# Between language correctness and domain knowledge in CALL

**Svetla Boytcheva[1], Ognian Kalaydjiev[2], Albena Strupchanska[2], Galia Angelova[2]**

[1] Faculty of Maths and Informatics, Sofia University "St. Kliment Ohridski",
5 J. Bauchier Blvd., 1164 Sofia, Bulgaria, svetla@fmi.uni-sofia.bg

[2] Linguistic Modelling Dept, Central Laboratory for Parallel Processing, Bulgarian Academy of Sciences,
25A Acad. G. Bonchev Str., 1113 Sofia, Bulgaria, {ogi, albena, galia}@lml.bas.bg

## Abstract

Building advanced CALL systems is a challenge; no universal solutions are attained so far regarding the most desired features of intelligent CALL like learner-system communication in Natural Language (NL), adequate processing of information about the learner's semantic errors, and adaptive strategies for choice of relevant tutoring materials. This paper presents work in progress in a larger CALL project, oriented to learners of written English as a second language. We focus on key issues dealing with proving the domain correctness of learner's NL utterances and planning the suggestion of relevant readings, dynamically collected from Internet. In this way we discuss the integration of NLP-techniques in CALL as an attempt to improve CALL coverage and performance.

## 1. Introduction and Background Research

Advanced educational systems are a challenging research goal, which would provide invaluable practical results. However, building an adaptive and intelligent educational application in language learning is not an easy task at all, although the field of intelligent CALL systems was established as a hot R&D area long ago. No universal solutions are attained so far regarding the most desired CALL features like learner-system NL communication, adequate processing of information about the learner's language reactions and adaptive strategies for structuring the NL tutoring materials. [Mur97] states that there are many (even commercially available) CALL applications which however "fall short when it comes to issues related to system's intelligence and adaptivity. ... As a result, CALL systems are often perceived by learners and teachers as dumb and inflexible, which is demotivating for the learner and restricts the independent use of CALL systems considerably".

Learner-system NL communication is particularly interesting for us. Language learning presupposes that students will type in free NL utterances since it looks rather artificial to acquire a new language by only selecting menu items and pressing buttons. However, providing free NL input requires integration of techniques for Natural Language Processing (NLP), esp. parsing and checking the correctness of the learner's NL answer. A number of systems try to cope with the non-trivial choice to support user-system dialogue in (almost free) natural language but according to [HKS95] (1995) "... so few of these systems have passed the concept demonstration phase". The systems overviewed in [HKS95] contain mostly modules for checking students' competence in vocabulary, morphology, and correct syntax usage (parsers); the most sophisticated semantic analysis is embedded in the system BRIDGE/MILT ([Sam95], [WGMM95], [DHBM95]) which matches the learner's utterance (a lexical conceptual structure) against the pre-stored expected lexical conceptual structures. This matching is done by following the intuitive notion of a *correct match*. [DHBM95] shows rather simple examples for semantic correctness when "Carlos killed Juan" is considered correct because it is syntactically correct and can be matched to the expected "Carlos murdered Juan". Some recent systems, built during the last five years, like CASTLE in RECALL project [Rec97] and SLALOM [MPC96], recently integrated in ICICLE [MMS01], focus as well mostly on spelling, morphological, and syntactic errors. The learner's input NL utterance is analysed and represented as a set of hierarchically ordered features. Features of similar complexity form stereotypical levels of language ability and system's inferences are based on these levels. In RECALL the semantic information, world knowledge, and inference only play a role to the extent that they are necessary to detect and diagnose spelling, morphological and syntax errors. Another rather sophisticated system with NL input, CIRCSIM-Tutor, has been developed since years and its new elaboration is presented in [Gla00]. The tutor's questions are closed and the expected answers are quite short. The input understander is as permissive as possible. It extracts whatever is needed from the student's input and ignores the rest. The input is analysed in sequential steps: lexicon lookup, spelling correction, partial parsing by finite state transducers, lookup in concept ontologies, and finally matching to the question.

To summarise, every intelligent CALL system has to decide how to analyse learners' NL input and check its correctness but the present solutions especially for semantic analysis are far from being perfect.

