

AUTOMATED REASONING ABOUT NATURAL LANGUAGE CORRECTNESS

Wolfgang Menzel
Zentralinstitut für Sprachwissenschaft
Akademie der Wissenschaften der DDR
Prenzlauer Promenade 149-152
Berlin, 1100, DDR

ABSTRACT

Automated Reasoning techniques applied to the problem of natural language correctness allow the design of flexible training aids for the teaching of foreign languages. The approach involves important advantages for both the student and the teacher by detecting possible errors and pointing out their reasons. Explanations may be given on four distinct levels, thus offering differently instructive error messages according to the needs of the student.

I. THE IDEA

The application of techniques from the domain of Automated Reasoning to the problem of natural language correctness offers solutions to at least some of the deficiencies of traditional approaches to computer assisted language learning. By supplying a specialized inference mechanism with knowledge about what is correct within fragments of natural language utterances, a flexible training device can be designed. It prompts the student with e. g. randomly generated sentence frames, where slots have to be filled in. The system then accomplishes two main tasks:

(1) It tries to diagnose possible errors in the students response in order to build up an internal model of the current capabilities of the student in terms of strictly linguistic categories.

(2) It gives an explanation of the diagnostic results to guide the student in his search for a correct solution.

In contrast to other approaches (c.f. Barchan et al. 1985, Pulman 1984, Schwind 1987) we concentrate our efforts more on the handling of fragmentary utterances, instead of trying to analyse the correctness of complete sentences. The enormous difficulties connected with the design of a universal error diagnosis for natural language sentences may only partially be

seen as a motivation for this restriction. Other, equally important justifications could be mentioned as well:

(1) The handling of only simple sentence fragments seems to be a more natural and transparent limitation compared with an ad hoc exclusion of important parts of the grammar from the rule system. Promising the student a universal sentence acceptor, the real capabilities of which are rather limited, may easily be misinterpreted as a kind of bluff, since the consequences of such a cut will always remain a mysterious thing to the student. Severe restrictions on the grammatical knowledge are inevitable at the moment, but probably nobody will ever be able to explain the language competence of a training system to a learner of a second language without totally confusing him. Hence, minimising the problem of grammatical coverage by accepting only fragments of sentences, drastically improves the prospects of finally achieving something like a "water-proof" solution. Nothing could be considered to be more harmful in a teaching environment than to blame a system's failure on the student.

(2) The concentration on small sub-fields of grammar makes the determination of very precise and detailed diagnostic results possible. This, of course, is not so much important if seen only for the purpose of direct explanation: An explanation overloaded with details is likely to irritate the student. Nevertheless, a very precise diagnosis is a sound basis for building up a model of the current capabilities of the student, which advantageously may be used to guide the further course of interaction.

(3) The approach allows a stepwise extension of the degree of sophistication while preserving the same basic principles on all levels. This enables a rather smooth accomodation to different performance classes of hardware as well as an easy adaptation to different paedagogical objectives. Indeed, there are good reasons to expect the very simple examples (e.g. the insertion of a correct German determiner) to be well suited for practical

training purposes.

(4) The focus on selected grammatical regularities facilitates a systematic training, which from a didactic viewpoint seems to be more promising than just the unspecified invitation: "Type in an arbitrary sentence!" with the always present risk to catch the system out. Here we prefer to guide the student in a rather unconstrained way by prompting him with carefully selected sentence frames or questions. To hide the limitations of the dictionary, as usual, the domain context of a simple exercise environment (a room, a shop, an airport etc.) is used.

In its diagnostic capabilities the presented approach shows a strong analogy to the basic concepts usually applied within a system of Automated Reasoning: a hypothesis is verified to be in accordance with a set of initial facts and a set of rules, which for our special purpose model the correctness conditions of a specific training exercise. The initial facts are given as a logical combination of syntactic and semantic features describing the grammatical properties of certain word forms in the system prompt. The hypothesis results from the student's response where word forms are internally represented by their associated features as well.

