Reference: *British Journal of Educational Technology BJET*, Vol. 29, No. 1, 1998, pp. 47-57

"Who is Who": Roles in an Intelligent System for Foreign Language Terminology Learning

Vanya Dimitrova, Darina Dicheva

Vanya Dimitrova is a PhD student in the area of intelligent tutoring systems. Darina Dicheva is at University of Sofia, Bulgaria. Her research interests include intelligent tutoring systems and learning environments, flexible and distance learning. Address for correspondence: Faculty of Math. and Informatics, University of Sofia, 5 James Bouchier Str, 1164 Sofia, Bulgaria. E-mail: darinad@fmi.uni-sofia.bg.

Abstract: ITELS is an intelligent system for foreign language terminology learning. It is aimed not only at assisting language learning focused on specific terminology but also at enhancing learner's conceptual knowledge in the subject area. The system could be tuned to aid learning in different terminological areas and supports three styles of tutoring: System-Initiated, Collaborative, and Learner-Initiated. Three agents take part in the instructional process supported by ITELS: the human teacher, the system, and the learner. Each of them is an active agent characterized by a set of specific roles with respect to the system functioning. In this paper we describe ITELS with emphasis on the roles of these agents.

Introduction

Learning subject area terminology is part of English instruction at Bulgarian universities. Experience shows that in the course of this instruction the students face difficulties not only with the foreign language but also with the concepts of the subject area. Since the English teachers who teach these courses are not usually specialists in the subject area (Computer Science, Economics, Chemistry, etc.) learning the required terminology is often not very successful. ITELS is an intelligent tutoring system that supports learning English focused on specific terminology as well as improving learner's conceptual knowledge in the subject area. The system could be tuned to support learning in different terminological areas. Computer Science terminology was chosen as a testbed.

ITELS is aimed at enhancing learners' skills in reading and comprehending English terminological texts as well as in understanding and using correctly subject area's terminology. Therefore, the system applies reading comprehension (Nuttal, 1982) and vocabulary learning (Carter and McCarthy, 1988) approaches to language instruction. Differently from most vocabulary learning systems which are built as learning environments (e.g. Ingraham et al., 1996; Swartz, 1992), ITELS supports three styles of tutoring:

- *System-Initiated*: The system is in complete control of the teaching process. It decides which teaching activity is most appropriate in the current situation.
- *Collaborative*: The system and the learner work in collaboration. A mixed-initiative approach is adopted where the system shares the instructional control with the learner. This allows the learner to insist on a preferred teaching activity or content. The system

could reject her/his suggestion if the learner model shows inability to self-estimation and lack of self-confidence.

- *Learner-Initiated*: The initiative is taken by the learner. The system behaves as a learning environment embodying different learning activities.

The system employs techniques from the areas of authoring systems, Intelligent Computer-Assisted Language Learning (ICALL), and knowledge-based machine-aided translation. It includes authoring facilities that allow the teacher to develop and change the material to be taught and unlike most Intelligent Tutoring Systems (ITSs) - the way in which material is taught.

As in the typical ICALL systems (Swartz and Yazdani, 1992; Holland et al., 1995) ITELS contains linguistic expert knowledge presented in a lexicon and grammar rules. In addition, the system applies morphological rules for generating word formation exercises and affixes related feedback. The latter explains the meaning of affixes and the way they are used to build new words. This is considered to be very useful for the learner in tackling new lexical items.

In ITELS the terms semantics is represented by using Conceptual Graphs (CGs) (Sowa, 1984). Such representation can be found in some knowledge-based machine aided translation systems which use these graphs for generating explanations, e.g. DB-MAT (Angelova and Bontcheva, 1996). Beside for explanation the CGs here are also used for diagnosing and instructional planning: as a source for detecting and correcting learner's misconceptions, for generating exercises and feedback, and for selecting the term to be next exercised.

A recent tendency in the ITS research is to encourage the learner's initiative by allowing her/him to take part in the diagnosing. The idea that the learner model needs not to be hidden but can be open and inspectable by the learner is suggested in (Cumming and Self, 1991). Such an idea is implemented in the system *Mr. Collins* (Bull et al., 1995) which emphasises on the shared learner modelling with a learner model open to negotiation. Similarly to *Mr. Collins,* in ITELS the learner could make a self diagnosis. Besides, s/he could be asked to help the system in diagnosing when it faces some problems. The learner can also point the way s/he prefers to be taught.

