EXPLORING SYNTACTIC STRUCTURES IN FIRST-LANGUAGE EDUCATION: EFFECTS ON LITERACY-RELATED ACHIEVEMENTS

REINOLD FUNKE*, REGINA WIELAND*, STEPHANIE SCHÖNENBERG* & FLORIAN MELZER**

** Pädagogische Hochschule, Heidelberg*
*** Staatliches Seminar für Didaktik und Lehrerbildung, Heidelberg*

Abstract
A quasi-experimental study with 256 German sixth-grade students examined the effects of exploring syntactic structures on literacy-related achievements. The core feature of the instruction was that students had the opportunity to deal with syntactic structures directly, not mediated by analytic operations. Instruction covered six lessons targeting two types of syntactic contrasts (syntactic category contrasts, noun case contrasts). The results indicate that students profited from the instruction in their ability to use a marker of syntactic structure in written German, the capitalization of nouns. As to the ability to use the same marker to interpret text read, students of different achievement levels seem to have been differently affected. No effects were found for reading comprehension in general and spelling at the word level. The findings suggest that students' recognition of syntactic categories was enhanced even in contexts other than those they were familiarized within the instruction. However, it was not enhanced when they worked on a task that was completely new to them.

Keywords: Category ambiguity, Grammar instruction, Reading comprehension, Spelling, Syntactic processing

Funke , R., Wieland , R., Schönenberg , S., & Melzer, F. (2013). Exploring syntactic structures in first-language education: effects on literacy-related achievements. L1-Educational Studies in Language and Literature, 13, *1-24. http://dx.doi.org/10.17239/L1ESLL-2013.01.03*
Corresponding author: *Reinold Funke, Pädagogische Hochschule Heidelberg, Im Neuenheimer Feld 561, D-69120 Heidelberg, Germany, e-mail: funke@ph-heidelberg.de*

1. INTRODUCTION

In first-language education, students may explore syntactic structures in various ways. They do so when teachers have them focus on syntactic structure in order to grasp it or to deliberately manipulate it. This includes activities such as discussing alternative ways of expressing a thought, resolving structural ambiguities in text and systematically studying sentence patterns. It may happen in a scheduled or an incidental fashion, and it may or may not encompass the use of linguistic nomenclature.

Linguists have repeatedly advanced the proposal that the grasp of syntactic structure is a precondition to successfully achieving literacy-related tasks such as comprehending written text (Glinz, 1963; 1993; Heringer, 2001), formulating thoughts in writing (Schleppegrell, 2004) and learning to spell (Maas, 1992). This idea flows from structural analyses of the demands of reading and writing. According to Maas (1992), for example, recording a message in writing involves 'fixating' its grammatical structure beyond what is given in oral records because of constraints set by the orthographic writing system.

The assumption that grasping syntactic features is, for structural reasons, a precondition to the accomplishment of reading and writing tasks might seem to suggest that exploring syntactic structures will enhance the literacy-related achievements of students. However, it does not necessarily imply that this is indeed the case. Empirical evidence is disparate and inconclusive. Work concerning the effect of grammar instruction in first-language education on the quality of students' writing has found no benefits resulting from it (see the overview by Andrews, Torgerson, Beverton, Freeman, Locke, Low, et al. 2006 and the meta-analyses by Hillocks, 1986; Graham & Perin, 2007 and Graham, McKeown, Kiuhara, & Harris 2012). It seems, however, to be possible to successfully work on narrowly defined writing problems by offering grammar instruction to students which is tailored to their individual needs (Rogers & Graham, 2008). In the following, we focus on studies pertaining to whether exploring syntactic structures affects reading comprehension. Such studies are rare and they vary in quality (see the short overview to be found in Bowey, 1993). However, considering them may serve to highlight questions that are relevant to writing and spelling as well.

Three strands of research may be distinguished. First, the impact of grammar instruction on reading comprehension has been investigated without specifically tailoring instruction to enhance comprehension and also without measuring reading outcomes by comprehension tasks in which syntactic skill is at stake. Studies of this type have failed to find effects when grammar instruction was the experimental condition (Elley, Barham, Lamb, & Wyllie, 1976; Layton, Robinson, Lawson, 1998; O'Donnell & King, 1974) as well as when it was the control condition (Straw & Schreiner, 1982; Trivelli, 1983; White, Pascarella, & Pflaum, 1981). Second, students have been trained to systematically manipulate word assemblies or senten-

ces to make them coherent wholes. This has been done with elementary and with learning-disabled students using sentence anagrams (Greenewold & Pederson, 1983; Weaver 1979; White et al., 1981), and with more advanced students by using sentence combining tasks (Gonsoulin, 1993; Morenberg, Daiker, & Kerek, 1978; Neville & Searls, 1985; Straw & Schreiner, 1982; Wilkinson & Patty, 1993). It is reported to result in gains in cloze comprehension tests, but there is only marginal evidence for gains in general comprehension (see the overview by Searls & Neville, 1988 and the meta-analytic reviews by Fusaro, 1992; 1993; Neville & Searls, 1991). Third, in psycholinguistic experiments, adult students have been individually trained to parse complex sentences belonging to specific patterns which they found hard (Chipere, 2003; Wills, Christiansen, Race, Achesen, & MacDonald, 2009). This was successful as far as the comprehension of the sentence pattern trained was concerned, but transfer to other comprehension tasks and to text comprehension was not examined.

From this overview, it seems that working on syntactic structures in first-language education has no or just a negligible effect on literacy-related abilities. However, one should qualify this conclusion by saying that, in many of the studies cited, working on syntactic structures has been not the experimental but the control condition. What type of instruction was actually administered is not well described. Instead, this instruction is in many cases characterized globally as 'traditional grammar'. It is possible that the instructional potential of working on syntactic structures has not been sufficiently explored under this condition. In any case, studies comparing different types of grammar instruction with regard to their impact on reading comprehension have led to results which seem interesting though they do not reveal a systematic pattern (Crews, 1971; Noyce & Christie, 1983). The same is true of studies which mix up syntactic awareness tasks with general language awareness tasks by having primary school children work on jokes and riddles in order to foster reading comprehension (Yuill, 1996; 2007; Zipke, 2008; Zipke, Ehri, & Cairns, 2009).

One consequence that one may draw from the studies cited is that research concerning the impact of exploring syntactic structures on reading comprehension might be taken a step further if the effects which are expected to result from it as well as the type of comprehension achievement it is hypothesized to serve were described in more detail. Theory should specify which reading processes are assumed to involve syntactic achievements, and in which reading situations these processes are likely to be needed to achieve the type of comprehension required by the situation.

Another point which becomes evident from the studies is that the effects of grammar instruction on reading that are claimed to exist are rather specific. That is, students display just the behavior trained or solve tasks resembling those occurring in training. This is similar in research concerning writing. As to reading, one may assume that instructional effects that are less narrowly constrained to what has

been taught would be more likely to influence reading practice because they are not tied to situations where specific tasks are given.