II. KNOWLEDGE REPRESENTATION

To formalize the correctness conditions of natural language constructs in a linguistically adequate manner we adopted two basic operators from a dependency grammar model (Kunze 1975):

constraints of the kind:

(** <destination> <condition>)

transmitters of the kind:

(<source> <destination> <category>)

Both of them operate on feature sets. A constraint reduces the feature set of a word form bound to the variable <destination> to its maximum subset which satisfies the given <condition>. Transmitters carry features belonging to a specific <category> from a <source> to a <destination>, changing the feature set at the destination according to a predefined agreement relation. Typical categories are the ordinary ones: GENDER, NUMBER, CASE, PERSON etc., but semantic or very language specific features (like INFLECTIONAL DEGREE for German, cf. Rüdiger 1975) may be used as well. Accordingly, by means of these operators the conditions for the morpho-syntactic correctness within a

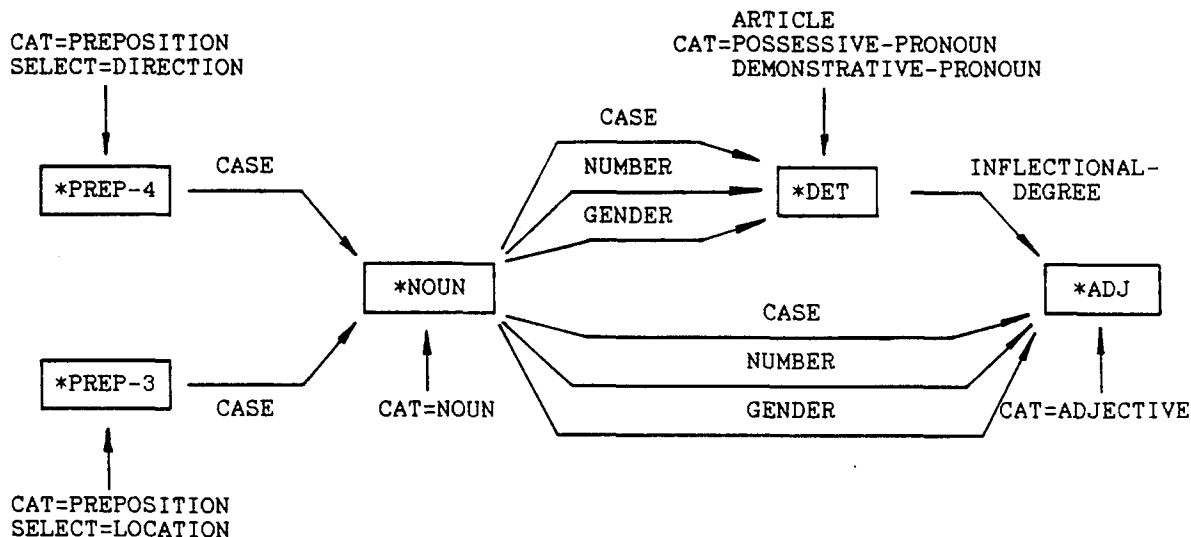


Figure 1: Correctness conditions for a special German prepositional phrase

simple German prepositional phrase of the type (PREP DET ADJ NOUN) may be coded as shown in figure 1.

The nodes in this graph denote variables, which have to be bound to single word forms. According to their value assignment mode two types of variables may be distinguished. Context variables belong to the sentence frame and receive their value (the feature set of a specific word form) already during the sentence generation process. The value of a slot variable, however, depends on the student's response and is established by a pattern matching procedure based mainly on word class information. The power of the pattern matcher used determines almost completely the flexibility of the system: A rather simple one, using obligatory slot variables only (hence, restricting the slot to a fixed length) will be sufficient under certain circumstances. The additional use of optional slot variables allows the implementation of more diversified exercises. Sometimes even a simple parser for sentence fragments may be required.

The transmitters obviously constitute the part of rules within the knowledge base. They can easily be interpreted as defining logical implications, semantically extended by two existential quantifiers for the variables <source> and <destination>. In a certain sense transmitters correspond to the well known

IF...THEN rules in a typical expert system.

The factual knowledge, on the other side, consists of constraints (which could be thought of to be transmitters with a nowhere-source, indicated by "***" in the rule set of figure 2) together with the feature combinations in the dictionary entries. Only from the point of view of explanation the factual information has a special status: one cannot ask for it by means of a why-question.

