

# Studying Model Ambiguity in a Language ITS

Brent Martin and Amanda Nicholas

Intelligent Computer Tutoring Group  
Department of Computer Science and Software Engineering  
University of Canterbury  
Private Bag 4800, Christchurch, New Zealand  
{brent, amn39}@cosc.canterbury.ac.nz

**Abstract.** Ambiguity is a well-known problem in student modelling, and in user modelling in general. In this paper we present the results of an experiment in the domain of German adjectives. We trialled a modified student interface that gathers more data during problem solving by requiring the student to perform a related subtask. There is evidence that the students who performed the subtask outperformed the control group on a post-test despite the extra task slowing them down, suggesting the extra effort required by the students to overcome ambiguity was worth the intervention.

**Keywords:** Student Modelling, Language learning, ITS

## 1 Introduction

Dealing with ambiguity is a serious problem in developing Intelligent Tutoring Systems for foreign languages [1]. Although the system can detect that the student has made an error, the source of this error may be difficult to determine. Menzel defines four sources of ambiguity: limited observability, polysemy, alternative conceptualisations of domain knowledge and structural uncertainty. In a domain with high ambiguity feedback messages can be difficult to determine. Good feedback should refer the student to the underlying domain principle. If it is not possible to determine which domain principle has been broken, correctly targeted feedback cannot be given. One approach to avoid ambiguity is to require the student to specify the intermediate steps they carry out mentally, however this may reduce transference [1] [2]. This research compares two constraint-based (CBM) tutors: one that matches the real world more closely, and one that decreases ambiguity as much as possible.

German adjective endings are a difficult topic for students to master. This is due to the number of endings that must be memorised, and the amount of knowledge required of the sentence to get the ending correct. Rogers studied the main areas of weakness in students with more than four years of experience learning German [3]. She states "...much anecdotal 'evidence' from teachers of German as a foreign language emphasises morphology as a major area of weakness (e.g. adjective endings...)". Her study showed that approximately 5% of errors made by advanced learners of German were errors in adjective endings.

German adjectives must agree with the nouns they modify. This means that the ending of an adjective varies based on the gender and case of the noun, and whether the noun is preceded by a definite article, indefinite article, or no article.

**Table 1.** Adjective endings when preceded by the definite article

Case	Masculine	Feminine	Neuter	Plural
Nominative	e	e	e	en
Accusative	en	e	e	en
Genative	en	en	en	en
Dative	en	en	en	en

For example, Table 1 lists the endings for the case where an adjective is preceded by an indefinite article. For example, take the sentence “Das graue Haus ist neu”. (The grey house is new). Here “Haus” is the noun, and its gender is neuter. The house is the subject of the sentence, and so it is in the nominative case. Das is the article, and it is the direct article. The adjective is “grau”, and it takes the ending “e” because, by consulting table 1, we see that adjectives preceding a neuter noun in the nominative case must end in “e”. It is important to note that the endings are not unique; the ending “e” appears in a number of situations, as does “en”. This is one reason why these endings are ambiguous.

Menzel identified four major sources of ambiguity that should be considered when creating CBM tutors, particularly for foreign languages [1]. These are: limited observability of internal variables of the problem domain; polysemy (multiple meanings) of symbols used in the problem domain; alternative conceptualisations of domain knowledge; uncertainty about the intended structure of the student’s solution. German adjective endings suffer from three of the four defined sources of ambiguity. Limited observability and polysemy are both present in the multiple possible meanings of a single ending. When the student incorrectly gives an adjective ending, it could be due to either a *rule* error or a *fact* error. If the student does not know the gender or the case of the noun, they have made a fact error. If the student has correctly determined the case, gender and article, and still gives the adjective ending incorrectly, they have made a rule error; they do not know the underlying grammatical principle that determines the adjective ending.

In the next section we summarise how constraint-based modelling was applied to the domain of German adjectives. Section 3 describes the experiment and presents the results. Finally, we conclude in Section 4.

