

Using Cooperation Scripts and Animations to Teach Grammar in the Foreign Language Classroom

In order to test the efficacy of using cooperation scripts in combination with animations to teach the passive in an advanced language classroom, we designed an intervention study with two groups. The first group received instruction with animations on the grammatical structure of the passive and the second group received the same instruction and animations as the first group, but they were given a cooperation script to use in their small group work. The study uses a quasi-experimental design with a post-test and delayed post-test. The learners were university students in advanced German who had received classroom instruction and spent time in a German-speaking country. Therefore, they had explicit instruction on the morphology and function of the passive as is customary in first- and second-year textbooks for English speaking learners. This paper reports on the quantitative results of the study with respect to our research questions.

1. Introduction

For many language learners and instructors, grammar is an integral part of the experience of learning in a classroom environment. Learners and instructors also note that grammar is one of the biggest challenges in language learning. Grammar is often viewed as an abstract and arbitrary system, with few hard and fast rules to rely on, but many exceptions that just have to be memorized. Materials for learners tend to present grammar with color coded markings, drawing and other visuals with the intent of students achieving the learning goal. Furthermore, the materials and activities often reflect the instructors' and learners' perception of grammar as an incoherent system in that grammatical structures are presented as stand-alone items with no reference to the metaphors that underlie the grammar. Materials for classroom use often lack a cognitive framework for presenting grammar that views grammatical structures as part of a larger system. In addition, grammar instruction often assumes prior knowledge of terms such as passive or accusative, even though the grammatical structures can be realized differently in another language (Croft 2003, Langacker 2008). In contrast to the approach outlined above, grammar instruction should build on prior knowledge and foster new conceptualizations for the learners.

The previous 30 years has seen a rise in Cognitive Linguistics (CL) in general and more specifically, Cognitive Grammar (CG). Within this framework, as well as in Construction Grammar, language is viewed as a realization of construal and is therefore, conceptually motivated (cf. Lakoff & Johnson 2003, Langacker 1987, 2008, Talmy 2000, Evans & Green 2006). The theoretical articulation of grammar as a usage-based and conceptually motivated system gave rise to a number of studies that described phenomena that are the underpinnings of a conceptually based grammar (e.g. force dynamics, Talmy 2000, Tyler 2008, transitivity Hopper & Thompson 1980, Rice 1987). There is also a considerable body of research on language specific grammatical structures from a cognitive perspective, such as pre- and postpositions (Smith 1987, Masuda 2013), case systems (Janda & Clancy 2002) and modality. These studies are important for the understanding of specific realizations of grammatical structures and a necessary first step towards a pedagogical grammar, but they are not always written with classroom use as the primary goal.

The first studies that outlined a cognitive approach towards grammar teaching discuss the advantages of conceptually based learning and the use of schemas to show the underpinnings of grammatical structures (e.g. Achard 2004, 2008, Bellavia 2007, Dirven & de Knop 2008). Building on that foundational work, there have been a number of recent classroom studies that assess the effectiveness conceptually based pedagogical grammars (e.g. Liamkina 2005, Lysinger 2009, Yasuda 2010, Tyler et al. 2011, Jacobsen 2015, Kohl-Dietrich et al. 2016). These empirical classroom studies show the efficacy of using schemata and the conceptual metaphors that underlie grammatical structures to teach grammar. The materials and methods used in language specific studies are often easily transferable to another language that has the same conceptualization of the relevant structure.

Most recently, there is a strand of research that assess the effectiveness of the visual materials used in the classroom (Bielak & Pawlak 2011, Reif 2012, Roche & Suñer 2014, Kanaplianik 2016, Suñer & Roche 2019). The general schema of a grammatical structure may be too abstract for the learner to easily understand. These recent studies argue that students should be presented with pedagogical metaphors that are part of their everyday world and that pedagogical metaphors are a bridge between a lack of understanding and the target language's conceptualization of a specific structure. Using an example from the passive, the grammatical construction under consideration in this paper, Langacker's (1982) billiard ball model can be

presented to the students as a representation of energy transfer and transitivity. Both of these concepts are necessary for understanding passivization. It is less clear in the billiard ball model, however, which element is the focus of the clause. Suñer (2013) uses animations of the billiard ball model to show energy transfer and a spotlight metaphor to highlight the salient participant in the clause. This is more visually rich for the learner than a static drawing with circles representing grammatical arguments and arrows showing the transfer of energy. A dynamic representation in the form of schematic and pedagogical metaphors leads to a conceptual system for the learner that is flexible and multifaceted in which ‘possible worlds’ can be reflected (Seel 2003).

We can draw three conclusions from the brief review of the relevant literature given above (Williams et al. 2013, Roche & Suñer 2016). First, in order to work successfully with conceptually based metaphors, instructors and learners need materials that highlight the conceptual basis of the grammatical structure and the opportunity to work with these materials in class. Animations are suited for this purpose because they present the grammatical metaphor in a visual way which gives the learner a conceptual model to build on (Seel 2003). Second, pedagogical metaphors should be taken from the learners’ culture and grounded in every day experience, e.g. English-speaking learners can easily grasp the billiard ball model as a representation of transitivity. Third, learner development of conceptual metaphors is connected with student-centered activities and active learning. In order to begin to understand a concept, students need to engage with the conceptual and pedagogical metaphors (Vygotsky 1986, Lantoff 2007, Negueruela 2009).