III. ERROR DIAGNOSIS

Commonly one tries to distinguish the field of Automated Reasoning from the development of expert systems by comparing a mean size of the knowledge base as well as the length of a typical inference chain. Normally, a system of Automated Reasoning is expected to have a rather limited number of rules but the ability to handle extremely long chains whereas the characteristics of an expert system include plenty of rules but very short inferences. In this respect, a system for foreign language training belongs to a third category, since both, the size of the knowledge base as well as the mean length of an inference path are comparatively small. Unfortunately, this simplicity doesn't result in a very simple design for the inference engine as well. Difficulties arise from a peculiarity of the language training task: On the one hand, facts and rules are given to describe the correctness of natural language constructs. On the other hand, explanations are required about the deficiencies of a student's solution. Probably the system is never asked to point out the reasons why a specific inference can be drawn, but it is expected to explain the reasons why a correctness proof can not be established. This, of course, requires a special diagnosis procedure which in the case of an error in the student's response searches for plausible alternatives which might have been leading to a correct solution.

The diagnosis is carried out in two steps (figure 3). Using a classical non-deterministic forward chaining algorithm the first step tries to show the correctness by successively applying constraints and transmitters on all the feature sets previously bound to variables. A transmitter can be applied, if its source doesn't appear to be a destination in any other

Constraints:

```
(*** *PREP-4 (CAT PREPOSITION))
(***) *PREP-4 (SELECT DIRECTION))
(***) *PREP-3 (CAT PREPOSITION))
(***) *PREP-3 (SELECT LOCATION))
(***) *NOUN (CAT NOMINAL))
(***) *DET (CAT ARTICLE
          POSSESSIVE-PRONOUN
          DEMONSTRATIVE-PRONOUN))
(***) *ADJ (CAT ADJECTIVE))
```

Transmitter:

```
(*PREP-4 *NOUN CASE)
(*PREP-3 *NOUN CASE)
(*NOUN *DET CASE)
(*NOUN *DET NUMBER)
(*NOUN *DET GENDER)
(*NOUN *ADJ CASE)
(*NOUN *ADJ NUMBER)
(*NOUN *ADJ GENDER)
(*DET *ADJ INFLECTIONAL-DEGREE)
```

figure 2: Rule set for the example in figure 1

transmitter waiting for application yet. This implies that cycles of transmitters are not allowed within the knowledge base, a configuration which actually doesn't occur in a natural language sentence, anyhow.

The application of a constraint or a transmitter fails, if it results in an empty feature set at the destination. Failures due to the missing of facts in the knowledge base may indicate an error in the students response, and all the categories, variables and values concerned are stored as failure points to be analysed in detail later. A sentence frame can be considered to be correctly completed by the student, if all the relevant constraints and transmitters have been applied successfully. If such a solution cannot be found (that is, a mistake of the student has been encountered), the second step resumes the analysis by investigating the consequences of assuming in each case just the complementary feature set at the failure point. By doing this, the diagnosis procedure in fact tries to simulate the ignoring of the corresponding rule by the student and aims at finding out all the resulting consequences.

To deliver the information needed by the second step of the diagnosis procedure requires to extend the capabilities of the basic routine for feature set comparison beyond the usual unification operations.

In addition to the normal intersection between the relevant features at the <source> and the <destination> the procedure determines the complement of the feature set at the <destination> (see figure 4). To achieve the desired high resolution of the diagnosis unification is always carried out for a single category. All the other features are left unchanged.

Given the case of an error in the students response the investigation of both alternatives, the intersection as well as the complement becomes necessary. That is, the diagnosis is confronted with an enormous number of analysis paths. Strong heuristic criteria are needed to restrict the size of the search space effectively. So far, an algorithm considering only paths with a minimum number of failure points has turned out to be sufficient in most cases.

IV. EXPLANATION COMPONENT

Usually, due to the often numerous morpho-syntactic readings of a word form the diagnosis component comes out with a couple of possible error interpretations, all of them can by no means be explained to a student without totally confusing him. Again, heuristic criteria are needed to reduce the number of interpretations in a sensible way.

