Foreign Language Tutoring and Learning Environment

Henry Hamburger¹ and Raza Hashim

George Mason University

ABSTRACT

In the FLUENT project, we contemplate and implement immersion-style computational environments for learning a foreign language. Immersion environments encourage the student to use the new language, not analyze or translate it. To this end, the student and the tutor engage in two-way, two-medium communication, where the two media are linguistic and spatial. Our design strategy is to attempt to capture in computational form some of the most effective aspects of the immersive methods that human teachers use with beginning students. The paper provides a rationale for immersion, presents a prototype system, and goes on to delineate component requirements for a full-fledged system.

A Significance and Objectives

The FLUENT² project is an attempt to specify and implement an effective computational environment for learning foreign languages. Specifically, it is an immersion-style foreign language tutoring and exploration system. Immersion means using the new language, not analyzing it, not consciously computing grammaticality, and not falling back into an already known language. It is suggestive that gradual immersion occurs and works in pure form for the child learning a first language (Hamburger, 1988). Of more direct relevance, immersion has also been used successfully as a second language technique, with human tutors/teachers. The foreign language classroom methods of "total physical response" and "the natural approach," in effect, immerse adults for the duration of each class period (Asher, 1977; Krashen and Terrell, 1983). With adults as with children, one tempers the demands of immersion by introducing new aspects of language only gradually. A prominent aspect of these methods is their use of visible entities, which are directly adaptable to images on the computer screen. Our design strategy is to attempt to capture some of the most effective aspects of immersive methods in computational form, to the extent possible. The system requires an internal knowledge representation that communicates in a tutorially sound way via linguistic and graphic interfaces. It also draws upon techniques of intelligent tutoring systems.

Substantial success for immersion should not be too surprising, in light of the following age-independent rationale. First, immersion, by its very nature, maximizes the time spent actually practicing the language to be learned, while minimizing potentially disruptive shifts between languages. A subtler but perhaps more crucial point is that immersion may undermine the often strong temptation to use the crutch of understanding by translating from the foreign language (L2) to one's own language (L1). Understanding by translation makes the comprehension of an

L2 utterance into a two-step mental process: translation to L1, followed by the usual process of interpreting an L1 input. This compound process would seem likely to foster errors of interference, especially the importation of L1 grammatical constructions. Moreover, with its extra step, it seems bound to be inefficient. Finally, such translation surely must impede what is often called "thinking in a language," regarded by many fluent L2 speakers as a significant breakthrough in their ability. McLaughlin et al. (1983) make a similar information-processing argument concerning explicit knowledge of grammatical rules. Immersion may yield acquisition rates that are slower in some ways than other methods. For example, it has been shown that certain mnemonic devices can speed up vocabulary learning, ostensibly by inducing a three-step process, one that is even more indirect than the two-step one above. But such results are not direct indicators of progress toward the more important long range goal of fluency, for which we believe immersion-style education will be more efficacious.

Adult language learners often have troubles that child first language learners do not share. Though it is possible that biological development and first language interference play a role, note also that adults may try too quickly to put their new language into service in abstract realms. Further, children's immersion is gradual: people speaking to children tend not to inundate the very young with the high seas of advanced syntax, but rather to use short, simple sentences. Both of these properties presumably keep down cognitive load, giving language learning a chance (cf., Sweller, 1988). An ironic related possibility is Newport's (1990) hypothesis that adults' greater short-term memory lures them into a combinatorial explosion of language-to-meaning mappings. Finally, close observation reveals that children rarely receive syntactic correction. Some initial tolerance of grammatical errors is also part of our approach, and is supported by observation and argument going back to Dulay and Burt's (1974) discussion of "goofing" in their comparison of child and adult second language learning. Consequently, although the system should always conform to standard grammar, we (along with most people talking to children and many foreign language instructors) accord perfect grammar a lower priority than the successful comprehension and communication of ideas at roughly normal speaking speed. The spirit of a naturalistic approach would certainly seem to demand beginning with spoken language before written, though to ease technical demands we have begun development in written mode.

The most distinctive characteristic of FLUENT, even in its current prototype form, is its fine-grain two-medium (graphic as well as linguistic) interaction between student and tutor. Within realistic everyday scenes, coordinated by everyday goal structures, both student and system can both discuss and manipulate objects in meaningful ways. For example, in some dialog modes, the student is empowered to move or alter objects, specifically to do things like making a character's hand turn a faucet on, pick up soap, get it wet, and wash, rinse and dry things, etc. These screen actions alter an internal model, and can lead the tutor to describe an action or its resulting state or to critique an action in terms of a common sense goal structure. These kinds of capabilities, together with an appropriate tutorial module, make possible the immersion approach, in which a student can extend his/her understanding to include a new language aspect that is introduced in a context that makes its meaning clear. Indeed context and interaction can enhance the effectiveness of visible objects so well that a skilled human teacher can convey abstractions, not just concrete notions, as Berlitz (1925, page 4) indicated quite awhile ago: "En introduisant des mots abstraits, il faut ... les employer de façon que le sens soit bien clair par le contexte."