In the grammar teaching method sketched above, students engage in teacher-fronted activities and group activities. In order to derive the most benefit from pair and small group activities, cooperation scripts can be given to the students to guide their discussion (Weinberger et. al. 2010, Compaoré 2018). More concretely, cooperation scripts prevent learners from spending too many mental resources in social interaction while learning collaboratively (intrinsic cognitive load, Sweller et al. 2011) and promote schema construction (germane cognitive load, *ibid.*). In other words, cooperation scripts help learners organize their discussion quickly and guide them through the learning process so they can focus on learning and not how to structure their learning. In this context, King (2007) differentiates between two types of cooperation scripts: Whereas so-called epistemic cooperation scripts provide learners with

content-related prompts, social cooperation scripts help learners sequence interactions in the group. Although educational research has consistently provided evidence for the usefulness of cooperation scripts in classroom settings (see literature review by Vogel et al. 2017), language teaching research has been slow in accepting these findings. In order to evaluate the use of collaboration scripts in grammar teaching, we focused on the following research questions:

- 1) Do cooperation scripts lead to short and/or longer-term gains in recognizing and producing the morphology and the semantic and pragmatic functions of the passive?
- 2) Does cooperation learning of conceptually based grammar instruction lead to short and/or longer-term gains than individual learning?

2. Methodology

This intervention study uses a pre-test, post-test and delayed post-test design. The subjects were an advanced German foreign language course with four class sections from a university in Ireland. Two classes received the treatment with animations, accompanying exercises and the cooperation script (COOP) and two classes received the same treatment with animations, but no cooperation script (INDI). The study design does not include a control group because of the limited number of potential participants and the structure of the course that has all four sections together in the first session each week and then two break out sessions. A control group was impossible because all of the students would receive the first phase of the treatment.

The sample size at the beginning of the experiment was 43 participants. 17 participants were eliminated during the course of the week-long treatment because they did not attend one or more of the scheduled classes. In addition, the delayed post-test reduced the number of participants to 22 because four students were missing on the day of the delayed post-test. This information is summarized in Table 1.

Class	Group	Pre-test	Post-test	Delayed Post-test
A (12)	} COOP (26)	24	12	12
B (14)				
C (12)	} INDI (21)	19	14	10

Table 1: Participant numbers for each test

2.1 Participants

The participants are students at an Irish university in their third and fourth years of study taking an advanced German class. The level of German competence is expected to be between B2 and C1 (CEFR). Given the language competence of the students, we know that they have received instruction about the passive in the past. The university curriculum includes the passive as a reoccurring topic after the second year. The passive is taught with an emphasis on the morphology of the statal and processual passive as well as the transformation from active into passive.

The students are taking a foreign language class in conjunction with another subject such as German literature, law, business, computer science, etc. Students enroll in the language course based on their previous instruction and free time available to take the class. Therefore, the students' placement in one of the four classes is independent of their chosen course of study or their ability in German. Given that we are doing a classroom study on the students enrolled in the course, randomization is neither practical nor expected. Mackey and Gass (2005) point out that using the students we have provides the advantage of validation of the study because the activities are done in the actual classroom, with the materials at hand and with the students that show up. Students are not self-selected by opting in, but rather they can opt out by not participating.

As part of the pre-test, students were given a survey that served as consent to participate in the study and elicited some necessary background information. All students stated that there were most comfortable speaking English, although three students listed Irish as a language they were equally comfortable speaking. Others listed Irish, French, Spanish as foreign languages as well as six other individual languages. With respect to age and length of time learning German, the groups were quite similar. All participants had spent several months in a German-speaking country as students at a university, language assistants in schools or another experience abroad. This information is shown in Table 2.

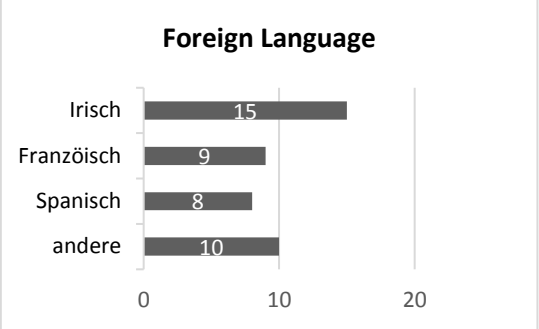
		COOP	INDI	Foreign Language 
Age	Range	20-23	21-24	
	Min.-Max.	21.8	21.7	
Length of German study	Average	9.4	10.2	
	Min.-Max.	6-11	6-17	

Table 2: Participant Data

It is also important to note here that Irish in the Republic of Ireland is an official language and is obligatory in primary and secondary schools. According to the most recent census, 39% of the population claims to know Irish, but only 4% respond that they use Irish on a daily basis.

2.2 Treatment

The treatment was conducted during regularly scheduled class meetings in the semester. The advanced German classes meet three times per week for 50 minutes. The treatment was integrated into the daily lesson plan and presented as part of the regular classroom instruction during the part of the course designated for the passive. In addition, the teaching activities for the treatment reflected the types of activities normally done in the structure of the class. For example, all students are together in the first session of the week and that session was the first day of the treatment and on the other two days of the treatment the larger group was split into four separate classes. The treatment was designed to fit this organizational structure. The class was led by one of the three language instructors in the department and the smaller sections were each taught by one instructor. The individual instructors were shown how to teach the treatment portion of their class by the researcher. The researcher also taught the whole group session on the first day of the treatment. The table below shows the design and time given to the treatment in each class session.