ITELS is an intelligent system relying on the active involvement of the human teacher and the learner. In this paper we describe ITELS with emphasis on the roles of the agents that take part in the instructional process: the human teacher, the system, and the learner. Each of these is an active agent characterized by a set of specific roles with respect to the system functioning.

The Teacher as Active Agent

In ITELS the human teacher is an active agent taking part in both preparing the material to be taught and course planning. ITELS supplies authoring tools that allow authoring at the domain level and at the strategic level according to her/his specific requirements and preferences. Thus the teacher acts as:

Reference: *British Journal of Educational Technology BJET*, Vol. 29, No. 1, 1998, pp. 47-57

A. *Course Materials Builder*

The human teacher is able to embody her/his subject matter and pedagogical expertise in the system by defining *Learning Blocks* (LB) of two kinds: *training* and *information blocks*. Each LB is to be built around a subject area term. Training blocks contain either a question or an exercise together with its level of complexity. Four basic types of questions are included: *multiple choice*, *multiple answer*, *fill-in-the-gap*, and *matching phrases*. The language-oriented exercises include sentence composition or translation (very simple sentences), changing the word order to get a correct sentence, changing the order of sentences to get a meaningful text. Each training block is aimed at mastering one of the following skills: understanding the main idea of a reading text, understanding a particular paragraph, mastering the forms of a word and their correct usage, acquiring terms definitions, and acquiring concepts and the relations between them.

B. *Help Information Provider*

The teacher could suggest what kind of help the learner should get when facing difficulties with a particular term by defining information blocks. The information blocks contain typical phrases, examples of correct use of the terms, etc. They include textual and/or visual information.

C. *Instructional Planner*

The human teacher's *planning* role in educational computer systems varies having its extreme cases in traditional authoring systems where a teacher can specify precisely the presentation order of the course material to the learners, and in ITSs, where s/he is not allowed to take part in the course planning. In ITELS the teacher specifies only the order of the course topics. In addition s/he determines the *key terms* in each topic and the terms relevant to the *goal term* in each learning block. This information along with the subject area conceptual knowledge stored in ITELS Knowledge Base (KB) is used by the system for instructional planning. Besides, in the beginning of each learning session the teacher (as well as the learner) can suppress the *goal* suggested by the system and choose a new one - a topic or a list of terms/concepts to be learned.

The System as Active Agent

ITELS is designed as an intelligent system which embodies the traditional ITS roles of an Expert, a Learner modeller, and an Instructor appropriately modified to accommodate three different tutoring styles: *System-Initiated, Collaborative,* and *Learner-Initiated.*

The System as Expert

Expert's Knowledge

The Expert modules of the ICALL systems typically include "some type of grammar and a lexicon for the target language (the expert knowledge) and a parser (the expert inference engine) to process language inputs" (Swartz and Yazdani, 1992). Since ITELS is an intelligent system aimed at assisting not only foreign language learning but also the enhancement of conceptual knowledge in the subject area, it is to act as both a *linguistic expert* and a *subject area expert*. Thus its Expert module in addition to the linguistic knowledge (grammatical and lexical) includes conceptual knowledge, i.e. knowledge about the subject area concepts and the relations between them and corresponding inference engine (Dicheva and Dimitrova, 1996a; Dimitrova and Dicheva, 1997). As in most ICALL systems the essential part of the linguistic knowledge is presented in a lexicon. It relates English words to their grammatical categories and their meaning in the native language. The lexicon contains entries for ordinary words and for terms. For each subject area term additional information is included: definition, pattern, meaning, and a link to the knowledge base. In order to avoid term ambiguity when a term has more than one meaning it is represented by several entries - one for each of its meanings. For example, the term "*Output*" which has three meanings (Dictionary of Computing, 1990) is represented in the lexicon by three entries with meanings "*Data*", "*Signal*" and "*Process*".