To place this work in computational perspective, it is well to begin with the pioneering artificial

intelligence system for foreign language learning by Weischedel et al. (1978). Though its pedagogy consists only of questions and answers about a small number of texts, the paper sets out many key issues, opportunities and challenges. Recent work sharing some of these objectives includes the sophisticated linguistic diagnosis in the intelligent tutor of Fum et al. (1988). Like other natural language processing systems, it has a strategy for constraining the subject matter to reduce ambiguity, relying on linguistic context, as in the task of replacing "write" in "I (write) the letter but I can't find a stamp." Here the intended response is "have written." On the basis of several responses the system analyzes the student's apparent misconceptions concerning tense and aspect. A more natural task, one that constrains subject matter by the use of pictures, is that of Underwood (1987), who has students of Spanish stating ways in which one picture differs from another. They may try to enter a whole sentence, like "Hay una ventana en casa," and are informed, in L2, of errors such as the missing article before "casa" in this example. Beyond pictures, there is the prospect of incorporating animation. In the FLUENT project, we have experimented with what we will call partial animation. Other projects have already gone to videodisc technology; see Murray (1987) and Swanson (1989). Sophisticated multi-medium work including language and graphics is being carried out by Bobrow et al. (1989), in work that also involves multiple applications. Graphical interaction in the educational arena is perhaps best exemplified by Munro and Pizzini (1990).

The visual and linguistic context strategies of constraint can of course both be put to good use, separately or together, focusing on whichever is most effective for a particular purpose. The argument here is only that pictures are especially advantageous for use with beginners, for whom linguistic context would present a bootstrap problem, and that it is indeed feasible to make use of pictures for a broader range of notions than one might casually think. As the student progresses, it gradually becomes possible to rely on linguistic context, and indeed it ultimately becomes important to do so, since it seems quite hopeless to convey pictorially the full range of, say, tense and aspect distinctions.

B The FLUENT Prototype

To make the discussion more concrete, we now briefly describe the prototype version, FLUENT-I developed by Tucker Maney, before laying out the planned architecture and conceptual structure in more detail in the next section. The current version, built as a feasibility study, embodies the pedagogical strategy propounded above and to some degree the capabilities that are considered in the next section. Although this work has not yet achieved the degree of breadth, generativity, modularity, student-sensitivity, and complexity of the planned version, the prototype is already sufficiently elaborated to give users a clear sense of the immersion experience and let them learn some Spanish in a memorable way. Users have found it highly engaging.

This version runs on either a Mac-II or a Mac-SE and is written in the language cT.³ A word about cT may be of interest, given that we found it useful for prototyping. Though intended for authors who need not be expert programmers, this language has the control structures, including recursion, that one generally associates with ordinary procedural programming languages. It is somewhat less satisfying in the data type department, providing strings and arrays, but no pointers (not to mention lists) and no records (or structures, not to mention objects). It does provide a convenient development environment. One of its strengths for educational authors is its capacity for establishing interaction with the learner. However, having our own agenda of interactive techniques, for example the use of partial animation, we made little use of these tools.

A key virtue for our purposes was the strength and elegance of the graphics portion of the language. Finally, cT does not provide easy access to sound, which clearly would be desirable.

Four kinds of domains have been developed to various degrees, introducing several of the objective concepts and attitude categories catalogued by Van Ek et al. (1980). Collectively these domains support instruction in many aspects of language. One of the domains deals largely with objects, their attributes and parts; another with location and directions; and a third with numbers and time. The most developed of the domains stresses actions with goal structures. In it, a human figure can purposefully manipulate objects in a washroom. As in all the domains, the system engages the student in a tutorially appropriate, two-medium dialog, using Spanish at various levels of linguistic difficulty. The domains themselves are not intended to present any new non-linguistic information or demand any but the most effortless reasoning. They embody only the most ordinary common sense knowledge, such as what faucets and clocks do, and the interface allows manipulation of the objects in very straightforward ways. Thus the student's cognitive effort can be focused on the new language, and new aspects of that language can be understood from the clear context.

The figure in the washroom domain has a movable hand connected to his shoulder by a straight arm that 'rubber-bands' with the hand as the hand moves about on the screen; see Figure 1. The hand can be manipulated by the student or the tutor, depending on the dialog type, to pick up and/or use the other objects in the scene, including faucets, soap, and a light switch. In one dialog type, the student first selects a goal, such as having the figure wash its own face. The student then goes on to carry out suitable moves by controlling the partial animation on the computer screen with a mouse, while the tutor makes relevant comments and suggestions, all in Spanish. This kind of dialog requires of the student only the recognition of language. There is another dialog mode, however, in which the student must exercise various degrees of language production ability, from a single word to a phrase to a very short and constrained sentence. The tutor's responses are usually in the form of language, but can include screen moves to perform actions specified by a student who has sufficient production capability to do so. In either case, the tutor is directly responsive to whatever the student has just done or said. The responses can vary according to different levels of general achievement the student may have attained, though no student model has been implemented yet.