The present investigation is based on the proposal that there are situations where readers profit from identifying syntactic structures. A reader is assumed to identify a syntactic structure when she or he is able to repeatedly gain access to it without losing sight of it or gliding into another syntactic pattern which happens to get activated. We hypothesize this to be the starting point from which readers detect how syntactic constructions signal features relevant to the interpretation of text, such as perspective (Dik, 1997; Kuno, 1987; Welke, 2005), aspect (Smith, 1997) and thematicity (Halliday, 2004). Doing so might be necessary in situations in which a close interpretation of text is required. Identifying a syntactic structure is different from parsing it in first-pass reading because the latter does not involve repeated access to the structure. We assume that going beyond first-pass reading will occur as part of normal reading practice if it is triggered by information generated in language processing itself, not just by the application of analytical operations extraneous to it. In summary, to use syntactic information in reading, readers have to learn to notice it as it occurs during language processing, and to resume access to it if necessary.

The instructional experiment described below is intended to probe whether it is possible to enhance literacy-related achievements by working on syntactic structures in first-language education according to these considerations.

2. THEORETICAL CONSIDERATIONS AND HYPOTHESES

The identification of a syntactic structure by a reader should not be equated with her or him being prepared to give a description of that structure. Identifying a syntactic structure means that when one stumbles over it once, one will be able to activate it another time without confusing it with other structures, thereby distinguishing it from them. This ability shows up in the disposition to judge whether two recurring construction tokens are ‘same’ or ‘different’ with regard to their structure (Fries, 1952, p. 74 and p. 294). It is a prerequisite for giving descriptions of constructions because the descriptions result from the application of linguistic operations on them, such as substitutions and elisions, which serve to discern their syntactic properties. When drawing on these operations, however, one has to make sure that their application on the construction does not change its structure. That is, one has to identify the structure of the construction before one can describe it. Note that this does not include a circularity because identifying a syntactic structure is conceived of as being independent from being able to describe it.

The resumption of a syntactic pattern which has just been activated does not place high demands on readers. It results from syntactic persistence, that is readers’ propensity to stick to syntactic templates that have got activated. Syntactic persistence has been observed in productive as well as in receptive speech behavior (Bock, 1986). If, however, a specific syntactic pattern has to be resumed, this

process is prone to get off track because the reader is likely to be influenced by concurring patterns which happen to be on scene. We assume that in order to reliably access a specific syntactic structure, the reader must dispose of a general scheme resonant with it which serves to identify it by strengthening and stabilizing it when it occurs. Such a scheme is likely to be acquired over time by linguistic experience (Funke, 2005; for the concept of scheme applying in this case, see Langacker, 1987).

Funke (2005) conducted a study with fifth- to seventh-grade students where subjects had to choose one sentence with a specific syntactic structure out of four sentences where the other three sentences shared another structure. He found that by far most subjects performed above chance on this task, but there was considerable interindividual variation among them as to their ability to solve it reliably. This conforms with older research suggesting that students are generally sensitive to syntactic information but may fail to reaccess it consistently (Claus-Schulze, 1966; Kilcher-Hagedorn, Othenin-Girard, & de Weck, 1987; Wittwer, 1958). The difficulty of reliably reaccessing syntactic information found in some students might result from working memory resources being limited (Crain & Shankweiler, 1988; Just & Carpenter, 1992). Alternatively (or, complementarily), it might result from a lack of learning experiences.

The identification of syntactic structures may be assumed to be related to the emergence of advanced literacy-related skills. As Maas (1992) put it, in orthographic writing, syntactic structures are 'fixated'. Their identification based on resonant schemes corresponding to their occurrence will enable individuals to use orthographic markers of syntactic features appropriately. We hypothesize that it will bear relevance to some advanced reading comprehension achievements as well. This applies, for example, to pondering text in detail by close reading, for doing so presupposes that the syntactic structure of sentences read remains present in readers' minds during the interpretive process.

Based on these considerations, an instructional program for sixth-grade German-speaking students was conceived which aimed at fostering students' ability to identify syntactic structures. The program did not include grammar instruction, that is, having students describe syntactic structures in an analytic fashion. Instead, students explored syntactic structures directly by being given multiple opportunities to recognize and to distinguish them and to become familiar with them through repeated exposure. The instructional objective was to make students aware of syntactic information present in them and to strengthen their access to it. The hypothesis was that the instruction would enhance orthographic writing as far as the use of an orthographic marker of syntactic structure, the capitalization of nouns which is prescribed in German orthography, was concerned. It was also expected that it would foster a close 'syntactic' reading where the capitalization marker had to be evaluated to interpret text. It was not hypothesized that students' reading comprehension as measured by general comprehension tasks would be affected (Funke

& Wieland, 2007). Also, no specific predictions were made as to the effect of the instruction on spelling single words without regard to syntactic context.

3. METHOD

3.1 Research design

The hypotheses were tested in a quasi-experimental study. From twelve sixth-grade classes which were arranged in pairs, one class out of each pair was randomly assigned to the experimental and one to the control condition. Experimental classes participated in six lessons giving students opportunities to explore syntactic structures. Instruction drew on students' direct acquaintance with syntactic structures which manifests in their ability to spontaneously recognize two recurring constructions as 'same' or 'different'. Nothing was undertaken to make them gain insight into structural features of sentences when direct access to these features did not seem to be present. However, some efforts were made to arrange materials to make a spontaneous recognition of syntactic features probable. No grammatical terminology and no analytical grammatical operations were introduced, though students could use both when they volunteered to do so. In the control classes, instruction in first-language courses proceeded as usual according to teachers' plans.

Pre- and posttests included measures of the use of the orthographic capitalization marker in writing and in reading. In addition, general measures of spelling and reading comprehension were enlisted. The pretest was administered immediately before the experimental instruction started. The posttest was given five weeks after instruction had been completed.

Details on the experimental instruction and on the pre- and posttest measures are given below.

3.1.1 Experimental instruction

The instruction in the experimental classes focused on two syntactic features: syntactic category and noun phrase case. Each feature was covered in three lessons and was brought to students' attention by contrasting constructions (noun vs. finite verb constructions and nominative vs. accusative case constructions).

The syntactic category feature was introduced to students by presenting them sets of four sentences which shared a homograph element serving as a finite verb in one sentence and as a noun in the other sentences (Figure 1). This type of material was called a 'black sheep task' because students were asked to sort out from each set the sentence which 'does not fit'.

In any single black sheep task, one may succeed in finding the odd sentence by focusing on semantic or pragmatic features instead of the syntactic category contrast. However, semantic or pragmatic features which set apart the odd sentence

vary from task to task. Drawing on them will not enable one to solve black sheep tasks above chance level on the long run. Funke (2005), in an empirical study with black sheep tasks done in grades 5-7, found that most students' scores surpassed chance level. Moreover, task difficulty did not depend on semantic features of the homograph element such as noun concreteness (when taken as a noun) or verb agentivity (when taken as a verb). Thus, grade 6 students may be assumed to draw on syntactic information when working on black sheep tasks.

