

## Constraint-based modeling and ambiguity

**Menzel, Wolfgang**

*Universität Hamburg, Fachbereich Informatik, Vogt-Kölln-Straße 30,  
D-22527 Hamburg, Germany*

*email: menzel@informatik.uni-hamburg.de*

*nats-www.informatik.uni-hamburg.de*

**Abstract.** Constraint-based modeling has been used in many application areas of Intelligent Tutoring Systems as a powerful means to analyse erroneous student solutions and generate helpful feedback. In contrast to domains where the structure of the problem under consideration allows a constraint to (almost) uniquely determine the possible cause of a particular student error there are other applications, where a multitude of competing error explanations has to be considered. In such cases constraint-based models alone hardly meet the requirements for a student model. Instead a constraint-based model clearly serves the purpose of error diagnosis and needs to be complemented by additional components for diagnosis selection based on general or individually tailored heuristics.

By investigating the apparent and strong parallelism between constraint-based modeling and model-based diagnosis, this paper identifies four major sources of ambiguity that need to be considered when using constraint-based modeling and describes options for dealing with situations in which alternative error descriptions are available. Examples are primarily drawn from the area of foreign language learning.

**Keywords.** constraints, diagnosis, model-based diagnosis, tutoring systems, foreign language learning, grammar, CALL

### INTRODUCTION

Constraint-based modeling (CBM) has demonstrated its high potential as a powerful approach for building Intelligent Tutoring Systems, which identify misconceptions of a student and possible shortcomings of her solutions as a basis for initiating suitable instructional actions. A number of systems have been built using this technology in various application domains (Mitrovic et al., 2001). Less clear, though, seems to be the precise role such a model plays within the overall architecture of a particular tutoring system. Here the individual perspectives vary between understanding CBM

- as a student model, as it was originally proposed by Ohlsson (1994),
- as a diagnostic approach (Mitrovic et al., 2003), or even

- as a domain model (Martin and Mitrovic, 2000).

On the one hand, all these conceptualizations seem to be well justified, but at the same time they also appear to be limited in some sense: viewing CBM as a student model is motivated by the fact that constraint violations indeed can tell us crucial information about a student's current level of achievement. Later, however, it will become apparent, that in application domains with a higher degree of ambiguity constraint violations alone are not sufficiently informative to provide a clear enough picture, from which an appropriate instructional response can be generated directly. Under such conditions, the view of CBM as a diagnostic engine seems more appropriate. It can feed a student model with analytic information about student solutions, but is no substitute for such a model. On the other hand, the potential of a CBM is not restricted to diagnosis alone. Under favourable conditions it can even be devised to serve as an expert module, which at least to some degree is able to solve a given problem entirely on its own. Unfortunately, this is not necessarily always the case, as the example of the balancing-bracket constraint taken from (Mitrovic and Ohlsson, 1999) can show:

If the code for a Lisp function has N left parentheses, there has to be N right parentheses as well (or else there is an error).

Although this constraint perfectly serves the purpose of *detecting* cases of unbalanced expressions in the student solution, it is not able to *localize* them properly or even produce an appropriate correction proposal. To compensate this deficiency Martin and Mitrovic (2002) extended the model by an additional problem solving component, which is able to derive correction proposals from the observed constraint violations and a given model solution.

In order to obtain a better understanding of the conditions under which CBM can make contributions with different degrees of diagnostic depth, we first will investigate the relationship between CBM on the one hand and *model-based diagnosis* on the other. It should become clear that whenever a constraint-based model can be cast in terms of the behaviour of informationally coupled components, a level of model adequacy is reached which allows it to be used immediately as an expert module. Unfortunately this is not always the case.

By analysing different examples for various types of model information from the domain of second language learning, we then draw our attention to *ambiguity* as a major challenge for more advanced CBM solutions. Ambiguity presents a particular problem to CBM, because it introduces yet another dimension into the space of possible constraint violations: in addition to the normal situation of having possibly several constraints violated simultaneously by one and the same student solution, we now are faced with the existence of *alternative views* on a conflict in the student's solution, where each of these alternatives can give rise to a perfectly valid error description and might again involve a number of distinct constraint violations.

Ambiguity is a defining property of a particular problem domain. It therefore cannot be completely avoided without losing essential characteristics of the domain to be taught. Using mainly examples from foreign language learning we will investigate the origin of ambiguity in more detail and discuss its consequences for the application of constraint-based reasoning techniques. In particular we will identify four areas which can be considered major sources for ambiguity, namely

- a limited observability of internal variables of the problem domain,
- polysemy of symbols used in the problem domain,
- alternative conceptualizations of domain knowledge, and
- uncertainty about the intended structure of the student's solution.

These problem classes then can be used to determine the level of sophistication an appropriate problem solution requires.

Ambiguity in the problem domain inevitably leads to diagnostic uncertainty, if erroneous solutions have to be considered. It is this uncertainty which prevents constraint violations from being used directly as a student model. Faced with ambiguity a constraint-based model usually provides a vast variety of diagnostic information, which needs to be filtered and condensed, before it can be used in a practical tutoring system. The section on hypothesis selection will be devoted to this issue.

## **CONSTRAINT-BASED MODELING**

Formally, constraints are pairs consisting of a relevance part and a satisfaction part (Ohlsson, 1994). They are used to describe conditions which must hold for every (partial) solution contributed by the student. Using the relevance part, constraints can be tailored towards specific exercise (types) and specific (structurally determined) configurations within a typical student solution. Additional requirements, which have to be fulfilled in that specific situation, are coded in the satisfaction part. From a formal point of view, constraints are universally quantified logical formulae consisting basically of an implication

$$\text{relevance part} \rightarrow \text{satisfaction part} \quad (1)$$

i.e. whenever the relevance part evaluates to true the satisfaction part must do so as well. If, however, a constraint is violated this indicates a shortcoming of the student solution and possibly hints to a misconception of the student. Thus, constraint violations can be used to trigger appropriate instructional actions.

CBM comes with a number of advantages. First of all, it obviates the need for a directly runnable expert module, which would be able to carry out the given problem solving steps on its own and is difficult to develop for many application areas. Instead, CBM aims at defining instructionally relevant equivalence classes, i.e. establishing abstractions of individual solution attempts. By not prescribing a particular sequence of solution steps, but checking for more general requirements a solution has to meet, CBM offers the student a remarkable degree of freedom to submit innovative solutions and even invites her to explore the space of possible solutions without any further restriction. If in the worst case constraints accidentally happen to be too weak to precisely restrict the space of admissible solutions, erroneous solutions can be left uncommented by the system. Nevertheless, such a behaviour compares favourably with alternative approaches, where model limitations inevitably cause the rejection of unusual solutions which are fully acceptable, but have not been expected by the system developer.

Furthermore, CBM appears to be neutral with respect to a specific pedagogical approach and can be used both in incremental scenarios, where the solution strategy of the student is monitored step by step and partial results have to be evaluated, and batch-like mode, which only considers complete solutions to be checked for errors. After all, the decisive question of using CBM in the design of tutoring systems is one of knowledge acquisition: is it possible to cast the problem domain in terms of sufficiently strong constraints, which also provide useful abstractions about the emerging skills of a student? To answer that question, we will investigate a number of subproblems from the area of foreign language learning, which introduce a number of interesting characteristics and might shed some light on the potential and the limitations of the CBM approach.

CBM is based on the idea that in many domains learning takes place as learning from performance errors, whereby errors result from declarative knowledge which has not yet been internalized. Second language acquisition in controlled learning environments, like in a common student-teacher setting, is a particularly instructive example for this kind of cognitive development:

Firstly, language is taught as a kind of declarative knowledge by providing the student with

- sample utterances,
- facts about the meaning of words and their grammatical properties,
- rules about how to combine words in a proper way, to eventually build up meaningful and well-formed complex language constructs, and
- guidelines about the appropriateness of certain utterances under particular circumstances.

Usually, teaching (part of) a grammar in such a way is by no means sufficient for the development of applicable language skills in a student. In any case, the learning process has to be complemented by intensive repetitive exercises. Such exercises allow the student to actively practise language use, thus helping her to internalize the acquired knowledge and making it accessible for communicative purposes.

Secondly, language production (the active and thus more important part of language use) is not a consciously guided process of reasoning, which can be learned by applying a particular predefined sequence of solution steps. In this respect natural language communication differs considerably from other cognitive tasks like algebra or computer programming. Under normal circumstances, the process of generating language utterances can hardly be observed, whereas its ultimate outcome definitely can. Therefore, the incremental analysis of partial results is not a primary goal in a language learning environment.

Thirdly, language communication is an activity primarily guided by the speaker's personal intentions. Although for practical purposes the space of possible intentions can be narrowed down to a certain degree (e.g. by formulating an explicit exercise task, by providing a textual or pictorial description of a real-world scenario, or by simulating a restricted kind of dialog), there is obviously no way to fully take control of a person's intentions without rendering the communicative setting completely unnatural. Again this is a remarkable difference to other learning tasks, where the aspired outcome of an exercise can clearly be specified by verbal or non-verbal means.

Finally, active language use is always a problem of choice between a large number of possible alternative expressions. Here, the analytical nature of CBM is of great advantage, since by not prescribing particular solutions it opens up possibilities for creative language use, instead of requiring the student to merely replicate predefined patterns.

CBM starts from the assumption that the observation of a constraint violation allows us to draw a direct conclusion about the current cognitive state of the student. However, experience with natural language data gives rise to the suspicion that this can be taken for granted only as long as the domain under consideration allows us to make some important simplifying assumptions, namely

- the preconditions of a constraint are always sufficiently strong to *uniquely* identify the relevance of the particular constraint and
- the satisfaction part of a constraint evaluates to a truth value that can be determined independently from the outcome of other constraint applications.

Both assumptions may be adequate for many of the situations handled in simple domains, like algebra, where both, the numerical value of a number and its role in a given problem are beyond any doubt. This situation, however, will change dramatically if different kinds of ambiguity have to be managed using CBM.

## MODEL-BASED DIAGNOSIS

To better highlight the differences between different application classes for CBM, we need to put the approach on a more formal foundation, something we can achieve using the descriptive framework of model-based diagnosis (MBD) (Reiter, 1987). MBD is based on the notion of informationally coupled components. Components receive their semantics via a type-sensitive description of their behaviour

$$\text{type}_i(X) \rightarrow \text{out}(X) = f(\text{in}(X)). \quad (2)$$

Note that formally there is a strong analogy between this expression and the formal description of a constraint in CBM: the premise simply contains a type check condition, whereas the conclusion states a (perhaps complex) input-output relationship which a component of the given type has to fulfill. To be able to perform diagnostic reasoning, component descriptions are additionally guarded by a "normal-behaviour" assumption

$$\text{type}_i(X) \wedge \neg \text{abnormal}(X) \rightarrow \text{out}(X) = f(\text{in}(X)) \quad (3)$$

which can be retracted as soon as the observed behaviour happens to be in conflict with the desired one. Thus, constraints are defeasible and can be violated.