Week/Session	Form	Length	Content
1/1		20 min	Survey, Pre-test
	Lecture (A, B, C, D)	30 min	Treatment Phase 1
1/2	(A), (B); (C), (D)	50 min	Treatment Phase 2
1/3	Discussion	50 min	Post-test
...			...
3/1		50 min	Delayed Post-test

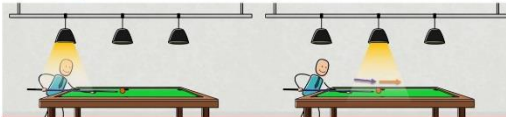
Table 3: Schedule of treatment

The grammar lecture with all the students present was devoted to the survey and pre-test and the first treatment phase. The first treatment phase had three components. The first component aimed at activating the students' prior knowledge of the passive. The students were given an authentic text with a variety of active and passive constructions and asked to mark the passive constructions. Then the instructor and students reviewed the formal aspects of the passive. The second component was intended to highlight the semantics of the passive and show the meaning differences between active clauses, statal passives and processual passives. At this point the animations showing these semantic distinctions were introduced. The animations use the billiard model to show the arguments of the verb and a spot light to indicate which participant has the focus (Suñer 2013). There are additional animations of a golfer, a baseball player and a barking dog causing a boy to get out of bed. The students can view the animations repeatedly and in any order. Figure 1 below shows screen shots of the animation for the processual passive. The third component of the day's treatment focused on the pragmatic aspects of the passive by returning to the text and analyzing why the passive was used in the text. The analysis showed the use of the passive construction for topical coherence and thematic development.

Bedeutung des Passivs

Aktiv: Der **Champion** spielt mit dem Queue die **Kugel**.

Passiv: Die **Kugel** wird vom **Champion** gespielt.



Fokalisierung: Der Fokus beim Vorgangspassiv liegt auf dem Patiens, wohingegen die anderen Partizipanten ausgeblendet werden können.

Topikalität



Topikwechsel

Topic 1: Der Hersteller des Fairphones ist ein soziales Unternehmen aus den Niederlanden, das 2013 aus einem Forschungsprojekt heraus gegründet wurde. Die Beteiligten

Topic 2: (implied from context)

Topikalisierung: Der Topik wird gewechselt, indem der Comment des ersten Satzes (HS) im zweiten (NS) zum Topik wird.

Fehlen des Agens im Passiv: Wer oder was das Unternehmen gründete ist hier irrelevant, unbekannt (oder evtl. absichtlich anonymisiert).

Figures 1: Screen shot of the processual passive animation and explanation.

The second phase of the treatment was conducted in the smaller classes. This phase focused on students working with animations that reinforced the explanation and activities from the day before. It is at this point that two classes receive the treatment in which they work with the passive animations as individuals (INDI) and two classes work with the animations in pairs and with a cooperation script (COOP). The materials used, the time spent on each activity, the sequencing of the activities as well as the animations were the same for each group.

The COOP classes began the second day of treatment with an instructor-fronted discussion of the semantic and pragmatic aspects of the passive. This discussion laid the groundwork for the next activity called a group puzzle (Hänze & Berger 2007). The class was divided into four groups and each group had two tasks and two texts each with a regular version and a manipulated version. In the regular text, the students identified the morphological, semantic and pragmatic aspects of the passive. The manipulated version had morphological, semantic or pragmatic irregularities as well as active sentences where passive should be used. The groups were given a cooperation script to guide them through the discussion of the text and the language of the second phase of treatment was German. After each group completed its analysis of the text, the groups were reconfigured so the students could discuss how to reconstruct the original text from the original version. Again, this discussion was guided by a cooperation script. During this time, the instructor was in an advisory role.

The INDI group received the same texts and the same tasks as the COOP group. However, they worked individually and approached the tasks sequentially. At the end of this

activity, the instructor followed up with a discussion of the texts with the whole class and moderated the discussion of each example.

2.3 Data Collection

The first post-test took place in the third class meeting of the week and the delayed post-test took place ten days after the treatment. There was no instruction or homework on the passive in the interim. The post-test and the delayed post-test were designed to measure performance on the formal, semantic and pragmatic aspects of the passive. In order to measure these aspects, a variety of formats were used in the test. To detect the different aspects of grammatical competence, the performance measurement consisted of different tasks for eliciting language that complemented one another as to format (closed, semi-open, open), range (competency-specific, integrative or global respectively), and activity (analytical, synthetic). Moreover, the variety of tasks aims to ensure the reliability and validity of the measurement procedure (Cohen 2001, Purpura 2004) and minimise factors that could influence the performance (Bachman & Palmer 1996). Table 4 below shows the formats used in the pre- and post-test.

Format	Task	Pre-test	Post-test	Delayed Post-test
A1 closed	Multiple choice	x	x	x
A2 Half open	Cloze text	x	x	x
A3 open	Explanation of semantics	x	x	
A4 open	Text production: Definition	x	x	
A5 open	Text production: Letter	x	x	

Table 4: Tasks and format

Tasks A1 and A2 are designed to elicit the formal aspects of the passive. The multiple-choice exercise (A1) asked the students to choose the correct passive sentence in the set and the cloze text (A2) had the student fill in the blank with the correct auxiliary for the statal or

processual passive. The third task (A3) asked the students to explain the difference in the roles of participants in two sentences that represented the same scene. The differences in the items in the task were perspective, focus and inclusive of participants. The students could express these differences in any way they chose, e.g. English description, drawing, etc. The last two tasks, A4 and A5, were oriented towards text production and the pragmatic function of the passive. The fourth task asked students to give a short description of how a product is made, for example Guinness or electric car, and the fifth task was a letter organizing a party at the end of term or discussing Germany as a multicultural society. Both choices had a list of helpful expressions and vocabulary and students were told in the instructions for the task to use passive constructions where necessary.