This paper presents relevant work in progress in a project[1] which aims at the development of the knowledge-based learning environment **STyLE** for teaching English terminology to adults, non-native English speakers. A central idea is that CALL of terminology requires checking the domain correctness (in addition to the linguistic one) of the learners' utterances in drills. The knowledge base (KB) is built using basic financial terms in financial markets. Terms and relations between them fix in a relatively natural way the choice and granularity of the formal concepts, so STyLE offers to its user lexical and conceptual units which correspond to the user's intuitive fragmentation of the domain. Similarly to approaches like the one described in [BMB99], the 'core' ontology encodes the educational content communicated to the learner. Once acquired, the ontology is applied in two essential ways: (*i*) the correctness of the student's answer can be evaluated against the KB and *(ii)* planning of moves in the system-user interaction can be guided by this information.

The paper is structured as follows. Section 2 sketches the project as a whole. Section 3 presents in more detail the prover for checking the correctness of the learner's utterances in free NL and an extended example. Section 4 describes the pedagogical agent planning "next reading material" according to its content relevance. Section 5 discusses current evaluation results and sketches further work.

## 2. LARFLAST Project Scenario

STyLE is a Web-based learning environment where the student accomplishes three basic tasks: *(i)* reading teaching materials, *(ii)* performing test exercises and *(iii)* discussing his/her own learner model with the system. The project is oriented to learners who need English language competence as well as expertise in correct usage of English financial terms. This ambitiously formulated paradigm requires the integration of some formal NLU techniques, allowing for analysis of the user's answers to drills where the student is given the opportunity to type in free NL text.

Fig. 1 represents the already elaborated NLP-related components of STyLE within the context of the system architecture. Student's performance in exercises is controlled by the *diagnostic module*. The response interpreter maintains answers to drills with fixed choices. This paper deals with the student's answers to exercises where the learner is given the opportunity to type in free NL sentences.

The system Parasite, developed at UMIST by Allan Ramsay, see e.g. [Ram94] and [RSe00], is already integrated in STyLE as an NLU machine for analysing learners' free utterances. Parasite provides checking of the morphological, syntactic and semantic correctness of the learner's utterances in especially designed drills, while the prover `STyLE-Parasite` checks the answers' correctness against the available domain knowledge and the expected answer. Other STyLE components provide learning materials to the student. Agents developed by project partners find texts in Internet and subsequent filters (an original implementation of Latent Semantic Indexing [DDSLH90]) juxtapose to each text some percentage of "relevance to STyLE financial terms". Although it is well known that information retrieval filters are not perfect, this technique seems to be the most natural Web-inspired solution, supplying STyLE with continuously

---

[1] **STyLE** (Scientific Terminology Learning Environment) is under development in the Copernicus'98 JRP LARFLAST (LeARning Foreign LAnguage Scientific Terminology), November 1999 - October 2001. Partners: CBLU Leeds, UK; UMIST Manchester, UK; LIRM Montpellier, France; Academy of Sciences, Romania; Simferopol University, Ukraine; Sofia University and Virtech Ltd., Bulgaria.

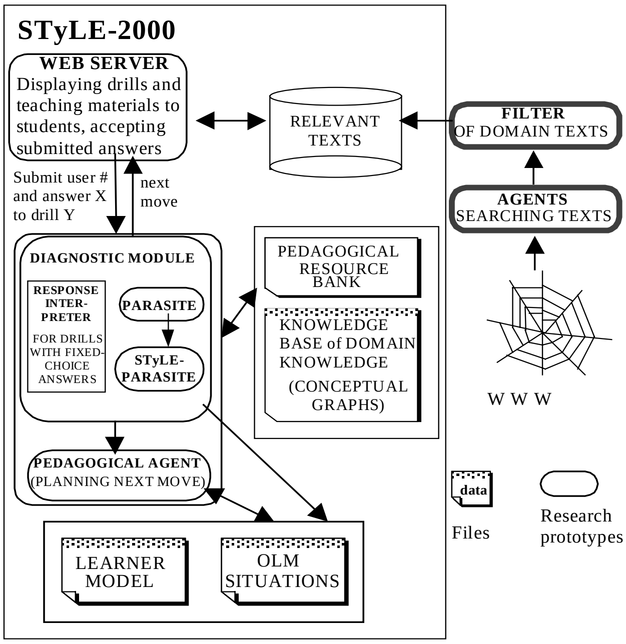


**Figure 1.** NLP-related modules in STyLE-2000

updated archive of texts relevant to the financial domain. Present experiments show that constraints on agent search in financial sites only reduces considerably the usual information retrieval noise, so in this paper we can assume that STyLE has a large, dynamically updated database of texts and the planning agent has to decide which text is to be recommended as a relevant reading in certain learning situations. Further STyLE components like the generator of www-pages WebGen [Tra00] and the Open Learner Model providing the learner-system dialog about students' misconceptions [DSB99] are not considered in this paper and not displayed at Fig. 1.