The Expert knowledge about *word formation* is represented in a table of affixes. Prefixes usually change the meaning of the word. Suffixes change the word from one part of speech to another. Each table record includes: affix, its kind, type, and meaning, for example: (*un*, *prefix*, '*negative/positive*', '*not, not good enough*'), (*ly*, *suffix*, '*adverb-forming*', '*in the manner of*'). The grammatical knowledge of the linguistic expert is very restricted. It consists of a few grammar rules about *articles*, *numbers*, and *cases*.

The knowledge of the subject area expert includes subject area terms and relations between them. It is classified into topics. The conceptual knowledge in each topic is organized into two parts representing the concept hierarchy (taxonomy) and the relationships between the concepts, correspondingly. The hierarchy of the concepts represents their level of generality and permits information inheritance. Part of the taxonomy of the "*Programming Language*" topic is shown in Figure 1.

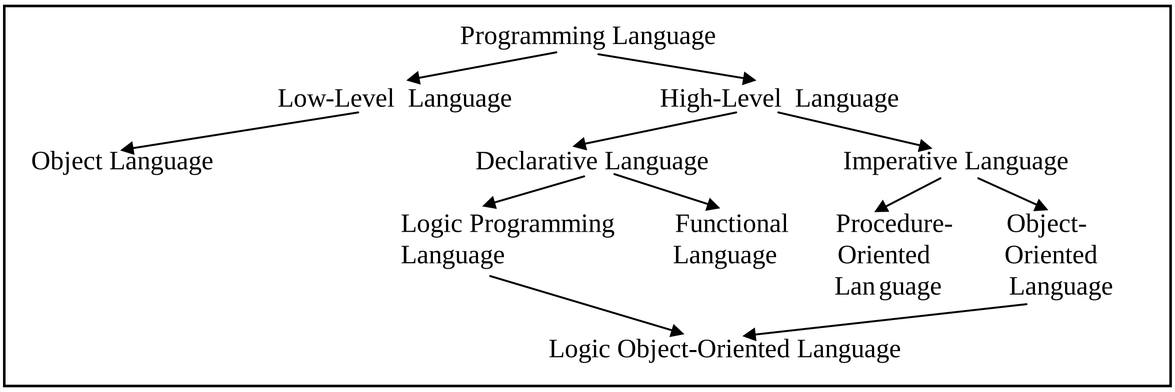

Figure 1: *Part of the taxonomy of the topic "Programming Language"*

The relationships between concepts are represented by conceptual graphs where each concept is determined by its concept type and its referent. CGs are formalism with direct mapping to the natural language allowing a convenient extraction of the meaning of a sentence (Sowa, 1992). An example of a CG that presents the sentence "*An object program is the translation of a source program into an object language*" is shown in Figure 2a. A conceptual graph connects concepts with conceptual relations. We use a set of basic conceptual relations which cover the most widely used relations in terminological areas (e.g. ATTR - '*attribute*', CHRC - '*characteristic*', PART - '*part*', AGNT - '*agent*', PTNT - '*patient*', RCPT - '*recipient*', INST - '*instrument*', RSLT - '*result*', etc.). It is possible new types and relations to be defined in terms of simpler ones. A new type is defined by specifying its *genius* (supertype) and *differentia* (a CG that allows the new type to be distinguished by the *genius*). Figure 2b shows an example of a new type definition.

```
[ACTION: 'Translate"] —
    (RSLT) → [Object Program]
    (PTNT) → [Source Program]
    (INST)  → [Object Language].
```

Figure 2a: *An examples of CG*

```
type Functional Language (x)
    genus: Declarative Language
    differentia
    [Declarative Language *x] —
        (INST) → [Function {*}].
```

Figure 2b: *A definition of a new type*

Roles of the Expert

The Expert can act as:

A. *Instructional Knowledge Provider*

The expert provides knowledge to be used in answering learner's questions, suggesting hints and generating exercises. The extraction of expert knowledge is discussed in more detail in (Dicheva and Dimitrova, 1996b).

- Providing help knowledge

When the Expert suggests hints or answers questions about the grammatical category, patterns, translation, or definition of a term it extracts linguistic knowledge from KB. In order to suggest terms *similar* to an anticipated answer it extracts conceptual knowledge. The *similarity* here includes both *hierarchical similarity* and *relational similarity*. Two terms are considered to be *hierarchically similar* if in the type hierarchy their concept types are either connected directly or have some common parents. For example, on Figure 1 the terms "*High-Level Language*", "*Logic Programming Language*", and "*Functional Language*" are hierarchically similar to "*Declarative Language*".

