gegend ...

☐ ... sind gerade mal halb so hoch.“

☐ ... zwei Garagen für denselben Preis.“

Jürgen erzählt: „Es passiert mir ja öfter, dass ich etwas verliere, aber seit meine Armbanduhr weg ist, finde ich keine Ruhe mehr. Die suche ...

☐ ... ich jetzt schon seit zwei Wochen.“

☐ ... danach dauert nun bald zwei Wochen.“

Norbert sagt: „Mittags kommt unser Hund sofort in die Küche, wenn er riecht, dass wir da Braten ...

☐ ... oder kochen.“

☐ ... für uns machen.“

Helga erzählt: „Neulich wurde unser Dackel Waldi von einem fremden Hund richtiggehend überfallen. Der biss von ihm ...

☐ ... verletzte Waldi am Ohr.“

☐ ... fast ein ganzes Ohr ab.“

Appendix B

Table B1 Features of the syntactic reading items (version A)

Critical unit	Item position (version B)	Position parameters		Category	Frequency class		Polarity index
		Pos. of critical unit	Pos. of solution		Noun	Verb	
ANTWORTEN	7	embedded	second	V	F1	F1	.178
DUSCHEN [b]	3	final	second	V	F0	F0	.504
VERSPRECHEN [b]	20	embedded	first	N	F1	F2	-.380
HUSTEN	11	final	second	V	F1	F1	.240
FRAGEN	15	embedded	second	N	F2	F3	-.814
TROPFEN	13	final	second	N	F1	F0	-.349
SCHADEN	9	final	first	V	F2	F1	-.457
PFEIFEN	5	final	second	V	F1	F2	.426
BLICKEN	8	final	first	N	F1	F1	-.008
KÄMPFEN	1	embedded	second	N	F2	F2	-.566
LÜGEN	2	embedded	first	V	F1	F1	.566
BÜRSTE	4	final	first	N	F1	F0	.225
SPIELE	6	final	first	N	F2	F4	-.271
FEIERN	19	embedded	first	V	F1	F2	.643
FRESSEN	10	embedded	second	N	F1	F3	-.147
TRÄUME	12	embedded	first	V	F1	F0	.225
MIETEN	17	embedded	first	N	F0	F0	-.132
SUCHE	14	final	first	V	F0	F3	.783
BRATEN [a]	18	final	second	N	F1	F1	-.209
BISS [a]	16	embedded	second	V	F1	F1	-.457

Note. Words marked as N (noun) were presented to subjects with a capitalized initial letter; words marked as V (verb) were presented with a lower-case initial letter. Noun and verb frequencies were taken from the word frequency list of spoken German compiled by Ruoff (1990) and were converted to frequency class values to be comparable across lists and languages. Each frequency class covers about 20% of the noun and verb tokens of the Ruoff database. Frequency class F1 is made up of the least frequent and frequency class F5 of the most frequent words. Words not contained in the Ruoff list were assigned F0. Polarity indices are positive if item polarity conforms to how the word was presented in version A and negative otherwise. Indices with absolute values above .173 are significant. Category values, position of solution, and polarity indices are inverted in version B.

[a] Item does not fulfill the definability condition because there were significantly more convergent (correct) solutions than expected even in subjects responding on chance level. [b] Item does not fulfill the definability condition because there were significantly more divergent (incorrect) solutions than expected even in subjects responding on chance level.