Despite the apparent similarity of formulae 1 and 2 there is a fundamental difference which should be noted: While under the CBM approach constraints are defined in a way that allows them to be evaluated independently from each other, here we might be faced with situations in which the truth value

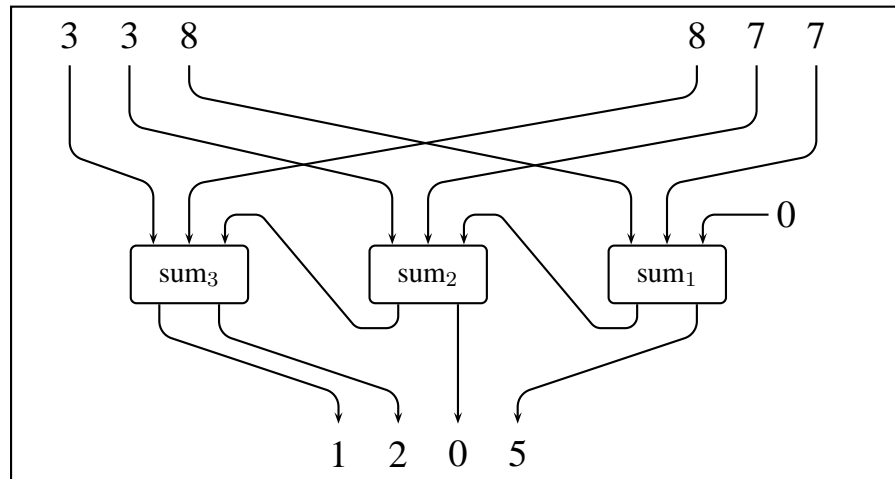


Fig.1. Component-based conceptualization of an addition problem

of the satisfaction or even the relevance part depends on *assumptions* about the behaviour of neighbouring components. As we will see, such dependencies are an immediate consequence of ambiguity in the diagnosis problem and require computationally more challenging solution procedures.

As a first approach to conceptualize a simple addition exercise for natural numbers, the semantics of a component for computing a (one-digit) sum 'out<sub>r</sub>' and a carry over 'out<sub>c</sub>' from three input digits can be described as

$$\begin{aligned} \text{type-sum}(X) \wedge \neg \text{abnormal}(X) \rightarrow \\ \text{out}_r(X) = (\text{in}_1(X) + \text{in}_2(X) + \text{in}_c(X)) \bmod 10 \\ \wedge \text{out}_c(X) = (\text{in}_1(X) + \text{in}_2(X) + \text{in}_c(X)) // 10 \end{aligned} \quad (4)$$

Components of this type can then be combined to build more complex models for multi-digit addition, by providing appropriate type declarations and connectivity statements (cf. Figure 1)

$$\begin{aligned} \text{type-sum}(\text{sum}_1) \wedge \text{type-sum}(\text{sum}_2) \wedge \text{type-sum}(\text{sum}_3) \\ \text{in}_1(\text{sum}_3) = 3 \wedge \text{in}_2(\text{sum}_3) = 8 \wedge \text{in}_c(\text{sum}_3) = \text{out}_c(\text{sum}_2) \wedge \dots \wedge \text{in}_c(\text{sum}_1) = 0 \end{aligned} \quad (5)$$

These axioms taken together comprise the *system description S*, a complete specification of the problem given to the student, e.g.

$$\begin{array}{r} \phantom{+} \phantom{+} \phantom{+} \\ \phantom{+} \phantom{+} \phantom{+} \\ + \phantom{+} \phantom{+} \phantom{+} \\ \hline ? \phantom{?} ? \phantom{?} ? \phantom{?} \end{array}$$

If now a student solution (an *observation O*) becomes available, i.e. in the example of Figure 1 the number 1205

$$\text{out}_c(\text{sum}_3) = 1 \wedge \text{out}_r(\text{sum}_3) = 2 \wedge \text{out}_r(\text{sum}_2) = 0 \wedge \text{out}_r(\text{sum}_1) = 5 \quad (6)$$

the complete set of axioms can be used to check the correctness of  $O$ . For that purpose, normality assumptions are kept as long as no counter-evidence is available. A *diagnosis*  $D$  then is a subset of retracted normality assumptions which together with the system description is sufficient to derive the observation

$$S \cup D \models O \quad (7)$$

and is minimal in the sense that there is no proper subset of  $D$  which also allows the derivation of the observation. Any set of abnormality assumptions which can be inferred under this definition can be said to provide a *minimal explanation* for the inconsistency of the axiom system. The empty diagnosis, of course, corresponds to an error-free observation.

In our example this kind of reasoning results in

$$D_1 = \{\text{abnormal}(\text{sum}_2)\}$$

being a diagnosis, since the student obviously was not able to properly add the two digits 3 and 7 in the second column plus a carry over of 1 from the rightmost one. A closer analysis, however, reveals that there is another minimal explanation namely

$$D_2 = \{\text{abnormal}(\text{sum}_1)\}$$

based on the assumption that the student was not able (or forgot) to determine the carry over in the rightmost column properly.

Two important lessons can be learned from this example.

- The diagnostic precision of the chosen axiomatization is fairly low. Although the procedure is able to roughly determine the place where the error occurred, so far it is not in a position to deliver a pedagogically useful diagnosis.
- In spite of its disappointingly weak diagnostic information, even such a simple axiom system systematically produces *alternative error interpretations* for a faulty system.

While I will return to the second issue in the section on limited observability, I would like to discuss the first one in more detail already now, because it touches the fundamentals of domain knowledge acquisition.

Raising the adequacy of diagnostic information for pedagogical purposes always requires the design of a more sophisticated axiomatization of the domain at hand. Two roughly equivalent approaches are at our disposal

- to introduce different fault-modes for the components in our model<sup>1</sup>, e.g. for not being able to add digits and for not being able to handle the carry over properly, or

---

<sup>1</sup>A corresponding axiomatization for a subtraction task is given in (Self, 1992)

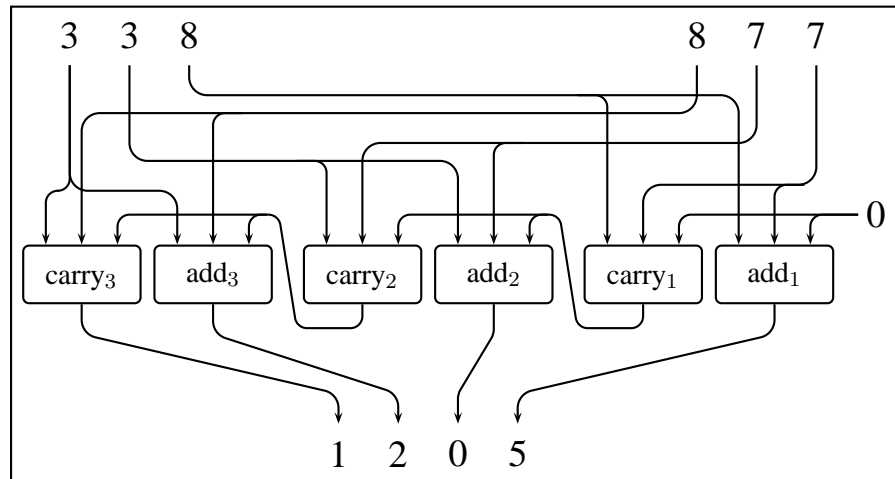


Fig.2. An alternative axiomatization of the addition domain

- to break down components into a more fine-grained description.

Following the second possibility, the bipartite definition for components of type 'sum' as given above already suggest a split into two sub-components for adding digits and for carry-over computation respectively

$$\text{type-3add}(X) \wedge \neg \text{abnormal}(X) \rightarrow \text{out}(X) = (\text{in}_1(X) + \text{in}_2(X) + \text{in}_c(X)) \bmod 10 \quad (8)$$

$$\text{type-carry}(X) \wedge \neg \text{abnormal}(X) \rightarrow \text{out}(X) = (\text{in}_1(X) + \text{in}_2(X) + \text{in}_c(X)) // 10 \quad (9)$$

An appropriate "wiring" for our addition problem is given in Figure 2. Now the two minimal explanations  $D_3 = \{\text{abnormal}(\text{carry}_1)\}$  and  $D_4 = \{\text{abnormal}(\text{add}_2)\}$  more precisely pinpoint the two possible reasons for the error. Note that none of the explanations seems to be more plausible, because (a) the correct handling of the carry over at the leading position indicates unquestionably that the student is well aware of the concept itself, and (b) all additions except the flagged one have obviously been treated properly.

Although the diagnostic ambiguity did not vanish, the new axiomatization provides us with a potentially useful criterion to select among competing error hypotheses. Imagine for that purpose, that now the student solution would have been 1105. In that case we would end up with three alternative diagnoses

$$\begin{aligned} D_5 &= \{\text{abnormal}(\text{carry}_1), \text{abnormal}(\text{carry}_2)\}, \\ D_6 &= \{\text{abnormal}(\text{add}_2), \text{abnormal}(\text{carry}_2)\}, \text{ and} \\ D_7 &= \{\text{abnormal}(\text{add}_2), \text{abnormal}(\text{add}_3)\}. \end{aligned}$$

A criterion for *error selection* could now be applied, based on a consistency assumption: error explanations are preferred if they exhibit a recurrent problem of the student. In the above case this would have



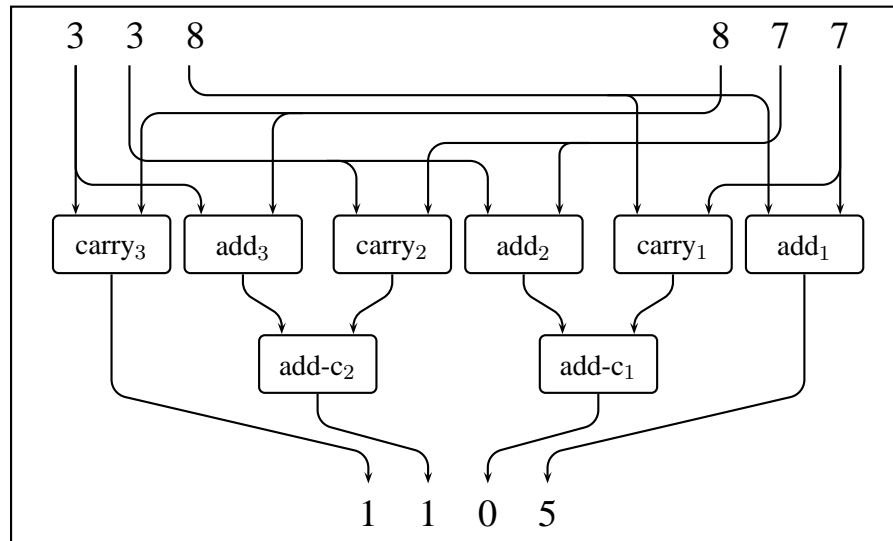


Fig.3. A third axiomatization of the addition domain

been diagnosis  $D_5$ , which addresses a problem of carry-over treatment at any but the leading positions. This possible error cause is even better highlighted in the third conceptualization of Figure 3, which is based on addition components for two digits, and treats the addition of the carry over separately.

$$\text{type-add}(X) \wedge \neg \text{abnormal}(X) \rightarrow \text{out}(X) = (\text{in}_1(X) + \text{in}_2(X)) \bmod 10 \quad (10)$$

$$\text{type-add-c}(X) \wedge \neg \text{abnormal}(X) \rightarrow \text{out}(X) = (\text{in}_1(X) + \text{in}_2(X)) \bmod 10 \quad (11)$$

$$\text{type-carry}(X) \wedge \neg \text{abnormal}(X) \rightarrow \text{out}(X) = (\text{in}_1(X) + \text{in}_2(X)) // 10 \quad (12)$$

Note that under this axiomatization the two types of addition components 'add' and 'add-c' have exactly the same semantics. The difference in their type assignment has only been introduced to better support a hypothesis selection based on a consistent-behaviour assumption. Thus, among the possible diagnoses

$$\begin{aligned} D_8 &= \{\text{abnormal}(\text{carry}_1), \text{abnormal}(\text{carry}_2)\}, \\ D_9 &= \{\text{abnormal}(\text{add-c}_1), \text{abnormal}(\text{carry}_2)\}, \\ D_{10} &= \{\text{abnormal}(\text{carry}_1), \text{abnormal}(\text{add-c}_2)\}, \\ D_{11} &= \{\text{abnormal}(\text{add-c}_1), \text{abnormal}(\text{add-c}_2)\}, \\ D_{12} &= \{\text{abnormal}(\text{add}_2), \text{abnormal}(\text{add}_3)\}, \\ &\dots \end{aligned}$$

there is only a single one (namely  $D_{11}$ ) which shows consistent behaviour for *all* components of a particular type.

