Extending the Scope of the Student Model

Bull, Susan; Brna, Paul; Pain, Helen

DOI:
10.1007/BF01101801

Document Version
Peer reviewed version

Citation for published version (Harvard):
Bull, S, Brna, P & Pain, H 1995, 'Extending the Scope of the Student Model', *User Modelling and User Adapted Interaction* , vol. 5, no. 1, pp. 45-56. https://doi.org/10.1007/BF01101801

Extending the Scope of the Student Model

Susan Bull
Dept. of Artificial Intelligence
University of Edinburgh
80 South Bridge
Edinburgh EH1 1HN

Paul Brna
Computing Department
SECAMS Building
Lancaster University
Lancaster LA1 4YR

Helen Pain
Dept. of Artificial Intelligence
University of Edinburgh
80 South Bridge
Edinburgh EH1 1HN

Abstract
In this paper we maintain that there are benefits to extending the scope of student models to include additional information as part of the explicit student model. We illustrate our argument by describing a student model which focuses on 1. performance in the domain; 2. acquisition order of the target knowledge; 3. analogy; 4. learning strategies; 5. awareness and reflection. The first four of these issues are explicitly represented in the student model. Awareness and reflection should occur as the student model is transparent; it is used to promote learner reflection by encouraging the learner to view, and even negotiate changes to the model. Although the architecture is transferable across domains, each instantiation of the student model will necessarily be domain specific due to the importance of factors such as the relevant background knowledge for analogy, and typical progress through the target material. As an example of this approach we describe the student model of an intelligent computer assisted language learning system which was based on research findings on the above five topics in the field of second language acquisition. Throughout we address the issue of the generality of this model, with particular reference to the possibility of a similar architecture reflecting comparable issues in the domain of learning about electrical circuits.

Key words
student model, intelligent learning environment, reflection, learning strategies, analogy, second language acquisition.

1. Introduction

Student models necessarily contain more flexible information than is normally found in other types of user model, as the aim of any learning environment is that a student should learn, and hence the student model should be continually changing to reflect successive changes in understanding. Furthermore, although stereotypical student models can be useful, in many cases a more individualised model is required if a learner is to be helped successfully. The model should contain information about domain knowledge (including errors and misconceptions), and also other learning issues, for example analogy, learning strategies and the promotion of student reflection. A method is needed to ensure that the model created is sufficiently detailed, but still as accurate and easily updated as possible.

In the area of electrical circuits there are various studies of misconceptions (e.g. Dupin, 1987; Psillos et al, 1987) but there are very few systems which apply the results of these studies[1]. Similarly, applied linguists have collected a great deal of data on errors and misconceptions, and also on other aspects of language learning, but few

[1] Exceptions include an approach to utilising such knowledge based on confronting misconceptions (Brna, 1988), and DIAG, a system for helping students to correct their own misconceptions (Boohan, 1993).

computer based systems have actually used this knowledge to provide support for language learners. For example, although some language learning systems are concerned with analogy, eg. Catt and Hirst (1990), Schuster (1986), Wang and Garigliano (1992) all describe systems which take account of learners' native languages, it appears impossible to find systems with a significantly wider scope. There are influential systems which do use information about errors and misconceptions. These include the ACT*-based tutors of Anderson and his colleagues (e.g. the Lisp Tutor (Anderson et al, 1984)), and PROUST (Johnson & Soloway, 1985). However while these do explicitly take advantage of such knowledge, they are not so concerned with knowledge about other learning issues such as uses of analogy or individual learning strategies. An increasing number of systems are being designed to promote reflection, though few if any learning environments take advantage of the variety of sources of knowledge about the student that can be derived from empirical research.

We hold that such information is valuable in terms of providing the basic substrate for student models. Here, foreign language learning is the initial domain in which our student model is implemented. We argue for the generality of our position with reference to the possibility of a similar basic architecture in the domain of learning about simple electrical circuits. Our student model has some stereotypical components (for aspects of learning common to most learners), but it also contains detailed information relating to individuals. This includes information both about domain knowledge and more general learning issues. Relevant information to be considered for this model was identified from the literature and empirical studies. Maintenance of the model is facilitated by enabling the student to contribute his beliefs about his learning directly to the model.

2. The Architecture of the Student Model

The student model we present here is called Mr. Collins (COLLaboratively constructed, INSpectable student model). The important issues for this model are:

1. domain knowledge and misconceptions;
2. acquisition order of the target knowledge;
3. analogy;
4. learning strategies;
5. awareness and reflection.

Thus Mr. Collins contains the usual information concerning a learner's knowledge of, and misconceptions about the domain. However an additional four issues outside the domain boundary are also handled, as these all constitute important factors in learning. Most student models do not consider this information, i.e. they are not concerned with analogy or learning strategies or the order in which material tends to be acquired by learners, and they do not attempt to use the model to increase reflection. Although these issues are dealt with on an individual basis in a small number of systems, there are none (to our knowledge) which tackle the majority. The architecture of Mr. Collins differs from that of a more conventional student model in order to take account of these additional issues (see figure 1). It is clear that these five considerations are not specific to language learning,

and therefore the architecture of Mr. Collins is general (though each instantiation of the model must be domain focussed[2]).

