Proceedings of the International Multiconference on
Computer Science and Information Technology pp. 553–559
ISBN 978-83-60810-27-9
ISSN 1896-7094

German subordinate clause word order in dialogue-based CALL

Magdalena Wolska, Sabrina Wilske
Computational Linguistics,
Saarland University,
Saarbrücken, Germany
{magda,sw}@coli.uni-saarland.de

***Abstract*—We present a dialogue system for exercising the German subordinate clause word order. The pedagogical methodology we adopt is based on focused tasks: the targeted linguistic structure is embedded in a naturalistic scenario, "Making appointments", in which the structure can be plausibly elicited. We report on the system we built and an experimental methodology which we use in order to investigate whether the computer-based conversational focused task we designed promotes acquisition of the form. Our goal is two-fold: First, learners should improve their overall communicative skills in the task scenario and, second, they should improve their mastery of the structure. In this paper, we present a methodology for evaluating learners' progress on the latter.**

I. Motivation

VERBAL communication in a foreign language, actual interaction, is for the language learner the ultimate site of language acquisition. Dialogue is a source of naturally occurring comprehensible input as well as useful negative feedback, reformulations or clarification questions, which may arise from communication problems and which draws attention to correct forms. What is crucial is that dialogue is an opportunity for learners to *produce language* as well as to *modify* their language in response to feedback. All these aspects of conversational interaction have been shown to promote learning [1], [2], [3], [4].

The communicative approch to language teaching advocates the use of goal-oriented realistic communicative activities, *tasks*, in the foreign language classroom, in order to encourage learners to use their developing language [5]. Important definitional characteristics of tasks are: focus on meaning, well defined communicative outcome, and free use of linguistic forms. If a specific grammatical structure is the target of instruction, the latter property of tasks turns out problematic: because learners are free to use any forms they want, it cannot be guaranteed that they will use the forms of interest. To remedy this, *focused tasks*, encouraging the use and processing of specific linguistic features, have been proposed [5]. Focused tasks combine focus on forms with the communicative approach to instruction by, among others, exploiting scenarios which are likely to elicit the structures of interest in an unobtrusive way, promoting thereby their incidental acquisition.

In this paper we report on a system we built and an experimental method designed to find out whether *computer-based* conversational focused tasks also promote acquisition of forms. The structure we targeted was the German word order in subordinate clauses. The goal of the computer-based communicative task was two-fold: On the one hand, learners should improve their overall communicative skills in the task scenario and, on the other hand, they should expand their mastery of the target structure. We have conducted a preliminary small-scale experiment with the system we built in order to assess the feasibility of an in-classroom evaluation. The learning gains results we have observed so far have, however, *not* been statistically significant. In this paper we concentrate on the system itself and the evaluation methodology in general.

The idea of computer-based dialogue activites for foreign langauge learning is not new: computer assisted language learning (CALL) has been an active research field for many years. With the progress in language technology, the number of intelligent CALL systems which allow learners to use natural dialogue has been growing. The study we present is at the intersection of two fields that are intuitively close: second language acquisition (SLA) and Natural Language Processing-enhanced CALL.

A CALL system can be evaluated in terms of learning gains it generates [6], [7], in terms of its usability (How did learners enjoy playing with the system? [8], [9], [10]) or its performance from an engineering perspective (Did it fail? [11]). We built a CALL system which implements two established SLA methodologies (focused tasks and conversational interaction) and attempt to evaluate whether iteracting with the system produces learning gain.

Outline The paper is organized as follows: In Section II we introduce the linguistic form of interest and the task scenario. In Section III we present two language learning activities we designed. The architecture of the system is outlined in Section IV. In Section V we summarise the setup of an experiment and the results. Section VI concludes the paper.

II. The target form and the task

For the focused communicative activity we selected a grammatical form and a task with the following considerations in mind: Firstly, we wanted a "non-trivial", demanding structure, i.e. a structure of certain complexity, one which is not aquired at the first stages of learning German. Secondly, a structure for

Fig. 1. System screenshot; task material (left), dialogue history and input entry field (right)

which we could find a scenario in which it is natural to use, so that we can create incidental opportunites for the learner to produce it. Thirdly, it has enough distinguishing features to be easily tested in a controlled experiment. Finally, in line with task-based teaching, we wanted a meaningful, realistic scenario and communicative task, useful for the learner.