FIGURE 1 ABOUT HERE

The language output approach in the current version, which is specialized to Spanish, is to select verbs and nouns that correspond to appropriate operators and operands. The latter hinge in turn on whatever progress the person in the figure has made on various direct or indirect subgoals of the current main goal. A typical main goal is getting washed, and one of its subgoals is to get the face soapy, which in turn requires picking up the soap. The treatment of article takes advantage of the limited context. Since there is only one towel, one piece of soap, etc., definite articles are appropriate throughout the interaction. This simplification allows the student to focus on other aspects of language, and it allows the implementer to focus on other components of the system. We will, however, import and adapt more sophisticated natural language processing capabilities for generation as well as for understanding. Context also helps with pronouns, since whatever the student's role in a dialog, he or she is always interacting with exactly one other agent, whether it is the tutor or the figure on the screen. Consequently, the referents of first person and second

person pronouns are always unique, as they would be in a simple real conversation. In a continuing related project with David Canody, we have developed in Common Lisp a more flexible natural language generator that produces descriptions in English for simple objects and states in a part-whole mini-domain. This system takes as its input a frame that is part of a reasoning system with forward and backward chaining, developed in another project (Hamburger et al., 1990).

How much of a language can we communicate clearly in the spatial mode? This question is crucial for the success of immersion and gets an encouraging preliminary answer from the experience with the prototype. We find that certain verbs and adverbials that one might not think would be easy to present graphically are made usable and clear by the existence of a common sense goal structure. The system is able, for example, to talk in a natural way about various actions one "needs" to do, or has "not yet" done, or that "help" with the stated current goal. The tutor's remarks, especially suggestions of what to do next, take into account what subgoals have already been achieved. One example of a structured goal is the one mentioned earlier, washing one's face. That particular goal has the subgoals of getting the face soapy, then rinsing and finally drying it. Each of these is further subdivided into primitive screen acts such as turning the hot or cold water faucet on, and later turning it off, picking up the towel and hanging it up when done.

The system keeps track of the location, surface condition, and other properties of objects. It recognizes pursuit of goals by any sequence of the primitive actions, so long as each is possible, and so long as any temporal constraints among subgoals are met. It also allows actions even if they do not progress toward the agreed goal, as long as they are possible, and makes comments accordingly. Once again the rich environment supports immersion-style learning of a surprising range of language constructs. For example, in the bathroom, one naturally encounters the reflexive Spanish verb, *lavarse*, used for washing one's own hands and face. In this same context but with English as the language we could introduce in a natural way the verb-particle construction with "put down" and "hang up," and adjectival complements to verbs, like "get <NP> soapy."

C Conceptual Structure of FLUENT

Although the FLUENT-1 system described in the preceding section exhibits many of the external properties called for in the introductory remarks, the ultimate success of the concepts will require a more flexible and complex internal structure, to which we now turn. Several important parts of the planned system architecture for the proposed revised version, FLUENT-2, arise in a natural way in the cycle of communication, beginning with input from the student. The system will carry out these tasks: (i) analyze the student's input linguistically or graphically; (ii) diagnose the linguistic input to update the student model and, in case of errors, to figure out what the student really meant, if possible, to keep the conversation going; (iii) make continuity possible by keeping track of the physical situation and the dialog structure; (iv) make tutorial decisions about what linguistic aspects to use, and what content and dialog structures will permit those linguistic aspects without upsetting continuity; and (v) generate appropriate output, whether linguistic, graphic, or both. These five capabilities are shown in Figure 2.

FIGURE 2 ABOUT HERE

The discussion is organized into three major sections: representation, communication, and tutoring. Section C.1 presents the representation of ideas in the three forms required by our commitment to immersion: language, pictures and an internal representation system. Section C.2 treats

communication, including both the translation among the three representations and an inventory of interactive patterns of communication that are suitable for the language learning experience. Section C.3 is a consideration of such tutoring aspects as construction of a syllabus, maintenance of a student model, and strategies for deciding what to do next. Some of these topics resemble those in generic discussions of intelligent tutoring systems (ITSs). Here, however, the focus is on natural language as the knowledge to be acquired, whereas for other domains of instruction natural language plays at most a role in the interface. The opposite situation holds for the conversational subject matter. This material would be the knowledge to be acquired in a non-language system, whereas here, though crucial for immersion, it is relegated to a lower status: it must be represented, but is not taught.