Here we focus on the language technologies implemented by the authors. We describe in more details the prover `STyLE-Parasite`, which checks the domain correctness of the learners' answers to especially designed drills (see section 3) and the pedagogical agent, which plans what is to be done next and offers, when necessary, some readings with relevant content (see section 4).

# 3. Proving domain correctness

Especially performed users' studies of terminology learning [Vit99] show that erroneous answers are usually caused by the following reasons:

- Language errors, e.g. spelling, morphological and syntax error;
- Question misunderstanding, which causes wrong answer;
- Correct question understanding, but absent knowledge of the correct English term, which implies usage of paraphrases and generalisation instead of the expected answer;
- Correct question understanding, but absent domain knowledge which implies specialisation, partially correct answers, incomplete answers and wrong answers.

In STyLE design we tried to cover all these cases. After analysing the free NL answer Parasite returns information about linguistic inconsistency of learner's utterances: morphological, syntax and semantic errors (in the later case no logical form is produced). Answers with correct linguistic semantics are subjects to further considerations of their domain relevance, proved by `STyLE-Parasite`.

The basic steps in `STyLE-Parasite`'s work are described in [BKNA00] (most generally, `STyLE-Parasite` compares the logical form of the learners' utterance to the logical forms of the predefined expected minimal and maximal answers and makes the necessary inferences). The prover makes forward chaining and looks for all possible bindings of variables. The diagnostics delivered after analysis of learners' utterance by Parasite and `STyLE-Parasite` are as follows:

• after Parasite's performance of the learners' answer: *(i)* ***unknown word***, *(ii)* ***morpho***, *(iii)* ***syntax*** and *(iv)* ***wrong***;

• after `STyLE-Parasite`'s post-processing of the learners' answer: *(i)* ***correct***, *(ii)* ***more_general***, *(iii)* ***more_specific***, *(iv)* ***paraphrase*** (usage of concept definition instead of the proper term), *(v)* ***incomplete***, *(vi)* ***partially_correct,*** *(vii)* ***wrong*** *and (viii)* ***combination of several mistakes***.

Figure 2 shows how `STyLE-Parasite` decides about the correctness of the input logical forms (which strongly depends on the lexical choices and the syntactic structure of the concrete answer). Since there might be many correct answers and their language expression varies considerably, it is not practical to compare the input logical form to a single predefined correct logical form. Rather, `STyLE-`

`Parasite` uses pre-stored maximal and minimal logical forms. The minimal correct answer has to be obligatory present in all the correct answers i.e. the minimal correct answer is the intersection of all correct answers. The maximal correct answer is the cover of all correct answers. Adding new terms to the maximal answer might be redundant or wrong. There might be several kinds of mistakes in the received answer, so learner's utterances are to be investigated with respect to all possible error types applying the above-described steps. `STyLE-Parasite` inference is sound [BKNA00], but it is not complete, because the conclusion "correct learner utterances" is indicated after the first correct binding of variables. Partially correct answers are indicated as well after the first occurrence of binding variables that fits to cases 3a, 4 and 7 (and the proving halts).

Table 1 gives examples of minimal and maximal expected answers (predefined) and system reactions to users' utterances in column 2.

The expected answers in STyLE are simple declarative sentences. However, Parasite handles complex sentences as well as simple discourse consisting of several sentences. Parasite allows as well for soft parsing, i.e. processing of sentences with "little mistakes" like wrong verb forms, irregular subject-verb co-ordination, a number of