Given the above criteria we opted to focus on the German word order in causal subordinate clauses and framed the usage of these structures in the context of a "Making appointments" scenario. We introduce the two parameters below.

A. Form: Subordinate clauses

Subordinate clauses are clauses that are dependent on another clause (main or subordinate). In German, subordinate clauses are characterized by a specific word order: the position of the finite verb in a subordinate clause is at the end of the clause. This position is obligatory. The canonical placement of the finite verb in the main clause is at the second position.

One of the subordinate clause types is an adverbial clause in which the subordinate clause, taking the function of an adverbial, qualifies the action expressed in the main clause by supplying additional information. An example of an adverbial clause is the causal clause which provides the reason for what is said in the main clause.

Examples (1) and (2) below show a causal clause introduced by the subordinating conjunction *weil* (Eng.: because) in the verb-final position and of a single main clause in which the finite verb **muss** is in second position:

(1) ...weil ich arbeiten **muss**.
because I work-inf must-fin.
'because I have to work'

(2) Ich **muss** arbeiten.
I must-fin work-inf
'I have to work'

The specific word order of subordinate clauses is problematic for learners and it has been shown that it is the last to be acquired by children and adult learners of German [12].

While subordinate clauses are a useful means to structure content and express relations between different propositions, they are used much less in oral communication than in written text. The proportion of subordinate clauses of all clauses ranges from 0.25 to 0.12 in corpora of spoken language, while it is 0.5 in written language [13],[14],[15], as cited by [16]. A possible reason for this dispreference is that subordinate clauses are more complex and thus require more effort to process, which is harder in spontaneous interaction.

B. Task: Making an appointment

In order to elicit causal clauses, we created a task in which the learner had to refuse a proposal and, for pragmatic, politeness considerations (a dispreferred second [17]) would likely provide a reason for the refusal.

The scenario in which we embed the focus on subordinate clauses is about arranging a meeting. The task for the learner is to make an appointment given a set of constraints on the available times: the learner is provided with a schedule with a set of occupied and free slots, as illustrated on the left side of Figure 1, and activites planned in the occupied slots (i.e. the reasons for refusals).[1]

[1]The agenda also includes conditionally busy slots, marked with *wenn* (Eng.: if) which can optionally serve to elicit conditional subordinate clauses characterised by the same verb-final word order.

```
get-user-input
if interpretation-found
  if no-justification-found
     elicit-justification
  else
     if TF-realized
        if TF-incorrect
           recast-TF
        else prompt-for-next-contribution
     else
        recast-justification-using-TF
else
  output 'Sorry, I didn't understand.'
```

Fig. 2. Dialogue strategy in the free production activity

The system would propose appointment times known to be occupied on the learner's schedule, thus expecting the learner to refuse the proposal and give a reason. However, keeping in mind that it is not obligatory to provide the reason at all and that a subordinate causal clause is an optional construction, in one of the activites we designed (described below) the system-side of the dialogue was modelled in such a way that it provided examples of the clauses of interest by embedding them in reformulation/paraphrasing utterances and giving them an appearance of implicit confirmation moves.

III. Task-based Activities

We designed and implemented two variants of a role-play type activity framed within the scenario described above: In both variants it involved a *type-written* dialogue with the system we built, with a goal of making an appointment. The system controls the interaction by means of a state-based dialogue model and explicitly implements form-focusing mechanisms: in one variant, this is done as part of the dialogue model, while in the other, by restricting the input mode.

The dialogue model encodes subdialogues which serve to *elicit the target forms* and it *provides feedback on forms* in case of learner form errors. The two variants of the activity differ in the extent of freedom of language production they offer and the realisation of form-focused feedback: one variant allows learners to freely formulate their dialogue contributions (free production) and provides implicit corrective feedback, while in the other learners are asked to produce only the target forms (constrained production) and the feedback merely informs whether the supplied form was correct. We elaborate on the properties of the respective system variants below.