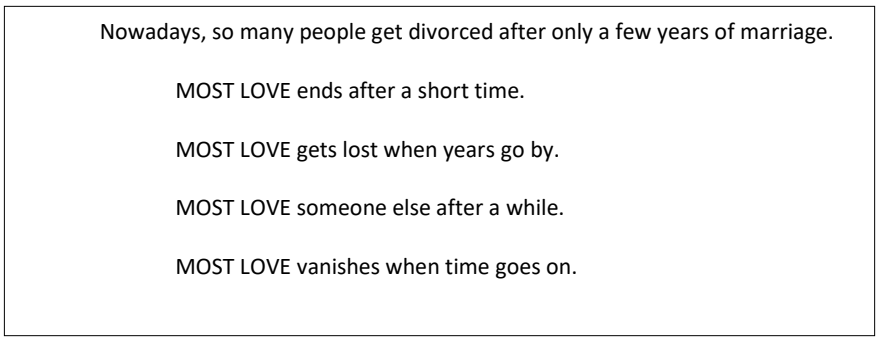

Figure 1 Example of a black sheep task for the syntactic category contrast, freely translated into English. Students are required to pick out the sentence which 'does not fit'.

After having completed some tasks of this type, students discussed sentences to share their observations. Later, they continued to work on the black sheep tasks in various ways.

Constructing black sheep sentence sets. This was done starting with incomplete sets where the part after the homograph element was missing in one sentence or two (black sheep or not) and students had to propose how to formulate it. The task is illustrated in Figure 2. In order to accomplish the task, students have to determine whether there is a black sheep among the sentences given. This means that they have to identify black sheep sentences not just as being the ones which 'do not fit' but as those which are 'like the odd ones in other tasks'.

Working with the problem box. The problem box consisted of file cards, each pertaining to a specific black sheep task and raising a question on it. The questions were not straightforwardly related to the syntactic contrast involved but rather to diverse aspects linked to it. For example, students were asked to say the sentences with the appropriate prosodic contour to others in an attempt to make the black sheep sentence distinguishable from the others, or to figure out which presuppositions were tied to sentences with differing syntactic structure. To illustrate, in the task displayed in Figure 1, the sentences containing *love* as a noun suggest that among the married partners there once had been love which has since disappeared, whereas the sentence containing *love* as a verb does not bear this presupposition.

In that little Celtic village, some people oppose the rule of Caesar.
ONE OF THESE REBELS even tries to expel the Roman officers.
ONE OF THESE REBELS _____.
ONE OF THESE REBELS _____.
ONE OF THESE REBELS even against the Roman officers.

Figure 2 Scheme for a construction activity. Students are required to complete the scheme to make up a black sheep task.

Speeding up recognition. Students were presented with sentences belonging to one set appearing sequentially on a computer screen. They were asked to press a button as soon as a black sheep sentence appeared. Reaction time latency and number of hits were indicated on the screen, and students were encouraged to speed up. The task was administered in three forms with different difficulty levels.

The syntactic case feature was introduced by black sheep tasks consisting of four sentences, each containing in initial position a noun phrase which seemed to be the same in all sentences on first sight because it was represented by a homograph string, but which turned out to have nominative case once and accusative case at other times (Figure 3). The problem is analogous to the syntactic category problem though in this case no systematic empirical evidence is available as to how grade 6 students might cope with it.

$\text{Milch}_{\text{NOM}}$ macht unseren Opa munter.	(Milk makes our grandpa energetic.)
$\text{Milch}_{\text{ACC}}$ macht mein Bruder gerne warm.	(Milk - my brother likes to heat it up.)
$\text{Milch}_{\text{NOM}}$ macht kleine Kälber groß.	(Milk makes little calves grow big.)
$\text{Milch}_{\text{NOM}}$ macht arme Bauern richtig reich.	(Milk makes poor farmers really rich.)

Figure 3 Example of a black sheep task for the case contrast. Students are required to pick out the sentence which 'does not fit'. Category indices and English translations have been added.

Work on these tasks began with discussing students' observations on the sentences and constructing new black sheep tasks of the case-contrasting type. It then proceeded with two other activities.

Coloring noun phrases. This activity went on from the construction task. Students were given complete black sheep tasks and asked to color noun phrases in

order to indicate the type of sentence: black sheep (highlight noun phrase with one color) or not (highlight noun phrase with another color). To make the assignment clear to them, we told them that they were to check each sentence in an attempt to make up a construction task. Which sentence type would be missing if one eliminated the sentence given? This activity served to smoothly shift attention from the sentence level (being different in structure) to the noun phrase level (being different in case).

Embedding sentences in stories. For this activity, sentences were used which were, when presented in written form, ambiguous as to the case of noun phrases (Figure 4). The ambiguity was discussed in class. Students then were asked to write short stories establishing contexts in which the sentences might be used in one version or the other.

Die Polizistin$_{NOM}$ übersah die Zeugin$_{ACC}$ komplett.
(It was the policewoman who completely missed the witness.)

Die Polizistin$_{ACC}$ übersah die Zeugin$_{NOM}$ komplett.
(It was the policewoman whom the witness missed completely.)

Figure 4 Example of a sentence with noun phrases which are ambiguous as to case when presented in written form and without context. Students are required to figure out both meanings. Category indices and English translations have been added.

From the example, it becomes evident that the versions differ in their information structure. When writing stories, students had, for each version, to figure out a situation where the contextual and pragmatic conditions for its use were met. In doing so they were supposed to concomitantly attend to the case contrast co-occurring with it. This conforms to a key objective of the experimental instruction, which is to have students alert themselves to syntactic information *in situ*, as it occurs to them during the writing process.

The instructional program for the experimental classes was tested in advance in three sixth grade classes from schools not participating in the main experiment. It was revised according to the experience gained. The test lessons were held by a member of the research team and videotaped by another member of the team. Videos were then discussed in the research team to detect deviations from the principles of the instructional approach and to make sure that the instructor accorded to these principles. The lessons delivered subsequently in the experimental phase were held by the same instructor. Additionally, a guide describing the role of the teacher as a discussion leader and a checklist serving to remind the teacher of the principles of the instructional approach were developed to ensure treatment fidelity.

The six experimental lessons were distributed over two weeks in each participating class. They also were videotaped by a member of the research team. For a

more detailed description of the instructional program and its implementation, see Melzer (2011).

3.1.2 Pre- and posttest measures

Pre- and posttests included three tasks from which four measures were derived.

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. In the following example sentence, students are given blanks with equal length in place of the underlined words and are requested to fill them in by dictation.

> *This is probably no good answer.*

Two measures were derived from the orthographic writing task. The first was capitalization. This refers to the special feature of German orthography that syntactic nouns have to be capitalized. It urges writers to continually consider the syntactic structure of the text written. Words dictated were selected from a published list of words often miscapitalized (Menzel 1985). No concrete nouns were included which might be capitalized correctly without regard to syntactic function. When determining the capitalization score, a word was taken to be written correctly if it was a noun capitalized or a non-noun not capitalized. The second measure was spelling. The spelling score was determined by counting all words which were written correctly without taking capitalization into account. This yields a measure of word-level orthographic writing skill. When evaluating the orthographic writing task, only the second and the third of the 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. As it seems advisable to restrict our analysis to the last two words when checking for capitalization, we did so for spelling as well. Thus, scores for capitalization and spelling run between 0 and 20.

Syntactic reading task. The syntactic reading task included eight written sentence fragments, each featuring a critical unit that could be regarded as either a noun or a verb. Whether or not the critical unit is capitalized resolves the ambiguity.

The following example illustrates this. It is freely translated to English but the critical unit is written according to German orthography.