As we have seen, model-based diagnosis is also a powerful tool for diagnosing student solutions. Its main advantage consists in only being based on *positive descriptions* of normal component behaviour. All diagnostic information is derived from conflicts between this positive knowledge and a particular observation. In this respect MBD shares its main advantages with CBM in general. In particular, no anticipation of possible errors is necessary.

MBD can always be applied, as long as a suitable decomposition into informationally coupled components can be found. Since the semantics of components is described by constraint formulae, from a superficial viewpoint MBD seems to be a special case of CBM. At this level of detail the main difference between the two approaches obviously consists in the guidelines they provide for specifying constraints. While CBM besides the general condition of independence imposes no further restrictions on form and content of a constraint, MBD requires constraints to precisely mirror the component structure of a domain. Whenever it becomes necessary to emphasize this difference in the sequel we will call the restricted models of MBD also component-based ones.

Since in MBD constraints are restricted to only local input-output relationships of components the space of possible conceptualizations is narrowed down considerably. Nevertheless these guidelines are not strong enough to uniquely determine a model structure in most cases. As we have seen already, different decompositions into model components are still possible.

As long as the observation is a complete and certain one, both approaches show indeed an almost identical behaviour. The fundamental differences only become visible if ambiguity is involved and MBD also has to establish a global consistency of the axiom system in order to be able to explain the observation in terms of component behaviour and error assumptions. On the one hand, it is this additional requirement that allows a model-based axiomatization to be used as an expert module. Being based on a model of correct behaviour it is not only able to derive error descriptions but can also generate correction proposals for the student directly. On the other hand, assumptions on the abnormal behaviour of components become sensitive to their context. This mutual dependence requires the propagation of the consequences of a local error assumption through the whole constraint net to check its compatibility with other assumptions elsewhere.

In contrast, the independence of constraints in CBM eases constraint checking considerably but makes it rather difficult to achieve a comparable diagnostic precision in domains with a high degree of ambiguity. Apparently, there is no fundamental difference between the two diagnostic approaches from a knowledge representation perspective. Both are clearly instances of a constraint-based model. Constraints, however, are differently processed by different constraint evaluation procedures. It is one of the main purposes of this paper to determine which kind of constraint satisfaction procedure will be appropriate for which class of applications.

## **SOURCES OF AMBIGUITY**

To identify why and under which conditions ambiguity is an important issue in a tutoring system, we first review the algebraic exercise of the preceding section again before turning our attention to a series of more challenging problems typically encountered when diagnosing errors in language utterances.

Natural language is notorious for its inherent ambiguity and therefore can serve as an ideal source

of examples to illustrate the difficulties of solving ambiguous diagnosis problems. We will determine four different domain characteristics which turn out to be major sources of ambiguity. Our goal for such an analysis will be threefold;

1. to provide a kind of checklist which can be used to find out whether other application domains will suffer from similar pitfalls,
2. to indicate solution methods which have been successfully applied in the corresponding application domains, and
3. to illustrate how increasingly ambitious system solutions for a complex problem domain like foreign language learning can be built by successively removing the limitations of earlier versions which, of course, have been imposed deliberately to avoid particular types of ambiguity.

### Limited observability

When considering the first axiomatization given in Figure 1 we have seen an explanatory ambiguity turning up, which prevented us from being able to simply compute the most plausible error position. This is the first type of ambiguity we have to consider: given a certain axiomatization one cannot uniquely infer from the data whether a constraint is violated or not. In our case the uncertainty has obviously been caused by the inability to directly observe carry-over values between the three columns of the addition problem. Thus, the two carry overs need to be considered *internal variables* of the model whose values are not restricted by any outside information. In all the examples discussed so far it remains unclear, whether the carry over was completely dropped, or the student actually took it into account, but failed to add it correctly at the position next to the left.

Although this explanatory ambiguity cannot be avoided without seriously modifying the original problem, different modeling approaches allow us to deliberately arbitrate between diagnostic precision on the one hand and diagnostic ambiguity on the other, depending on which kind of domain knowledge can be provided and how effective the available selection criteria are.

The more limited the visibility of certain model information is, the more explanatory ambiguity has to be expected. This can easily be confirmed by comparing the axiomatization of Figure 2 with that of Figure 3, where the doubling of internal variables gives rise to a proportional increase in the output ambiguity.

Proposing to additionally include "slipper" components, Self (1992) certainly introduces an extreme case of limited visibility into the axiomatization. "Slipper" components have to be inserted between the algebraic components proper and the observation, in order to model unpredictable slips of the student

$$\text{type-slipper}(X) \wedge \neg\text{abnormal}(X) \rightarrow \text{out}(X) = \text{in}(X) \quad (13)$$

Thus, arbitrary substitutions of elements of the student solution are licensed. The "real" result, i.e. the one the student actually had in mind, is made highly invisible to an outside observer. Accordingly, a huge number of spurious error hypotheses (exponential in the number of "slipper" components) will be generated by the diagnostic inference, which can only be compensated for by very powerful selection

criteria.

From a tutoring perspective it is important to note that the ambiguity resulting from limited access to system-internal variables prevents instructional feedback to be immediately invoked by individual constraints. In such cases CBM can no longer be considered a student model *per se*. Usually several competing error explanations will be available, which need to be combined with additional evidence from other sources, e.g. additional observations or selection preferences. Although sometimes, as in the addition example above, students can be asked to provide the missing information, in many cases such an interaction renders the exercise task somehow unnatural.

Unlimited observability is only granted, if all conditions inside the satisfaction part of a constraint can be evaluated immediately. That means not only that their truth value can be directly derived from the observation, but also that in combinatorial problem spaces the evaluation should restrict itself to only locally available information. Both conditions are violated in the above mentioned addition exercises, where the most plausible diagnosis does not only depend on the observation of a single component but also on hypotheses about the fault mode of neighbouring ones.

## Polysemy

Polysemy is not only a well known phenomenon in human communication but also happens to be a popular technique in the design of programming languages: one and the same symbol receives several different meanings, among which the actual one has to be selected according to the context in which the symbol appears. Take as an example the '+'-sign, which in addition to its normal semantics for adding numerical values, could also be used to concatenate strings or even append lists.

An artificial kind of polysemy is introduced if "spelling" errors are also considered. In many cases they can be corrected in quite different ways and therefore require to maintain precise assumptions about their origin, which might cause other constraints to fail in quite different ways (Menzel, 2004).

From a diagnostic perspective polysemy makes the analysis problem considerably more difficult, because now two highly intertwined tasks have to be carried out simultaneously: (1) to select the most plausible meaning of the observed symbols (a process usually called disambiguation) and (2) to determine possible constraint violations given the disambiguated meaning.

The notion of polysemy is used here in its widest possible sense and as such is abundant in natural language. For instance, the English word *fish* carries both number features, singular and plural, but is disambiguated if being placed in an appropriate context: *these fish*. On the semantic level words like *bank* denote completely different things and with respect to its pragmatic function one and the same utterance can convey quite different speaker intentions. In the following we will start to consider simple examples from the area of morpho-syntax and try to extend the approach to more ambitious problems later on.

Even though in most cases ambiguous word forms are sufficiently disambiguated by their context, their ambiguity may easily resurface if a diagnosis of erroneous constructions is attempted. Consider the example sentence

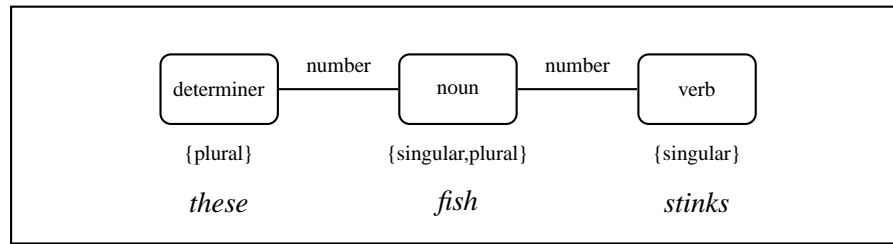


Fig.4. A component-based model for a simple agreement problem

*These fish stinks.*<sup>2</sup>

Here we are faced with an agreement problem involving two components for number agreement, which has to hold between determiner and noun on the one hand ('number<sub>det-noun</sub>') and noun and finite verb on the other ('number<sub>noun-verb</sub>'). Using a set-based representation for feature handling, set membership can be used to denote the assignment of (possibly alternative) feature values. Thus, the number slot of the word form *fish* is mapped to the feature set '{singular,plural}'. Such a representation is not only advantageous for an efficient implementation by means of *constraint propagation techniques*, but also facilitates an alternative conceptualization as discussed in the following section.

For the simple agreement problem two components are required, which are informationally coupled at the noun (c.f. Figure 4)<sup>3</sup>

$$\text{type-number}(\text{number}_{\text{det-noun}}) \wedge \text{type-number}(\text{number}_{\text{noun-verb}}) \quad (14)$$

$$\text{out}(\text{number}_{\text{det-noun}}) = \text{in}(\text{number}_{\text{noun-verb}}) \quad (15)$$

For agreement to hold between two word forms the following conditions have to be satisfied:

- the intersection of the two feature sets at the input and the output of a component has to be non-empty, and
- all values not in the intersection have to be excluded from further consideration in other (neighboring) constraint evaluations.

The latter condition takes into account, that agreement is a transitive relationship and the procedure for constraint checking needs to compute the transitive closure. This is ensured by introducing an update

<sup>2</sup>English is morphologically a rather poor language, where "speaking" examples are hard to find. We will discuss more plausible applications in other languages later on.