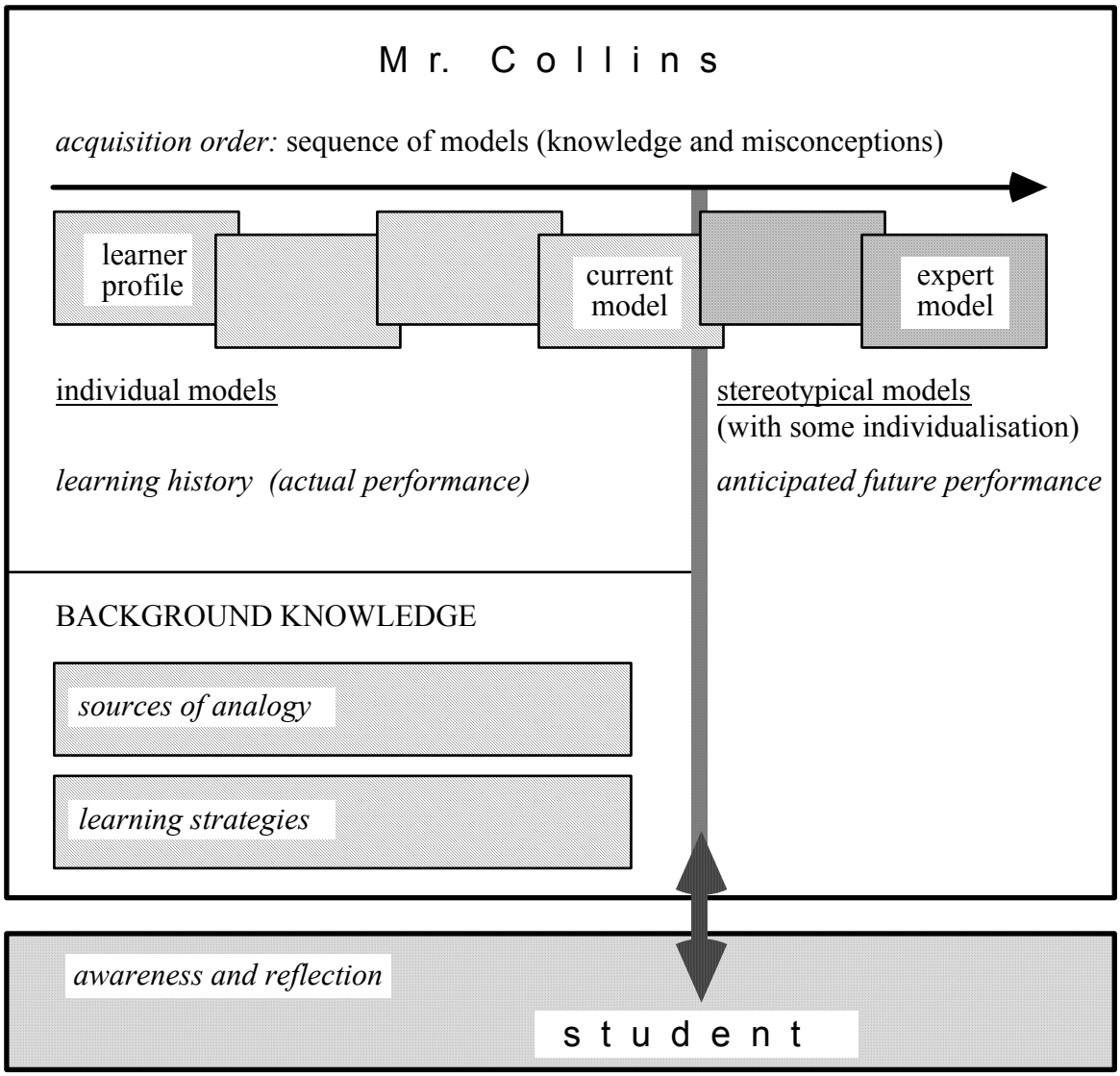

Figure 1: The Architecture of Mr. Collins

Mr. Collins is structured in the following manner. The *acquisition order* of the target knowledge is represented in the form of a sequence of student models. The initial part of this sequence (from the initial learner profile to the current version of the student model) is formed from a series of individualised models based on a learner's current and previous interactions. The final part of the acquisition order (the anticipated future performance) is represented by stereotypical models based on the typical acquisition sequence observed across many learners. (These stereotypical models may be enhanced by some individual information depending on the contents of the earlier models. There is no individual information in the final expert model, as this contains only knowledge as applied by an expert - in this case a native speaker of the target language.) The more traditional components of *knowledge and misconceptions* of the domain (based on

[2] As stated previously, we are here using second language learning as the domain, but in this paper we also consider applying the student model in an alternative domain: electrical circuits.

individual and typical performance) are integrated into the acquisition sequence. Potential sources of *analogy* (here: other languages) and *learning strategies* are represented as appropriate for each student in the individualised section of the student model. *Awareness and reflection* are promoted by encouraging the student to inspect all aspects of the model, and to negotiate the contents of his own current student model. This increased reflection is therefore not part of the student model itself, but occurs through learner interaction with the model. Actually representing this reflection in the model is an issue for further research.

3. An Implementation of the Student Model

The foreign language learning system in which Mr. Collins is currently implemented is aimed at facilitating the acquisition of clitic pronoun placement in European Portuguese[3], and also reducing the occurrence of other pronoun-related errors. This implementation of the student model has been based both on research in the general second language acquisition literature, and specific domain targeted research resulting from this initial consideration of the claims of applied linguists. In order to identify the various types of error occurring for Portuguese pronoun usage, research was undertaken to produce a corpus of actual errors, thereby providing a true reflection of possible problems without increasing the likelihood of predicting improbable errors. This is intended to aid the diagnostic capabilities of the system. The corpus has been used to classify errors according to their degree of difficulty, co-occurrence and incompatibility (i.e. where the existence of one type of error may prevent the occurrence of another). Similarly, actual student problems with the positioning of pronouns (regardless of whether the correct form of the pronoun was used) were recorded in order to estimate a likely acquisition sequence for the conditions governing pronoun placement.

Student reference to other languages in their learning of Portuguese has also been investigated in order to tackle the issue of language transfer (or analogy). Language learning strategies are also important; those relevant to the intelligent computer assisted language learning (ICALL) environment have been included, and students' use of these is traced by the system. Language awareness is encouraged by promoting learner reflection through making all system information available for learner scrutiny; not only domain information, but also transfer issues, learning strategy usage and the representations in the student model. It is anticipated that this open student model will encourage reflection by provoking the student to think about his own beliefs, and to defend himself if he disagrees with any representation in the model. Because this is a more personal perspective on learning (i.e. the student's beliefs about the domain are challenged; it is not simply the domain itself which can be presented and discussed), it is expected that learners will more readily enter into negotiation with the system (for examples see Bull & Pain, 1995).

In the case of our ICALL system, *awareness* is expressed in terms of language awareness. Applying the student model to other domains such as that of electrical circuits requires some consideration as to the meaning of this issue and its significance. We return to this in section 8. The other issues described are more straightforward. There are therefore grounds for believing that the ICALL student model is an instance of a generic student model. While still focussing on the language learning system as

[3] The discussion is here limited to European Portuguese. The situation is different in Brazil.

encompassing an instance of this student model, the issue of the generality of the model is addressed throughout the paper in terms of its applicability to learning about electricity in the context of electrical circuits. Thus, although Mr. Collins is currently contextualised in one system, this should be viewed simply as an example.