## 2 Constraint-Based Modelling

CBM[4] is a modelling approach based on the theory of learning from performance errors [5]. It models the domain as a set of state constraints, where each constraint represents a declarative concept that must be learned and internalised before the student can achieve mastery of the domain. Constraints represent restrictions on solution *states*, and take the form:

If <relevance condition> is true for the student’s solution,  
THEN <satisfaction condition> must also be true

The following is an example of a constraint from the German adjectives domain (\*IS\* and \*SS\* refer to the ideal and student solutions, respectively):

```
(21 "When the article is an 'ein-word', the only possible
endings are -e, -en, -er and -es."
```

```
(and (match *IS* ("I"))(match *SS* (?anything ?*)))
(or (match *SS* (?* "e")) (match *SS* (?* "e" "n"))
    (match *SS* (?* "e" "r")) (match *SS* (?* "e" "s"))))
```

This constraint checks that the student has used an appropriate ending for an “ein-word”. The relevance condition first checks that the article for this problem is indefinite (“I”), and that the student has attempted an answer. The satisfaction condition then checks that the adjective the student typed ends with one of the valid endings. If this is not true, the feedback message is given. The student model consists of the set of constraints, along with information about whether or not each has been successfully applied, for each attempt where it was relevant.

### 3 Experiment

We hypothesised that forcing the students to supply information about their problem-solving process and providing feedback based on that information would enable the system to give them better instruction, and thus they would be better able to learn the domain. We tested this hypothesis by building two versions of an ITS for German adjectives, where the two systems differed in the interface used and the underlying domain/student model (constraints).

The tutors were developed using the WETAS tutoring shell [6]. The problem set comprised of 55 problems, which was identical for both tutors. Some were obtained from existing sources [7, 8], however, most problems were written especially for this ITS. An example of one of the problems in the tutor is

Die ? Blumen gefallen mir. (bunt) (*I like the colourful flowers*)

The two tutors shared a very similar interface. In the centre of the screen was an area for the student to answer the question. The problem was displayed in the form of a sentence with a gap left where the adjective should be, and the adjective to be inserted was given in brackets at the end of the sentence. Below there was a selection box that allowed the student to choose the desired feedback level, and a button to submit their answer for evaluation. Feedback messages appeared at the bottom of the screen.

Students using the experimental system were asked to fill in the gender and case of the noun, the article type, and the adjective with its ending. The possible answers for gender, case and article were all given in combo boxes. This ensured that there would not be problems with students referring to the same concept by a different name, or misspelling names. Below the combo boxes there was a text field for the student to fill in the appropriate form of the adjective. Students using the control were asked to fill in the correct adjective form only. Domain constraints were sourced from a number of German textbooks (e.g. [8]), which contain advice on how students can remember the endings more easily.

An evaluation study of the two tutors was conducted on the 6th of September 2006 at the University of Canterbury, Christchurch. Students enrolled in a beginning German course used one of the two systems over one 50-minute period. The students had been taught adjective endings previously in class. The class was divided randomly into two even groups, and the students were first asked to complete a pre-test. They then used the tutoring system for as long as time permitted, or until they finished all 55 questions. Afterwards they completed a post-test. Each test contained six questions in the same format as described previously. The final three questions also asked the student to specify the gender and case of the noun present in the sentence, and the type of article preceding the noun. The experiment was carried out in two streams. To allow for any difference in the difficulty of the pre- and post-tests, Test 1 was used as the pre-test for Stream A, and the post-test for Stream B; Test 2 was used as the post-test for Stream A and pre-test for Stream B.

23 students took part in the evaluation. 12 students used the experimental tutor and 11 students used the control. Students using the control system solved more problems with fewer attempts than those using the experimental tutor. This result is unsurprising, because students using the control only had to fill in one value correctly, whereas students using the experimental tutor had to supply four values. Students using the experimental tutor also saw more feedback messages. This is also unsurprising; their task was larger so there were more opportunities to make mistakes.