Two terms are considered to be *relationally* (*conceptually*) *similar* if they are included in an existing CG or such a graph can be obtained by applying the operations *copy*, *restrict*, *join*, and *simplify* (Sowa, 1984) on CGs from the KB. For example, "*Object Program*", "*Source Program*", and "*Object Language*" are relationally similar to "*Translation*" (see Figure 2a).

- Providing knowledge for generating exercises

The Expert provides knowledge for generating linguistic exercises or exercises related to concept learning by request from the system Instructor. It could supply information about word formation, term patterns, and term definitions for linguistic exercises as well as conceptual knowledge from the concept taxonomy and CG KB for exercises related to term understanding.

B. *Answer Evaluator*

The expert compares the learner's answers against the correct answers. This includes spell checking, phraseology checking, and concept checking. The latter determines the *concept distance* between the suggested term and the correct term. Two terms are *near* if they are *similar* (*hierarchically* or *relationally*). Two terms are *far* if they are not *near*. For example, terms "*Object-Oriented Language*" and "*Object Language*" are *far*.

The System as Learner Modeller

This role of the system is related to constructing and maintaining a learner model which allows for more adaptive instruction. The learner model includes information that describes the learner's domain knowledge (grammatical and terminological) and the learner's personality (psychological profile and individual preferences).

A. *Learner's Domain Knowledge Modeller*

The learner's grammatical knowledge is modelled by using the issue based approach (Polson and Richardson, 1988). The different issues (e.g. ability of correct uses of cases, articles, etc.) are represented by grammar rules. A counter pair is attached to each issue indicating its correct and faulty uses.

The learner's terminological knowledge is diagnosed by using both overlay and bug catalogue approaches (Polson and Richardson, 1988). The completeness of the learner's terminological knowledge is represented as an overlay on the terms representations in the lexicon and on the conceptual relations in the CGs. Four states are considered: *completely known*, *probably known*, *probably unknown*, *completely unknown*. The incorrectness of the terminological knowledge is represented by using a bug catalogue. The following types of errors are considered: *Spelling errors*, *Translation errors*, *Phraseology errors*, *Taxonomy errors*, *Conceptual errors*. In the learner model each error is represented by an entry including the correct term, the erroneous term, the type and the cause of the error.

B. *Learner's Personality Modeller*

As part of the learner model the system maintains a learner profile characterizing the learner's personality. It consists of two parts: *category* and *characteristics part*. The category part contains information about the level of the learner's knowledge in both English and the subject area (*beginner*, *intermediate* and *advanced*). The second part contains information about the learner's psychological characteristics and individual preferences (preferred tutoring style and activities). The following identity characteristics are included:

- ability to deal with negative states;
- having self-confidence;
- ability to self-estimation;
- taking initiative.

The system determines the learner's category and self-estimation ability through pre-tests and the rest of the identity characteristics as well as the preferred teaching style and activities - by observing the learner's work.

The System as Instructor

In ITELS the system can play the following Instructor's roles:

Reference: *British Journal of Educational Technology BJET*, Vol. 29, No. 1, 1998, pp. 47-57

A. *Instructional Planner*

The strategic decisions on the next term/concept to be taught and on the nature of the next interaction could be taken either by the system or by the learner depending on the selected tutoring style. When responsible, the system chooses the next term and the most appropriate teaching activity to perform. In all cases it determines the LBs relevant to the selected terms.

- Selecting terms to be taught

Deciding on the next term/concept depends on the way in which the domain knowledge base is to be traversed during a tutorial session. For each selected course topic the system maintains a list of terms to be taught/exercised (organized as a stack). This list contains initially all key terms of the topic. The system updates the list after each learner's answer. In case of an erroneous answer it includes new terms in the list using information from two sources - the KB and the currently used LB - depending on the error type. Assume, for example, that in a LB the learner had suggested the erroneous term "*Object Program*" which is *conceptually near* to the correct term "*Source Program*" (as shown in Figure 2a). The system will then include in the terms list all terms from the CG containing both terms, namely "*Object Language*" and "*Translation*". If both terms - the correct one and the erroneous one are *near* in the taxonomy since they have common parents these parent terms will be included in the terms list. If both terms are found to be *far* the terms from the corresponding LB would be included in the terms list.