Figure 1 (section 2) indicates how the components of Mr. Collins are linked. However, in order to demonstrate the importance of each aspect in its own right, these will be discussed separately in the following sections.

4. Performance in the Target Domain

This section discusses the more traditional aspect of the student model; i.e. that part concerned with modelling a learner's knowledge and misconceptions. If a system is to successfully model learner performance, it should know the likely errors and where possible also their causes in order to manage the interaction in the most appropriate manner. Regardless of the theoretical approach to student modelling adopted in a system, in the early stages an empirical study of students' difficulties in the domain should be referred to in order to ensure the validity of the student model and explanations offered. Such domain-based research has been described for a variety of subjects; see, for example, Dumont (1993) for fractions, Escott and McCalla (1988) for programming, and Brna (1988) and Dupin (1987) for electrical circuits.

The example domain described in this paper is a particularly complicated area in European Portuguese; there are a variety of different problems which can arise. In order to identify the range of pronoun errors committed, a study of 47 undergraduate students of Portuguese was undertaken. The errors identified were included in the student model. The study was carried out over 5 weeks, and was based on (paper and pencil) multiple choice, translation and sentence transformation tasks. The tasks were given as homework, therefore all errors still occurred despite free access to the class grammar notes (covering all information necessary for correct performance). The study was then supplemented by a revision test taken by 13 of the 47 students. The learners were beginners in their fourth week of Portuguese when the study began. (The course consisted of 3 taught hours per week - 1 hour of grammar and 2 hours of conversation/exercises.) Most learners were native English speakers who knew Spanish, and many had also previously learnt French. This study and subsequent test are described in greater detail in Bull et al (in press).

Four main types of error were identified. These involved:

1. hyphens;
2. sentence structure;
3. the pronouns themselves (including both incorrect pronoun selection, and errors in phonetic contractions);
4. problems of pronoun placement.

The ICALL system currently deals with errors of types 1, 3 and 4. For discussion, the example of phonetic contractions will be focussed on in this section, as these appear to cause particular problems for students. Also difficult is the placement of pronouns. This will be dealt with in the following section.

Results indicate a lack of predictability for error types; i.e. the existence of one type of error does not necessarily exclude another type from being made by the same student, even in cases where the two errors appear contradictory. The following example of phonetic contractions with pronouns illustrates this point. The forms of the direct object pronoun are:

1st person singular:	me	1st person plural:	nos
2nd person singular:	te	2nd person plural:	vos
3rd person singular:	o/a	3rd person plural:	os/as

In cases where the direct object pronoun follows the verb, if the verb form ends in *r, s* or *z* and the pronoun is a third person pronoun, the final *r, s* or *z* is omitted and *l* is affixed to the front of the pronoun. In addition, when the stress is on the final syllable, or the verb form consists of only one syllable, a written accent is usually required.

Thus	**faz-o*	becomes	*fá-lo*	(he does it)
and	**quer saber-o*	becomes	*quer sabê-lo*	(he wants to know it).

The most common type of error observed amongst students who did not remove the final *r, s* or *z* before a third person direct object pronoun was the omission of the accent. This occurred most often in combination with correct addition of the *l,* though in about one third of cases the *l* was also omitted. It might be expected that in combination with the omission of the accent, a student would consistently either remember to add the *l*, or never do so (at least within one exercise). This type of behaviour would make assessment by the system easier, as the same combinations for each student could always be checked. Although this is indeed usually the case, i.e. most students committing final letter errors add the *l* but omit the accent, and another group omit both components, in one (sentence transformation) exercise four students sometimes added the *l* and sometimes did not. Therefore, although as a general rule a system can look for certain groupings of errors, it cannot be assumed that a student will always be consistent. Thus in the ICALL system presented here, the system must know the relative frequency of different errors eg. in the above example of non-omission of verb-final *r, s* or *z,* nearly all students will fail to provide the additional accent. Most will however remember to add the *l* to the pronoun, though a still significant proportion will not do this. A small percentage will sometimes add the *l,* and sometimes forget it. Until the system has obtained sufficient input to determine a particular learner's usual preference (if he has one), these error types will be sought in the order presented above (for greater computational efficiency).

A second example, similar to that described above, is the case of missing accents vs accents provided in inappropriate contexts (eg. in *unstressed* final syllables) vs the wrong type of accent (eg. *é* instead of *ê*). It could be hypothesized that a learner who often omits the accent will not provide the wrong accent or provide an accent in an inappropriate context, because in general he tends to not use these in obligatory contexts. However, this is not necessarily the case! 12 of the 47 students sometimes miss out the accent, while a near equal number (13) provide an accent in an inappropriate environment. 5 students provided examples of the wrong type of accent. For individual learners there were 6 co-occurrences of missing accent + accent in inappropriate context; about half the cases of missing and inappropriately produced accents therefore occurred in the same students. There were 3 combinations of missing and wrong accent types (and 2 cases of inappropriately provided and wrong type of accent - this latter combination of

error types cannot be regarded as incompatible). It can be seen from this that the student model of the ICALL system must allow the existence of such incompatibilities.

To summarise: actual student errors in the target domain of object pronoun use in European Portuguese have been researched to enable the system to identify probable misconceptions while at the same time avoiding anticipating or describing unlikely errors[4]. As occurs in other domains, it has been discovered that what appear to be incompatible error combinations can occur, and a means of representing this possibility is therefore essential. Separate *student and system belief measures* in the student model are calculated and used to predict the likelihood of a student making a particular error or error combination, based on recent input (see Bull, 1994b).

Similarly, in the context of students learning about electricity, studies of student performance have demonstrated a pattern of errors which indicates extensive and systematic misunderstandings. These errors have been frequently observed, and appear to occur within diverse approaches to teaching about electricity (Shipstone et al, 1988). An analysis of the classes of errors can be found in Brna (1988). Such information about both typical and more individual learner performance would be beneficial to the student modelling component of an environment for learning about electricity.

The remainder of this paper is concerned with modelling issues which are not directly related to the correctness or incorrectness of domain knowledge but which are nevertheless important in learning, though often ignored in student models.

5. Acquisition Order of the Target Knowledge

The order in which concepts should be taught has been a subject of interest for many years. Following Gagne's (1977) approach, instructional designers used to devise teaching sequences for concepts based primarily on the formal structure of the domain. However, this is not necessarily the way in which students acquire concepts. For example, Raven (1968) showed that a concept of momentum was often acquired prior to the concepts of mass and velocity, despite the fact that mass and velocity are the components of momentum. This illustrates the point that paths taken by students to acquire concepts is an important aspect for those modelling students.