A. Free language production

In the free-production system, the learner is able to type their utterances freely without any restrictions on the language used. The system implements two input interpretation strategies: one based on a grammar with mal-rules, and a fall-back strategy based on keyword matching; details follow in Section IV. It classifies the learner's input into one of the three categories ("TF" stands for "target form"): TF-realized-correct, TF-realized-incorrect, TF-not-realized. The high-level dialogue and feedback strategy is summarised as pseudo-code in Figure 2.[2]

The system provides implicit feedback in case of learner errors in the TF by reformulating (*recasting*) the learner's utterance (or parts thereof). Recasts are realised in a way so as to give them an appearance of implicit confirmation type of grounding moves, as in **S1**, below, which corrects the error made in **L1**:

(3) **L1:** *Nein, ich kann nicht, weil ich muss arbeiten.
'No, I can't because I have to work.'
S1: Ah, du kannst nicht, weil du arbeiten musst.
'Ah, you can't because you have to work.'

The dialogue model encodes three strategies of eliciting causal clauses if the learner does not use them spontaneously: (**A**) If the learner gives a reason for refusal, but does not produce a subordinate clause the system will recast the refusal into a subordinate clause and put emphasis on the conjunction **weil** by setting it in bold face as illustrated below:

(4) **L2:** Nein, ich kann nicht, ich muss arbeiten.
'No, I can't, I have to work.'
S2: Ah, du kannst nicht, **weil** du arbeiten musst.
'Ah, you can't **because** you have to work.'

(**B**) If the learner fails to give a reason in their refusal the system will ask for one explicitly:

(5) **L3:** Nein, am Montag um 15 Uhr kann ich nicht.
'No, I can't make it on Monday at 3.'
S3: Warum kannst du denn nicht?
'Why can't you make it?'

(**C**) In order to present an example of a causal clause not as part of a recast, but as an original refusal-reason pair the system will refuse any learner-initiated proposal with a reason formulated as a causal clause. If the learner does not initiate a proposal the system will try to elicit one by asking the learner what day and time would suit them.

B. Constrained production

In the constrained system the learner's production is restricted to supplying the target form by putting a set of words in the correct order creating a dialogue turn in this way. The words are given in a random order, as in the example below:

(6) **S4:** Kannst du am Montag um 10 Uhr?
'Are you available on Monday at 10am?'
L4: Nein, ich kann nicht, weil (arbeiten | muss | ich)
'No, I can't because I have to work.'

The learner is allowed three attempts to produce the correct form. In case an invalid form is supplied, the system signals it with a message 'That was wrong!' and subtracts one point

[2]We omit some system turns signalling non-undertanding due to unknown words to simplify the presentation.

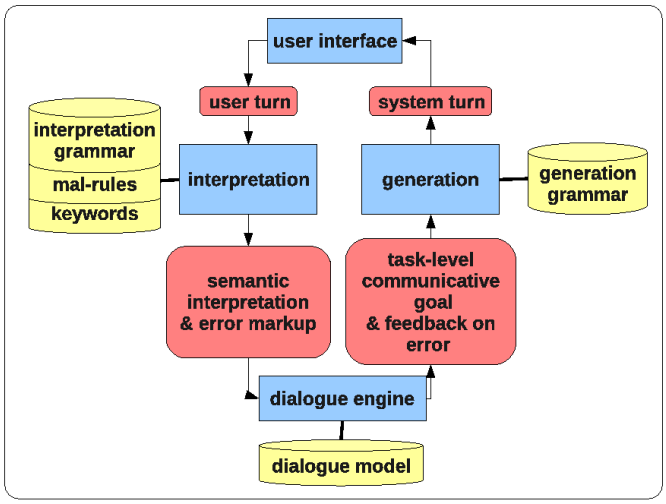

Fig. 3. The system architecture

from a learner's "score" on the activity; correct forms increase the score by one. The feedback and the score are displayed in a designated feedback area. After the third unsuccessful attempt the correct utterance is appended to the dialogue. The system then generates its next turn based on the dialogue model.