- Choosing a LB

For each term from the terms list the system selects an appropriate LB. The type of the LB depends on the cause of the last learner's error, the learner's profile, and the learner's knowledge of the goal term. For example, if the learner is a beginner in English and is not confident about the term patterns, the system would suggest a LB related to term formation (such a LB is shown in Figure 3).

- Choosing a teaching style and activities

The system chooses the most appropriate tutoring style and teaching activities using the learner model and a number of *pedagogical* production rules. In choosing a tutoring style the identity characteristics of the learner profile are used. The following teaching activities are included: suggesting hints, giving examples, explaining, and dialogues.

B. *Help Provider*

- Question answering

The system could answer questions about: grammatical categories and patterns of terms, translation of answers supplied in Bulgarian, term definitions, and terms similar (*hierarchically* and *relationally*) to the expected answer. The system could suggest either *general* explanation or *tuned* explanation depending on the current state of the learner's knowledge. In the first case all similar terms from the KB would be used and in the second - only the terms which the learner already knows.

- Hinting

When a student answer is *near* to the correct answer (*conceptually* or in the *taxonomy*) the system suggests a hint either by explaining the difference between both terms or by providing information about all terms similar to the correct term. The Expert explains the difference between a concept type and its *genus* by using the *differentia* from the type definition. For example, for "*Declarative Language*" and "*Functional Language*" the expert generates the following explanation "*The functional language is a declarative language which operates with functions*" (see Figure 2b). In order to explain relational similarity the Expert generates a sentence from the corresponding conceptual graph.

C. *Course Materials Builder*

If in a given situation the system cannot find an appropriate LB predefined by the teacher it constructs such a block. Figure 3 shows two questions about the term "*Object Program*" generated by the system. The first one is generated by extracting term patterns from the lexicon and the second one by constructing a sentence from the CG shown in Figure 2a. The Expert's ability to generate exercises is very useful when the learner faces regularly difficulties with particular terms.

term: *object program* **kind**: *multiple answer* **type**: *term formation* **complexity**: *easy* **author**: *system*	**content**: Find the terms which have similar pattern with the term ***object program***: machine architecture data structure programming language object-oriented language structured programming

term: *object program* **kind**: *fill_in_the_gaps* **type**: *term understanding* **complexity**: *easy* **author**: *system*	**content**: Fill in the blank the correct term: "An .... is the translation of a source program into an object language."

Figure 3: *Two examples of questions generated by the system*

The Learner as Active Agent

The importance of metacognition and of learner initiative is widely recognized (Cumming and Self, 1991). In ITELS the learner is expected to be an active participant in the instruction not only in learning but also in controlling the overall teaching process. S/he is allowed to take the initiative and choose topics for study, to ask questions, to suggest her/his self-diagnosis and also the way s/he prefers to be taught. Thus the learner undertakes the following roles in the system:

A. *Active Tutee*

- Asking questions

The learner is allowed to ask questions about linguistic and conceptual characteristics of terms.

Reference: *British Journal of Educational Technology BJET*, Vol. 29, No. 1, 1998, pp. 47-57

- Choosing a tutoring style

The learner is enabled to choose a tutoring style for each instructional session. When working in a *learner-initiated* mode s/he has full control over the teaching process.

B. *Instructional Planner*

- Specifying preferred teaching activities

In ITELS the learner could specify her/his own view on teaching. S/he can identify situations and the activities which should take place when these situations are recognized.

- Choosing a teaching activity

The learner is allowed to choose a teaching activity at each step of her/his work.

- Selecting terms to be learned/exercised

The learner is allowed to choose either a course topic or a list of terms to be learned.

C. *Learner Modeller*

In ITELS the learner model is open for inspecting and changing by the learner. The learner is provided with tools that help her/him to observe the learner model and to suggest some corrections. The system accepts or rejects the learner's self-diagnosis depending on the learner profile: if the learner is not self-confident and does not have ability to self-estimation the system rejects her/his suggestion.