Although there is still some debate, a natural acquisition sequence for the rules of a foreign language is generally acknowledged, and this acquisition order should tend to hold across learners. This is reflected through (overlapping) stages on a developmental sequence (see Pienemann, 1989).

The study introduced in the previous section also sought to identify whether it was possible to determine a typical acquisition sequence for the subset of Portuguese pronoun placement rules in the ICALL system described in this paper, even though this deals with a small area of Portuguese. If so, this information would be useful to the system in both the prediction of probable difficulties at each stage of a learner's development, and in the sequencing of material. The study cannot be widely generalised as the group involved was homogeneous; as stated in the previous section, all learners were either native or near fluent speakers of English, and the majority also had a high

[4] For further examples of errors see Bull (1994a).

level of Spanish. Nevertheless it was possible to observe some patterns which at least suggest an order of acquisition for the type of learner involved in the study, who is assumed to be typical of most UK students of Portuguese (see Bull et al, in press).

Although an acquisition order has been tentatively identified, some rules have not been distinguished within particular stages; for example, although the positioning of pronouns within negative clauses and open questions appears to be acquired before the positioning of pronouns in alternative contexts, from the data available it was not possible to distinguish the order of acquisition of these two rules from each other - some learners achieved better performance using negative constructions, while others initially had greater success with open questions. The acquisition order does not coincide with the order in which these students were taught the rules. Neither does it depend on the actual position of the pronoun in the sentence; i.e. pre-verbal and post-verbal pronoun placement alternate in the sequence, it is not the case that all rules for pre-verbal placement are acquired before those for post-verbal placement of the pronoun, and vice versa.

Having taken steps to identify an acquisition order for the target rules, it is useful to incorporate this information into the student model. The student model of the ICALL system is therefore not concerned only with the learner's current state, but also traces a student's developmental sequence (his learning history), and future performance is anticipated based both on this previous performance and also the typical acquisition order of rules. Thus the student model is a continuum, as depicted below:

learning history	-->	CURRENT STUDENT MODEL	-->	anticipated future performance	-->	expert (domain) model[5]

The current model represents the learner's present state (which, together with the learning history, is the more conventional component of the student model). However, in addition to the usual information on performance in, and understanding of the domain, this current representation also includes the learner's knowledge of other languages and the learning strategies used (see following sections), and also both user and system confidence levels in the student's performance.

It has been shown in this section that information about the acquisition sequence for the target rules has been incorporated into the design of the student model; this information may be used by the system to inform ordering of material in teaching and also to aid diagnosis. However, although an acquisition order can be postulated for the rules of pronoun placement, it is not possible to anticipate a typical order for the appearance of (non-position) pronoun-related errors, as learners can be inconsistent in their production of such errors. For example, the study of undergraduate students of Portuguese revealed that initially learners often omit the hyphen connecting a post-verbal pronoun to the verb (this may be because in this sample students were never explicitly informed that the hyphen is necessary - though examples were provided with hyphens clearly included). Nevertheless, some learners correctly use the hyphen in earlier stages, but later sometimes omit it. This may be due to carelessness, but whatever the cause, it is

[5] The component usually represented as a separate domain model is in this system located on the expert extreme of the student model continuum. Thus it may be used by the system as an autonomous model of the domain, or can also be viewed simply as the final stage of the learning process.

necessary for the system to be able to model such behaviour. Furthermore, the fact that a student is experiencing difficulty with a particular rule of pronoun placement does not allow associated non-placement error types to be anticipated. Therefore the student model continuum described above is based on the typical acquisition sequence for the rules of pronoun placement, but a large variety of individual non-placement errors is allowed for at each stage of the learning process.

This type of sequence model could also be useful in other domains. White and Frederiksen (1987) use a similar approach to that of our ICALL system, comprising a progression of mental models for electrical circuit behaviour based on the desired (expected and correct) evolution of knowledge. Although the two approaches are similar, a major difference is that White and Frederiksen's modelling is expert-based, as they claim that through the design of appropriate model progressions and associated problems, incorrect model transformations will be improbable. Our approach differs in that we do not believe that learning can be guaranteed to occur without the formulation of misconceptions. This is certainly true in the domain of electrical circuits. Many researchers have reported the common occurrence of misconceptions arising in student learning about electricity (e.g. Brna, 1988, Dupin, 1987).

In language learning, students also make (consistent) errors. Thus the student model of our system is more learner-centred; i.e. its focus is on student progression through the domain (which includes a consideration of misconceptions), rather than a straight sequence of domain-centred models. In the domain of learning about electricity, research has provided evidence of patterns in the development of understanding about electricity (Osborne, 1981). Researchers have observed shifts in the conceptual understanding connected with the notion of current. There is evidence that there is a general move from a naive model of current to a more acceptable one via one or two other incorrect models (over a period of several years). A student model such as Mr. Collins is clearly relevant to this domain.

The previous section was concerned solely with the (more usual) representation of domain knowledge and misconceptions. However, representation of the typical learner's route of acquisition of the target material goes beyond representing actual domain performance to include external factors, which, although domain-related, do not strictly form part of the domain. We continue in the following sections by describing those components of Mr. Collins which are clearly outside the domain boundary.

6. Analogy

As with the previous issues included in our student model, analogy is also relevant in many domains. For example, Escott and McCalla (1988) emphasise the role of incorrect analogy in novice LISP programmers. Gentner and Gentner (1983) claim that inference patterns in electricity are likely to vary according to the analogy used (water-flow or moving-crowd). This is obviously similar to the position in language learning, where different base languages may lead to differing performance in the target language. In the field of second language acquisition, language transfer is a term which is used to refer both to a process of transfer and to the product of such a process. Here we assume that the process of language transfer is an analogical process (which is sometimes performed consciously, and sometimes not).