2.4 Data Scoring

Tasks A1 and A2 were marked as either correct or incorrect with an answer key provided by three highly proficient German speakers. Tasks A3, A4 and A5 are open ended so we used an evaluation rubric and each task was scored by two independent raters. In the few cases in which the raters disagreed, the scores were averaged. Task A3 was scored on the students' ability to convey the semantic difference between the two sentences and it was either correct or incorrect. For tasks A4 and A5 we used a rubric that assessed formal, semantic and pragmatic components of the text.

- I)** Formal aspects: the verbal complex in the active or passive voice often / sometimes / seldom contains mistakes, and case marking is seldom / sometimes / often applied appropriately.
- II)** Semantic aspects: the participants vary seldom / sometimes / often, and their discrete salience is seldom / sometimes / often used.
- III)** Pragmatic aspects: The various pragmatic possibilities of the passive are seldom / sometimes / often used, and the use is seldom / sometimes / often appropriate.

The total score was a combination of all five scores and each task was weighted equally.

3. Results

The data were analysed using descriptive and inferential statistical methods. A Shapiro-Wilk test ($p > 0.5$) (Shapiro & Wilk 1965, Razali & Wah 2011) and a visual inspection of the histograms, normal Q-Q plots and boxplots showed that the dependent variable, i.e. the results of the post-test for both independent variables, was approximately normally distributed. With regard to the distribution, the skewedness for the COOP students at -1.008 (SE = 0.637) and the curvature at 1.528 (SE = 1.232); and for the INDI students 0.539 (SE = 0.597) and -0.619 (SE = 1.154) respectively (Cramer 1998, Cramer & Howitt 2004, Doane & Seward 2011). The z-score values are thus within the critical range of -1.96 and +1.96. Since the normal distribution of the results is characteristic of the sample, a parametric test procedure is used below for statistical significance.

The results of the pre-test show that the COOP and INDI groups were at the same level in their understanding of the formal aspects of the passive as tested by tasks A1 and A2. The results of the pre-test of INDI and COOP differ only slightly by 0.4% in their means. It is also noteworthy that in the pre-test, both groups managed to achieve just under 75% of the total score, i.e. 30 out of a potential maximum of 40 points. On the post-test, the INDI group saw an increase of 4.2% and the COOP students, 9.9%. In the case of the COOP students, the increase is so high that a statistical significance ($p \leq .05$) is present and the effect size reaches a high value. Cohen's (1998) guidelines were used to interpret effect size, according to which $d \geq 0.2$ represents a small, $d \geq 0.5$ a medium and $d \geq 0.8$ a large effect.

This increase in performance is put into perspective by the delayed post-test of both groups, with the results decreasing again and even falling almost to the level of the pre-test. Yet despite the trend being the same in both groups, there are differences between them in the post-test. The results of the COOP students immediately after the treatment is 5.3% higher than that of the INDI students. This advantage is largely preserved over the period between the immediate to the delayed post-test. Here the difference is still 4.6%, but the results are not or no longer statistically significant compared to the initial values of the pre-test. Noteworthy in this respect is that in this competency-specific pair of tasks, the maximum score was reached nine times in all three tests. In the multiple-choice exercise (A1) the maximum was achieved in 37.8% and in the fill-in-the-gaps exercise (A2) in 23.0% of the cases.

To summarize, on the post-test, the COOP students perform better than the INDI students and on the delayed post-test both groups perform worse than the post-test, but the COOP students' results are still better than the pre-test, whereas the INDI students have returned to pre-test levels. The results of tasks A4 and A5, which were not part of the pre-test, show that the COOP students do better than the INDI students on the post-test and the delayed post-test, but the INDI students' performance improves from the post-test to the delayed post-test. The results are shown on the graph below in Figure 2.