Step I: CORRECTNESS PROOF

Step II: INVESTIGATION OF INFERENCE FAILURES

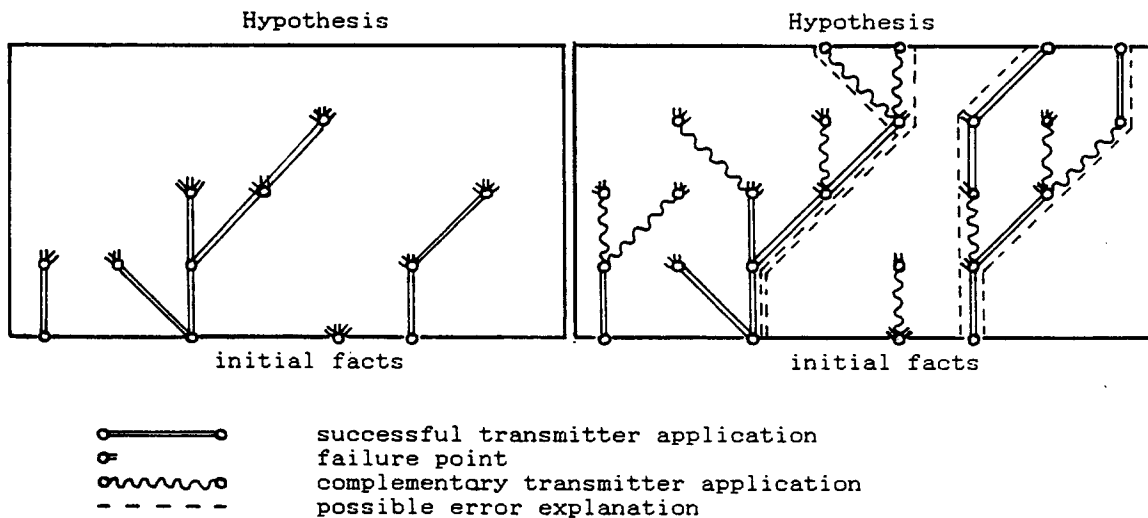


Figure 3: Two step diagnosis

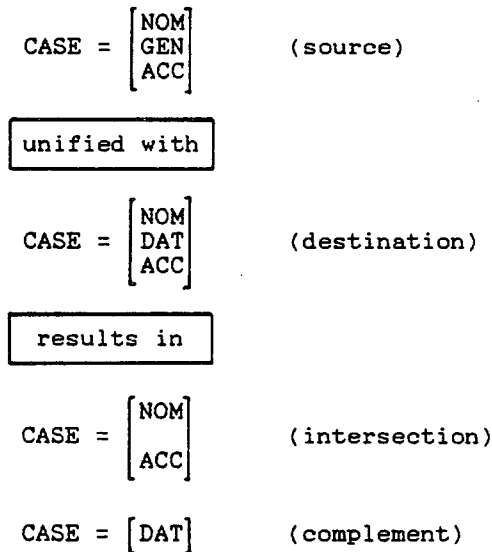


Figure 4: Example for the extended feature set unification

To select an appropriate (that is, helpful from the student's point of view) error description the diagnostic results have to be ordered by an estimated explanatory power. So far, the following criteria have been taken into consideration:

(1) A category preference, which chooses a certain transmitter function (e.g. GENDER) as a more probable one. This is a simple but obviously crude and unreliable criterion.

(2) The distance between the complementary transmitter application and the hypothesis, whereby errors "higher up" in a sentence structure are preferred. For example, it is more likely that the case governed by a preposition has been mistaken than that the agreement within the prepositional phrase is violated.

(3) In a multiple error diagnosis a category common to most of the alternatives could be taken for the explanation. Given the very frequent error combination

(CASE and GENDER) or (NUMBER and GENDER)

missing gender agreement should be a reasonable explanation.

A good heuristics certainly has to include the structure of the dictionary entries and the rule set in its investigation of possible alternatives. If there is indeed a second reading with respect to

one of the hypothesised error reasons then probably the student overlooked this possibility. Here further investigations are necessary.