The following section summarizes the architecture and the implementation of the system.

IV. The System

Both dialogue activities are implemented on the same system architecture; we concentrate on the components required for the free production activity because the constrained production activity is its simplified variant.

The system maintains a dialogue with the learner by following a dialogue strategy outlined in Section III (see Figure 2). This involves interpreting the learner's input, responding to the learner by selecting a communicative goal according to the dialogue model and the paedagogical strategy, and realizing the goal as a surface string. Specifically for the learning context, the system has to recognize errors in the learner input and generate feedback on them.

Figure 3 shows the system's architecture: the modules and the flow of information between them. We describe each of the functions below.

A. The task and dialogue model and the dialogue engine

The dialogue model represents the sets of possible turn transitions: alternating turns produced by the user and the the system. Task-related parameters, the information about the slots in the time-table, are encoded in an external data structure which is imported into the dialogue model.

The dialogue model is implemented as a state machine using State Chart XML (SCXML) as an underlying representation. We use the Java implementation of Apache SCXML.[3] The framework also provides a dialogue execution engine which receives input interpretations and triggers the system responses according to the given model.

B. Interpretion of learner's input

In general, interpreting the user input involves mapping a surface string of an utterance to a meaning representation. As typical in small-scale dialogue systems, we implement the system's language model (the set of linguistic expressions) as a context free grammar with semantic tags. For parsing, we use the Java Speech API implementation of the CMU parser which is part of the Sphinx system.[4] The semantic tags encode two types of information: first, the symbolic meaning of utterances, and second, information on violations of grammatical constraints; more on error handling below.

1) Fuzzy matching for unknown words: In order to ensure robustness with respect to typos and spelling errors the system first identifies unknown words in the input and tries to map them to known words by calculating the Levenshtein distance between the unknown word and known words. Candidates for replacement of out-of-vocabulary words are those known words which have a Levenshtein distance within a certain range normalized by word length.

2) Grammatical error handling: Since the system interacts with learners, i.e. non-native speakers of German, their input is likely to contain other errors apart from misspellings, in particular errors in the target structure. An essential requirement of the system is to recognize those errors and give feedback on them. One strategy to deal with errors is to explicitly integrate anticipated errors into the grammar in the form of so called mal-rules, i.e. grammar productions which are outside of the standard rules of the given language. Erroneous utterances are parsed using mal-rules and the parse result contains information about the error.

The drawback of this approach is that it is hard to anticipate all possible errors that might occur. Therefore, our system also implements a fall-back strategy based on keyword spotting: If no parse is found for an utterance, we create a semantic interpretation based on content words, using a keyword lexicon. We encoded a set of mal-rules based on informal prior pre-testing of the system with beginner learners.

C. Generation of system responses

The system output realization is performed using a template-based approach. The output is produced by generating a dialogue move selected according to the dialogue model using a context free generation grammar. The grammar associates atomic keys representing communicative goals with sets of possible realizations. Slots in the generation templates are filled using feature-value pairs passed as arguments to the templates along with the communicative goals to be realized.

D. User interface

The user interface is implemented as a Java applet embedded in a website. The applet displays the task material,

[3] http://commons.apache.org/scxml

[4] Available at http://cmusphinx.sourceforge.net

an input field for learner, the dialogue history and additional buttons for editing the input utterances and selecting specific tasks. Figure 1 shows the graphical interface including the task-material (the agenda; left) and a part of a dialogue history (upper right) with an example of a system recast (as in (3)) and of an elicitation strategy (as in (5)).

V. Evaluation

In this section we present an evaluation methodology which we use to evaluate learning gains produced by the system variants discussed above. The evaluation is based on in-classroom activities with the system we built.

We have conducted a preliminary small-scale experiment using the experimental design introduced below, however, the results we have obtained so far have not been statistically significant. Therefore, we present the current results only very briefly in Section V-B.