> *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 example begins with a sentence fragment which contains the critical unit *Trust*. It must be taken as a noun because it is capitalized. Accordingly, the sentence fragment may only be accomplished by the noun alternative which is presented in second position in the case given. If the critical unit were written with a lower-case initial letter, one would have to take it as a verb. In that case, the proper continuation of the sentence would be the verb alternative, which is, in the case given, the one being presented first. Subjects were asked to mark the alternative 'which fits' without their attention being directed to capitalization. From their response it can be seen whether capitalization was interpreted correctly. An item is scored to be correct if the noun alternative is marked for a critical unit capitalized and the verb alternative for a critical unit not capitalized. Each subject is assigned the sum of correct solutions as his or her score. This score may take values between 0 and 8.

Reading comprehension task. In the reading comprehension task, a text on archeologists was given with ten information and inferential questions to be answered. The items were mostly taken from a collection of reading tests for grades five and six published by Kühn & Reding (2004). They were used with the permission of the Auer Verlag, Donauwörth, Germany. Questions had to be answered by marking one out of four (or, in some cases three) options. The number of correct solutions for the reading comprehension task varied between 0 and 10.

For all tasks, one version for the pretest and one version for the posttest were devised. The orthographic writing pretest version was adopted from a previous study with fifth- to seventh-grade students (Funke & Sieger, 2012). Its posttest version was constructed by preserving sentence frames but changing the lexical material. To make sure that the new words inserted had the same difficulty as those they were substituted for, misspelling counts from the Menzel (1985) list and frequency counts taken from Ruoff (1990) were used. The syntactic reading items were selected from a pool of items used in the study cited (Funke & Sieger, 2012). Pre- and posttest difficulty were balanced by drawing on the data collected in that study. For the reading comprehension task, no information on item difficulties was available in the source (Kühn & Reding, 2004). Thus, the items were tested in advance in a pilot study with 454 sixth grade students. They were then split up to obtain pre- and posttest versions of the reading comprehension task which were balanced according to difficulty. Three self-conceived items were added to each version.

Reliability was estimated by computing Cronbach's α based on the data described. It was .78 for capitalization and .65 for spelling (pretest versions). For syntactic reading, the median reliability of the pre- and posttest versions was estimated to be .41 based on the data from the previous study. For reading comprehension, the median of the reliability estimates taken from the pilot study was .64. Because the time allotted for the tests was limited, only small numbers of items could be included. The small size of the reliability indices results from this. In the case of the syntactic reading measure, it is also likely to be caused by chance varia-

tion in low achieving subjects. The low reliability of this measure has to be taken into account when interpreting the results of the study.

3.1.3 Participants

Classes participating were from all educational levels common in German schools (Gymnasium, Realschule, Hauptschule). For each of the three levels, two schools were found which ran at least two classes at that level. One of the classes was assigned to the experimental and one to the control condition. Schools were located in small rural towns in southern Germany.

Altogether, 277 subjects (133 experimental, 144 control) participated in the study. Only data from students for whom complete pre- and posttest data could be gathered were retained for analysis. Consequently, the final sample comprised 256 subjects (119 experimental, 137 control). Though more experimental subjects than control subjects were lost, the rate of subjects dropping out during the experimental period did not differ significantly in the experimental and the control group. 5.1% of the students in the final sample indicated that they had begun to learn German after entering kindergarten (5.9% in the experimental and 4.4% in the control classes). Among these were two students, both belonging to experimental classes, who started to learn German only after being enrolled in first grade. Compared to common German classes, the ratio of subjects with a first language other than German in our sample is rather low (cf. Chlosta & Ostermann, 2008), probably due to the rural location of the schools participating. We refrained from eliminating these subjects from the database.

Table 1 Classes and subjects participating

Educational level	Nr of schools	Nr. of classes		Nr. of students	
		Experimental	Control	Experimental	Control
Gymnasium	2	2	2	41	50
Realschule	2[a]	2	2	50	51
Hauptschule		2	2	28	36
Whole sample	4	6	6	119	137

[a] *Classes of Realschule and Hauptschule type were run by the same schools.*

Pre- and posttests were administered by members of the project team. The syntactic reading task was given first, followed by the orthographic writing task and the reading comprehension task. Students' participation was based on parental and personal consent. Table 1 shows the number of classes and subjects in each condition for whom complete pre- and posttest data are available.

4. RESULTS

Descriptive statistics for the four outcome measures are provided in Table 2. Pretest achievement did not differ in the experimental and the control group according to Mann-Whitney U-Tests applied on the four measures ($p > 0.278$ in all cases). The raw score distributions from which the statistics in Table 2 were computed are roughly symmetric in the case of syntactic reading and reading comprehension, but skewed to the left with modal values at the maximum score in the case of capitalization and spelling. The table shows that, nevertheless, average scores increased from pre- to posttest for capitalization and spelling and decreased for syntactic reading and reading comprehension.

Table 2 Means and standard deviations (in brackets) for the outcome variables

	Whole sample		Experimental group		Control group	
	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test
Variable	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Capitalization	17.5 (2.54)	18.1 (2.23)	17.4 (2.64)	18.3 (2.21)	17.6 (2.45)	18.0 (2.26)
Spelling	18.8 (1.29)	18.9 (1.45)	18.9 (1.17)	19.1 (1.17)	18.8 (1.39)	18.8 (1.64)
Syntactic reading	5.1 (1.57)	4.8 (1.70)	5.2 (1.59)	4.7 (1.77)	5.0 (1.56)	4.9 (1.63)
Reading comprehension	7.5 (1.83)	7.2 (1.90)	7.5 (1.85)	7.1 (1.83)	7.5 (1.81)	7.2 (1.87)

When subjecting the data to statistical tests, it seems advisable to note that most of them are more appropriately taken to be categorical than quantitative. This is especially true for the syntactic reading measure because it is made up of just eight tasks which may be solved by chance in one out of two cases each. Little room is left for reliable quantitative differentiation above chance level among subjects. By the same token, though less compelling, it would appear that the reading comprehension and the capitalization measure yield categorical rather than quantitative data. For this reason, raw scores of all measures were transformed to categorical data by assigning scores to intervals.

For capitalization, syntactic reading and reading comprehension, intervals can be defined consistently across measures because the probability of gaining a specific score for individuals responding on chance level can be computed by assuming that scores are binomially distributed in these individuals. A first cut-off point was

defined to be the score expected for subjects responding by chance (chance level score). A second cut-off point was the score which may not be reached by students simply guessing unless they obtain a result with a probability lower than the significance level accepted, which was .05 (criterion score). A third cut-off point was set to characterize a perfect or nearly perfect achievement, that is, obtaining the maximum score or one point below it (mastery score). In an attempt to transfer the scheme to spelling, the first cut-off point was set at the score which could be obtained by just writing based on regular phoneme-to-grapheme conversions. The second cut-off point was defined to be the score which could be gained by additionally observing morphemic relations between words. The third cut-off point was the same as in the other measures. Based on these definitions, four categories of subjects were established: *non-responding* ($0 \leq x \leq$ chance level score), *transitional* (chance level score $< x <$ criterion score), *responding* (criterion score $\leq x <$ mastery score) and *mastery* (mastery score $\leq x \leq$ maximum score). In the case of syntactic reading, the last two intervals coincide. For the other measures, there were no or hardly any subjects belonging to the first interval, so this interval was combined with the second one. With these modifications, the scheme was applied to the data, yielding three categories for each measure (see Table 3). The agreement index defined by Subkoviak (see Meyer 2010) was, when computed based on the data of the control group, at least twice as high for the categorized data than for the raw scores in all four measures.