Cross-linguistic influences have been identified as important factors influencing second language acquisition (Odlin, 1989); learner errors may therefore not only have their cause in the overgeneralisation of target forms, but may also be a result of transfer from another language. Consequently, in order to be effective, an ICALL system should broaden its knowledge by looking beyond the actual target domain to also take account of other languages known by the student in order to consider the influence of these on performance in the target language. This includes not only the student's native language, but also other foreign languages known. The importance of this can be seen below, in relation to the ordering of pronouns:

target language: *Portuguese*

1. verb, pronoun.	eg. Compra-os.
2. neg, pronoun, verb.	eg. Não os compra.
3. aux, pronoun, verb.	eg. Tem-nos comprado

other language: *English*

1. verb, pronoun.	eg. He buys them.
2. neg, verb, pronoun.	eg. He does not buy them.
3. aux, verb, pronoun.	eg. He has bought them.

other language: *Spanish*

1. pronoun, verb.	eg. Los compra.
2. neg, pronoun, verb.	eg. No los compra.
3. pronoun, aux, verb.	eg. Los ha comprado.

other language: *Catalan*

1. pronoun, verb.	eg. Els compra.
2. neg, pronoun, verb.	eg. No els compra.
3. pronoun, aux, verb.	eg. Els ha comprat.

other language: *French*

1. pronoun, verb.	eg. Il les achète.
2. neg1, pronoun, verb, neg2.	eg. Il ne les achète pas.
3. pronoun, aux, verb.	eg. Il les a acheté.

The rule applicable for the position of pronouns in affirmative statements in Portuguese dictates that the pronoun should follow the verb (rule 1). The same is true in English; however Spanish, Catalan and French differ, requiring the pronoun to be pre-verbal. In negative sentences (rule 2), the rules for pronoun placement are similar in Portuguese, Spanish, Catalan and French, but different in English. There are also situations in which the positioning of the pronoun in Portuguese is not analogous to the position of the pronoun in any of these other languages, as occurs in the perfect tense (rule 3). In Portuguese the pronoun is placed between the auxiliary and past participle, while English places it in final position, after the past participle. In Spanish, Catalan and French the pronoun occurs before the auxiliary. From these three examples it can be seen that errors of pronoun placement, although possibly a result of overgeneralisation of other Portuguese rules, may instead be based on analogy with another language. Moreover, it is not necessarily an easy task for a system (or a teacher, for that matter) to determine which of a learner's background languages is the most probable cause of such an error; a learner who knows both Spanish and French, for example, could be comparing

(consciously or not) with either of these languages when placing the pronoun pre-verbally in affirmative main clauses.

Accordingly, in addition to the Portuguese rules, the ICALL system also has a knowledge of equivalent rules for pronoun placement in English, Spanish, Catalan and French in order that the system may recognise transfer problems from these languages. Clearly other languages could also be included, however the above are sufficient for the purpose of evaluating Mr. Collins (English, Spanish and French are the main languages common to the target users). As stated above, difficulties may have their source in any of the languages which are known by a student. Thus it is important to determine which of a learner's background languages is the most likely cause of transfer. Research suggests that proficiency in a background language may affect the extent of transfer (see Ringbom, 1983). However, this is not the only factor involved; Kellerman (1977) states that learners tend to transfer more from languages which they perceive to resemble the target language than from those they consider more distant. These claims have been further investigated in the context of the ICALL system described here; a questionnaire study involving 20 learners of Portuguese was undertaken. Students were asked about languages known, their proficiency in these languages and whether they consciously compared Portuguese to any of these other languages. (This excludes the possibility of considering non-conscious comparison, though it has been assumed that where this occurs, it may take place in similar circumstances.) The results are as follows:

Every student of Portuguese who also knows *Spanish* (15 in total), compares Portuguese with Spanish (all students were either advanced or near fluent in Spanish[6]).

Only two students compare to *English*. One knows no Spanish, and perhaps for this reason looks to English, and the other compares to all languages known (English, Spanish and French). All members of the group are native/near fluent speakers of English.

Nearly twice as many learners transfer from *French* as not; the tendency is for transfer to occur from advanced level upwards. All three non-Spanish speakers transfer from French (possibly because this is the most similar alternative).

No-one transfers from *German*, despite a high level of knowledge of this language by all speakers. (This is probably due to the relative dissimilarity of Portuguese and German.)

A few other languages not perceived as closely related to Portuguese are not consciously referred to: *Danish, Polish, Dutch, Irish* and *Ancient Greek*. Apart from the Danish speaker, speakers of these languages had a low level of proficiency. (The Danish speaker compares Portuguese to both Spanish and French, and maybe for this reason does not compare with the more distant Danish.) One learner compares Portuguese with *Japanese* (in which he is a beginner). This last example does not fit the pattern of similarity and proficiency leading to transfer.

---

6 It would of course be interesting to see the extent of reference to Spanish by speakers of lower proficiency in this language.

It was not possible to observe a pattern for *Italian* (possibly because of the low level of proficiency of speakers of this language.)

The above results indicate a clear tendency for similarity of, and proficiency in a background language to lead to comparison (and probably transfer). Proficiency alone is not sufficient (see English/German), and similarity alone is also not sufficient (see French/Italian). Proficiency + Similarity is a much better predictor (though this is not infallible). Only 11 out of 73 possible cases do not fit the pattern of proficiency + similarity leading to transfer:

i. 4 of these 11 cases fulfil similarity + proficiency, but do not result in (conscious) comparison.
ii. 2 of these 11 cases fulfil only proficiency, but comparison is reported despite dissimilarity.
iii. 4 of these 11 cases fulfil only similarity, but comparison is reported despite low proficiency.
iv. 1 of these 11 cases is comparison with a dissimilar language known only at a basic level.

Of these 11 exceptions, 6 meet one half of the requirements for the assumed case of proficiency + similarity leading to transfer. Only 5 (under 7% of total: i & iv above) seem completely atypical.

These results indicate the need to consider students' other languages when trying to model their learning of a new foreign language. A learner profile (forming the initial point on the student model continuum) is constructed co-operatively by the student and system at the start of the first interaction. This profile includes information on a learner's knowledge of other languages, i.e. his proficiency and perception of similarity to Portuguese of each. (Recall that it is a learner's *perception* of similarity between languages which is thought to influence the extent of transfer, rather than some objective measure, or some assessment of similarity made by the system.) The above information on the importance of similarity and proficiency as a reasonable predictor of transfer has been used to calculate the most appropriate order in which rules of an individual's background languages are parsed by the ICALL system when it is seeking the cause of transfer problems in order to ensure that the most likely source of an error is tested first, and the most probable representation is the one recorded in Mr. Collins (for cases in which there may be doubt).