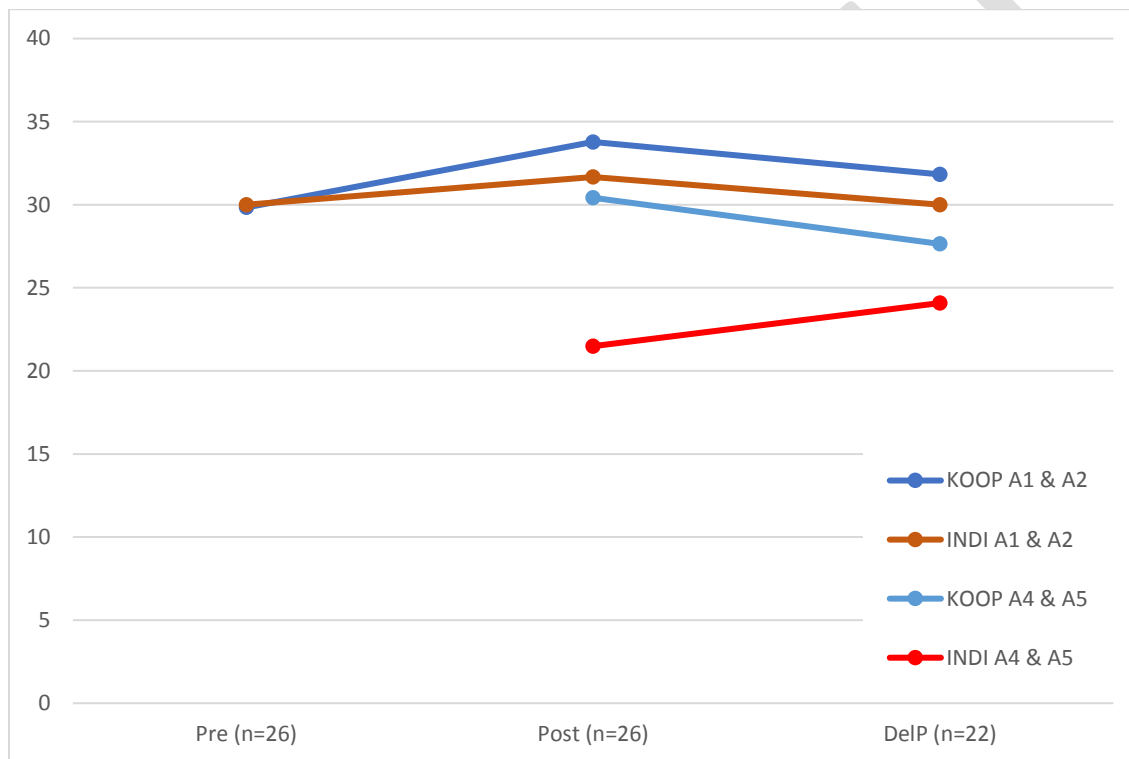


Figure 2: Average results of Tasks A1, A2, A4 and A5

The text production tasks in 4 and 5 present a different picture from Tasks A1 and A2. As mentioned above, Tasks A3, A4 and A5 were not part of the pre-test and because the open-ended tasks are more complex and, as expected, the percentages are not as high as for Tasks A1 and A2. There is a 22% difference on the post-test between the COOP students and the INDI students and an 8.9% difference on the delayed post-test. On both the post-test and the delayed

post-test, the COOP students outperform the INDI students, although the performance of the INDI group on the delayed post-test is better than on the post-test.

Table 5 below shows the results of a t-test for Tasks A1 and A2. The t-tests compared the performance on Tasks A1 and A2 between the pre- and post-test, the post-test and the delayed post-test and the post-test and the delayed post-test. The table shows that the only significant difference is between the pre- and the post-test for the COOP students.

Group	Pre-test		Post-test		Delayed Post-test		Test-pair	p	Effect Size (d)
	M	SD	M	SD	M	SD			
COOP (n=12)	29.83	4.89	33.78	4.92	31.83	4.52	Pre-Post	.02	.8
							Post-Del	.17	.4
							Pre-Del	.15	.5
INDI (n=14)	30.00	6.59	31.67	7.03	30.00	6.81	Pre-Post	.37	.3
							Post-Del	.31	.3
							Pre-Del	.92	.1

Table 5: Results of t-tests comparing pre-, post- and delayed post-tests

One way to visualize the results of Tasks A4 and A5 is the box-plot graph in Table 6. The results of the post-test and delayed post-test show that the median of the COOP students and the INDI students is farther apart than the average (marked with an X). This means that the COOP students not only achieved better results in the average, but the number of above average individual students was higher among the COOP students than the INDI students. This tendency is also shown in the average of the COOP students in both post-tests, which is above the top quarter of the INDI students.

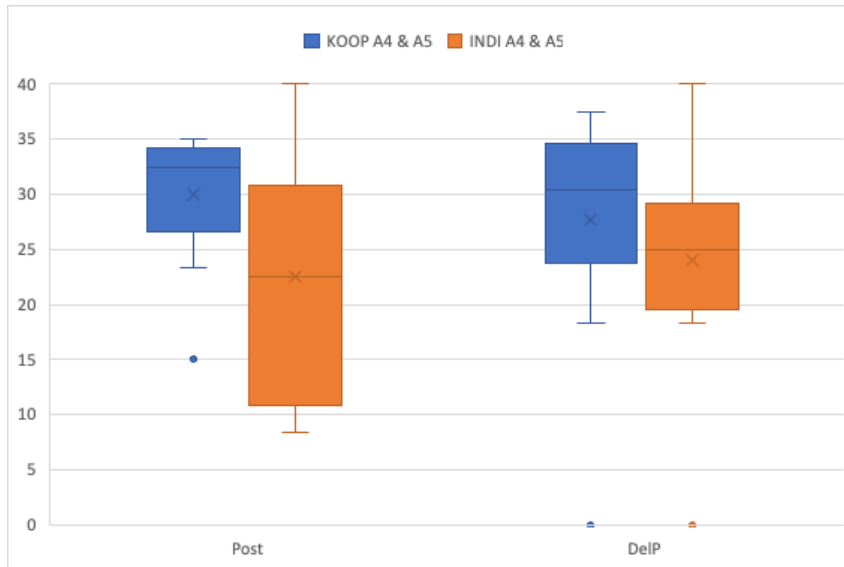


Table 6: Tasks A4 and A5 on the post- and delayed post-test

The graph also shows how wide the spread of INDI students is in the post-test in comparison to the COOP students. In the delayed post-test the variation is not so striking because the lowest part of the INDI grouping is not so far below the lowest part of the COOP grouping.

The results of all of the tasks individually and cumulatively show that the COOP students almost without exception have the higher mathematical mean and smallest standard deviation. Task 3, however, shows a different pattern. The mean of the INDI group is marginally higher than the COOP students and the standard deviation is higher, which means that there is more variation in the COOP students.