Table 3 Range of scores defining categories in the ordinal regression model

	Capitalization		Spelling		Syntacti reading		Reading comprehen-sion	
	Range	N	Range	N	Range	N	Range	N
Non-responding/ transitional	0-14	47	0-17	34	0-4	93	0-5	37
Transitio-nal/ responding	15-18	87	18	50	5-6	110	6-8	140
Mastery	19-20	122	19-20	172	7-8	53	9-10	79

Note. Range – Range of scores, N – Number of subjects.

An appropriate statistical model to test hypotheses based on categorized raw scores is the cumulative link or ordinal regression model (Agresti, 2002; 2010). In ordinal regression, the effect of one or more factors measured categorically on an

outcome variable measured with ordered categories is tested. The model computes beta weights indicating the influence of each factor on the outcome variable. Model fit may be checked by computing the likelihood ratio statistics *LR,* which compares the likelihood which is ascribed to the data when one includes the factors to the likelihood ascribed to the data by a baseline model which does not include any factors. The ordinal regression model makes use of a link function to ensure that the probability the model computes for each data point does not run outside the range between 0 and 1.

In the case given, for each of the four measures an ordinal regression model was computed with the pretest values (*category 1, category 2, category 3*, as defined above) and group (*experimental, control*) as factors, and posttest values (*category 1, category 2, category 3*, taken as being ordered) as outcome variable. Unfortunately, no link function was suited to model all four outcome measures. Goodness-of-fit statistics as well as the test of parallel lines indicated that for spelling and reading comprehension, the logit link led to the best fit, whereas for capitalization and syntactic reading, it was the complementary log log link. A likelihood ratio test which, according to Agresti (2010, p. 129f.), yields approximate evidence, suggested that the fit may be assumed to be significantly better for the logit link in the case of reading comprehension ($LR(1) = 8.421$; $p = .004$) and for the complementary log log link in the case of syntactic reading ($LR(1) = 5.308$; $p = .021$). Since, in our study, capitalization and syntactic reading were the variables of main concern, we used the complementary log log link. As far as statistical significance is concerned, the results which were obtained by using this link were no different from those which had been obtained by using the logit link.

With regard to the model specified, the hypotheses were formulated referring to the beta weights assigned to the group factor. Since the test statistic for the beta weights is, in the case of the complementary log log link, not suited for one-tailed testing, the null hypothesis was set to be $\beta = 0$ for all measures. However, this hypothesis was expected to be rejected for capitalization and syntactic reading and to be accepted for spelling and reading comprehension.

The ordinal regression analyses revealed that, for all four measures, the fit was better for the model adopted than for the baseline model ($p < .001$ in each case according to the likelihood ratio statistics *LR*). However, only in the case of capitalization, excluding group from the model resulted in a significant loss of model fit ($LR(1) = 7.157$; $p = .008$). The beta weights and the test statistics for all outcome measures are provided in Table 4. Additionally, the *Pseudo-*R^2 is included which may be used to estimate to what extent the variability found in the data is explained by the model.

Given model fit, tests on the beta weights may be performed by using the Wald statistic as displayed in Table 4. For capitalization, the null hypothesis that $\beta = 0$ for the beta weight assigned to the group factor is rejected ($p = .009$). However, for syntactic reading ($p = .822$), spelling ($p = .080$) and reading comprehension ($p = .231$), the hypotheses are maintained.

Table 4 Parameters and statistics estimated by the ordinal regression model

Variable	Group parameter		Test statistic			Pseudo-R^2
	β	*SE*	Wald χ^2	*df*	*p*	
Capitalization	0.596	0.227	6.862	1	0.009	0.536
Spelling	0.461	0.263	3.065	1	0.080	0.250
Syntactic reading	-0.033	0.145	0.051	1	0.822	0.091
Reading comprehension	-0.193	0.161	1.436	1	0.231	0.247

Note. β - beta weight for the group factor (experimental vs. control group), SE - standard error of the beta weight, Wald χ^2 - Wald test statistic for the beta weight, df - degree of freedom, p - probability assigned to the Wald statistic under null hypothesis, Pseudo-R^2 - measure of the strength of association (Nagelkerke's R^2).

In the case of capitalization, spelling and reading comprehension, the strength of the group effect was not dependent on the pretest results. This is indicated by the fact that including the pretest × group interaction term in the model did not significantly change model fit ($LR(2) = .186$; $p = .911$ for capitalization; $LR(2) = 3.246$; $p = .197$ for spelling; $LR(2) = 1.402$; $p = .496$ for reading comprehension). In the case of syntactic reading, however, a significant interaction of pretest results and group was found to exist ($LR(2) = 10.288$; $p = .006$). The interaction resulted from the fact that experimental students belonging to the non-responding and the responding pretest levels were favored by the instruction, whereas students from the transitional pretest level were disadvantaged by it (see Figure 5 as compared to Figure 6). This means that a straight answer to the question of whether the hypothesis is true or not will, in the end, be impossible for syntactic reading because a presupposition upon which the hypothesis was based must be rejected.

Since hypothesis testing on syntactic reading did not lead to results which may be interpreted easily, it may be legitimate to add some observations which serve but exploratory objectives. First, in the case of syntactic reading, the ordinal regression model obviously fails to include relevant factors. This is evidenced by the low *Pseudo-*R^2 value, which is 0.091 and just rises to 0.132 when the interaction term is included. Second, the Wald statistic presented indicates that belonging to the experimental or the control group generally did not influence syntactic reading gains. This is due to the fact that the Wald statistic is computed using the standard error of the beta weight. Statistics which do not take into account parameter estimates and their error may even lead one to conclude that the control group outstripped the experimental group in syntactic reading (Melzer, 2011). This is because more students of the experimental group belonged to the transitional level than to each

of the other levels. Third, differences concerning the syntactic reading results of the experimental and the control group are found not only when pretest achievement is taken into account, but also when one considers the amount of change from pre- to posttest. In the control group, small changes prevail which may be assumed to result from chance fluctuation. In the experimental group, more cases are observed where students made great gains. There are more students in the experimental group than in the control group who moved from the lowest to the highest achievement level (p = .031 according to a Fisher-test). In summary, there is more to consider about the syntactic reading data than just 'whether they are significant or not'.