A. Methodology

Design The experimental design we use is a nonrandomized pretest multiple-posttest design involving students from German language classes at the university, taught by different teachers. The classes are split randomly into two sub-groups: one assigned to the free production condition, and the other to the constrained production condition. We are interested in two questions: 1) whether the interactive activities produce learning gains, and 2) whether the free production condition, which requires more computational effort (e.g. in interpreting the learners' turns), produces more gains than the activity in which the learners' production is restricted.

Procedure At the first session, time 1., both groups complete a pretest, then interact with one of the system variants (repeating the exercise twice), and subsequently complete an immediate post-test (posttest1). At the next session a week later, time 2., the groups again perform an in-classroom exercise with the system in a different configuration of the schedule and complete another post-test at the end of the session (posttest2). Finally, time 3., after a couple of week's break (five weeks in the pilot study) the groups complete another post-test (delayed posttest).

After the second session the participants fill out a demographic/learning history survey (anonymous) and a feedback questionnaire on the interaction with the system; we ask about the usability, usefulness, interest in future use, etc.

We provide two different variants of the task for each participant. The basic scenario frame, a weekly time schedule, is kept, but the character of the system is changed: In the first variant, the system takes the persona of a fellow student, whereas in the second it is introduced as a learner's supervisor. The motivation for the latter is that an interaction with a superior might produce behaviour which more closely conforms to the politeness norms, in this case, providing reasons for declining a proposal and hence also the target forms. Also the set of days and times which the system proposes is different for each exercise. In the first repetition of the activity, the system makes 5 different proposals, for the second it makes 4 proposals, thus we expect 9 uses of causal subordinate clauses according to our dialogue script.

Tests We use two types of assessment tests: a timed grammaticality judgment test, targeting implicit knowledge, and an untimed sentence construction test, targeting explicit knowledge. Implicit knowledge refers to knowledge accessible through automatic processing and which learners are intuitively aware of, while explicit knowledge is knowledge accessible through controlled processing [18].[5]

a) Timed grammaticality judgment: Following Ellis [19] we designed a timed grammaticality judgment test to measure implicit knowledge. The test items include causal subordinate clauses of different complexity. The complexity varies as to the amount of additional material present in the clause, e.g. objects, modal verbs, negations or additional modifiers.

The test consists of 6 grammatical, 6 ungrammatical test items and 9 grammatical and 9 ungrammatical distractor items, including a subset of other subordinate clauses. We set the time-limit for the test to 10 seconds per item. This is roughly twice the maximum time a native speaker used. Ellis timed his test at 20% above the average time native speakers needed [19]. Han and Ellis used 3.5 seconds as the time constraint in [20] based on pretesting the items, while Bialystok used an even shorter time limit [21]. Based on our own pretest with native speakers, we performed, already the threshold of 3.5 seconds would have excluded a couple of slow native speakers. Since we are not aware of research which explicitly addresses the issue of the time limit on the timed judgement tasks, we opt for a more generous time-limit.

Each correctly judged item is scored at 1 point, each incorrectly judged item is scored at 0.

b) Sentence construction: For the explicit knowledge test, participants are asked to complete sentences given the beginning of a sentence and a set of unordered uninflected phrases or words as in the example below:

Item:	Ich kann nicht (weil, arbeiten, müssen, ich)
Solution:	Ich kann nicht, weil ich arbeiten muss.
	'I can't because I have to work.'

The test consists of 6 test items for causal conditional clauses. There is no time-limit. The items are scored at 1 point if the word order is correct, 0 otherwise. All form errors other than those in the target structure are neglected.

We created four versions of the tests described above to be administered at the four times of assessment (pretest, posttest1, postest2, delayed postest). The versions differ in the combinations of lexemes, but are otherwise comparable with regard to complexity of the lexical items used. The assignment of a test version to a time varies between participants in order to compensate for unintended differences between versions. Within each test, items are presented in random order.

[5]The tests are prepared and administered using Webexp Experimental Software (`http://www.hcrc.ed.ac.uk/web_exp/`).