Post-test	COOP		INDI		t	p	Effect size (d)
	(n=12)		(n=14)				
	M	SD	M	SD			
A1	16.00	4.18	15.71	4.83	0.160	.874*	.1
A2	17.77	2.60	15.95	4.75	1.187	.972*	.5
A3	14.38	7.70	14.46	5.21	-0.035	.247*	.0
A4	15.63	2.95	10.66	6.59	2.541	.02*	1.1
A5	14.17	5.10	8.93	7.44	2.117	.045*	.8
Total	78.57	11.61	67.62	17.39	1.853	.076	.8

Table 7: Means and t-tests for all tasks on the post-test

The total score mean for all tasks is 10% higher for the COOP students than for the INDI students and the standard deviation is one third lower. This difference can be accounted for by results of Tasks A4 and A5 as well as A2, although in a more limited way. The consequence of this this difference is a high effect size, which is 1.1 in Task A4. This means that the results of one student in the COOP group is on average one standard deviation above any individual student in the INDI group, that is the mean from the COOP students is higher than the mean of 86% of the students in the INDI group. Task A2 also shows a size effect, although it is a middle one. The p-values of the subtasks (marked with *) must be treated with caution, because due to the repeated comparison, the significance level has to be adjusted and is no longer at $p \leq .05$, but after a Bonferroni-correction at $p \leq .01$. In this respect, there is in any case no statistical significance, which is why the null hypothesis cannot be refuted.

In the delayed post-test, the COOP students generally show better values than INDI group, i.e. higher average values and lower standard deviations. This applies to the overall result as well as the individual task results, except for Task 2, where the INDI group performs better, although the difference is minimal. By contrast, the mean values of the text productions (A4 and A5) are somewhat more constant. Although the distance between the groups is reduced in the delayed post-test, the COOP group again achieves the higher scores on average. By contrast, the standard deviations are closer together, and the lower values are no longer found exclusively in the COOP students. In the overall result, the difference of the averages is slightly lower than in

the post-test and reaches only 6.73 points, rather than 10.95. The standard deviation, however, increases from 5.78 to 9.61 points. Significant contributors to the INDI group's high variance may have been the results of two subjects in the group who once obtained the minimum and once the maximum possible score in the text productions.

4. Discussion

The current study addresses two research questions. The first question concerns short- and long-term learning effects through the combination of collaborative learning methods and cognitive linguistic approaches to grammar teaching. Overall, the test results of the COOP group in Tasks 1 and 2 show that the learners have improved as a result of the treatment. The inferential statistical analysis, however, shows that this improvement is only significant in the post-test and thus indicates a short-term learning effect. Since the difficulty level of the tasks was, in many cases, too low for the participants, it is quite possible that a ceiling effect prevented an exact measurement and that the actual learning gain is even higher than that observed in the study.

A similar picture emerges for Tasks 4 and 5 with regard to the sustainability of the learning effects (delayed post-test), namely a drop-in performance can be observed between the post-test and the delayed post-test in the COOP group. The decrease in the learning outcomes over time corresponds with our expectations, since it can be assumed that storing what has been learned in long-term memory requires constant cognitive activation, for example through exercises (Sweller et al. 1998). Such exercises between post-tests were not part of the treatment and it can be excluded with high probability that collective passive training outside the regular classroom led to an increase in the performance only for the INDI group.

Another explanation as to why the results of the text productions (A4 and A5) were below those of the competence-specific ones (A1 and A2) could be the different degree of complexity in the task. Thus, not only is the grammatical competence of the passive in its entirety required in text production, its application becomes a secondary matter with the multitude of aspects that need to be considered when mastering these complex tasks. In view of the fact that no statistical significance was achieved in the two text productions (A4 and A5), the low p-values are nevertheless very striking compared to the others and also deserve attention in

the context of the heterogeneous mean values and standard deviations as well as the high effect sizes. Furthermore, the results in Task 3 show a slightly different pattern in both groups: While the learning performance in the COOP students remains relatively constant and even increases slightly over time, the performance of INDI group drops by more than one point after the treatment. It can thus be suggested that social modeling while learning collaboratively leads to deeper processing and thus more sustainable learning effects (King 2007).

The second research question concerns the efficacy of collaborative learning methods within the framework of cognitive linguistic approaches to grammar teaching in comparison to individual learning methods. For this purpose, the results of the two groups were compared in the different tasks. The analysis shows that the COOP students outperformed the INDI group in all tasks both in the post-test and in the delayed post-test except for Task 3 in the post-test and Task 2 in the delayed post-test, where the INDI group performed better on average than the COOP students. Even if the differences between the groups are not significant after a correction of the p-value, the overall picture of the results suggests a higher efficacy of the collaborative learning methods compared to the individual learning method. This positive learning effect is also due to the use of cooperation scripts, which have engaged learners in dealing with relevant aspects in the animations. In the future, however, the effect of cooperation scripts should be investigated by more differentiated study designs including a third group that uses collaborative learning methods, but without cooperation scripts. We would expect the group with collaborative learning methods and cooperation scripts to achieve better learning outcomes than the other two groups. The additional cognitive load of the collaborative learning method, which results from the interaction in the group and the processing of all relevant aspects of the learning material, can be compensated by scripting collaborative learning. It is also important to note that the use of cooperation scripts over time is intended to trigger self-regulation in collaborative learning and thus internalize the process of successful social construction of knowledge (cf. King 2007)