It might be thought that these results are somewhat specific to second language acquisition. However, there has been relatively little comparable research as to how learning a new domain is influenced by multiple sources of knowledge. Some recent research has moved towards investigating this issue, for example, Johsua & Dupin (1993) demonstrated how analogies drawn from different domains were separately useful for teaching about electricity. They showed the benefits of models of current based on 1. an analogy with a chain of train carriages forming a continuous loop around a track (a model which is not so useful for parallel circuits) and 2. an analogy with heat flow in a refrigerator. Some recent evidence exists for the classroom effectiveness of combining knowledge sources (Alexander et al, 1994). However, to our knowledge, few systems actually try to exploit the possibility of different sources of transfer. It is believed that the above results could have significance for many domains.

7. Learning Strategies

Another issue which is not actually part of our target domain, but is nevertheless related to it, is that of learning strategies. The consideration of this aspect is similar in one sense to the discussion of the learner's background languages in the previous section, as, in common with other languages, a student also brings his own learning strategies to the task. Therefore it is useful for the ICALL system to know about a student's approaches to learning in order to model an individual's learning effectively. However, although the learning strategy component is not strictly part of the domain knowledge, it is nevertheless important, as one of the objectives of the system is to promote efficient learning strategy usage within this domain[7].

A description of the distinction between the target Portuguese domain and the consideration of language learning strategies is useful here. The target domain comprises the expert extreme of the student model continuum. Perturbations and variations of this knowledge will tend to occur across the whole student model continuum, while the representation of correct knowledge steadily increases along the continuum. Learning strategies are not represented through the entire student model; they form a separate module. The appropriate strategies for a particular student are contained at each stage in the historical and current sections of the student model.

The following learning strategies (drawn from O'Malley & Chamot's classification of language learning strategies, 1990) were selected as being both relevant to, and implementable in the ICALL environment:

Metacognitive Strategies:
organizational planning (of strategies), self-monitoring (production/visual/ strategy/double checking), self-evaluation (performance/ability/strategy use).

Cognitive Strategies:
resourcing, note-taking, grouping, summarization, deduction, substitution, translation, transfer, inferencing.

Social Strategies:
co-operation, question for clarification.

An investigation of language learners' strategy use was undertaken using Oxford's Strategy Inventory for Language Learning (SILL version 1.5, Oxford, 1990). This study (described in Bull et al, 1993) revealed that different students used a wide range of language learning strategies. These strategies could not be predicted from a learner's language background, language learning experiences, etc. A separate investigation of 20 students' approaches to the completion of a specific short Portuguese multiple choice exercise on pronoun placement indicated widespread unprompted use of the following strategies: resourcing, studying material prior to attempting exercise, guessing, checking work both during and after task completion. The fact that students use various learning strategies (even for short exercises) justifies their inclusion in the ICALL system and representation in the student model. Moreover, the 16 students who took the SILL evaluation were without exception interested in improving their use of learning strategies,

---

7 Transferability of learning strategies is not dealt with here.

even those students who were not particularly interested in learning strategies as a topic in itself.

Mr. Collins records a student's use of each type of learning strategy from the O'Malley and Chamot classification presented above (for details see Bull et al, 1993), thereby enabling the system to not only tackle specific domain issues, but also to deal with problems relating to carrying out the task (an issue often ignored from the student modelling perspective, unless the domain itself is one of problem solving).

Learning strategies are, of course, not only relevant to language learning, but are also used (spontaneously or after prompting) in other domains. Nisbet and Shucksmith (1986), in a consideration of the use of learning strategies in general, list six of the most frequently mentioned strategies as: asking questions, planning, monitoring, checking, revising, self-testing. The similarity to strategies in language learning is evident.

It is easy to see how many of the strategies used in the ICALL system could be represented for other domains, for example, a student learning about electrical circuits may undertake resourcing in a similar manner to the foreign language learner, but rather than seeking translations in a dictionary, or inspecting grammar rules or example sentences he would be consulting descriptions or diagrams of the functioning of electrical circuits, or example simulations of their use. Like the foreign language learner, he may also make notes related to his new knowledge, uncertainties, etc. Even a strategy which may at first appear to be very language specific can have an equivalent in another domain; consider translation. As a language learning strategy this is defined as more or less word-for-word translation. In a domain such as electrical circuits this could be conceived of as deliberate attempts to reuse concepts from one area in another, without trying to modify the concepts to fit this new context.

8. Awareness and Reflection

Collins and Brown (1988) emphasise the advantages of using the computer to promote learner reflection. In language learning, reflection can in part be related to language awareness. This issue has recently been much discussed in applied linguistics. One interpretation of the term language awareness considers the learner's awareness of patterns in language, leading to attempts to develop learner consciousness of such patterns by making these more explicit.

The concept of 'language awareness' itself is not one that transfers very easily to other domains. Consider, for example, the notion of 'electrical circuit awareness' - what might it mean? There is an equivalent, eg. thinking about electricity in system terms as advocated by Haertel (1982). Haertel draws on Cohen et al's (1983) work on misconceptions of electricity where it is argued that many errors are caused by students using 'local reasoning' and not following through the implications of alterations made to a circuit. The advocation of 'system thinking' is also at the root of Johsua & Dupin's (1993) recent work on electricity. Given that there is an equivalent concept to language awareness in the domain of electricity, it is nevertheless still the case that much more work is needed to clarify the meaning for researchers in this and other domains.

Although there is disagreement about the extent of its importance, there is evidence that greater awareness of language form is facilitative of language learning

(Schmidt, 1990, van Baalen, 1983). Ellis (1992) concludes that although practice may have some value, what is more important is the development of explicit knowledge. He states that consciousness-raising leads indirectly to the acquisition of implicit knowledge, i.e. it has a delayed effect. This takes place firstly by facilitating the student's general awareness of, and comparison of features which will enable integration to take place, and secondly the explicit knowledge gained will be available for when the student is ready to process it.

Therefore the ICALL system, in addition to providing opportunity for practice in the traditional sense, seeks to enhance learner awareness of language in that all rules for pronoun placement are made explicit[8], as are the phonetic changes often required. Moreover, these rules may also become the object of discussion between the student and system; an additional source for the development of learner awareness. Language awareness is not explicitly represented in the student model itself, but should occur as a direct result of the openness of the model (and the system in general). Furthermore, it is not only the target rules (and transfer, learning strategies and acquisition order information) which are inspectable by the student: even the student model is open to learner scrutiny, and in cases where the learner disagrees with the contents of the model, repairs may be negotiated by the two parties[9] (see Bull & Pain, 1995). This should lead to a more accurate student model as the student is a good source of information regarding his own beliefs, but more importantly, rather than placing an additional burden on the student, this act of co-operative maintenance of the student model is aimed at promoting learner reflection, thereby leading to enhanced awareness and improved learning.