C.1 Three Representations: Language, Spatial, and Internal

We first state a strategy for tapping into existing achievements in natural language processing. A longer subsection is then devoted to the more unusual aspect of FLUENT, the spatial considerations. Finally we discuss how the spatial concerns influence the internal representation.

C.1.1 Natural Language

Natural language understanding and generation can be monumental tasks if not somehow constrained. Compounding the difficulty of understanding is the learner's propensity to err. There are, however, several sources of constraint that can serve to reign in the problem. First, FLUENT-2 will rely on the natural language processing (NLP) software of the Athena Language Learning Project (ALLP). The intent is to tie FLUENT domains to ALLP interlingual semantic structures in such a way as to benefit from certain potential efficiencies. (ALLP itself has paid significant attention to efficiency; see Malone and Felshin, to appear). By providing a limited domain at any particular time, FLUENT can eliminate much lexical ambiguity. By providing a context constrained by the graphical interaction and previous discourse, it can cut down referential ambiguity, smoothing the path from logical form to final meaning in context. Such ambiguity reductions are usually pursued for the sake of the understanding side, but are relevant to generation too, since what one generates for the learner should be uniquely interpretable.

Another contributor to manageability of the natural language problem is that in the early stages of learning, one only needs a sharply curtailed grammar and vocabulary, though these do increase with progress. Even for the errors, there is a source of containment, namely the propensity of various learners to make similar mistakes, which can therefore be to some extent anticipated. These errors may be oversimplifications of the target language or inappropriately imported aspects from a particular native language. The ALLP system already deals gracefully with some of these, by means of a best-first parsing strategy that tends to favor parses containing popular errors. Even fluent speakers make mistakes; Carbonell and Hayes (1983) provide a useful taxonomy of strategies for ill-formed input.

Some styles of interaction also diminish the problems. For example, yes/no questions require only a word of response and wh-questions only a noun phrase; commands even accept a purely graphic response. By selecting these particular dialog modes in the early stages, the tutor can minimize not only the demands on the learner's language production but also on the system's ability to understand. On the generation side, one aspect of the natural language processing problem, that of deciding what to say, can be decided essentially by fiat, by choosing a particular

dialog mode, such as Movecaster (section C.2.2), in which the tutor describes student actions and resulting state changes.

C.1.2 Spatial representation

At first glance, a mouse and a computer screen may not seem up to the task of providing the rich spatial interaction we need for immersion. The above description of the prototype may give some confidence on this matter. Still, a few remarks are in order about the power of pictures and mouse maneuvers to convey movement and change in continuing everyday scenes. First, the spatial medium, like the linguistic, is bi-directional, with objects manipulated within scenes by both learner and tutor, enabling the two to maintain mutual understanding. From the mouse primitives of movement and buttoning, with suitable conventions about how to interpret them, one can build up a spatial communication repertoire that permits the student to express such notions as: selecting an object or a part of one; connecting two objects or parts to each other; operating on an operand; and moving something to a point or to a region. We plan to investigate the usefulness of recognizing movement according to certain types of trajectories, such as circumnavigating or passing through something, moving back and forth, or following a route. In the prototype system we used a simple but highly effective form of partial animation: letting a hand perform movements while its arm behaves as a rubber band from the shoulder.

Turning from the mouse to the content of pictures, we now try to show that the range of concepts suitable for spatial communication is sufficiently broad to make FLUENT work. We will mention a few lexical categories and syntactic aspects and look at the promise and problems of conveying each of them visually. The categories of noun, adjective and verb all have members that submit to visual presentation. There are straightforward techniques for proper as well as common nouns. Reasonably straightforward visual treatment is also possible for the functionality of spatial prepositions, plurals, and possessives.

Consider first the common nouns, which should be the simplest category to present. A recognizable object typically conveys a common noun that represents a natural type to which it belongs. (On this and related points, see Sowa, 1984). Moreover, although there is some flexibility about how high in the type hierarchy to look, there is often a convention about the choice. For example, if the system presents a picture of a dog, presumably that is to convey the common noun "dog," even though what the picture shows is in fact more precisely a terrier or less a precisely a member of the canidae, a mammal, an animal, etc. However, a dog and a cat together can create the conditions for speaking about a higher type to which both belong. Again there is a conventional level at which to look, here "animals" rather than "mammals" or "physical objects." So pictures would seem to have broad potential for common nouns, provided that the new language disagrees only rarely with the learner's first language on what is in the type hierarchy and on the conventions about where in that hierarchy to look when picking appropriate common nouns for talking about ordinary objects.