<sup>3</sup>To simplify the presentation of more complex models for all other natural language examples, from now on components are represented as arc labels, whereas boxes correspond to variables, i.e. word forms filled in from either the exercise context, or the student response.

semantics into our axiomatization (which is needed for an efficient implementation as a constraint propagation procedure anyhow), thus changing the nature of constraints from a passive condition checker to a mechanism of active value assignment

$$\text{type-number}(X) \wedge \neg \text{abnormal}(X) \rightarrow \text{in}(X) := \text{out}(X) := \text{in}(X) \cap \text{out}(X) \neq \emptyset \quad (16)$$

The observation is derived from the input sentence by means of a dictionary, which maps input word forms to logical combinations of morpho-syntactic features. Replacing word forms by their dictionary entries we come up with the following value assignments<sup>4</sup>

$$\begin{aligned} \text{in}(\text{number}_{\text{det-noun}}) &= \{\text{plural}\} \wedge \text{out}(\text{number}_{\text{det-noun}}) = \{\text{singular}, \text{plural}\} \\ &\wedge \text{out}(\text{number}_{\text{noun-verb}}) = \{\text{singular}\} \end{aligned} \quad (17)$$

This example illustrates that the transitive nature of agreement is indeed crucial for an adequate behaviour of the diagnosis: both agreement conditions in Figure 4 are satisfied locally (either in the singular reading or the plural one). Global consistency, however, can only be established assuming one of two alternative error hypotheses

$$\begin{aligned} D_{13} &= \{\text{abnormal}(\text{number}_{\text{det-noun}})\} \\ D_{14} &= \{\text{abnormal}(\text{number}_{\text{noun-verb}})\} \end{aligned}$$

Note that any attempt to exclude certain interpretations from the space of alternatives *a priori* does not result in a satisfactory solution of the diagnosis problem, since the different error interpretations usually have different degrees of plausibility in different contexts. In our example the plausibility probably will depend on how many fish have to be considered in a particular scenario.

Unfortunately, the necessity to establish global consistency has an immediate consequence: the outcome of a constraint violation no longer can be determined locally, but also needs to consider the outcome of constraint applications at neighboring components. Since local agreement is not a sufficient condition to infer the absence of an error, all possible combinations of error hypotheses have to be checked for global plausibility, which in complex constraint systems might already incur a substantial computational effort.

Obviously the existence of word forms with different morpho-syntactic readings confronts us with the same type of ambiguity as we have already observed in the addition example: considering a component-based constraint in isolation we cannot determine uniquely whether it holds or not. However, this uncertainty has quite different reasons now. While it was caused by a limited observability in the addition example, it is an inherent property of the observable data in the agreement task. In a broader sense, morpho-syntactic ambiguity of word forms could also be considered a special case of limited observability: although word forms are fully visible to the diagnostic procedure, the value assignments established by them are not. In contrast to the carry over of an addition exercise, the intended selection of a grammatical feature can not easily be elicited from the student without heavily disturbing the natural conditions

---

<sup>4</sup>Although agreement relations are in fact symmetrical, we keep the traditional notation and refer to the two connection points of a component as 'in' and 'out'.

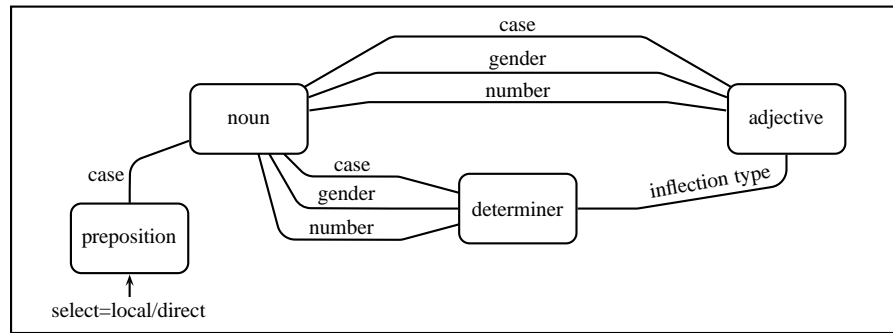


Fig.5. A component-based model for the agreement within a German prepositional phrase

of human language production.

Of course, alternative axiomatizations to the ones suggested by MBD are possible. As long as we are concerned with rather simple agreement problems as the one in Figure 4, where only a single two-valued feature is involved, the usual local constraints

$$\begin{aligned} & \text{agree}(\text{determiner}, \text{noun}, \text{number}) \\ & \text{agree}(\text{noun}, \text{verb}, \text{number}) \end{aligned} \quad (18)$$

could be complemented by an additional non-local constraint like

$$\text{ambiguous}(\text{noun}, \text{number}) \rightarrow \text{agree}(\text{verb}, \text{determiner}, \text{number}) \quad (19)$$

It requires verb and determiner to also agree with respect to number if there is a number ambiguity at the noun.<sup>5</sup> Apparently the advantage of this alternative modeling consists of all constraints being completely independent from each other. Thus, they can be checked in isolation and, if violated, directly invoke an appropriate error message. Unfortunately, the approach does not provide a general solution to the ambiguity problem, since it simply moves the difficulties from the algorithmic level into the composition of the constraint set. It fails already if slightly more complex agreement problems need to be considered.

The reason is that MBD always establishes global consistency for a particular value assignment and a particular set of error assumptions. If global consistency is abandoned, diagnosis results in incomplete and often misleading error descriptions. To avoid this problem when designing independent constraints would require to combine local constraints into global ones taking *all combinations* of component failures for *the same value assignment* into account. Only then can the necessity to check the satisfaction part independently from the outcome of other constraint applications be properly addressed. We will discuss a similar axiomatization within the MBD framework in the next section.

<sup>5</sup>T. Mitrovic, personal communication.

Situations in which global consistency is crucial but cannot be easily modeled with independent constraints occur frequently in languages with a fairly rich morphology. Figure 5 shows an example for the correctness conditions of a German prepositional phrase. The model is based on agreement components and similarly defined ones for value restriction, which are needed to model e.g. case government. It also includes selectional restrictions imposed by the verb, thus distinguishing local PPs from directional ones. For an utterance like

<i>Es</i>	<i>liegt</i> local	<i>hinter</i> local,dat direct,acc	<i>der</i> masc,nom,sg all,gen,pl	<i>Tische</i> masc,nom,pl masc,gen,pl masc,acc,pl
-----------	-----------------------	--	---	--

*(It is lying behind the tables)*

a MBD produces the correct double error

$$D_{15} = \{\text{abnormal}(\text{case}_{\text{prep-noun}}), \text{abnormal}(\text{case}_{\text{noun-det}})\}.$$

whereas the axiomatization with the above given non-local constraint is not even able to detect an error, since the three word forms agree pairwise but, unfortunately, not all three together. Similarly Holland (1994) complains that typical foreign language tutoring systems diagnose a sentence like

<i>*Wir</i>	<i>stehen</i> local	<i>auf</i> local,dat direct,acc	<i>die</i> femin,nom,sg femin,acc,sg all,nom,pl all,acc,pl	<i>Berg.</i> masc,nom,sg masc,dat,sg masc,acc,sg
-------------	------------------------	---------------------------------------	--	---

*(We are standing on the mountain.)*

into a single error description

$$D_{16} = \{\text{abnormal}(\text{gender}_{\text{det-noun}})\}.$$

which is clearly inappropriate, because the verb requires a local prepositional phrase, which in turn translates into a dative case requirement for the noun and subsequently also for the determiner. Simply replacing the determiner with the corresponding masculine form wouldn't help either because it triggers a contingent case error. Therefore, a double error assumption is required to describe the situation correctly to the student. Again, the problem is caused by either an inadequate model or an insufficiently general procedure for constraint solving in domains with a high degree of ambiguity. If the general CBM approach should be maintained, but the loss of diagnostic precision cannot be tolerated, constraints will become extremely complicated.

The example of a German prepositional phrase can serve as a good illustration for the degree of sophistication which can be achieved by means of a purely feature-based approach. Although a considerable degree of flexibility is already available, exercises are still restricted to the particular phrase type



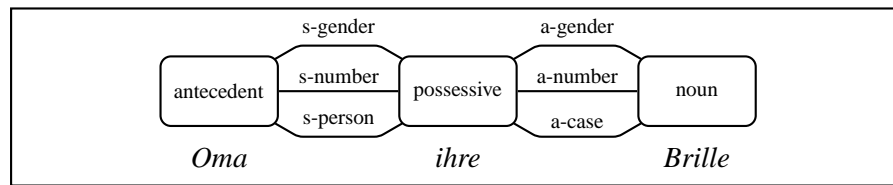


Fig.6. A model for exploring the use of German possessive pronouns

for which the model has been designed. Therefore the solution is still far from what a natural communication would require. Nevertheless, even simple dialogues can be simulated already, as long as the student is made aware of the system limitations

- System: *Wo hast du Omas Brille gefunden?*  
*(Where did you find grandma's glasses?)*
- Student: *Auf dem kleinen Tisch.*  
*(On the small table.)*

In limited domains, such a model can be made fairly "water-proof" which even suggests inviting the student to experiment with it, i.e. to explore the space of possibilities with the goal of inferring the underlying regularities. A good example for such a domain is the correct use of the German possessive pronoun, which requires obeying different agreement constraints with the antecedent *and* within the corresponding noun phrase (stem vs. affix inflection). Figure 6 shows an outline of the model structure. It can easily be integrated into a slot-filling exercise like

- System: *Oma* <sub>antecedent</sub> *sucht* ... *Brille* <sub>governor</sub>.  
*Grandma is looking for ... glasses.*

where only the possessive can be chosen by the student, whereas the other two variables (governing noun and antecedent) are controlled by the system.

Obviously, the application of a feature-based diagnosis is not limited to the agreement between complete word forms. Word forms can be decomposed along their morphological structure and the combinability of morphemes is again modeled by means of feature agreement components. This way at least parts of the inflectional regularities of a language can be included into the range of possible exercises. Irregular inflectional forms like *\*goed* can be analysed and properly explained to the student. This leads to interesting applications in inflectionally rich languages, e.g.