Although we stated above that the language awareness expected to result from this explicit exposure to rules and discussion of the student model may not necessarily be easily transferred to alternative domains, the reflection which is caused by such model negotiation is certainly relevant. An example of a similar attempt to encourage a form of reflection in the domain of electrical circuits can be found in QUIMON (Quest Instructional Monitor, Feurzeig & Ritter, 1988). Nevertheless, their focus is more on obtaining increased information for the student model about a student's goals, plans and hypotheses, than on promoting reflection. However the authors do state that this interaction 'helps shape a student's thinking', but it is also stated that a student is not expected to analyse all his actions, and further, it is suggested that better procedures could be designed to enable student information to be requested by the system less often. Although this is less intrusive, it also limits opportunities for encouraging reflection. Given that information is explicitly requested from the learner, it could help the student if this model construction process were focussed on more specifically as an important aid to learning in its own right.

Inspection and negotiation of Mr. Collins occurs through menu selection; presentation of information to the student is based on templates producing text explanations. Discussion is based on separate belief measures reflecting the (possibly different) viewpoints of the student and system. The student's belief measures are based on the learner's own assessment of his confidence in his input (entered together with this input). The system's belief measures (relating to system assessment of learner performance) are based on the average result (i.e. correct/incorrect and error type) of the

---

8 Examples are also provided.

9 Implementation of the discussion of target rules is not yet complete. Discussion and negotiation of the contents of the student model is currently possible.

student's last five attempts to use a grammatical rule (see Bull, 1994b). Any discrepancies in these two sets of belief values are made explicit, and should provoke argumentation (explanation, justification, defence by both agents), this resulting also in learner reflection. The representation of the student's beliefs may be changed easily by the learner. However, the system's beliefs may only be altered as a result of changes in learner behaviour, or through valid argumentation by the student. Because the student is presented with the system's assessment of his own beliefs this becomes more personal, and it is therefore anticipated that the student will be more motivated to defend himself. We believe he is more likely to argue if he disagrees with an expert's representation of his own views, than he would about an expert's view of the domain.

In summary: a learner's awareness of language form and approaches to learning is enhanced by enabling student inspection and discussion of all aspects of the system. This is aimed at fostering reflection, leading to improved learning. Additionally, learner input to the student model will result in the construction of a more accurate model from which the system may draw further inferences. Clearly this collaborative approach to student modelling is also relevant to learning in other domains. An example of this is the work of Kay and Crawford (1991). They suggest a similar co-operative approach to modelling the learner, using as an example the learning of an operating system. However, there are differences between the two approaches; Kay and Crawford propose allowing learners (and teachers) to directly change the student model as they feel appropriate, and Kay (1994) describes a system to coach the use of a text editor, stating that it may override user-provided information in cases of conflict. In contrast, our ICALL system retains two perspectives on the student; one easily alterable by the student, and the other under the control of the system - reflecting the (possibly conflicting) beliefs of the system about the student. (Both belief measures are accessible to student viewing.) It remains to be seen which approach is better in practice, though our approach of maintaining two separate views of the student opens avenues for user/system discussion of the student model, which will potentially further promote learner reflection.

9. Conclusion

The need for student models to contain information beyond representations of the learner's current understanding, history of problem solving and stereotypical learner characteristics has been argued, and Mr. Collins, a student model currently contextualised in an ICALL system has been presented in order to illustrate the possibilities of such a model. The student model is necessarily domain specific, as issues related to language learning in general, and the learning of Portuguese in particular are considered important, and therefore should be represented in the student model if the aim is to develop an effective model. The issues involved were selected as a result of research undertaken in the field of second language acquisition, and consist of the following: performance in the target domain, acquisition order of the target rules, analogy or transfer, learning strategies and language awareness. The first four of the above are explicitly represented in the student model, and language awareness results from the general transparency of the system and the learner model. The openness of the student model has provided the basis for a co-operative model construction process, which benefits both the system (by providing a more accurate student model), and the learner (through promoting reflection).

The discussion has here centred on a system for foreign language learning. References to other work, and in particular generalisation to knowledge about electrical

circuits have shown the applicability of modelling similar aspects outside the target material in other domains. We have shown that the architecture of Mr. Collins is generalisable, though we do not claim that each of the issues described here will be equally important in all domains, or that this set of issues will be exhaustive for every domain. However we believe that our approach of considering more general learning issues is relevant in other areas, and it would not be difficult to adapt the architecture of Mr. Collins to accomodate the needs of a different domain.

Acknowledgements

This is an extended version of a paper presented at the International Conference on User Modeling, 1994.
We would like to thank David Frier for providing the student exercises on which the learner performance and acquisition order research was based, and for assisting in the interpretation of student errors.
Work on the ICALL system described here was funded by the Economic and Social Research Council, studentship no. R00429124033.

References

Alexander, P.A., Kulikowich, J.M. & Schulze, S.K. How Subject-Matter Knowledge Affects Recall and Interest, *American Educational Research Journal*, vol. **31**, no. 2, 1994, pp. 313 - 337.

Anderson, J.R., Farrell, R. & Sauers, R. Learning to Program in LISP, *Cognitive Science,* **8** (2), 1984, pp. 87 - 129.

Brna, P. Confronting Misconceptions in the Domain of Simple Electrical Circuits, *Instructional Science*, vol. **17**, 1988, pp. 29 - 55.

Boohan, R. Using Computer-Based Questionnaires to Diagnose Students' Models of Electricity, in Caillot (ed), Learning Electricity and Electronics with Advanced Educational Technology, *NATO ASI Series F,* Vol. **115**, Springer-Verlag, Berlin, 1993, pp. 173 - 195.

Bull, S. Learning Languages: Implications for Student Modelling in ICALL, *ReCALL* vol. **6**, no. 1, 1994a, pp. 34 - 39.

Bull, S. Student Modelling for Second Language Acquisition, *Computers and Education* vol. **23**, no. 1/2, 1994b, pp. 13 - 20.

Bull, S. & Pain, H. "Did I say what I think I said, and do you agree with me?": Inspecting and Questioning the Student Model, to appear in *Proceedings of World Conference on Artificial Intelligence and Education,* Washington DC, August 1995.