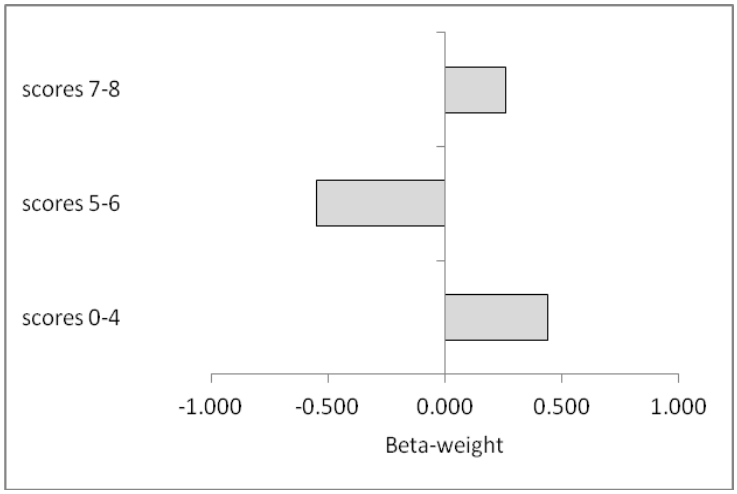

Figure 5 Effect of group (experimental vs. control) on syntactic reading gains at different achievement levels.

Finally, it should be noted that although the experimental subjects belonging to the transitional group fell off in their posttest syntactic reading scores, they did not lag behind in their posttest capitalization scores. This was substantiated by including the syntactic reading pretest level (*transitional* vs. *not transitional*) in the ordinal regression model for capitalization. There was no increase in model fit compared to the model without the syntactic reading level factor ($LR(1) = 0.650$; p = .420) and also no interaction of group (*experimental* vs. *control*) with syntactic reading level ($LR(1) = 0.009$; p = .924).

To sum up, in capitalization the experimental students improved significantly more than the control students did. In syntactic reading, different effects emerged depending on the point students started from at the outset of the instruction. There was a group of students who, according to their pretest scores, seemed to oscillate between sensitivity and insensitivity to the capitalization cue. Students

belonging to this group fell off when exposed to the instruction as compared to students not exposed to it. This was a specific effect restricted to syntactic reading. The same students did apparently respond well to the instruction as far as capitalization was concerned. The remaining students of the experimental group seemed to progress better in syntactic reading than control students with similar starting points, though the significance of this effect can only be established *post hoc*.

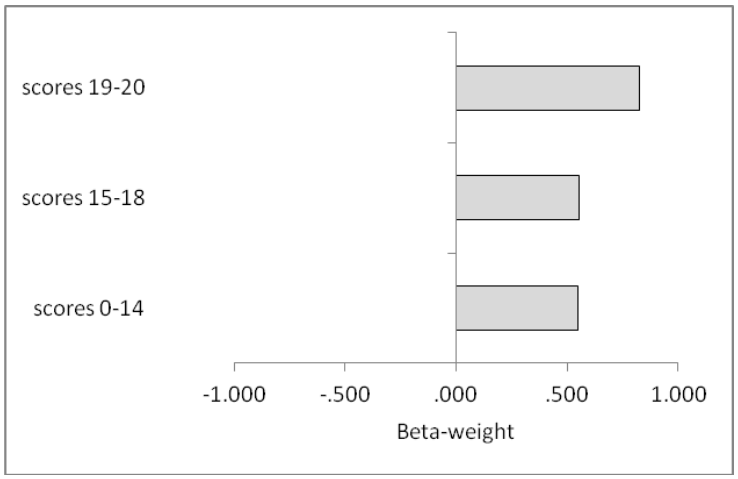

Figure 6 Effect of group (experimental vs. control) on capitalization gains at different achievement levels.

5. DISCUSSION

In the global measures of literacy-related achievement, that is to say spelling and reading comprehension, no instructional effects were found. As for reading comprehension, this does not come as a surprise. The ability measured by general comprehension tests is very globally defined. In particular, it is assumed to be causally related to content-specific knowledge and reading motivation (see, for example, Oakhill & Cain, 2004; Perfetti, Landi, & Oakhill, 2005). Both factors may hardly be fostered by exploring syntactic structures. Spelling is assumed to depend on syntactic skill based on theoretical considerations (Maas, 1992). However, the spelling measure used in the present study refers to word-level orthographic skills. Relations of these skills to syntactic skill, if they exist, might be indirect and hard to capture by empirical tools not designed specifically for this sake.

Though no effects of the experimental instruction on the global measures were expected, exploring syntactic structures would be of little educational interest if it turned out not to influence reading and writing in at least some relevant aspects. The rationale of the present investigation is that the less instructional effects are restricted to the type of the tasks trained, the more it is plausible to assume that

they will influence reading and writing practice. Research on grammar instruction has frequently found specific learning effects, that is effects consisting of subjects only learning to solve tasks which come close to those trained during instruction. We assumed that if working on syntactic structures in first-language education is to enhance literacy-related achievements, its effects have to go beyond specific learning of that type.

If the effects of the experimental instruction in the present study were narrowly specific, it would seem odd that capitalization was improved. During the experimental instruction, students dealt with black sheep tasks where syntactic categories were presented in contrasting constructions. However, when they had to decide whether to capitalize or not in the orthographic writing task, as a rule no category contrast was present. One may wonder why the experimental instruction fostered capitalization in this situation nevertheless.

The instructional effect on capitalization was not very marked, but it seems to have occurred over a wide range of achievement levels, and it was observed even five weeks after completion of the instruction. Notably, it was accomplished although transfer to capitalization was not trained during the instruction. Thus, it seems likely that by exploring syntactic category contrasts students learned to identify syntactic categories, not just to distinguish them when they were presented in the context of category contrasts as used in the instruction. We assume that they were alerted to how they notice occurrences of syntactic categories even when these were not tied to opposing constructions such as those they had worked on.

If so, one should expect that the instruction would enhance syntactic reading as well. One condition for a reader to solve a syntactic reading item correctly seems to be that syntactic category information has become accessible to him or her during language processing itself instead of being searched for in subsequent considerations concerning how to respond to the task (Funke & Sieger, 2012). From the findings on capitalization, one should conclude that this condition was fulfilled more frequently in experimental subjects than in control subjects due to the instruction. Thus, the chance to succeed on the syntactic reading items should rise in the experimental group as compared to the control group. One wonders why in experimental subjects belonging to the transitional group, syntactic reading scores fell off rather than improved instead.

An explanation for this might start from the assumption that, among subjects belonging to the transitional group, there was a greater proportion of students who are prone to letting themselves be influenced by the ambiguity which seems to reside in the syntactic reading items. This may have enabled them to respond above chance level when working on the syntactic reading pretest, leading them to consider both possible meanings of the item, not just one, more frequently than others. Because of their propensity to get involved in ambiguity, however, the instruction may have sensitized them to the complexities of the syntactic reading items even more. This may have prompted them to check each item systematically

when taking the posttest, looking for a clue for how to solve the ambiguity. In syntactic reading items, doing so is not conducive to success because the only valid clue, the capitalization of the critical unit, is hard to find if one searches the whole sentence instead of selectively considering the critical unit. Systematically checking an item's text results in high demands on working memory and analytical processing. Thus, it is possible that students with marked sensitivity to ambiguity were slightly favored in the syntactic reading pretest but adversely affected by the instruction with regard to the syntactic reading posttest. This explanation is in line with observations made in a syntactic reading study suggesting that some subjects were strongly confused by the potential ambiguity of the task (Funke & Sieger, 2009).