- missing stem inflection (umlaut) in German

*die Gabel* ... *(mit) den Gabeln* (*the fork/with the forks*)  
*der Apfel* ... *\*(mit) den Äpfeln* (*the apple/with the apples*)  
 → 'Äpfeln' requires umlaut in plural

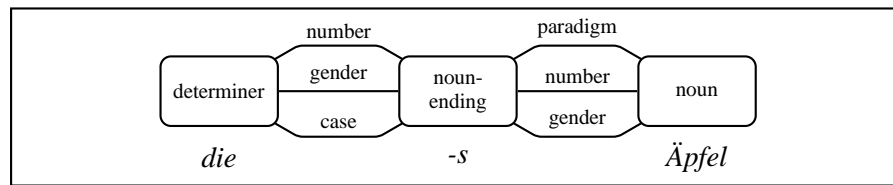


Fig.7. A component-based model for German noun inflection

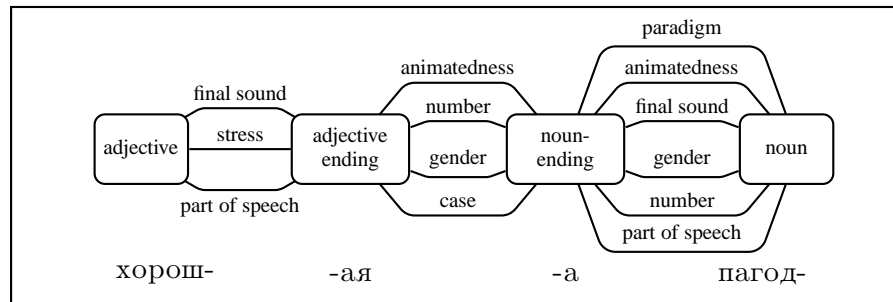


Fig.8. A component-based model for Russian noun phrase inflection

- inappropriate choice of gender specific inflectional paradigms in German

*die Schachtel* ... *die Schachteln* (the box/the boxes)  
 \**die Apfel* ... \**die Äpfeln* (the apple/the apples)  
 → 'Apfel' is masculine, not feminine

- phonologically inappropriate inflectional paradigms in Russian

новый дом but большой дом (new house/big house)  
 \*большый дом  
 → большой has word stress on the ending, which therefore must change from -ый to -ой

- phonetically inappropriate inflectional paradigms in Russian

новый дом but маленький дом (new house/small house)  
 \*маленький дом  
 → the root маленьк- ends with a hard consonant but the ending -ый can only be used with a soft one

The corresponding models are given in Figure 7 and 8 respectively.

The approach can easily be extended to also include ordering requirements (linear precedence: lp) and even dominance restrictions like obligatoriness ( $\wedge$ ), conditional obligatoriness ( $\rightarrow$ ), optionality ( $\vee$ ),

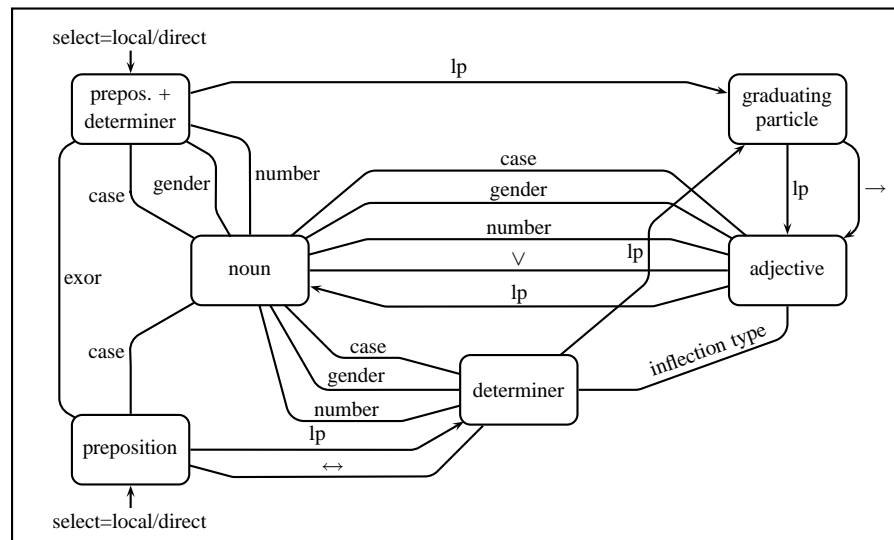


Fig.9. A component-based model for German prepositional phrases

mutual incompatibility (exor), or mutual necessity ( $\leftrightarrow$ ) of constituent components (Menzel, 1990). Figure 9 shows a corresponding model for a German prepositional phrase. The additional constraints contribute to a considerably increased syntactic flexibility for the student, which now covers

- optional adjectival modifiers: *unter dem [linken] Stuhl* (*under the [left] chair*)
- optional modification of adjectives: *unter dem [ganz] linken Stuhl* (*under the left[most] chair*)
- nominalization of adjectives: *unter dem linken* (*under the left one*), and even
- phonological contraction: *unterm Stuhl* (*under the chair*),

and make possible the diagnosis of a broad range of additional error types, among them reorderings, missing obligatory elements, etc.

### Alternative conceptualizations of domain knowledge

Diagnosing morpho-syntactic errors by means of agreement components results in error hypotheses reflecting a particular view on an erroneous situation: if a certain constraint has been found to be violated, this diagnostic result corresponds to the assumption that the student did not obey (or was not aware of) the underlying grammatical regularity, i.e. made a *rule error*. Thus, the system presumes a lack of (applicable) generalised grammar knowledge. There exists, however, a complementary view on agreement errors, which considers the student being fully aware of the underlying constraints, but instead making wrong assumptions on the actual lexical value assignments in the dictionary of a language. Under

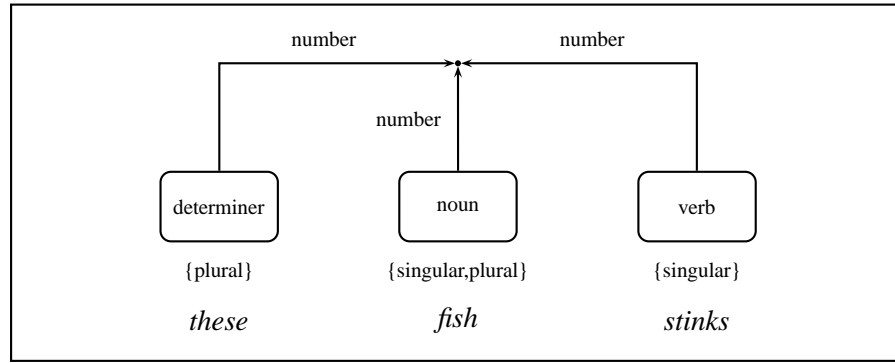


Fig.10. A component-based model for factual errors diagnosis of agreement problem

this perspective the student committed a *fact error* by perhaps not knowing that the word form *these* is actually plural, not singular.

Both perspectives address completely different cognitive problems of the student, require specific instructional actions to be taken and are differently plausible in different exercise contexts. This introduces a second important class of ambiguity into the diagnostic problem. Even if a unique value assignment would allow the violation of a constraint being determined uniquely, we end up with several competing error descriptions reflecting different points of view.

Like rule errors, their fact-based counterparts can also be diagnosed using the model-based approach. They require, however, components which select a *subset* of the features from the dictionary to be checked for equality

$$\text{type-number-lex}(\text{number}_{\text{det}}) \wedge \text{type-number-lex}(\text{number}_{\text{noun}}) \wedge \text{type-number-lex}(\text{number}_{\text{verb}}) \quad (20)$$

$$\text{out}(\text{number}_{\text{det}}) = \text{out}(\text{number}_{\text{noun}}) = \text{out}(\text{number}_{\text{verb}}) \quad (21)$$

$$\text{type-number-lex}(X) \wedge \neg \text{abnormal}(X) \rightarrow \text{out}(X) := \text{subset}(\text{in}(X)) \quad (22)$$

The observation is identical to that of the original modeling. Figure 10 shows the corresponding model layout. Note that the choice of a subset from a given set is a non-deterministic operation, which again requires search to find the optimal combination of component retractions. Such an approach is only feasible because the selection is controlled by the very strong equality requirement from the connectivity axioms.

Choosing appropriate value subsets amounts to determining a particular value assignment among the possible ones. Thus morpho-syntactic ambiguity is successively removed during the course of computation. Having eventually a unique assignment available, would even allow unrestricted constraints in the original CBM sense to be applied. However, the advantage of a simpler constraint application procedure is traded for the necessity to enumerate all possible value combinations.

Fact-based diagnoses for lexical value assignments differ from rule-based ones, primarily in that they are more closely connected to a correction possibility. They are therefore perfectly suited to precisely identify those positions in a student's utterance where a repair would be necessary in order to remove a constraint violation. While a single rule error in many cases can be repaired at both connection points of the violated agreement component, a factual error always pinpoints a single word form, which needs to be exchanged. Consider, for instance, the English noun phrase *\*this apples*, which causes a single rule error diagnosis (missing number agreement) but can be repaired in two different ways: *these apples* or *this apple*. These two options are then directly expressed by means of two alternative fact error descriptions. Hence, fact diagnoses are also inherently ambiguous and therefore cannot be transformed into feedback immediately.

Since it is strictly correction oriented, an axiomatization based on lexical value assignments turns out to be an ideal choice for an expert module: in addition to being able to precisely determine possible positions for a repair, the feature assignments inferred by the diagnostic procedure can be used directly to retrieve appropriate word form substitutions from the dictionary. Therefore, correction possibilities can be offered to the student, which transform her solution into a correct one using the lexical choice of the original utterance. Thus, model-based diagnosis of lexical value assignments combines four major advantages of constraint-based modeling;

- It offers the student a high degree of flexibility to creatively construct individual exercise solutions within the limits of the model and the dictionary.
- It provides diagnoses which closely reflect the correction-oriented viewpoint of a student.
- It is capable of a precise error localization.
- It produces diagnostic results precise enough to generate correction proposals from them on demand.

Accordingly, diagnoses of lexical value assignments seem to optimally meet the requirements of training scenarios for second language acquisition. These advantages, however, do not obviate the necessity of rule-based diagnoses, which due to their grammar-centered perspective in some cases much better reflect the general picture of a constraint violation by e.g. being able to summarize a well-formedness condition for a complete phrase. If, for instance, the utterance

*auf die kleinen Tische (onto the small tables)*

which happens to be a perfectly correct prepositional phrase for a directional adverbial (*Putting something there.*), has been used by mistake in an environment which requires a local one (*Something being there.*) the error can be described to the student either by means of at least two simultaneous fact-error explanations

The determiner and the noun have wrong case (accusative instead of dative).

which correspond to the two necessary corrections (*die* → *den* and *Tische* → *Tischen*). A rule-based explanation, on the other hand, can express the same information in a more general way

The prepositional phrase is of wrong case (accusative instead of dative),  
i.e. it is a directional instead of a local one.

because both the preposition and the adjective are ambiguous with respect to the dative/accusative distinction.

If desired, both axiomatizations could even be combined within a single model. Experience has shown, however, that such a merge not only results in an unacceptably high number of competing error hypotheses, but also that mixed diagnoses consisting of rule and fact errors are usually too difficult to be understood by a language learner (Menzel, 1992).

### **Structural uncertainty**

Up to now, all modeling was done based on the simplifying assumption that components and their types can be determined unambiguously *prior* to the diagnosis itself. While the algebraic exercises used above satisfy this precondition in a natural way, natural language training requires restricting the possible tasks somewhat artificially for that purpose. Similar approaches have also been used for dealing with programming languages. Although their syntax is highly deterministic and the degree of polysemy rather low, programming languages still offer a great number of solution possibilities to a particular problem. In the SQL-tutor (Mitrovic, 1998), for instance, a query is prestructured by means of a specifically designed user interface, where the student has to fill in her solution into distinct slots for the different clauses.