Except in the case of common nouns, however, a picture out of context does not convey a unique word or phrase, so tutoring is not just a matter of something so simple as presenting picture-phrase pairs. To get a quick idea of how crucial context can be for presenting verbs, for example, note that a hand moving back and forth can be an instance of wiping, waving, drawing or combing (etc.) depending on what the hand is holding, if anything, and on what its owner's goals are known to be. That a context with goals will be helpful should be clear from the earlier discussion

of the verbs in the bathroom scene. Interestingly, a type ambiguity problem similar to that just discussed for common nouns also shows up with verbs and also must be dealt with by the use of context and/or multiple uses. If, for example, John is shown "going" somewhere, he will necessarily be going in some particular way, like walking. This ambiguity can ultimately be resolved if the learner's processing is complex enough to take advantage of seeing a ball, as well as a person, executing "going" at various times, since a ball does not walk or run (though it may roll or bounce). "Go" is more abstract than "walk" in two ways: it tolerates a broader class of arguments and it covers different forms of locomotion.

Having discussed a couple of lexical categories, we now touch briefly on some exercises for syntax in FLUENT-1. Few if any grammatical aspects are directly picturable, so the issue is how effective indirect methods can be. Take the relationships between adjectives and their head nouns, with respect to agreement and relative order. First of all, note that some adjectives even in linguistic isolation can be conveyed pictorially by a variant of the "cat plus dog equals animals" technique mentioned above, for example, presenting "black" with black objects of disparate type. Once some adjectives and nouns are known, FLUENT-1 presents by picture a concept that in language is represented by an adjective and a noun. Once again context plays a key role: to make the use of the adjective pragmatically necessary, the system requires the student to distinguish objects of the same type, hence requiring the same noun, but with differing attributes, distinguishable by adjectives. With a sufficient variety of such pictures, each associated, in context, with corresponding language, we hypothesize that an adult will learn, as a child does, to conform to the grammatical constraints, and to do so fluently, unencumbered by conscious grammatical computation, even though an adult may learn grammar explicitly as a back-up capability.

C.1.3 Internal representation

It is appropriate to use an object-oriented representation, not only to benefit from the reduced data entry effort and storage efficiency that inheritance permits, but more significantly because the system needs the appropriate knowledge of how objects are related in order to converse properly. FLUENT-2 will have one hierarchy for things and another for activities, with type constraints on arguments. Two such hierarchies are related in that an event can change the state of an attribute of an object. Most clearly in need of representation are the physical objects whose images appear on the screen. These objects have visible attributes like size, color, and image shape, as well as non-visible ones such as surface condition and movability. Some of these properties have unchanging values, but others may vary. The surface condition of an object, for example, can be in various states like dry, wet and soapy at different times. Interestingly, the surface condition attribute itself has an attribute with respect to visibility, specifically that it is unchangingly invisible. In addition to the attributes of individuals and of attributes themselves, there are relationships between objects, like part-whole and the ability of one to support another. As noted, common sense goal structures can be useful for coordinating learner-tutor dialog. Thus we need and/or goal trees and some temporal ordering constraints. At the leaves of a goal tree are simple activities some of which correspond to primitives in the graphics.

C.2 Communication

Communication is a fundamental activity in FLUENT; indeed from the viewpoint of the learner, it is really the only activity. We discuss communication at two levels. To communicate in two media, the system translates between them, via its internal representation, as discussed in C.2.1.

These capabilities are building blocks, from the higher level viewpoint of the tutorial component of the system, which must arrange that the translations be performed at appropriate times, that is, according to tutorially useful patterns of interaction, taken up in C.2.2.

C.2.1 Translation among representations

Four kinds of translation are needed: in both directions between the internal knowledge representation and each of the external representations, linguistic and graphic. Translations of various subsets of English to model-theoretic semantics have been presented by various authors; especially relevant here is the work of Crangle and Suppes' (1987) because of its attention to context and spatial relations. Graphical representations must also be translated to and from the internal representation.

To begin on the graphics side, our general strategy for graphics output will be to attach simple graphic procedures to the lowest level actions in the goal structures alluded to above. For example, drying one's hands, though a subgoal in the example above, includes subgoals of its own, like picking up a towel, which includes moving a hand to the towel, a graphics primitive. Turning to the input or understanding side of graphics, for the washroom scene we have designed a rule system with over twenty generic rules for updating the internal representation in response to events that the learner instigates on the screen. Seven of these rules concern the surface condition of objects like dirty hands, a soapy face, or a dry towel. The one that concerns wetting things under the water spout looks like this:

Preconditions:	Mobile-hand : holding = X
	X unequal Soap
	either Hot-faucet : activity = on
	or Cold-faucet : activity = on
Action:	X visits Water-spout
Results:	Mobile-hand : surface = —w→
	X : surface = —w→

The preconditions and results all concern the internal representation. The action of the hand is an abstract screen event, but is defined in terms of concrete (quantitative) variables like mouse states and the numerical locations and sizes of the objects. The "—w→" notation lets the rule be generic. The "w" mapping specifies how to update the value of the surface attribute from each possible state, say from dirty to muddy or from soapy to wet. There are other mappings of this sort. This notation enables us to simplify the preconditions, which otherwise would need information about the surface state of both the object (X) and the mobile hand. The consequence would be a proliferation of rules and loss of clarity.