Thus, the capitalization and the syntactic reading results seem to lead to conflicting answers as to whether the effects of the instruction were restricted to the specific instructional tasks or not. The conclusion which one might draw from this is that the idea from which the study started, that is to say, the distinction between specific and non-specific instructional effects, is vague and needs to be elaborated upon. The capitalization task was relatively far from the instructional tasks in its content, but it was familiar to the students in its type. The syntactic reading task was relatively close to the instructional content, but its demands were highly unusual in their type. If one considers instructional outcomes, one has to distinguish between effects constrained to specific contents and effects tied to specific tasks.

When drawing conclusions based on the results of this study, one should bear in mind its methodological limitations. First, classes rather than students were randomly assigned to experimental and control conditions. Nevertheless, due to the small number of classes participating, inferential statistical procedures were employed to student level instead of class level data. Second, no special instructional program was run in the control classes. This could have led to a Hawthorne effect favoring the experimental group. Third, the reliability of the measures, especially the syntactic reading measure, was low. The small *Pseudo-*R^2 value found in the ordinal regression model for syntactic reading suggests that this influenced the results.

Although these limitations reduce the force of arguments based on the data, it is unlikely that they invalidate the key observations of the study as a whole. First, the instruction seems to have favored experimental classes as well as individual experimental students as far as capitalization is concerned. In three out of six cases, the experimental class made progress in capitalization during the course of the experiment whereas the corresponding control class did not (where progress in capitalization is defined to have occurred when more than half of the students improved). These classes were distributed across all three educational levels represented in the study (Hauptschule, Realschule, Gymnasium). The opposite case – a control class making progress in capitalization but the corresponding experimental class not doing so – was never observed. Second, if the experimental effects were simply of the Hawthorne type, one would expect them to show up across outcome

measures. However, in spelling and reading comprehension, no such effects were found. Third, the syntactic reading measure, though not very reliable, seems to have been sensitive to the experimental condition as opposed to the control condition. If it is true, as we have assumed, that the low reliability of this measure results from chance variation in the syntactic reading scores, it seems that enough room was left nevertheless for systematic variation in these scores to allow them to serve as indicators of experimental effects.

In sum, the data suggest that it is possible to sensitize students to syntactic information by having them explore syntactic structures directly, without referring them to analytical grammatical operations. More specifically, students seemed to notice syntactic information beyond what was presented in the instruction, though this did not imply that the information became equally accessible to them in reading situations of all types. The results may encourage researchers to explore the effects of having students deal with syntactic structures directly over a longer instructional period than allowed for in this study and by working on these structures in a more distributed fashion during that time.

AUTHORS' NOTE

This study has been funded by a research grant from the Pädagogische Hochschule Heidelberg.

REFERENCES

Agresti, A. (2002). *Categorical data analysis*. 2nd edition. Hoboken, NJ: Wiley. http://dx.doi.org/10.1002/0471249688

Agresti, A. (2010). *Analysis of ordinal categorical data*. 2nd edition. Hoboken, NJ: Wiley. http://dx.doi.org/10.1002/9780470594001

Andrews, R., Torgerson, C., Beverton, S., Freeman, A., Locke, T., Low, G., et al. (2006). The effects of grammar teaching on writing development. *British Educational Research Journal, 32*, 39-55. http://dx.doi.org/10.1080/01411920500401997

Bock, K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*, 355-387. http://dx.doi.org/10.1016/0010-0285(86)90004-6

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.

Chipere, N. (2003). *Understanding complex sentences. Native speaker variation in syntactic competence.* Hampshire, UK: Palgrave Macmillan. http://dx.doi.org/10.1057/9780230005884

Chlosta, C. & Ostermann, T. (2008). Grunddaten zur Mehrsprachigkeit im deutschen Bildungswesen [Basic data on multilingualism in the German educational system]. In B. Ahrenholz & I. Oomen-Welke (Eds.), *Deutsch als Zweitsprache* [German as a second language] (pp. 17-30). Baltmannsweiler, Germany: Schneider.

Claus-Schulze, A. (1966). *Ergebnisse des Grammatikunterrichts in ihrer Beziehung zur Denkleistung der Schüler* [Results of grammar instruction and their relation to students' cognitive achievements]. Habilitationsschrift. Potsdam: Pädagogische Hochschule Potsdam.

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.

Crews, R. (1971). A linguistic versus a traditional grammar program: The effects on written sentence structure and comprehension. *Educational Leadership, 5*, 145-149.

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

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

Elley, W., Barham, I., Lamb, H., & Wyllie, M. (1976). The role of grammar in a secondary school English curriculum. *Research in the Teaching of English, 10*, 5-21.

Fries, C. C. (1952). *The structure of English*. New York, NY: Harcourt, Brace and Comp.

Funke, R. (2005). *Sprachliches im Blickfeld des Wissens* [The expliciteness of linguistic knowledge]. Tübingen, Germany: Niemeyer. http://dx.doi.org/10.1515/9783110924701

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. & Sieger, J. (2012). Continued access to syntactic information in reading. *Reading and Writing: An Interdisciplinary Journal, 25*, 1769-1794. http://dx.doi.org/10.1007/s11145-011-9341-8

Funke, R. & Wieland, R. (2007). *Die Wirksamkeit sprachreflexiven Unterrichts im Bereich der Grammatik für die Ausbildung von Lese- und Rechtschreibfähigkeiten. Projektantrag* [The effects of exploring syntactic structures on reading and writing abilities. Grant proposal]. Heidelberg: Pädagogische Hochschule Heidelberg.

Fusaro, J. A. (1992). Meta-analysis of the effect of sentence-combining on reading comprehension when the criterion measure is the test of reading comprehension. *Perceptual and Motor Skills, 74*, 331-333. http://dx.doi.org/10.2466/pms.1992.74.1.331

Fusaro, J. A. (1993). A meta-analysis of the effect of sentence-combining on reading. *Reading Improvement, 30*, 228-231.

Glinz, H. (1963). Der Sprachunterricht im engeren Sinne oder Sprachlehre und Sprachkunde [Language instruction or system-oriented and phenomenon-related approaches to teaching first language]. In A. Beinlich (Ed.), *Handbuch des Deutschunterrichts im ersten bis zehnten Schuljahr. Erster Band* [Handbook of German studies in grades 1-10. First volume] (pp. 225-323). 3rd edition. Emsdetten, Germany: Lechte.

Glinz, H. (1993). Eine neue deutsche Grammatik in der Schule – Überblick [New German grammar in the school—an overview]. In H. Glinz, *Sprachwissenschaft und Schule. Gesammelte Schriften zu Sprachtheorie, Grammatik, Textanalyse und Sprachdidaktik 1950-1990* [Linguistics and the school. Collected papers on language theory, discourse analysis, and language education 1950-1990] (pp. 277-297). Zürich, Switzerland: Sabe.

Gonsoulin, K. W. (1993). The effects of sentence-combining curricula on descriptive writing performance, reading achievement, and writing apprehension: A study with high-risk secondary students. Dissertation. Gainesville, FL: University of Florida.

Graham, S. & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*, 445-476. http://dx.doi.org/10.1037/0022-0663.99.3.445

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. H. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology, 104*, 879-896. http://dx.doi.org/10.1037/a0029939

Greenewald, M. J. & Pederson, C. (1983). Effects of sentence organization instruction on the reading comprehension of poor readers. In J. A. Niles & L. A. Harris (Eds.), *Searches for meaning in reading/language processing instruction. Thirty-second Yearbook of the National Reading Conference* (pp. 101-103). Rochester, NY: NRC.