ITELS could face some problems in its diagnosing, e.g. when the correct term and the suggested term are *far*. In this case the learner could be asked to help. Assume, for example, that in a LB the learner has suggested the erroneous term "*Object Language*" which is *far* from the correct one "*Object Oriented Language*". Then the learner will be presented with the terms similar to each of the above terms and will be asked to select the ones which s/he is not familiar with.

Conclusion

In this paper we discussed the roles of the human teacher, the system, and the learner in the instructional process supported by ITELS - an intelligent system for foreign language terminology learning.

ITELS is being implemented in C++ for Windows. It has a multi-agent architecture which uses a blackboard model to implement the communication between the agents. The pilot version of the system is concerned with teaching Computer Science terminology and its KB currently includes terms from the topic "*Programming Languages*". In order to avoid natural language processing problems some simplifications are made, e.g. for detecting spelling errors only simple pattern matching is used and for generating sentences from CGs a very simple version of the Sowa's algorithm (Sowa 1984) is applied.

The future work on ITELS will be focused on further development of the diagnostic and the instructional planning modules. This will cover developing graphical tools for observing and changing the learner model and tools for generating word formation exercises.

Reference: *British Journal of Educational Technology BJET*, Vol. 29, No. 1, 1998, pp. 47-57

The pilot version of the system will be experimented in the Autumn of 1997 in the regular English course for Computer Science students at Shumen University as well as at the specialized Shumen mathematical school where the subject Informatics is delivered in English. The experiments will enable a proper evaluation of the system.

Acknowledgements
This work is partly funded by NSF Project No. MY-MM 6/95 and NSF Project No. I-502/95. The authors wish to thank the anonymous referees who provided helpful comments on the paper.

References

Angelova G and Bontcheva K (1996) NL domain explanations in knowledge based MAT in *Proceedings of COLING'96*, Copenhagen, Denmark.

Bull S, Pain H and Brna P (1995) Mr Collins: student modelling in intelligent computer assisted language learning *Instructional Science* 2 65-87.

Carter R and McCarthy M (1988) *Vocabulary and Language Teaching* Longman, London.

Cumming G and Self J (1991) Learner modelling in collaborative intelligent educational systems in Goodyear P (ed) *Teaching Knowledge and Intelligent Tutoring* Ablex Publishing Corporation Nowood, New Jersey.

Dicheva D and Dimitrova V (1996a) Knowledge representation in intelligent computer assisted language learning in *Proceedings of TAINN'96* Istanbul, Turkey.

Dicheva D and Dimitrova V (1996b) Representation and extraction of terminological knowledge in ITELS - an intelligent system for foreign language terminology learning in *Proceedings of the Joint Conference of Knowledge-Based Software Engineering* Sozopol, Bulgaria.

Dictionary of Computing (1990) Oxford University Press, Market House Books Ltd.

Dimitrova V and Dicheva D (1997) Learning terminology in a foreign language in *Proceedings of RANLP'97* Tzigov Chark, Bulgaria.

Holland M, Kaplan J and Sams M (1995) (eds) *Intelligent Language Tutors: Theory Shaping Technology* Lawrence Erlbaum Associates, New Jersey.

Ingraham B, Chanier T and Emery C (1994) CAMILLIE an European project to develop language training for different purposes in various languages on a common hypermedia framework *Computers and Education* 23 (1-2) 107-115.

Nuttal C (1988) *Teaching Reading Skills in a Foreign Language* Heinemann, Oxford.

Polson M and Richardson J (1988) (eds) *Foundations of Intelligent Tutoring Systems* Lawrence Erlbaum Association.

Sowa J (1992) Conceptual graphs summary in Nagle T, Nagle G, Gerholz L and Eklund P (eds) *Conceptual Structures: Current Researches and Practice* Ellis Horwood.

Sowa J (1984) *Conceptual Structures: Information Processing in Mind and Machine.* Addison-Wesley, MA.

Swartz M (1992) Issues for tutoring knowledge in foreign language intelligent tutoring systems in Swartz M and Yazdani M (eds) *Intelligent Tutoring Systems for Foreign Language Learning* Springer-Verlag.

Swartz M and Yazdani M (1992) (eds) *Intelligent Tutoring Systems for Foreign Language Learning* Springer-Verlag.