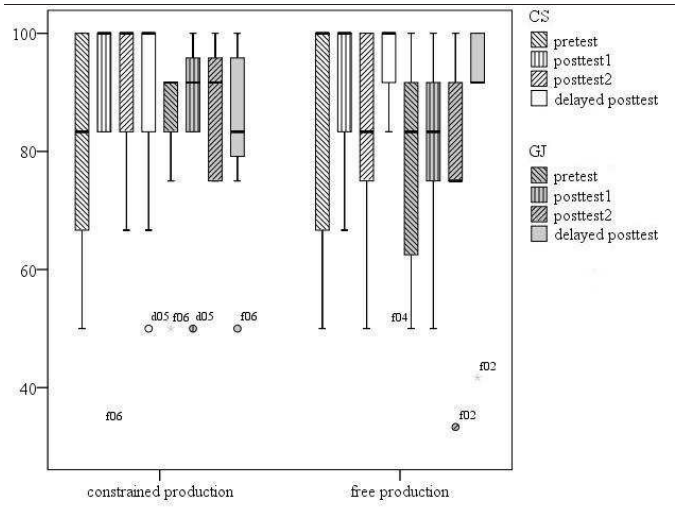

Fig. 4. Overall results for sentence construction (SC; white) and grammaticality judgement (GJ; gray) for both conditions

B. Pilot experiment

We conducted a small-scale experiment, using the setup described above, in order to assess the feasibility of an in-classroom evaluation and in order to get an impression of whether the learners benefit from the system(s). As we had mentioned earlier the results we have obtained so far have *not* been statistically significant in neither of the groups, however, the usability questionnaires and the feedback we got from the participants of the study encourage us to pursue the free-form exercises further. We are therefore planning to conduct another analogous experiment in the coming fall semester, however, we will invite learners from German courses at a lower proficiency level than the groups we had access to for the pilot study. Below, we briefly outline the current results.

Participants For the pilot experiment we had access to 26 learners from two German language courses. The participants came from different language backgrounds, were both male and female, with an average age of 25 years, and had been learning German for an average of about two years prior to experiment.[6] The courses met twice a week for 90 minute sessions. The experiment started 6 weeks (ca. 15 instruction hours) into the course.

The subjects participated in two sessions of in-classroom exercises with one of the system variants with one week's break between the sessions. Each session consisted of at least two repetitions of the activity in different configurations of the task material (the time schedule) as described in Section II. To complete the activity the participants took between 5 and 25 minutes in the free condition and between ca. 2 and 10 minutes in the constrained condition.

With the experiment spanning over a few weeks, subject drop-out was inevitable. Due to a high course drop-out rate (42%), at this point we have data for only 15 subjects for all the four assessment points for both tests: 7 subjects in the free production condition and 8 in the constrained production. We removed a set of data for one subject in the constrained condition who obtained full scores at all the assessment points obtaining a data set with two groups of 7 subjects.

[6]Their German proficiency level was classified as ranging from A2 to B1+ CEF level, based on scores on an initial course placement test.

TABLE I
NUMERICAL RESULTS OF THE SENTENCE CONSTRUCTION TEST: MEANS (M) AND STANDARD DEVIATIONS (SD) FOR PERCENTAGE SCORES

		Pretest		Posttest 1		Posttest 2		Delayed Posttest	
	N	M	SD	M	SD	M	SD	M	SD
Percentage scores									
Constrained	7	80.95	20.25	85.71	24.40	90.48	16.26	88.10	20.89
Free	7	83.33	21.52	90.48	16.26	83.33	19.24	90.48	18.90

TABLE II
NUMERICAL RESULTS OF THE TIMED GRAMMATICALITY JUDGMENT TEST: MEANS AND STANDARD DEVIATIONS FOR PERCENTAGE SCORES

		Pretest		Posttest 1		Posttest 2		Delayed Posttest	
	N	M	SD	M	SD	M	SD	M	SD
Percentage scores									
Constrained	7	83.33	15.96	85.71	17.16	86.90	11.64	83.33	17.35
Free	7	77.38	19.07	80.95	17.16	77.38	22.42	88.10	20.89

Analysis Because of the small sample size and because of the violation of parametric assumptions[7] we perform non-parametric analyses: in order to compare within subject differences we use the Friedman test.[8] For between groups comparisons we use the Mann-Whitney U test. We set the significance level at 0.05.