If we now talk about natural language in a more unrestricted sense, however, such methods are not easily applicable. Human language (even on a beginner's level) is characterized by a great variety of lexical *and* structural formulation alternatives, and a tutoring system must be able to deal with them appropriately, if a communicatively meaningful use of language is aspired to. This variety makes the disambiguation problem even more challenging: in addition to selecting among alternative value assignments for the components of a diagnostic problem, possible ways of combining them need to be decided upon.

Unfortunately, even the very existence of a component (e.g. for agreement) hinges upon a syntactic relationship being established between particular word forms. So, for instance, the head noun of a noun phrase has to be nominative and must agree in number and person with the finite verb of a sentence, if it is meant to be its subject. In case of an object, however, no agreement is required at all, but instead the verb may govern the case of the noun phrase (usually accusative or dative). Hence yet another type of ambiguity is introduced into the diagnostic problem: in addition to ambiguous value assignments and different diagnostic perspectives now we have to deal with the problem that not even the relevance of a constraint can be determined uniquely!

Computing the syntactic structure of a sentence, called *parsing*, is not a trivial task. Especially for languages with a relatively free word order, like German or Russian, it relies heavily on the very same conditions we would like to retract during our model-based diagnosis. Thus we seem to be trapped in a vicious circle: to diagnose constraint violations in an utterance, we need its syntactic structure, but we cannot determine the syntactic structure, if we permit arbitrary constraint violations. Any attempt to do so would inevitably lead to an unmanageable explosion in the size of the hypothesis space. This vicious circle can only be broken if both processes (parsing and constraint retraction) are integrated into a unified solution procedure and additional constraining information from independent sources is made available to compensate for the loss of guiding information in case of constraint violations.

Such an integration can be achieved by means of a parsing system based on *weighted constraints*. Instead of the binary retraction scheme we used so far, now a weight describes the *degree* to which a constraint violation can be accepted by the system. Constraints are not directly used to describe the possible combination of word forms in the observation, but restrict the space of structural interpretations (i.e. trees) into which the observations can be organized. Given a natural language utterance, the system is expected to determine (1) its optimal syntactic structure and (2) the constraints violated by the optimal structure, where optimality is defined again by constraint violations. Parsing, therefore, becomes a *constraint optimization procedure*, which tries to find a structural interpretation for the input sentence violating as few and as least important constraints as possible (Foth et al., 2005a) (Foth et al., 2005b) (Schröder et al., 2000).

Among the available solution methods for constraint optimization transformation-based algorithms have been most successful. Starting from an initial dependency structure (usually one that satisfies at least the unary constraints) they try to repair constraint violations by successively modifying the current structural interpretation. The search is guided by the score, the available diagnostic information, and the effort spent so far. Although the optimal solution cannot be guaranteed, such procedures obtain good approximations in many cases (Daum et al., 2003). Their particular advantage is the availability of a full structural description at an arbitrary point in time which can be improved further in subsequent repair steps. The computation can be terminated if no further improvement is expected or a prespecified time limit is exceeded.

To determine the optimal structure a weight aggregation scheme has to be defined, which in our case will be the product of constraint weights for all constraints assumed to be violated by a structure. Under this multiplicative combination scheme a number of constraint violations for comparatively weak constraints can override a relatively strong one and *vice versa*. Note, however, that absolute constraints with a weight of zero cannot be retracted at all.

The availability of gradually retractable constraints allows the system developer to differentiate between knowledge sources of different strength. As usual, there are normal well-formedness conditions like agreement, government, valency requirements, and strong linear ordering requirements, which should be enforced by the system and reported to the student if violated. Additionally, various kinds of *preferences* can be incorporated, which are helpful (1) to guide the search for the optimal interpretation, (2) to decide in case of otherwise unresolvable ambiguity, and (3) to arbitrate between different sources of evidence. This way, a huge body of *uncertain knowledge* is exploited for disambiguation, which in a

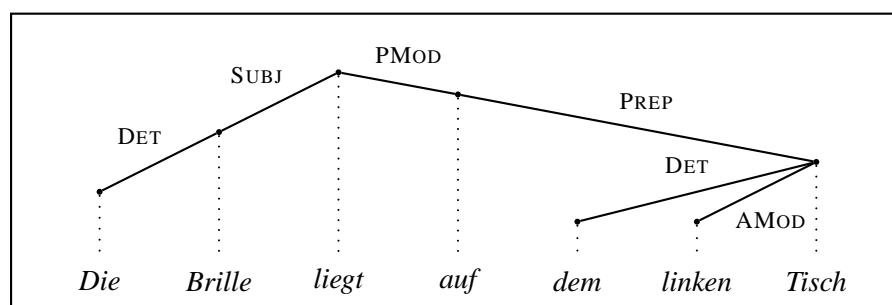


Fig.11. A dependency tree for a simple German sentence (*The glasses are on the left table.*)

crisp grammar without weights would not be available.

Constraints are defined over dependency relations holding between two word forms and relations are annotated with a label (cf. Figure 11). A suitably defined set of constraints forms a *weighted constraint dependency grammar* (WCDG), which assigns a dependency structure to the word forms of a sentence.

For reasons of efficiency constraints should be restricted to only access locally bounded substructures consisting of at most two dependency edges (i.e. four word forms) simultaneously. As usual constraints consist of a relevance part and a satisfaction part. Retractability is not explicitly expressed, because it can be indirectly inferred from the constraint weight.

In addition to the use of preferences, separate description levels can be established, as another way to compensate for the weak restrictive power of retractable constraints. Thus it becomes possible to model relationships originating from semantics, world knowledge, and the specifics of the application scenario independently of the syntactic structure. Defeasible mapping constraints facilitate a (bidirectional) information flow between the description layers. In such an architecture world knowledge can be used immediately during parsing without, however, making parsing depend too strongly on the availability of this knowledge under all circumstances.

Here, the exploitation of weighted constraints is especially important, because the syntax-semantics interface is dominated by strong preferences, which sometimes might be deliberately violated by a speaker to achieve a certain communicative effect. This happens e.g. in case of selectional restrictions, which usually are a powerful means to disambiguate an utterance, especially if it contains syntactic errors. However, sometimes they might need to be retracted e.g. in case of metaphorical language use: e.g. only living beings can drink, but sometimes a car is said to be drinking as well.

Constraints in WCDG are strictly passive. They only check whether the satisfaction part is fulfilled, whenever a dependency edge or a pair of them satisfies the relevance part. No value assignments can be carried out and no structure building as in unification grammars is possible. In this respect constraints in WCDG again resemble those of CBM. They differ, however, in that (1) WCDG constraints are weighted, and (2) they are not evaluated directly on properties of the student input, but on a huge space of structural hypotheses derived from it.



WCDG-constraints consist of five components delimited by colons

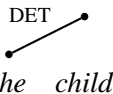
- an application pattern which can be seen as belonging to the relevance part, since it specifies whether the constraint is defined for a single dependency edge (unary: ' $\{X:SYN\}$ ') or two edges (binary: ' $\{X:SYN, Y:SYN\}$ ') and the level the constraint has access to (here in both cases the syntactic description),
- a name, which is used to uniquely identify the grammar regularities violated by the student,
- a class, which the constraint belongs to,
- the constraint weight, i.e. a number between zero and one, and
- a logical formula describing the structural requirements for well-formed utterances.

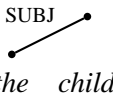
A constraint has access to the lexical and positional information at both the dominating word form of an edge (' $X\uparrow$ category') and the dominated word form (' $X\downarrow$ category'). Moreover, the label of a dependency edge can be retrieved using the function ' $X.label$ '.

Constraints can roughly be divided into three classes

- *hard constraints* which ensure that elementary requirements of syntactic modeling are obeyed, e.g. a unary constraint licensing the modification of a noun by a determiner with label 'DET'

$\{X:SYN\}: \text{det\_noun\_modification\_1} : \text{np} : 0.0 :$   
 $X\uparrow\text{cat}=\text{noun} \wedge X\downarrow=\text{det} \rightarrow X.\text{label} = \text{DET}$

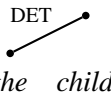
allowed: 

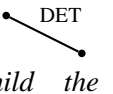
disallowed: 

- *error constraints*, which can be violated by the student, e.g.

- a unary constraint restricting the position of a determiner to positions left of its head noun

$\{X:SYN\}: \text{det\_noun\_modification\_2} : \text{np} : 0.1 :$   
 $X.\text{label} = \text{DET} \rightarrow X\downarrow\text{pos} < X\uparrow\text{pos}$


allowed: 


penalized: 

- a binary constraint, making determiner and adjective within a noun phrase agree with respect to their number feature

{X:SYN,Y:SYN}: np\_number\_agreement : np : 0.2 :

$X.\text{label} = \text{DET} \wedge Y.\text{label} = \text{AMOD} \wedge X\uparrow\text{pos} = Y\uparrow\text{pos} \rightarrow X\downarrow\text{number} = Y\downarrow\text{number}$


allowed:   
*das kleine Kind*  
*the small child*


penalized:   
*das kleinen Kind*  
*the small child*

- or a binary constraint selecting the adjective inflection (weak vs. strong) according to the type of determiner used

{X:SYN,Y:SYN}: np\_inflection\_type : np : 0.2 :

$X.\text{label} = \text{DET} \wedge Y.\text{label} = \text{AMOD} \wedge X\uparrow\text{pos} = Y\uparrow\text{pos} \rightarrow X\downarrow\text{itype} = Y\downarrow\text{itype}$

allowed:   
*ein kleines Kind*  
*the small child*

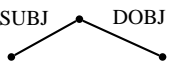
penalized:   
*ein kleine Kind*  
*the small child*

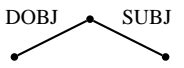
- *preferential constraints*, guiding the search by weak ordering regularities, e.g.

- a weak constraint for the tendency of the subject to precede the object in German clauses

{X:SYN,Y:SYN}: subj\_obj\_preference : clause : 0.7 :

$X.\text{label} = \text{SUBJ} \wedge Y.\text{label} = \text{DOBJ} \wedge X\uparrow\text{pos} = Y\uparrow\text{pos} \rightarrow X\downarrow\text{pos} < Y\downarrow\text{pos}$

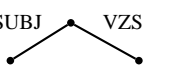
preferred:   
*Ich trinke Milch*  
*I drink milk*


dispreferred:   
*Milch trinke ich*  
*Milk drink I*

- or a distance constraint, which prefers short attachments (here for the detachable German verb prefix) over longer ones by calculating the constraint weight dynamically depending on the current structural configuration

{X:SYN}: short\_attachment\_pref : clause :  $0.9 * (1 - 1/\text{abs}(X\downarrow\text{pos} - X\uparrow\text{pos}))$  :

$X.\text{label} = \text{VZS} \rightarrow \text{abs}(X\downarrow\text{pos} - X\uparrow\text{pos}) = 1$

preferred:   
*Ich reise ab*  
*I depart*

dispreferred:   
*Ich reise heute ab*  
*I depart today*

Usually, even the optimal structure of a correct sentence will contain many constraint violations. They are mainly caused by weak ordering and distance constraints and need not be reported to the student. An instructional action need only be initiated for constraints which are explicitly marked as error constraints.