We now turn from graphics to language. Section C.1.1 touches on the input or understanding side, specifically the pervasive issue of disambiguation and how some aspects of this application

may assist with it. Here we will comment only on the output side. Natural language generation (NLG) can be divided into (at least) two parts: deciding what to say and determining how to say it. The ALLP software can already carry out the second of these, provided it is given as input an idea expressed in terms of its interlingual semantic structures. The core of this representation is the predicate-argument structure, but information about time, reference and mood is also encoded. Consider what FLUENT-2 will have to do to provide that semantic input when, for example, it is describing to the learner what s/he has just done in the washroom scene. There are various levels in the goal structure at which to view an act. Each level can have its own semantic predicate-argument structure. The high level activity that is ultimately expressed in English as "get X clean" (where X is some object) includes lower level ones expressed as "get X soapy," "rinse X" and "dry X," that in turn have finer grain parts, ultimately associated with specific state transitions like that from soapy to wet. The selection of which of these levels to express will have to be determined within FLUENT-2, by reasoning about context and goals.

Further, the states and transition in the FLUENT-2 knowledge representation will have to make contact with the ALLP lexicon and semantic representation system. In the bathroom example, the states need predicates that are typically be associated with adjectives like "wet" or "dry." More subtle is the role of words like "make" and "get" that express the idea of doing some unspecified activity to bring about the resulting state. Important NLG information also comes from the particular current dialog mode, which determines the mood (question, imperative, etc.) of the output and, in the case of a command, which entity fills the syntactic role of second person. For English, the sentences that describe the transitions and the resulting states take forms like those listed just below, where <agent> is the screen character whose hand is performing the actions on some <theme> or object, possibly itself or some part of its body.

<agent> "make" <theme> <Adj>

<theme> "get" <Adj>

<agent> <Verb> <theme>

C.2.2 Interactivity

A pedagogy can be characterized in part as a collection of dialog modes and a way of choosing among them according to circumstances. A Socratic dialog, for example, uses a dialog mode that involves questions from the tutor. In contrast, the foreign language pedagogy of "total physical response" requires dialog modes that call for visible physical action, by either the teacher/tutor or the learner. Dialog modes differ in their difficulty for the learner, suitability for different cognitive styles and the clues they can supply about what the learner knows. In one (partially implemented) interaction mode, Movecaster, the tutor is like a sportscaster telling what is happening on the field of play. Here, the field of interest is the computer screen and the events are the moves by the learner, who can make arbitrary legal moves. The Movecaster then makes a comment, which may be simply a statement of what action has just occurred. Alternatively, the comment may describe a changed aspect of the situation, or specify how the change relates to a previously established goal. The Movecaster mode has the desirable property that it keeps language flowing, and keeps it clearly synchronized with meanings. Also, it gives the opportunity for the beginner to practice comprehension skills before having to produce language. Movecasting

provides even non-beginners the opportunity to explore a new domain, to gain familiarity with ways of expressing its new relationships, before going on to more demanding dialog modes. Finally, this mode does not require the NLG system to make micro decisions about what to say.

In the prototype, we have experimented, using template-style NLG, with limited versions of several other dialog modes that we call Presentation, Command, Execute, Delegate, Query, and Answer. Each brings together ideas and the corresponding language. For Presentation mode this is done directly by presenting a phrase or sentence in L2, along with a visual representation of its meaning, both chosen by the tutor. A Command interaction begins with the tutor telling the learner, in L2, to do something by manipulating screen objects. The converse activity is Delegate mode, in which the learner has the initiative. In this case a palette of verbs and physical objects is available to help to learner compose her thoughts and equally to constrain the natural language understanding capability that the system will require. Delegating begins with the tutor commanding a high level goal which the learner is to decompose into subgoals and give as lower level commands to a screen character.

C.3 Tutoring

C.3.1 Syllabus

Syllabus issues include most prominently the language aspects themselves; we shall return to these shortly. Also of concern is the conversational subject matter via which the student can possibly be exposed to the language. The subject matter, though not the central issue, must nevertheless be chosen with care, so as to include useful topics as well as support the introduction of the crucial grammatical constructs. The Waystage project (van Ek et al., 1980) has outlined a single set of concepts intended to suffice for a variety of languages, and has catalogued the key grammatical constructs of English that express these concepts. That project does not endeavor to impose an order on either the concepts or the corresponding aspects of English grammar. Still, it does turn out to be a highly suggestive resource, cataloguing general objective concepts (temporal and spatial notions, existence, quantity, attributes, relations, etc.), speaker attitudes toward those concepts (assistance, suggestion, possibility, likelihood, emotional stance, etc.), and a listing of some particularly appropriate domains.