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

Heringer, J. (2001). *Lesen lehren lernen* [Reading—teaching—learning]. 2nd edition. Tübingen, Germany: Niemeyer.

Hillocks, G. (1986). Research on written composition. Urbana, IL.: ERIC.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122-149. http://dx.doi.org/10.1037/0033-295X.99.1.122

Kilcher-Hagedorn, H., Othenin-Girard, C., & de Weck, G. (1987). *Le savoir grammatical des élèves* [The grammatical knowledge of students]. Bern, Switzlerland: Peter Lang.

Kühn, P. & Reding, P (2004). *Lesekompetenz-Tests für die Klassenstufen 5 und 6* [Reading comprehension tests for grades 5 to 6]. Donauwörth, Germany: Auer.

Kuno, S. (1987). *Functional syntax. Anaphora, discourse, and empathy*. Chicago, IL: University of Chicago Press.

Langacker, R. (1987). *Foundations of cognitive grammar*. Vol. 1: Theoretical prerequisites. Stanford, CA: Stanford University Press.

Layton, A., Robinson, J., & Lawson, M. (1998). The relationship between syntactic awareness and reading performance. *Journal of Research in Reading, 21*, 5-23. http://dx.doi.org/10.1111/1467-9817.00039

Maas, U. (1992). *Grundzüge der deutschen Orthographie* [An outline of German orthography]. Tübingen, Germany: Niemeyer. http://dx.doi.org/10.1515/9783111376974

Melzer, F. (2011). Direktes Umgehen mit syntaktischen Strukturen: ein Unterrichtsansatz für die Sekundarstufe I [Direct exploration of syntactic structures: an instructional approach for secondary school classes]. In C. Noack & J. Ossner (Eds.). *Grammatikunterricht und Grammatikterminologie* [Grammar instruction and grammatical terminology]. Osnabrücker Beiträge zur Sprachtheorie 79 (pp. 159-180). Duisburg, Germany: Universitätsverlag Rhein-Ruhr.

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

Meyer, P. (2010). Reliability. Oxford, UK: Oxford University Press.

Morenberg, M, Daiker, D., & Kerek, A. (1978). Sentence combining at the college level: an experimental study. *Research in the Teaching of English, 12*, 245-256.

Neville, D. D. & Searls, E. F. (1985). The effect of sentence-combining and kernel-identification training on the syntactic component of reading comprehension. *Research in the Teaching of English, 19*, 37-61.

Neville, D. D. & Searls, E. F. (1991). A meta-analytic review of the effect of sentence-combining on reading comprehension. *Reading Research and Instruction, 31*, 63-76. http://dx.doi.org/10.1080/19388079109558072

Noyce, R. M. & Christie, J. F. (1983). Effects of an integrated approach to grammar instruction on third graders' reading and writing. *Elementary School Journal, 84*, 63-69. http://dx.doi.org/10.1086/461343

O'Donnell, R. C. & King, F. J. (1974). An exploration of deep structure recovery and reading comprehension skills. *Research in the Teaching of English, 8*, 327-338.

Oakhill, J. V. & Cain, K. (2004). The development of comprehension skills. In T. Nunes & P. Bryant (Eds.), *Handbook of children's literacy* (pp. 155-180). Dordrecht, The Netherlands: Kluwer. http://dx.doi.org/10.1007/978-94-017-1731-1_9

Perfetti, C. A., Landi, N., & Oakhill, J. V. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading*. *A handbook* (227-247). Malden, MA: Blackwell. http://dx.doi.org/10.1002/9780470757642.ch13

Rogers, L. A. & Graham, S. (2008). A meta-analysis of single-subject design writing intervention research. *Journal of Educational Psychology, 100*, 879-906. http://dx.doi.org/10.1037/0022-0663.100.4.879

Ruoff, A. (1990). *Häufigkeitswörterbuch gesprochener Sprache* [Frequency dictionary of spoken language]. 2nd ed. Tübingen, Germany: Niemeyer. http://dx.doi.org/10.1515/9783110916058

Schleppegrell, M. J. (2004). *The language of schooling. A functional linguistics perspective*. Mahwah, NJ: Erlbaum. http://dx.doi.org/10.1017/S0047404504263059

Searls, E. F. & Neville, D. D. (1988). An exploratory review of sentence-combining research related to reading. *Journal of Research and Development in Education, 21(3),* 1-13.

Smith, C. (1997). *The parameter of aspect*. 2nd edition. Dordrecht, The Netherlands: Kluwer. http://dx.doi.org/10.1007/978-94-011-5606-6

Straw, S. B. & Schreiner, R. (1982). The effect of sentence manipulation on subsequent measures of reading and listening comprehension. *Reading Research Quarterly, 17*, 339-352. http://dx.doi.org/10.2307/747523

Trivelli, E. A. (1983). A study of the effects of sentence combining on eighth-grade students' written syntactic ability and reading comprehension. Dissertation. Akron, OH: University of Akron (DA 83, 14388).

Weaver, P. A. (1979). Improving reading comprehension. Effects of sentence organization instruction. *Reading Research Quarterly, 15*, 129-146. http://dx.doi.org/10.2307/747435

Welke, K. (2005). *Deutsche Syntax funktional. Perspektiviertheit syntaktischer Strukturen* [Functional syntax of German. Perspective in syntactic structures]. 2nd edition. Tübingen, Germany: Stauffenburg.

White, C. V., Pascarella, E. T., & Pflaum, S. W. (1981). Effects of training in sentence construction on the comprehension of learning disabled children. *Journal of Educational Psychology, 73*, 697-704. http://dx.doi.org/10.1037/0022-0663.73.5.697

Wilkinson, P. A. & Patty, D. (1993). The effects of sentence combining on the reading comprehension of fourth grade students. *Research in the Teaching of English, 27*, 104-125.

Wills, J. B., Christiansen, M. H, Race, A. S., Achesen, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*, 250-271. http://dx.doi.org/10.1016/j.cogpsych.2008.08.002

Wittwer, J. (1958). *Les fonctions grammaticales chez l'enfant. Sujet – objet – attribut* [Syntactic functions in children's cognition. Subjects—objects—predicatives]. Neuchatel, Switzerland: Delachaux & Niestlé.

Yuill, N. (1996). A funny thing happened on the way to the classroom: jokes, riddles, and metalinguistic awareness in understanding and improving poor comprehension in children. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties* (pp. 193-220). Mahwah, NJ: Erlbaum.

Yuill, N. (2007). Visiting Joke City: How can talking about jokes foster metalinguistic awareness in poor comprehenders? In D. S. McNamara (Ed.), *Reading comprehension strategies* (325-346). New York, NY: Erlbaum.

Zipke, M. (2008). Teaching metalinguistic awareness and reading comprehension with riddles. *The Reading Teacher*, *62*, 128-137. http://dx.doi.org/10.1598/RT.62.2.4

Zipke, M., Ehri, L., & Cairns, H. S. (2009). Using semantic ambiguity instruction to improve third graders' metalinguistic awareness and reading comprehension. *Reading Research Quarterly*, *44*, 300-321. http://dx.doi.org/10.1598/RRQ.44.3.4