Results and discussion The overall results are shown in Figure 4. Because of the small sample size, the skewed distribution of scores (as seen in the figure) we cannot draw any ultimate conclusions: In both groups the repeated measures statistic turned out to be not significant. The reason for this is likely to be the fact that both of our groups started off with relatively high scores. There was no significant between-group difference on the pretest, according to Mann-Whitney U test, i.e. the groups started off at the same level. However, this pretest level was at an average of 81% and 83% of the total scores on the sentence construction test in the constrained and free production condition respectively, and 83% and 77% in the free production. That is, the subjects were perhaps too familiar with the target structure.

Table I and Table II show the percentage scores' means and standard deviations for the pretest, posttest1, posttest2, and delayed posttest for the *sentence construction test* and the *grammaticality judgement test*. The general pattern in the scores on both tests is the same: Both groups increased accuracy in the use of subordinate clause word order from pretest to posttest1, however, while the constrained production group further increased between posttest1 and posttest2, the

[7]According to Shapiro-Wilk and Levene tests both the normality assumption and the assumption of homogeneity of variance were violated on at least some of the within-subject and/or between-subject variables on either tests in the pilot study.

[8]Since we did not obtain a statistically significant result, we did not perform post doc tests at this time.

free production group declined. The accuracy then declined between posttest2 and delayed posttest in the constrained condition, but improved in the free production condition. As mentioned above, these differences were, however, not statistically significant at the level we had set. Neither were the differences in between-group comparisons of the scores. Some of the differences were, however, marginally significant at a more liberal level of 0.1.

As mentioned at the beginning of this section, we do not draw ultimate conclusions as to the learning gains based on the pilot study results. We believe that the experiment is definitely worth re-running in a course at a lower proficiency level than the one we were working with at this time. We are planning such an experiment for the coming fall. The free production system was rated significantly higher on the questionnaire than the constrained production system and some of the lower scoring learners did declare that they would like to use such a system for at-home exercises, were it available.

VI. Conclusion

We presented an architecture of a dialogue system for interactive computer-based exercises for the German subordinate clause word order. The exercises are designed based on an established communicative-teaching methodology of focused tasks. The system implemets elicitation mechanisms which cue the learner on using the target structure of interest. In one exercise variant the target structures are ellicited in an unobtrusive way, while the other exercise has more of a drill-like character. We also presented an evaluation methodology for assessing learning gains upon interaction with the system.

While our small-scale pilot study was not conclusive, we believe our appoach is worth pursuing further based on encouraging feedback from the participants of the pilot study. One conclusion we might perhaps draw is that the exercise mode we propose is more suitable at an earlier stage of acquisition of the form we target, i.e. around the time when the learners are first introduced to the subordinate clause word order, rather than being familiar with it already. However, we would certainly need another experiment to confirm this and we do intend to conduct a larger scale study using the methodology we described.

Based on the same architecture and a modified implementation, we also built another dialogue activity for exercising the German locative use of the Dative case, framed in a "Directions giving" scenario. In an analogous pilot study, using the same evaluation methodology as described in Section V, we found significant improvement in the use of Dative upon exercising with our system [22].

Acknowledgements

We would like to thank Dr. Kristin Stezano Cotelo and Meike van Hoorn, language instructors at the German courses of the International Office at the Saarland University, as well as their students who participated in the experiments, for the help in conducting this study.

Magdalena Wolska's position at Saarland University is funded through the INTERREG IV A project ALLEGRO (Project No.: 67 SMLW 1 1 137).

References

[1] M. H. Long, "Input, interaction and second language acquisition," in *Native language and foreign language acquisition*. Annals of the New York Academy of Sciences, 1981, vol. 379, pp. 259–78.