Bull, S., Pain, H. & Brna, P. Student Modelling in an Intelligent Computer Assisted Language Learning System: The Issues of Language Transfer and Learning Strategies, *Proceedings of the International Conference on Computers in Education*, Taipei, Taiwan, 1993, pp. 121 - 126.

Bull, S., Pain, H. & Brna, P. Mr. Collins: a Collaboratively Constructed, Inspectable Student Model for Intelligent Computer Assisted Language Learning, *Instructional Science*, in press.

Catt, M. & Hirst, G. An Intelligent CALI System for Grammatical Error Diagnosis, *CALL* vol. **3**, 1990, pp. 3 - 26.

Cohen, R., Eylon, B. & Ganiel, U. Potential Difference and Current in Simple Electrical Circuits: A Study of Students' Concepts, *American Journal of Physics*, vol. **51**, no. 5, 1983, pp. 407 - 412.

Collins, A. & Brown, J.S. The Computer as a Tool for Learning Through Reflection, in Mandl & Lesgold (eds), *Learning Issues for Intelligent Tutoring Systems,* Springer-Verlag, New York, 1988, pp. 1 - 18.

Dumont, B. A Diagnostic Knowledge-Based System for Fractions, in Nwana (ed), *Mathematical Intelligent Learning Environments*, Intellect Books, Oxford, 1993, pp. 189 - 198.

Dupin, J-J. Conceptions of French Pupils Concerning Electric Circuits: Structure and Evolution, *Journal of Research in Science Teaching*, vol. **24** no. 9, 1987, pp. 791 - 806.

Ellis, R. *Second Language Acquisition and Language Pedagogy,* Multilingual Matters Ltd., Clevedon, 1992.

Escott, J. A. & McCalla, G. I. Problem Solving by Analogy: A Source of Errors in Novice Lisp Programming, *Proceedings of ITS-88*, Montreal, 1988, pp. 312 - 319.

Feurzeig, W. & Ritter, F. Understanding reflective problem solving, in Psotka, Massey, & Mutter (eds), *Intelligent Tutoring Systems: Lessons Learned*, Lawrence Erlbaum Associates, Hillsdale, NJ, 1988, pp. 435 - 450.

Gagne, R.M. *The Conditions of Learning*, Holt-Saunders, 1977.

Gentner, D. & Gentner, D. R. Flowing Waters or Teeming Crowds: Mental Models of Electricity, in Gentner & Stevens (eds), *Mental Models*, Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1983, pp. 99 - 129.

Haertel, H. The Electric Circuit as a System: A New Approach, *European Journal of Science Education*, vol. **4**, no. 1, 1982, pp. 45 - 55.

Johnson, W.L. & Soloway, E. PROUST: Knowledge-Based Program Understanding, *IEEE Transactions of Software Engineering,* **SE-11** (3), 1985, pp. 267 - 275.

Johsua, S. and Dupin, J-J. Using "Modelling Analogies" to Teach Basic Electricity: A Critical Analysis, in Caillot (ed), Learning Electricity and Electronics with Advanced Educational Technology, *NATO ASI Series F*, vol. **115**, Springer-Verlag, Berlin, 1993, pp. 229 - 249.

Kay, J. *The UM Toolkit for Reusable, Long Term User Models*, Report 94/3/36.2, Systems Centre, University of Sydney, 1994.

Kay, J. & Crawford, K. *Interactive Learner Models as a Co-operative Learning Tool*, Report 91/3/5.1, Systems Group, University of Sydney, 1991.

Kellerman, E. Towards a Characterisation of the Strategy of Transfer in Second Language Learning, *Interlanguage Studies Bulletin* (**2**)1, 1977, pp. 58 - 145.

Nisbet, J. & Shucksmith, J. *Learning Strategies*, Routledge and Kegan Paul, London, 1986.

Odlin, T. *Language Transfer, Cross-Linguistic Influence in Language Learning*, Cambridge University Press, Cambridge, 1989.

O'Malley, J.M. & Chamot, A.U. *Learning Strategies in Second Language Acquisition*, Cambridge University Press, Cambridge, 1990.

Osborne, R.J. Children's Ideas about Electric Current, *New Zealand Science Teacher*, vol. **29**, 1981, pp. 12 - 19.

Oxford, R. *Language Learning Strategies, What Every Teacher Should Know*, Heinle & Heinle Publishers, Boston, Massachusetts, 1990.

Pienemann, M. Is Language Teachable? Psycholinguistic Experiments and Hypotheses, *Applied Linguistics* vol. **10**, no. 1, 1989, pp. 52 - 79.

Psillos, D., Koumaras, P. & Valassiades, O. Pupils' Representations of Electric Current before, during and after Instruction on DC Circuits, *Research in Science and Technological Education*, vol. **5**, no. 2, 1987, pp. 185 - 99.

Raven, R.J. The Development of the Concept of Momentum in Primary School Children, *Journal of Research in Science Teaching* vol. **5**, 1968, pp. 216 - 223.

Ringbom, H. Borrowing and Lexical Transfer, *Applied Linguistics,* vol. **4**, no. 3, 1983, pp. 207 - 212.

Schmidt, R. W. The Role of Consciousness in Second Language Learning, *Applied Linguistics,* vol. **11**, no. 2, 1990, pp. 129 - 158.

Schuster, E. The Role of Native Grammars in Correcting Errors in Second Language Learning, *Computational Intelligence,* vol. **2**, 1986, pp. 93 - 98.

Shipstone, D.M., Rhoneck, C., Jung, W., Karrqvist, C., Dupin, J.J., Johsua, S. & Licht, P. A Study of Student Understanding of Electricity in Five European Countries, *International Journal of Science Education*, vol. **10**, no. 3, 1988, pp. 303 - 316.

van Baalen, T. Giving Learners Rules: A Study into the Effect of Grammatical Instruction with Varying Degrees of Explicitness, *Interlanguage Studies Bulletin,* **7** (1), 1983, pp. 71 - 100.

Wang, Y. & Garigliano, R. An Intelligent Language Tutoring System for Handling Errors caused by Transfer, *Proceedings of ITS-92*, 1992, pp. 395 - 404.

White, B.Y. & Frederiksen, J.R. *Causal Model Progressions as a Foundation for Intelligent Learning Environments*, Report No. 6686, BBN Laboratories, Cambridge, Massachusetts, 1987.