Unfortunately, the study revealed the pre- and post-tests were not of comparable difficulty. To overcome this, we compared the results for test 1 only, and compared the outcome for pre- and post-test regardless of which stream the students belonged to. This is not strictly valid because the samples are different; it relies on the assumption that the students in the two streams (and using the same tutor) were comparable, and this cannot be easily measured. Using this assumption, a t-test of the score for producing the correct adjective ending showed no significant difference between the test 1 pre-test scores for the two tutors (mean = 4.8 and 4.6 for the control and experimental groups respectively, SD = 0.8 and 1.6,  $p > 0.7$ ). When test 1 is used as a post-test however, there is a larger difference between the two groups, with the experimental tutor achieving a score of 5.7 compared to 5.0 for the experimental group, although the result is not statistically significant ( $p = 0.15$ ).

We also compared the performance of the two groups in terms of their ability to perform the subtask (determine case and gender). Again there was no significant difference on pre-test score between the control and experimental groups (5.0 versus 4.9). For the post-test, the experimental group again outperformed the control group, scoring an average of 5.7 compared to 4.8 for the control group. The result was statistically significant ( $p < 0.05$ ). This is what we would expect, given that the experimental group practised this specific task. An analysis of learning curves [9] also showed a better “power curve” for the subtask constraints.

Finally, the students were asked to fill in a subjective survey at the end of the study. Responses from were overwhelmingly positive to both versions of the tutor. Further, the staff from the German department indicated they would like to pursue this technology further, because the students had reacted so positively. They also commented that the results for the formal adjectives test were considerably higher than in previous years, which they attributed to the tutoring systems.

## 4 Conclusions

Tutoring systems that teach natural languages are susceptible to the problem of ambiguity in student answers, making it difficult to apportion blame appropriately and thus provide effective feedback. Even a highly constrained domain such as German adjectives exhibits this problem. Requiring the student to supply additional information is often frowned upon because it reduces the similarity with “real world” problems and may thus negatively affect transfer.

We examined this problem in the domain of German adjectives by providing two versions of a simple ITS; the control required the students to complete the original task only (and thus suffered from ambiguity) while the experimental group forced them to also complete a subtask that disambiguated their response. The results were not conclusive because of problems with the pre- and post-test difficulties. However, there was evidence from these tests that the experimental group performed better on both the original task and the subtask despite having solved considerably fewer problems because of the additional time needed to complete the subtask. This suggests that far from detracting from the students’ ability to complete the main task, the extra disambiguation benefited their learning. Further, when the domain models were analysed (via learning curves), the additional constraints required for the subtask appeared to enhance the performance of the model in capturing what was learned, suggesting that the subtask was an integral part of the main task.

This study has shown that adding extra task requirements to overcome ambiguity in language learning is not always a bad thing, and can in fact be advantageous. This is a positive outcome that encourages us to further explore how constraint-based models may support language learning.

## References

1. Menzel, W., *Constraint-based modeling and ambiguity*. International Journal of Artificial Intelligence in Education, 2006. 16(1): p. 29-63.
2. Anderson, J.R., Corbett, A.T., Koedinger, K.R., and Pelletier, R., *Cognitive Tutors: Lessons Learned*. Journal of the Learning Sciences, 1995. 4(2): p. 167-207.
3. Rogers, M., *On major types of written error in advanced students of german*. International Review of Applied Linguistics in Language Teaching, 1984. 22(1): p. 1-39.
4. Ohlsson, S., *Constraint-Based Student Modeling*, in *Student Modeling: The Key to Individualized Knowledge-Based Instruction*, J. Greer and G. McCalla, Editors. 1994, Springer-Verlag: New York. p. 167-189.
5. Ohlsson, S., *Learning from Performance Errors*. Psychological Review, 1996. 3(2): p. 241-262.
6. Martin, B. and Mitrovic, A. *Authoring web-based tutoring systems with WETAS*. in *International conference on computers in education*. 2002. Auckland. p. 183-187.
7. Werner, G., *Langenscheidts Grammatik-training Deutsch*. 2001: Langenscheidt KG.
8. Kahlen, L., *Interactive German Made Easy*. 2006: McGraw-Hill.
9. Newell, A. and Rosenbloom, P.S., *Mechanisms of skill acquisition and the law of practice*, in *Cognitive skills and their acquisition*, J.R. Anderson, Editor. 1981, Lawrence Erlbaum Associates: Hillsdale, NJ. p. 1-56.