Although the current study substantially adds to our understanding of how grammar animations can best be leveraged by collaborative learning methods, the following methodological limitations need to be considered in future research. First, the relatively small sample size led, in many cases, to large standard deviations and consequently to a higher degree of variability and less representativity. Thus, larger groups of participants will provide more accurate measures and allow stronger claims. Second, given the fact that we did not test the

participants' prior knowledge on the same tasks other than in the post-test and delayed post-test, we cannot exclude the existence of differences between the groups before the treatment. In addition, including the same tasks in all three measurement points would have allowed us to look into whether performance in text production between the pre-test and post-test increased or leveled off. Finally, it is worth mentioning that further research is needed on the difference between epistemic and social cooperation scripts. Whereas the former focuses on the proper processing of the content while learning collaboratively, the latter aims to assist learners in their interaction within the group. In this context, the combination of content- and interaction-based prompts in the same cooperation script might well be a useful way for learners to engage in collaborative learning that is worth being investigated.

References

- Achard, Michel. 2004. Cognitive grammar and communicative grammar instruction. *Cognitive linguistics, second language acquisition and foreign language teaching* (Studies on language acquisition 18), ed. by Michel Achard and Susanne Niemeier, 167–97. Berlin: Mouton de Gruyter.
- Achard, Michel. 2008. Teaching construal: Cognitive pedagogical grammar. *Handbook of cognitive linguistics and second language acquisition*, ed. by Peter J. Robinson and Nick C. Ellis, 432–55. London: Routledge.
- Bachman, Lyle F., and Adrian S. Palmer. 1996. *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Bellavia, Elena. 2007. *Erfahrung, Imagination und Sprache: Die Bedeutung der Metaphern der Alltagssprache für das Fremdsprachenlernen am Beispiel der deutschen Präpositionen*. Tübingen: Gunter Narr Verlag.
- Bielak, Jakub, and Mirosław Pawlak. 2011. Teaching English tense and aspect with the help of cognitive grammar: An empirical study. *Studies in second language learning and teaching* 1(3). 365–400.
- Cohen, Andrew D. 2001. Second language assessment. Teaching English as a second or foreign language, 3rd ed., ed. by Marianne Celce-Murcia, 515–34. Boston, MA: Heinle & Heinle.
- Cohen, Jacob. 1988. *Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum.
- Compaoré, Clément. 2018). *Evaluation kollaborativer eLern- und Lehrprozesse. Instruktionsdesign zum Einsatz kognitionsbasierter Grammatikanimationen in virtuellen Klassen. Reihe: Transkulturelle Kommunikation*. Münster: Lit Verlag.
- Cramer, Duncan. 1998. *Fundamental statistics for social research: Step-by-step calculations and computer techniques using SPSS for Windows*. London: Routledge.
- Cramer, Duncan & Dennis Howitt. 2004. *The SAGE dictionary of statistics: A practical resource for students in the social sciences*. London: SAGE.
- Croft, William. 2003. *Typology and universals*, 2nd ed. Cambridge: Cambridge University Press.
- Dirven, René, and Sabine de Knop. 2008. Motion and location events in German, French and English: A typological, contrastive and pedagogical approach. *Cognitive approaches to pedagogical grammar: A volume in honour of René Dirven* (Applications of cognitive

- linguistics 9), René Dirven, ed. by Sabine de Knop and Teun de Rycker, 295–324. Berlin: Mouton de Gruyter.
- Doane, David P. and Lori E. Seward. 2011. Measuring Skewness: A Forgotten Statistic? *Journal of Statistics Education* 19(2). 1–18.
- Evans, Vyvyan and Melanie J. Green. 2006. *Cognitive linguistics: An introduction*. Mahwah, NJ: Lawrence Erlbaum.
- Hänze, Martin and Roland Berger. 2007. COOPeratives Lernen im Gruppenpuzzle und im Lernzirkel. *Unterrichtswissenschaft* 35(3). 227–40.
- Hopper, Paul and Sandra Thompson. 1980. Transitivity in grammar and discourse. *Language* 56 (2). 251–299.
- Jacobsen, Natalia. 2015. A cognitive linguistic analysis of English conditionals in English for academic purposes (EAP) instruction: Implications from sociocultural theory. *Cognitive linguistics and sociocultural theory: Applications for second and foreign language teaching* (Studies in second and foreign language education 8), ed. by Kyoko Masuda, Carlee Arnett and Angela Labarca, 103–26. Berlin: Mouton De Gruyter.
- Janda, Laura and Steven Clancy. 2002. *The case book for Russian*. Columbus, OH: Slavica.
- Kanaplianik, Katsiaryna. 2016. *Kognitionslinguistisch basierte Animationen für die deutschen Modalverben: Zusammenspiel der kognitiven Linguistik und des multimedialen Lernens bei der Sprachvermittlung* (Transkulturelle Kommunikation Band 7). Berlin: Lit Verlag.
- King, Alison. 2007. Scripting collaborative learning processes: A cognitive perspective. *Scripting computer-supported collaborative learning: Cognitive, computational, and educational perspectives*, ed. by Frank Fischer, Ingo Kollar, Heinz Mandl, and Jorg M. Haake, 13–37. New York: Springer.
- Kohl-Dietrich, Dorothee, Constance Juchem-Grundmann and Wolfgang Schnotz. 2016. Conceptual Motivation as a tool for raising language awareness in the English as a foreign language classroom – Does it enhance learning outcomes? Insights from an empirical study. *Special issue of the yearbook of the German cognitive linguistics association on “Cognitive approaches to second language learning and teaching”*, ed. by Juliana Goschler and Susanne Niemeier, 193–209. Berlin: Mouton de Gruyter.
- Lakoff, George and Mark Johnson. 2003. *Metaphors we live by*. Chicago, IL: University of Chicago Press.