Turning to the core of the syllabus, the language aspects to be taught, one needs a substantial array of capabilities, even for a modest system. It seems reasonable to demand explicit representation of lexical categories, lexical features, particular lexical items, phrase structure (dominance and ordering) rules, verb subcategorization and case marking, various kinds of constraints, including agreement, and rules about the semantic properties and consequences of the foregoing. Given the known complexity of devising all these correctly, it is tempting to think of letting the role of syllabus be filled by the language knowledge of an existing NLP system, taken off the shelf. It is instructive to see why it's not that simple; we note several obstacles. First, experience with intelligent tutoring systems suggests that the knowledge in a performance-oriented expert system is not sufficient for communicating knowledge to a student (Clancey, 1987). This is true even when the rules are in non-technical (glass box) form. The problem is worse if the rules are in technical (black box) form; see the distinction made by Chanier et al. (this volume). Next, unlike other systems FLUENT does not teach grammatical and morphological rules explicitly, though it does need to have them in declarative form for student modeling. A further difficulty arises if the application end of the chosen off-the-shelf system is specialized for a different

knowledge representation. Finally, the syllabus must include for each language a partial ordering of its rules.

Such an ordering is important for any subject, since it helps the system to introduce material at a time when it is pedagogically reasonable to do so. One justification for introducing an aspect relatively early is that it is logically a prerequisite of other aspects. A related but distinct property is how intrinsically easy an aspect is to learn, as determined by studying the natural course of acquisition, in L2 where possible; cf. Broselow (1988). A third consideration, noted above, is an aspect's practical role in expressing needed concepts. Finally, of special concern here is the possibility of constructing picturable phrases and establishing graphical interaction. Textbooks that teach pictorially, such as Ibarra (1962) may be credited with implicitly attempting to strike a balance among all these criteria. Unfortunately, such books provide little explicit pedagogical reasoning, but we have been able to do some reverse engineering. In partially pictorial books, like Sacks da Silva (1982), the grammatical intent is clear, but the many pictures are seldom pedagogically independent, relying instead on explanation and heavy use of cognate vocabulary.

C.3.2 Individuation

Once the syllabus is established, the tutoring system needs to move around in it in response to the apparent abilities of a particular student. To do this, the tutor also needs to choose a scene that supports use of the specific language aspects the student would work on. We have devised a method of scene selection that can roughly be characterized as inverting an index to scenes, though it is somewhat more complex than that. Within a scene, a tutor can, even without recourse to a student model, be responsive to the student's current response. In effect, the tutor reasons, "Since you said that, you need to know (or do) this next." Such immediate responsiveness is all we expect to achieve in the near term, and the prototype does not contain an individual student model. Generally, of course, a defining aspiration of ITSs is to gain a longer term picture of what a student is up to, in order to respond to general tendencies or to put the current response into a broader perspective, the better to understand the underlying bases of a superficial misperformances. A relevant idea that emerged in the 1960s among child language acquisitionists is that the child has her/his own grammar, which may include rules not to be found in a correct adult grammar, but that may in some cases be simplifications of adult rules. This approach, perhaps best exemplified in Bowerman (1973), may be regarded as a precursor of the ITS work on bugs and malrules. One must of course allow for the fact that L2 learners share some kinds of L1 bugs, typically avoid others, and bring new ones related to their own L1.

C.3.3 Tutoring Principles

A tenet of ITSs favors the declarative expression of tutorial principles and techniques as an explicit set of rules accompanied by an interpreter that reasons with them to reach tutorial decisions in particular situations. We mention two techniques that can be formulated as sets of rules. One is the constraint of introducing only one new thing at a time, for reasons discussed by Hamburger and Wexler (1975) in the context of language learning and Van Lehn (1987) in terms of machine learning. The consequences of this principle are strongest when the rules are of fine grain, as in Hamburger and Crain's (1982) delineation of twelve steps on childrens' paths to a correct relative clause. Another technique that appears to facilitate language learning is the use of expansion of a learner's sentence or partial sentence into an appropriate longer form (Nelson, 1975). Some tutorial principles can, however, most appropriately be built directly into the conceptual

organization of the system. For example, our emphasis on the pairing of linguistic and visual communication is built into several of the dialog modes. Also essentially built-in are the pedagogical decisions to avoid explicit discussion of grammar and to downplay dependence on the student's first language.

FOOTNOTES

1. We are happy to acknowledge the helpful insights of Tucker Maney, who developed the implementation discussed in section B at the Navy Center for Applied Research in Artificial Intelligence, a part of the Naval Research Laboratory, which has supported much of this work.
2. FLUENT is an acronym for Foreign Language Understanding Engendered by Naturalistic Techniques.
3. cT is a trademark of Carnegie Mellon University. The language was developed principally by Judith Sherwood and Bruce Sherwood, and is now distributed by Falcon Software. A good source is Sherwood (1988).