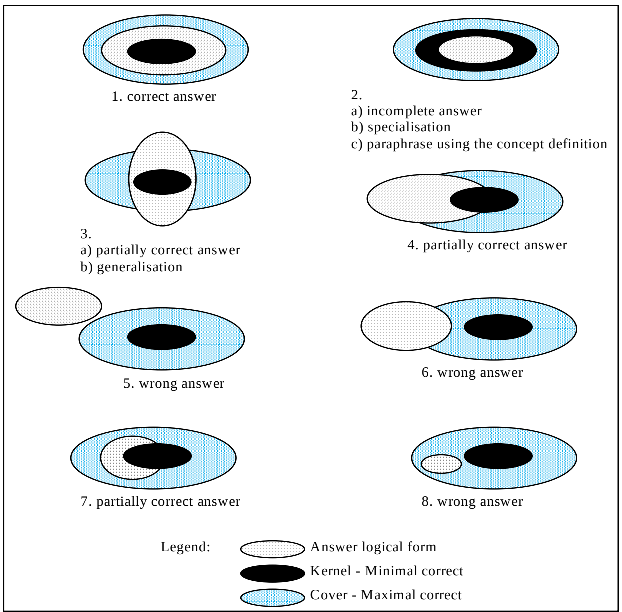


Fig. 2. Comparison and inference with answers' logical forms

cases with wrong word-order etc. Such mistakes are considered as typical for second language learning. In this way STyLE can be used by learners with intermediate English, who often make language mistakes and who will benefit by the elaborated language diagnostics delivered by Parasite. We believe the above-described scenario is the most perfect one for implementation of CALL with free user utterances, although it is clear in advance that complete NL understanding of arbitrary sentences is a rather complicated task with cannot be solved in the foreseeable future.

## 4. Suggesting relevant readings as a step towards personalised IR

This section describes a recently implemented module providing information filtering in STyLE. This module operates within the planning agent [Kal01]; here we focus on the personalised presentation of tutoring materials only. The idea is inspired by the Web-context, where many financial sites expose and frequently update relevant texts. So, if we want to show to the learner readings containing most relevant information, we have to support and continuously update a database of relevant texts (the collection itself is performed by Web spiders). The planer thus operates with:

- a text DB, containing financial texts collected from Internet;
- a relevance measure, showing for each text the percentage of its relevance to the domain terms under consideration. These terms are juxtaposed to KB items. The relevance measure is associated automatically to each text by an LSA-module (original implementation of Sofia team [Nak00]).

The goal of the planner is to select which text is most relevant to be displayed as a tutoring material (reading) at the particular learning situation. At each situation, the Learner Model (see Fig. 1) keeps track of the concepts which are unknown or known wrongly by the learner. The text with higher relevance to all these terms has to be selected. This is done as follows:

The learning situation is estimated with respect to the terms $T_{N1}$, $T_{N2}$, ..., $T_{Nm}$ which appear in the Learner Model as *not-known* or *wrongly-known*.

| Case | Sample of learner's utterance | Discussion |
|---|---|---|
| Kernel (predefined minimal answer) | **Primary market is a financial market that operates with newly issued debt instruments and securities.** | The logical form is pre-stored in the system as a Kernel. |
| Cover (predefined maximal answer) | *Primary market is a financial market that operates with newly issued debt instruments and securities and provides new investments and its goal is to raise capital.* | The logical form is pre-stored in the system as a Cover. |
| 1.Correct answer | **Primary market is a financial market that operates with newly issued debt instruments and securities** *and provides new investments.* | This logical form is between the Kernel and the Cover. |
| 2a) Incomplete answer | **Primary market is a financial market that operates with newly issued securities.** | Missing Kernel term: **debt instruments.** |
| 2b) Specialisation of concepts from the definition | **Primary market is a financial market that operates with newly issued** bonds. | Bond is a specialisation of **security;**<br>Missing: **debt instruments.** |
| 2c) Paraphrase using the concept definition | **Primary market is a financial market that operates with** new emissions of stocks, bonds and other financial assets. | New emissions **= newly issued;**<br>stocks, bonds and other financial assets **= debt instruments and securities.** |
| 3a) Partially correct | **Primary market is a financial market that operates with newly issued debt instruments and securities** for instant delivery. | Wrong: for instant delivery. |
| 3b)Generalisation of concepts from the definition | **Primary market is a** market **that operates with newly issued** financial instruments. | Market is a generalisation of **financial market;** Financial instruments are generalisation of **debt instruments and securities**. |
| 4. Partially correct | **Primary market is a financial market that operates with newly issued securities** for instant delivery *and provides new investments.* | Wrong: for instant delivery;<br>Missing: **debt instruments.** |
| 5. Wrong answer | **Primary market is** an organisation in which the total worth is divided into commercial papers. | Wrong: an organisation in which the total worth is divided into commercial papers;<br>Missing: **financial market that operates with newly issued debt instruments and securities.** |
| 6. Wrong answer | **Primary market** *provides new investments* for instant delivery. | Wrong: for instant delivery;<br>Missing: **financial market that operates with newly issued debt instruments and securities;** |
| 7. Partially correct | **Primary market is a financial market that operates with newly issued securities** *and provides new investments.* | Missing: **debt instruments.** |
| 8. Wrong answer | **Primary market** *provides new investments.* | Missing: **financial market that operates with newly** |

| | | |
|---|---|---|
| | | **issued debt instruments and securities.** |

Table 1. Style-Parasite decisions according to the cases shown in Fig. 2

Actually we operate with the KB concepts, juxtaposed to these terms. The estimation is unique for the current learning situation, it is calculated for each term $T_{N1}$, $T_{N2}$, ..., $T_{Nm}$ and represents for each term the sum of:

- predefined weight of the concept in the KB hierarchy, an integer between 1 and 10;
- closeness of the focused concept $T_{Ni}$ to the concepts $T_{N1}$, ..., $T_{Ni-1}$, $T_{Ni+1}$, ..., $T_{Nm}$. All pairs ($T_{Ni}$, $T_{N1}$), ..., ($T_{Ni}$, $T_{Ni-1}$), ($T_{Ni}$, $T_{Ni+1}$), ..., ($T_{Ni}$, $T_{Nm}$) are considered and the values "close-distant" (respectively "1-0") are summed. Two concepts are "close" if they are either linked as child-parent in the hierarchy, or they are sisters according the same partitioning perspective. Otherwise the concepts are considered as distant.

After calculating the sums $S_1$, $S_2$, ..., $S_m$ for the terms $T_{N1}$, $T_{N2}$, ..., $T_{Nm}$, the integers $S_1$, $S_2$, ..., $S_m$ are sorted in decreasing order. Let the obtained list is $S_{D1}$, $S_{D2}$, ..., $S_{Dm}$ and corresponding terms are $T_{D1}$, $T_{D2}$, ..., $T_{Dm}$. For each term $T_{D1}$, $T_{D2}$, ..., $T_{Dm}$ the planner finds the set of relevant texts, available at the moment in the text DB. Starting from $T_{D1}$ to $T_{Dm}$, the planner looks for texts maximally relevant to all terms. In this way the planner proposes for readings texts, that provide "maximal relevance" to the unknown terms, taking into consideration the estimation of terms' weights.

Since Larflast project is entering the final evaluation phase, the planner and its strategy for choosing relevant text will be soon evaluated too. Therefore small modifications of the above-described heuristics might be expected.

## 5. Evaluation and further work

Currently, STyLE offers to its learners 30 test units, covering about 150 basic English terms in financial markets. The corresponding KB contains about 220 concepts and 30 relations, plus increasing number of domain facts (represented at present as more than 300 conceptual graphs). This is the pedagogical and knowledge environment where STyLE maintains the learner model, proves the correctness of user's answers and plans what is to be done next. The NLU components operate with the following parameters: Parasite maintains a lexicon of 300 words and a KB of 150 meaning postulates. Its performance is very fast: average speed of syntax and semi-compositional semantic analysis is less than few seconds for a short extended discourse of 3 – 4 English sentences. This speed of analysis allowed us to integrate both Parasite and `STyLE-Parasite` (in SICStus Prolog) within a web-based learning environment with on-line performance of drills in real time.

The final user validation will take place in October 2001, so for the moment we have only partial learners' assessment of our ongoing work. The main disappointment of learners and teachers is that such a system cannot answer *why*-questions, i.e. Parasite and `STyLE-Parasite` provide extremely comprehensive diagnostic about the error type but not the error reason. It seems to be difficult to explain to non-specialists that AI techniques have their specific features and limitations. Anyway the attempt for providing complete NL diagnostics is (up to our knowledge) the most serious in CALL at present.

Current results and the user studies allow for the evaluation of some other important characteristics of the final product:

(i) The relatively complete ontology of 150 basic terms in financial markets provides a good background for covering educational content in terminology learning;

(ii) The on-line integration of Parasite can support maintenance of a larger dictionary of common English lexica, which is sufficient for learning English as LSP;

(iii) Planning helps essentially in guiding the learner within a rich environment where the learner is offered many choices, including free Web surfing.

Due to these reasons, we evaluate the system architecture and performance speed as a good CALL solution. The future work includes final integration of the whole system STyLE and corresponding further user study and evaluation.

## References

**[BKNA00]** Boytcheva, Sv., O. Kalaydjiev, A. Nenkova and G. Angelova. *Integration of Resources and Components in a Knowledge-Based Web-environment for terminology learning*. Proc. AIMSA-2000, Varna, Bulgaria, Sept. 2000, LNAI 1904, Springer, pp. 210-220.