[2] S. D. Krashen, *Input Hypothesis: Issues and Implications*. London: Longman, 1985.

[3] M. Swain, "Communicative competence: Some roles of comprehensible input and comprehensible output in its development," in *Input in second language acquisition*, S. Gass and C. Madden, Eds. MA: Newbury House, 1985, pp. 235–53.

[4] ——, "Three functions of output in second language learning," in *Principle and practice in applied linguistics: Studies in honour of H.G. Widdowson*, G. C. . B. Seidlhofer, Ed. Oxford: Oxford University Press, 1995, pp. 125–144.

[5] R. Ellis, *Task-based Language Learning and Teaching*. Oxford University Press, 2003.

[6] V. M. Holland, J. D. Kaplan, and M. A. Sabol, "Pre-liminary tests of language learning in a speech-interactive graphics microworld," *Calico Journal*, vol. 16, no. 3, pp. 339–359, 1998.

[7] W. Harless, M. Zier, and R. Duncan, "Virtual dialogues with native speakers: The evaluation of an interactive multimedia method," *Calico Journal*, vol. 16, no. 3, pp. 313–37, 1999.

[8] C. Wang and S. Seneff, "A Spoken Translation Game for Second Language Learning," in *Proceedings of the Conference on Artificial Intelligence in Education*, 2007, pp. 315–322.

[9] T. Lech and K. de Smedt, "Dreistadt: A language enabled moo for language learning," in *Proceedings of the ECAI-06 Combined Workshop on Language-enabled Educational Technology and Development of Robust Spoken Dialogue Systems*, 2006, pp. 38–44.

[10] W. L. Johnson and S. Wu, "Assessing aptitude for learning with a serious game for foreign language and culture," in *Intelligent Tutoring Systems*. Springer Berlin / Heidelberg, 2008.

[11] S. Seneff, C. Wang, and J. Zhang, "Spoken conversational interaction for language learning," in *Proceedings of the InSTIL Symposium on Computer Assisted Language Learning*, 2004, pp. 151–154.

[12] M. Pienemann, *Language Processing and Second Language Development: Processability Theory*. Benjamins, 1998.

[13] K. Bayer, "Verteilung und funktion der sogenannten parenthese in texten gesprochener sprache," in *Forschungen zur gesprochenen Sprache und Möglichkeiten ihrer Didaktisierung*, Goethe-Institut, Ed. Kemmler, 1971, pp. 200–214.

[14] U. Engel, "Syntaktische besonderheiten der deutschen alltagssprache," in *Gesprochene Sprache*. Pädagogischer Verlag Schwann, 1974.

[15] J. Weijenberg, *Authentizität gesprochener Sprache in Lehrwerken für Deutsch als Fremdsprache*. Groos, 1980.

[16] F. Bubenheimer, "Grammatische besonderheiten gesprochener sprache und didaktische konsequenzen für den daf-unterricht," Published online at http://www.deutschservice.de/felix/daf/gesprkom.html.

[17] S. C. Levinson, *Pragmatics*. Cambridge University, 1983.

[18] R. Ellis, S. Loewen, and R. Erlam, "Implicit and explicit corrective feedback and the acquisition of L2 grammar," *Studies in Second Language Acquisition*, vol. 28, pp. 339–368, 2006.

[19] R. Ellis, "Modelling learning difficulty and second language proficiency: The differential contributions of implicit and explicit knowledge," *Applied Linguistics*, vol. 27, no. 3, pp. 431–463, 2006.

[20] Y. Han and R. Ellis, "Implicit knowledge, explicit knowledge and general language proficiency," *Language Teaching Research*, vol. 2, pp. 1–23, 1998.

[21] E. Bialystok, "Explicit and Implicit Judgements of L2 Grammaticality," *Language Learning*, vol. 29, pp. 81–103, 1979.

[22] M. Wolska and S. Wilske, "Form-focused task-oriented dialogues for computer assisted language learning: A pilot study on german dative," in *Proceedings of the Interspeech-2010 Workshop "Second Language Studies: Acquisition, Learning, Education and Technology"*, 2010, To Appear.