REFERENCES

- Asher, J.J. (1977) *Learning another language through actions*. San Jose, CA: Pajaro Press.
- Berlitz, M.D. (1925) *Deuxieme livre pour l'enseignement des langues modernes: Partie francaise pour adultes*. New York: Berlitz.
- Bobrow, R. J., Resnik, P. and Weischedel, R. (1989) Multiple underlying systems: Translating user requests into programs to produce answers. Paper presented at BBN Natural Language Symposium, Cambridge, MA, Nov. 29 - Dec. 1.
- Bowerman, M. (1973) *Early syntactic development: A cross-linguistic study with special reference to Finnish*. Cambridge, England: Cambridge U. Press.
- Broselow, E. (1988) Second language acquisition. In Newmeyer, F.J., *Linguistics: The Cambridge survey, v.III*. Cambridge: Cambridge University Press.
- Carbonell, J. and Hayes, P.J. (1983) Recovery strategies for parsing extragrammatical language. *American J. of Computational Linguistics*, 9, 3-4, 123-146.
- Clancey, W. J. (1987) *Knowledge-Based Tutoring: The GUIDON Program*. Cambridge, MA: MIT Press.
- Crangle, C. and Suppes, P. (1987) Context-fixing semantics for instructable robots. *Int. J. Man-Machine Studies*, 27, 371-400.
- Dulay, H. C. and Burt, M. (1974) You can't learn without goofing: An analysis of second language 'errors'. In Richards, J.C. (Ed.) *Error Analysis: Perspectives on Second Language Acquisition*. London: Longmans Group Ltd.
- Fum, D., Giangrandi, P. and Tasso, C. (1988) ET: An intelligent tutor for foreign language teaching. Proceedings of the Conference on Intelligent Tutoring Systems, Montreal, 462-468.
- Hamburger, H. (1988) FLUENT: Foreign language understanding engendered by naturalistic techniques. Internat. Wkp. on ITS for Foreign Language Learning. Trieste, Italy, November 28-29, 1988.

- Hamburger, H. and Crain, S. (1982) Relative acquisition. In *Language development: Syntax and semantics*. S. Kuczaj (ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hamburger, H., Lodgher, A., Maney, T., Jardine, C. and Siff, F. (1990) A Decentralized AI System for Transcript Applications. *Proceedings of IEEE/AISIG, May 9-11*.
- Hamburger, H. and Wexler, K. (1975) A mathematical theory of learning transformational grammar. *Journal of Mathematical Psychology*.
- Ibarra, F. (1962) *Look and learn Spanish*. New York: Dell Publ. Co.
- Krashen, S.D. and Terrell, T.D. (1983) *The natural approach*. Oxford: Pergamon Press.
- Malone, S. and Felshin, S. (to appear) GLR parsing for erroneous input. In Tomita, M. (Ed.), *GLR Parsing*.
- McLaughlin, B., Rossman, T. and McLeod, B. (1983) Second language learning: An information-processing perspective. *Language Learning*, 33, 135-158.
- Munro, A. and Pizzini, Q. A. (1990) Direct manipulation authoring of object behavior in an interactive graphical modeling environment. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, Los Angeles, November 4-7.
- Murray, J. (1987) Humanists in an institute of technology: How foreign languages are reshaping workstation computing at MIT. *Academic Computing*, September.
- Nelson, K.E. (1975) Facilitating syntax acquisition. Paper presented to the Eastern Psychological Assn New York (April).
- Newport, E.L. (1990) Maturational constraints on language learning. *Cognitive Science*, 14, 1, 11-28
- Sacks da Silva, Z. (1982) *Persona a persona*. New York: Macmillan Publ. Co.
- Sherwood, J. (1988) *The cT reference manual*. Champaign, IL: Stipes Publ. Co.
- Sowa, J. (1984) *Conceptual Structures*. Reading MA: Addison Wesley.
- Swanson, C.W. (1989) Deutsch in Deutschland: repurposing a foreign language videodisc. *Academic Computing*, 4, 1.
- Sweller, J. (1988) Cognitive load during problem-solving: Effects on learning. *Cognitive Science*, 12, 2, 257-286.
- Underwood, J. (1987) When the music stops: A review of "Using computers in language learning." *Calico Journal*, 4, 3, 35-50.
- Van Ek, J.A., Alexander, L.G. & Fitzpatrick (1980) *Waystage English*. Oxford: Pergamon Press.
- VanLehn, K. (1987) Learning one subprocedure per lesson. *Artificial Intelligence*, 31, 1, 1-40.
- Weischedel, R.M., Voge, W.M. and James, M. (1978) An artificial intelligence approach to language instruction. *Artificial Intelligence*, 10, 3, 225-240.

Options

CT

Bien. Ahora obten el jabon, por favor.

FLUENT PROCESS

S = Student
I/O = Input/Output

TA = Tutorial Analyst
SR = Situation Reasoner
TS = Tutorial Strategist