- Langacker, Ronald W. 1982. Space Grammar, analysability, and the English passive. *Language* 58(1). 22-80.
- Langacker, Ronald W. 1987. *Foundations of cognitive grammar*. Stanford, CA: Stanford University Press.
- Langacker, Ronald W. 2008. Cognitive Grammar as a basis for language instruction. *Handbook of cognitive linguistics and second language acquisition*, ed. by Peter J. Robinson and Nick C. Ellis, 66–88. London: Routledge.
- Lantolf, James P. 2007. Sociocultural source of thinking and its relevance for second language acquisition. *Bilingualism: Language and Cognition* 10(1). 31–3.
- Liamkina, Olga. 2005. The role of explicit meaning-based instruction in foreign language pedagogy: Applications of cognitive linguistics to teaching the German dative case to advanced learners. Washington DC: Georgetown University dissertation.
- Lysinger, Diana. 2009. *Teaching case to L2 students of German and Russian*. Davis, CA: University of California dissertation.
- Mackey, Alison and Susan M. Gass. 2005. *Second language research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum.
- Masuda, Kyoko. 2013. Applying cognitive linguistics to teaching Japanese polysemous particles. *Sophia Linguistica* 60. 103—120.
- Negueruela, Eduardo. 2009. A conceptual approach to promoting L2 grammatical development. *Conceptions of L2 grammar: Theoretical approaches and their application in the L2 classroom* (Issues in language program direction 2008), ed. by Stacey K. Bourns and Johanna Watzinger-Tharp, 151–71. Boston, MA: Heinle Cengage Learning.
- Purpura, James E. 2004. *Assessing grammar* (The Cambridge language assessment series). Cambridge: Cambridge University Press.
- Razali, M. Nornadiah and Yap B. Wah. 2011. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics* 2(1). 21–33.
- Reif, Monika. 2012. *Making progress simpler? Applying cognitive grammar to tense-aspect teaching*. Frankfurt am Main: Peter Lang.
- Rice, Sally Ann. 1987. *Towards a Cognitive Model of Transitivity*. University of California at San Diego dissertation.

- Roche, Jörg and Ferran Suñer. 2014. Kognition und Grammatik: Ein kognitionswissenschaftlicher Ansatz zur Grammatikvermittlung am Beispiel der Grammatikanimationen. *Zeitschrift für interkulturellen Fremdsprachenunterricht* 19(2). 119-145.
- Roche, Jörg & Ferran Suñer. 2016. Metaphors in grammar teaching. *Yearbook of the German cognitive linguistics association*, ed. by Anatol Stefanowitsch and Doris Schoenefeld, 89–112. Berlin: Mouton De Gruyter.
- Seel, Norbert M. 2003. *Psychologie des Lernens*, 2nd ed. München, Basel: E. Reinhardt.
- Shapiro, Samuel S. and Martin B. Wilk. 1965. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* 52(3/4). 591–611.
- Smith, Michael. 1987. *The semantics of dative and accusative in German: An investigation in cognitive grammar*. San Diego, CA: University of California/San Diego dissertation.
- Suñer, Ferran. 2013. Bildhaftigkeit und Metaphorisierung in der Grammatikvermittlung am Beispiel der Passivkonstruktion. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 18(1).
- Suñer, Ferran and Jörg Roche. 2019. Embodiment in concept-based L2 grammar teaching: The case of German light verb constructions. *IRAL: International Review of Applied Linguistics in Language Teaching*.
- Sweller, John, Jeroen J. G. van Merriënboer and Fred G. W. C. Paas. 1998. Cognitive architecture and instructional design. *Educational Psychology Review* 10(3). 251–96.
- Sweller, John, Paul Ayres and Slava Kalyuga. 2011. *Cognitive load theory*. New York: Springer.
- Talmy, Leonard. 2000. *Toward a cognitive semantics: Volume 1: Concept structuring systems* (Language, speech, and communication). Cambridge, MA: MIT Press.
- Tyler, Andrea. 2008. Cognitive Linguistics and second language instruction. *Handbook of cognitive linguistics and second language acquisition*, ed. by Peter J. Robinson and Nick C. Ellis, 456–88. London: Routledge.
- Tyler, Andrea, Charles Mueller and Vu Ho. 2011. Applying cognitive linguistics to learning the semantics of English to, for, and at: An experimental investigation. *Vigo International Journal of Applied Linguistics* 8. 181-205

- Vogel, Freydis, Christof Wecker, Ingo Kollar & Frank Fischer. 2017. Socio-Cognitive Scaffolding with Computer-Supported Collaboration Scripts: a Meta-Analysis. *Educational Psychology Review* 29, 477–511.
- Vygotsky, Lev S. 1986. *Thought and language*. Cambridge, MA: MIT Press.
- Weinberger, Armin, Karsten Stegmann and Frank Fischer. 2010. Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Computers in Human Behavior* 26(4): 506–515.
- Williams, Lawrence, Lee B. Abraham and Eduardo Negueruela-Azarola. 2013. Using concept-based instruction in the L2 classroom. Perspectives from current and future language teachers. *Language Teaching Research* 17. 363-381.
- Yasuda, Sachiko. 2010. Learning phrasal verbs through conceptual metaphors: A case of Japanese EFL learners. *TESOL Quarterly* 44(2). 250–73.