**[BMB99]** Breuker, J., Muntjewerff A. and B. Bredeweg. *Ontollogical Modelling for Designing Educational Software*. In Proc. AI-ED Workshop on Ontologies for Intelligent Educational Systems, Le Mans, France, July 18-19, 1999.

**[DHBM95]** Dorr, B., Hendler, J., Blanksteen, S., and B. Migdaloff. *On Beyond Syntax: Use of Lexical Conceptual Structure for Intelligent Tutoring*. In [HKS95], pp. 289-310.

**[DDSLH90]** Deerwester S., Dumais S., Furnas G., Laundauer, T. and Harshman R. *Indexing by Latent Semantic Analysis.* Journal American Society for Information Sciences 41(1990), pp. 391-475. Also http://lsa.colorado.edu (1990-99); USA patent 4839853 http://www.patents.ibm.com.

**[DSB99]** Dimitrova, V., J.A. Self and P. Brna. *The interactive maintains of Open Learner Models*. In Lajoie and Vivet (eds.), Proc. 9th Conf. AI in Education, Frontiers of AI and Applications, Vol. 50, IOS Press, pp. 405-412, 1999.

**[Gla00]** Glass, M. *Processing Language Input in the CIRCSIM-Tutor Intelligent System*. AAAI 2000 Fall Symp. on Building Dialogue Systems for Tutorial Applications. http://www.csam.iit.edu/ ~circsim/index.html

**[HKS95]** Holland, V. M., Kaplan, J. and M. Sams (eds.) *Intelligent Language Tutors: Theory Shaping Technology*. Lawrence Erlbaum Associates, UK, 1995.

**[Kal01]** Kalaydjiev, O. *An Intelligent Pedagogical Agent in CALL*. In Proc. Doctoral Consortium, 8th Int. Conf. of User Modelling, Germany, July 2001, pp. 283-285.

**[MMS01]** Michaud, L., McCoy K. and L. Stark *Modeling the Acquisition of English: an Intelligent CALL Approach* In Proceedings of The 8th International Conference on User Modeling, July 13-17, 2001, Sonthofen, Germany, pp.14-23.

**[MPC96]** McCoy, K., Pennington, C.A., and Suri, L.Z. *English error correction: A syntactic user model based on principled "mal-rule" scoring*. In Proc. Fifth Intern. Conf. on User Modelling, 1996, pp. 59-66.

**[Mur97]** Murphy M. and M. McTear. *Learner Modelling for Intelligent CALL*. In: A. Jameson, C. Paris and C. Tasso (Eds.). Proc. User Modelling 97, Springer Wien New York, 1997. (http://um.org)

[**Nak00**] Nakov, Pr. *Getting Better Results with Latent Semantic Indexing*. Proc. Student Presentations at ESSLLI-2000, pp. 156-166, Birmingham, August 2000.

**[Ram94]** Ramsay, A. *Meaning as Constraints on Information States*, in Rupp, Rosner, Johnson (eds.) Constraints, Language and Computation, 1994, Academic Press, London: 249-276.

**[Rec97]** *RECALL, a Telematics Language Engineering project* (1997), see http://iserve1. infj.ulst.ac.uk/~recall. (CASTLE was developed at the IBM Centre in Heidelberg, Germany.)

**[RSe00]** Ramsay, A. and H. Seville. *What did he mean by that*? Proc. AIMSA-2000, Bulgaria, Sept. 2000, LNAI 1904, Springer, pp. 199-209.

**[Sam95]** Sams, M. (1995) *Advanced Technologies for Language Learning: the BRIDGE Project within the ARILanguage Tutor Program*. In [HKS95], pp. 17-21.

[**Tra00**] Trausan-Matu, St. *Metaphor Processing for Learning Terminology*. Proc. AIMSA-2000, Varna, Bulgaria, Sept. 2000, LNAI 1904, Springer, pp. 232 - 241.

**[Vit99]** Vitanova, I. *Learning Foreign Language Terminology: the User Perspective*. Larflast report 8.1, August 1999, submitted to the European Commission.

**[WGMM95]** Weinberg, A., Garman, J., Martin, J. and P. Merlo. *A Principle-Based Parser for foreign Language Tutoring in German and Arabic*. In [HKS95], pp. 23-44