Due to the limitations imposed upon the power of constraints, a WCDG cannot be used as an expert model, which would need to be able to generate correct and meaningful sentences in response to an exercise. Although constraints are strong enough to diagnose a broad range of grammatical phenomena, they are far too weak to effectively restrict the vast space of possibilities resulting from all the available lexical items in all their possible combinations. For a similar reason, constraints of a WCDG do not represent a universal means for diagnosing arbitrary errors in natural language utterances. Especially their local nature prevents them from being able to check the closure of a transitive relationship, an ability which is a particular strength of MBD. Moreover, a WCDG is restricted to the diagnosis of rule-violations, and therefore does not support the computation of different error perspectives (retraction of rules vs. retraction of lexical facts).

To overcome these serious limitations, we have coupled the WCDG-based parser with a component for MBD in a three stage architecture (Menzel and Schröder, 1998b) (Stockfleth, 2000), where

- in the first stage the WCDG-based parser tries to obtain a structural description of the input sentence together with some classes of constraint violations (most notably, missing or superfluous material and linear precedence violations),
- in the second step a constraint net for model-based diagnosis is generated from the optimal structural description of the input sentence, and
- the third step carries out a model-based error diagnosis, paying particular attention to a proper treatment of transitive condition chains, and a full account of different error perspectives.

All diagnostic results obtained in stage one and three are then subjected to an error explanation component, which tries to select the most plausible hypotheses and decides about possible instructional actions to be taken.

## **HYPOTHESIS SELECTION**

Polysemy, alternative error perspectives, and structural uncertainty are the most important sources of ambiguity a component for grammar diagnosis is confronted with. Together they contribute to a great variety of diagnostic results, which even for short erroneous sentences can easily amount to several dozens of alternative error interpretations from which a small subset has to be selected which finally can be reported to the student. The problem of error selection happens to be even more complex, since each hypothesis in turn may consist of a combination of several elementary error descriptions, which in many cases cannot easily be separated from each other, without potentially misleading the student.

To deal with these problems, we have implemented an error selection based on two main principles: error explanations need to be as *plausible* and as *comprehensible* as possible. Neither of these criteria is directly measurable. Hence, they have to be approximated.

Since comprehensibility is highly influenced by the simplicity of error descriptions, the general subset-based selection criterion of model-based diagnosis can be reinforced to a cardinality-based one, resulting in the most simple error explanations being preferred. Since we have seen in the section on alternative conceptualizations that sometimes a single rule violation corresponds to a combination of two fact-error descriptions, the minimality criterion should be applied to the different error perspectives separately.

Special care has to be taken if the minimality criterion is applied to fact error diagnoses. Sometimes fact error descriptions do not correspond to an easily available word substitution, e.g. if a diagnosis requires changing the gender of a noun. In such cases, error explanation will need to resort to alternative error hypotheses, even if they are not minimal in the strong sense.

Even if selection is restricted to the most simple error explanations, it might end up with several competing interpretations. In such cases other criteria will be required to further narrow down the space of alternatives. Among the easily available ones is a preference for more plausible explanations which can be based on the likelihood of an error type. Such a likelihood can be estimated and exploited in quite different ways depending on how much information is available about the student

- a general likelihood for all kinds of prospective students. Among language teachers e.g. the likelihood of a gender problem is usually considered much higher than one of case or number, since gender is of lexical origin and almost arbitrary. It therefore needs to be learned individually for each new noun, while other categories can also be derived from more general heuristics.
- a likelihood specific to the mother tongue of the student, which is increased e.g. for the gender feature, if the gender of a noun in the first and second language deviate from each other.
- a student-specific likelihood, which needs to be derived from a long-term observation of the student's performance and her individual learning problems.

Now, student modeling in the narrower sense comes into play. The information collected about the student can for instance be used to tailor the hypothesis selection to the proficiency level of the student. Alternatively, even the weights of the WCDG can be adjusted, in order to better reflect the individual habits of a particular student.

How fine-grained the information about a student could be, depends mainly on the available amount of individual observations. Although e.g. conditioning the likelihood of error types on their lexical or syntactic environment would certainly be desirable, it obviously incurs a substantial knowledge acquisition bottleneck, because hardly enough observations will be available to estimate meaningful probabilities therefrom, as would be required for such a detailed model.

All selection strategies based on grammar-specific criteria share a common drawback: they try to determine the most plausible error hypothesis still relying on a kind of similarity-based comparison metrics. Even though in the case of a likelihood-based selection similarity is biased by plausibility estimations, the selection criterion is still geared towards minimum cardinality hypotheses. As a consequence, hypothesis selection always attempts to find an error explanation, which transforms the student solution



Fig.12. A static micro-world used in a (web-based) language-driven computer game.

into the *most similar one* among all the correction possibilities. In many cases this also corresponds to a proposal for repair with minimal effort.

Unfortunately, sometimes such a strategy is misleading, because it completely ignores the *communicative intent* of the student. Consider again a situation where an utterance can be transformed both into singular or plural (c.f. the examples for polysemy as a source of ambiguity). Which of the two alternatives should be proposed to the student by no means depends on the error type or its individual likelihood, but rather on the exercise environment, i.e. how many objects the student is likely to talk about. A similar problem occurs whenever the minimum error explanation suggests exchanging the gender of a noun, which only in exceptional cases is possible without completely changing the propositional content of an utterance.

Certainly, the most simple approach to control the communicative intent of a student is to provide her with a picture of a particular scenario (micro-world), like the one in Figure 12, taken from a prototype implementation for a web-based language learning system. Here, the student is asked to describe the subject to a blind artist in order to find out a falsification among a series of paintings. In our example the student's input sentence was *Auf der Tisch liegt die Banane*. (*The banana lies on the table*). Master Albrecht, the painter, tries to remember if, according to his memories, he ever painted such a still. The parrot comments on possible language errors, here pointing out the place of the error in the student's utterance: *der Tisch*.

The knowledge underlying such a picture can not only be used to determine the truth value of the student's proposition (in terms of Master Albrecht's memories) but also to narrow down the space of possible diagnoses. This is achieved by either selecting plausible hypotheses, e.g. in case of cardinality restrictions like  $\text{card}(\text{pear}, 2)$ ,  $\text{card}(\text{apple}, 1)$  etc. as described above, or by integrating the ax-

ioms as additional constraints into the WCDG parser itself, establishing a separate knowledge representation level, which is initialized with appropriate spatial relationships like `behind(bottle, glass)`, `left_of(apple, pear2)` etc. (Menzel, 1998). Earlier experiments have shown that, if this additional representation level is appropriately coupled with the syntactic one, the additional evidence might help to guide the diagnosis towards the most plausible interpretation using the pictorial information from the knowledge base (Menzel and Schröder, 1998a).

Obviously, by providing just a static micro-world the communicative intent of the student can only be inferred approximately. If a student feels the desire to develop the given scenario, she might want to signal her intentions non-verbally. For that purpose, an additional communication channel can be established by means of an interactive graphical environment (Hamburger, 1995). Available non-verbal clues can then be dynamically integrated into the WCDG parsing to bias the search for an optimal set of constraint violations based on the probably intended propositional content.

If in cases of insufficient language proficiency the verbal communication with the system completely fails, the graphical interface could even serve as an *alternative* communication channel. By directly manipulating objects of the virtual micro-world the student can try to illustrate her intentions thus giving the system additional clues for guiding the diagnostic procedure.

## RELATED WORK

Constraint relaxation techniques have been used for quite a long time as a means to increase the robustness of natural language systems against ill-formed input (Weischedel and Black, 1980). Given the case that the normal analysis of an utterance fails to deliver a spanning structural interpretation, a second attempt is made with certain constraints of the grammar being retracted.

With respect to error handling the language learning setting differs from other more traditional applications of natural language technology

- in the need to precisely diagnose errors instead of simply tolerating them,
- a highly increased error probability, and
- a strong bias towards errors caused by the interference of linguistic structures from the first language.

Nevertheless, the same technique has been successfully used in a number of systems for language learning. Schwind (1995) e.g. extended a unification grammar based on context-free productions with a special purpose feature unification, which in case of a feature clash produces appropriate error descriptions. The procedure is guided by general error-type preferences to always determine a single, most plausible error interpretation. Since constraint checking is done only within the limits of a constituent, sometimes the most plausible reading is difficult to find. In case of the German sentence

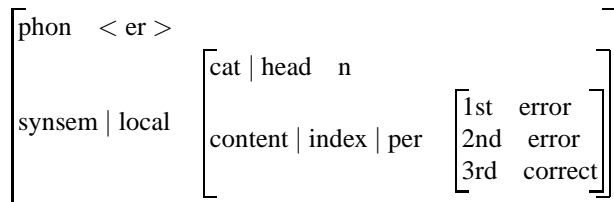
<i>*Der</i>	<i>Götter</i>	<i>zürnen.</i>	<i>(The gods are angry.)</i>
masc,nom,sg	masc,nom,pl	pl	
all,gen,pl	masc,gen,pl		
	masc,acc,pl		

the parser does not detect an error within the noun phrase *der Götter* (which is a perfect genitive NP) and therefore signals a violation of the nominative case restriction for subjects, although there would have been a far more plausible alternative error interpretation with a number disagreement between determiner and noun.

Not only does Schwind's system treat ambiguities in a heavily restricted way, it also does not support alternative error perspectives. Structural deviations (e.g. linear ordering problems, omissions, insertions etc.) are modeled by means of special grammar rules, i.e. cannot be dealt with by means of constraints. To handle these error categories dedicated mal-rules describing possible faulty structures have to be included into the grammar.

Vandeventer (2000) incorporated a similar approach into a grammar of French and was able to demonstrate that a substantial amount of agreement errors in real student sentences was correctly identified by the system. Reuer (2003) used a unification grammar to additionally express existence requirements for syntactic constituents by means of a unification grammar. Thus, he is able to diagnose omissions in a similar way as agreement requirements are handled. He also introduced a special purpose mechanism for relocating sentence constituents during the analysis: superfluous words from the student's input are stored on a heap for possible reinsertion at a later position in the sentence. This way, linear ordering errors can be diagnosed and reported to the student and even a precise correction proposal is available.

With the goal of systematically treating the inherent ambiguity of morpho-syntactic descriptions, Heift (1998) implemented a diagnosis component using a constraint-based grammar formalism. Although she did not make the constraints explicit, in principle her approach is also based on constraint retraction. To model constraint violations by means of a normal feature unification she exploits a particularly inspired method to code all possible ways to violate a particular constraint directly into the lexical entries of the system



In case of a feature clash the corresponding slot will directly contain an error flag, which allows a full reconstruction of the conditions leading to the problem situation. In the terminology of this paper, the diagnostic results obtained from such a description provide a strictly lexical (i.e. fact-based) view on the error. More general (rule-based) error descriptions cannot be derived. A similar encoding scheme has also been invented for the possible verb positions in a German sentence offering the possibility to also diagnose linear ordering problems in a student solution.