From a paedagogical point of view it would be desirable to explain the diagnostic results (detected errors and their possible reasons) on differently instructive levels, selecting the right one according to previous results or current desires of the student. The following four levels seem to be appropriate and theoretically motivated:

(1) right/wrong answer without further explanation

(2) explanation on the level of rules (e.g. "missing gender agreement between xxx and yyy")

(3) explanation on the level of facts (e.g. "xxx is a feminine noun, hence you should take a feminine determiner")

(4) explanation on the level of examples using the inverted dictionary as a data base to retrieve appropriate word forms by means of the inferred feature sets.

The verbalization of an explanation is done on the basis of sentence schemata, which have to be defined together with the correctness conditions. On demand, the actual categories, values or examples are inserted and minor surface smoothing operations are carried out.

V. DIALOG CONTROL & USER MODELLING

By carefully investigating a series of responses a model of the current capabilities of the student can be build up. Based on this model the system autonomously may vary different aspects of the dialog behaviour. The most simple example is the selection of one of the explanation levels. The system switches over to a deeper level of explanation if the student either repeatedly fails to find the correct solution or signals his inability for understanding the previous error message. It goes back to a higher level if consecutive successes of the student justify this.

A series of responses may contain hints about where the weaknesses of the student actually lie. Thus, in addition to the criteria of section IV another heuristics for the selection of diagnostic results is available: Continued repetition of one and

the same error type will cause the explanation to focus on this category. Furthermore, the collected information can be used to guide the training strategy. Exercise generation may be controlled to just concentrate on the weak points of the student or even to alter the degree of exercise difficulty.

VI. EXPERIMENTATION

To study some selected problems (especially the exploitation of heuristic rules within the diagnosis and explanation components) in greater detail, a first prototype has been implemented. Currently the system includes a random sentence generator to supply the system prompts, a simple pattern matcher for obligatory slot variables, the two step diagnosis described above and an explanation component up to the level of facts.

The training examples studied so far have mainly been taken from the area of German noun phrase inflection (indeed an intricate subject from the foreigner's point of view). The experiments confirmed that simple versions of training exercises may run already on very cheap type of hardware (i.e. 8-bit micros).

VII. DISCUSSION

The design of foreign language training systems based on fundamental techniques of Automated Reasoning exhibits several important advantages as compared with an immediate implementation of the almost trivial scheme a Pattern Drill Book is based upon:

(1) Automated Reasoning allows more flexibility. Not the one correct solution is asked for. The student may choose his solution within the limitations of the dictionary (expressed by the exercise environment). Dialog situations may easily be simulated. Experimentation becomes possible.

(2) In addition to the right/wrong diagnosis further three levels of explanation are available. A correct solution can be generated just for the particular word samples chosen by the student.

(3) It becomes possible to include rather complex regularities between context and slot variables. Nevertheless,

the explanation mostly points out the location of the error rather precisely.

(4) A model of the student's capabilities is built up and the teacher is supplied with a statistics in terms of linguistic categories even in the case of very complex or mixed exercises.

(5) Instead of explicitly listing them, exercises can be generated automatically, thus achieving a variety which almost excludes repetition even in the case of extremely long or repeated training sessions.

Limitations for the application domain mostly result from the feature based approach to knowledge representation. It first of all predestines the solution for the training of morpho-syntactic regularities (esp. agreement relations). To handle problems of e.g. usage or style in a sufficiently general manner seems to be far beyond the current possibilities.

REFERENCES

- Barchan, J.; Woodmansee, B. and Yazdani, M. (1985) Computer Assisted Instruction using a French Grammar Analyser. Research Report 126, Department of Computer Science, University of Exeter.
- Kunze, J. (1975) Abhängigkeitsgrammatik. studia grammatica XII, Akademie-Verlag, Berlin.
- Pulman, S.G. (1984) Limited Domain System for Language Teaching. Proceedings Coling 84, Stanford: 84-87.
- Rüdiger, B. (1975) Flexivische und Wortbildungsanalyse des Deutschen. Linguistische Studien, Reihe A, Sonderheft 1975, Berlin.
- Schwind, C.B. (1987) Prototyp eines Sprachtutorsystems für Deutsch als Fremdsprache, KI-Rundbrief 44, Januar 1987: 42
- Wos, L.; Overbeek, R.; Lusk, E. and Boyle, J. (1984) Automated Reasoning. Prentice Hall, Englewood Cliffs.