While the approach makes possible a thorough treatment of alternative readings, it also neutralizes valuable morpho-syntactic information, which is no longer available for guiding the analysis. Accordingly, a great number of spurious interpretations can be expected as the complexity of utterances or the coverage of the grammar grows. This casts serious doubts on the potential for scaling up the approach

beyond the rather controlled exercises for which it has been used so far.

## CONCLUSIONS

We have investigated a number of problem areas in which constraint-based modeling can be used to diagnose erroneous student input and which are characterised by the presence of ambiguity on a massive scale. By comparing these applications we were able to identify four major sources of diagnostic ambiguity, namely

- the inability to evaluate the satisfaction part of a constraint independently from the outcome of other constraint applications, because either
  - certain variables of the problem domain are not directly observable (as in the example of addition exercises), or
  - relevant features of the observation cannot be determined uniquely prior to the constraint application (a situation which is typical for e.g. agreement problems in natural language),
- the necessity to consider different error perspectives which can highlight different kinds of misconceptions leading to the same error (e.g. the distinction between rule and fact knowledge in natural language), and finally
- the inability to uniquely evaluate the relevance part of a constraint because the precise structure of the particular diagnostic problem is unknown (which is typical for unrestricted natural language).

Due to the fact that ambiguity is pervasive in natural language and that this ambiguity is mirrored in a multitude of diagnostic alternatives we came to the conclusion that constraint violations alone do not provide us with the desired clues about the current capabilities of the student. Instead, we also need sophisticated mechanisms for ranking and selecting error hypotheses according to certain plausibility criteria, before a decision about an appropriate instructional action can be taken.

From that perspective constraint-based models clearly belong to the domain model of a tutoring system. In some restricted areas a high level of model adequacy can be achieved that is sufficient to directly use constraints as an expert module, which is capable of solving the problem itself. For unrestricted language, however, constraint-based models are definitely too weak to serve that purpose.

Obviously, different application areas differ with respect to the strength of the constraining information available to solve the diagnostic problem. According to our experience the following problem classes can be distinguished;

1. The available constraints are strong enough to describe the space of admissible solutions exactly. Model-based diagnosis techniques can be applied and, since the model is able to solve the problem on its own, it can derive correction proposals starting from the sample data provided by the student.
  - (a) The available constraints are strong enough to restrict the space of potential diagnostic results to a unique error hypothesis. In such a case, constraints based on a two-valued logic



will suffice and no error selection becomes necessary. The distinction between domain and student model vanishes.

- (b) Otherwise the student model needs to be separated from the domain model and error selection heuristics have to be used.
  - i. Either heuristics are available which can be directly built into the diagnosis procedure, e.g. restricting it to minimum cardinality hypotheses, or
  - ii. a *post hoc* selection has to be devised based on e.g. error likelihoods.
2. Otherwise preferential constraints have to be applied which can be used to directly integrate a great variety of uncertain knowledge into the decision process on the optimal solution. Even dynamic contributions derived from the context of an utterance or from non-verbal interactions with the student can be made use of. An enumeration of all possible error interpretations is infeasible and alternative error perspectives are difficult to accommodate. The domain model can no longer be used as an expert module.

Even the simple algebraic problem discussed above turned out to be an instance of class 1b), i.e. without a properly designed hypothesis selection a residual diagnostic uncertainty cannot be avoided. The simple agreement problems in the section on polysemy also belong to this problem class. They differ from the algebraic exercises both in the much higher degree of ambiguity and the availability of different error perspectives. Unrestricted natural language input clearly is an example of a class-2 problem. Without the immediate integration of all the available evidence directly into the analysis procedure no proper solution to the diagnostic problem can be found at all.

Using these guidelines to individually tailor constraint-based techniques to different problem classes in natural language made it possible to reliably diagnose corresponding errors in a wide variety of learners' utterances. Of course, the high degree of ambiguity we encounter in this particular domain will always require us to make some simplifying assumptions. Fortunately such simplifications are not only welcome from a system developers' viewpoint but also agree with the common didactic goal to hide the enormous complexity of language from the learner at least in the early phases of second language acquisition. Nevertheless, the task remains a challenging one, since every approach to the diagnosis of ungrammatical natural language will suffer from its intrinsic fundamental dilemma: consistency checks have to be carried out in order to determine the structure of an utterance, whereas it is the same (still unknown) structure which determines the consistency conditions that must hold. Even worse, consistency conditions might also be violated by the student.

Using weighted constraints on word-to-word relationships, we have been able to extend the idea of defeasible constraints also to problems with an *a priori* unknown structure. An integrated decision procedure was devised which overcomes the fundamental dilemma by fully integrating natural language parsing with error diagnosis and therefore opens up promising perspectives for the design of communicatively meaningful language learning scenarios.

Combining a range of techniques from computer science and computational linguistics a system has been developed which is based on defeasible constraints throughout and uses constraint violations

as the only means for diagnosing student errors. It provides a rich body of high quality diagnostics which in cooperation with a student model enables the tutoring system to select situation specific error hypotheses, generate appropriate feedback and provide specifically tailored follow-up exercises for the student. Despite a multitude of alternative diagnostic results, hypothesis selection criteria are available to generate informative and individually tailored feedback in many cases of ungrammatical input.

The approach lends itself to stand-alone solutions in limited purpose exercises, but can also be scaled up to more ambitious communication settings. By integrating it into a complex multi-modal interaction environment it might help one of the dreams of computer-based language learning come true: a system capable of dealing with unrestricted language and communicatively relevant discourse, ready to explain errors or make correction proposals, and being available whenever needed.

## REFERENCES

- Daum, M., Foth, K., and Menzel, W. (2003). Constraint-based integration of deep and shallow parsing techniques. In *Proc. 11th European Conference of the ACL, EACL-2003*, pages 99–106, Budapest, Hungary.
- Foth, K., Daum, M., and Menzel, W. (2005a). A broad coverage parser for German based on defeasible constraints. In *Constraint Solving and Language Processing*, volume 3438 of *Lecture Notes in Artificial Intelligence*. Springer-Verlag, Berlin.
- Foth, K., Menzel, W., and Schröder, I. (2005b). Robust parsing with weighted constraints. *Natural Language Engineering*, 11(1):1–25.
- Hamburger, H. (1995). Tutorial tools for language learning by two-medium dialogue. In Holland, V. M., Kaplan, J. D., and Sams, M. R., editors, *Intelligent Language Tutors*, pages 183–199. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Heift, T. (1998). *Designed Intelligence: A Language Teacher Model*. PhD-Thesis, Simon Fraser University.
- Holland, M. V. (1994). Lessons learned in designing intelligent call: Managing communication across disciplines. *Computer Assisted Language Learning*, 7(3):227–256.
- Martin, B. and Mitrovic, A. (2000). Tailoring feedback by correcting student answers. In *Proceedings 5th International Conference on Intelligent Tutoring Systems, ITS-2000*, pages 383–392, Montreal, Canada. Springer-Verlag.
- Martin, B. and Mitrovic, A. (2002). Automatic problem generation in constraint-based tutors. In Cerri, S., Gouarderes, G., and Paraguacu, F., editors, *Proc. 6th Int. Conf on Intelligent Tutoring Systems, ITS-2002*, number 2363 in *Lecture Notes in Computer Science*, pages 388–398, Biarritz, France.
- Menzel, W. (1990). Anticipation-free diagnosis of structural faults. In *Proceedings 13th International Conference on Computational Linguistics, Coling-1990*, volume III, pages 422–424, Helsinki, Finland.
- Menzel, W. (1992). *Modellbasierte Fehlerdiagnose für Sprachlehreysteme*. Sprache und Information. Niemeyer, Tübingen.
- Menzel, W. (1998). Sehen und Verstehen: Der Beitrag bildlicher Information zum Sprachverstehen. In *Tagungsband GI-Jahrestagung 1998*, pages 161–170, Magdeburg. Springer-Verlag.

- Menzel, W. (2004). Errors, intentions, and explanations: Feedback generation for language tutoring systems. In *Proc. Int. Conf. InSTIL/ICALL-2004*, pages 75–82, Venice, Italy.
- Menzel, W. and Schröder, I. (1998a). Constraint-based diagnosis for intelligent language tutoring systems. In *Proceedings IT&KNOWS, IFIP World Congress*, pages 484–497, Wien and Budapest.
- Menzel, W. and Schröder, I. (1998b). Model-based diagnosis under structural uncertainty. In *Proceedings European Conference on Artificial Intelligence, ECAI-1998*, pages 284–288, Brighton, UK.
- Mitrovic, A. (1998). A knowledge-based teaching system for SQL. In *Proceedings ED-MEDIA/ED-TELECOM-98*, pages 1027–1032, Freiburg, Germany.
- Mitrovic, A., Koedinger, K. R., and Martin, B. (2003). A comparative analysis of cognitive tutoring and constraint-based modelling. In *Proceedings 9th International Conference on User Modeling, UM-2003*, pages 313–322, Johnstown, PA. Springer-Verlag.
- Mitrovic, A., Mayo, M., Suraweera, P., and Martin, B. (2001). Constraint-based tutors: a success story. In *Proceedings 14th International Conference on Industrial and Engineering Applications of AI and Expert Systems, IEA/AIE-2001*, pages 931–940, Budapest, Hungary. Springer-Verlag.
- Mitrovic, A. and Ohlsson, S. (1999). Evaluation of a constraint-based tutor for a database language. *International Journal of Artificial Intelligence in Education*, 10:238–256.
- Ohlsson, S. (1994). Constraint-based student modeling. In Greer, J. E. and McCalla, G. I., editors, *Student Modeling: The Key to Individualized Knowledge-based Instruction*, pages 167–189. Springer-Verlag, Berlin.
- Reiter, R. (1987). A theory of diagnosis from first principles. *Artificial Intelligence*, 32:57–95.
- Reuer, V. (2003). *PromisD: Ein Analyseverfahren zur antizipationsfreien Erkennung und Erklärung von grammatischen Fehlern in Sprachlehrrsystemen*. Dissertation, Humboldt-Universität zu Berlin.
- Schröder, I., Menzel, W., Foth, K., and Schulz, M. (2000). Modeling dependency grammar with restricted constraints. *Traitement Automatique des Langues (T.A.L.)*, 41(1):97–126.
- Schwind, C. B. (1995). Error analysis and explanation in knowledge based language tutoring. *Computer Assisted Language Learning*, 8(4):295–324.
- Self, J. (1992). Cognitive diagnosis for tutoring systems. In *Proceedings European Conference on Artificial Intelligence, ECAI-1992*, pages 699–703, Vienna, Austria.
- Stockfleth, N. (2000). *Integration von CDG-Parsing und Fehlersimulation*. Diplomarbeit, Universität Hamburg, Fachbereich Informatik.
- Vandeventer, A. (2000). Diagnostic d’erreurs grammaticales par relâchement de contraintes dans le cadre de l’ELAO. In *Proceedings TALN-2000*, pages 357–366.
- Weischedel, R. M. and Black, J. E. (1980). Responding intelligently to unparsable inputs. *American Journal of Computational Linguistics*, 6(2):97–109.