# The challenges in adapting traditional techniques for modeling student behavior in ill-defined domains

**Amy Ogan, Ruth Wylie, Erin Walker**
Human Computer Interaction Institute, Carnegie Mellon University
aeo, rwylie, erinwalk@andrew.cmu.edu
5000 Forbes Ave.
Pittsburgh, PA 15213
(412) 268-1208

**Abstract.** Designing cognitive tutors and modeling behavior for ill-defined domains require innovative methods and techniques. We combine a top-down, theoretical approach with a bottom-up, empirical approach to develop a student model for the selection of aspect in French verbs. In performing this task, we design a new representation applicable to feature-driven ill-defined problem spaces and utilize tutoring scaffolds in order to elucidate the student thought process. We then evaluate and refine our model based on empirical data collected through student think-alouds. We plan to use our preliminary results to design and evaluate a fully-developed cognitive tutor, and hope to generalize our tutor development process to other ill-defined domains.

**Keywords:** Ill-defined domains, student modeling, passé compose and imparfait

## MOTIVATION

A completely well-structured problem is a problem in which the starting statement contains all relevant information, and there exist a limited number of relatively easily-formalized rules used to reach an unambiguously correct or incorrect solution [Jonassen, 1999]. These problems are often found in mathematical domains; to an expert, the simple arithmetic question "26+38=?" has a clear set of steps from the initial state to the goal state. Most real-world problems are not so well-structured. In particular, domains like design and cultural education contain problems that are not so tidy. When the start state, rules, or goals of a problem are not easily formalized, the problem is ill-structured [Ormerod, 2006]. For example, writing an essay is completely ill-structured: the start-state is underspecified, there is no predefined set of rules for completing the task, and it is difficult to know when a satisfactory result has been attained. Not all domains are so ill-defined. Some domains lie somewhere in the middle. Their problems may have well-structured start and goal states, but ill-structured rules because 1) there are multiple representations of knowledge with complex interactions and 2) the ways in which the rules apply vary across cases nominally of the same type [Spiro, 1991]. Unfortunately, when rules and the conditions in which they are applied are difficult to formalize, it is also difficult to form a model of student and expert performance for that domain.

This modeling issue is particularly problematic when trying to build a cognitive tutor for an ill-defined domain. A cognitive tutor is an intelligent tutoring system that compares student action to a model of correct and incorrect behavior and provides context-sensitive feedback and problem selection. Cognitive tutors have been effective at increasing student learning in real-world settings by as much as one standard deviation over traditional classroom instruction [Koedinger, 1997]. However, most successful tutors have been limited to well-defined domains like algebra and physics. Ill-defined domains may also benefit from cognitive tutoring, but this area is only beginning to be explored. Cognitive tutors rely on the existence of a formal domain model as the basis for other stages of tutor development like identifying problem content or designing tutor feedback, thus the absence of a model for a given ill-defined domain makes developing a tutor problematic. Alternative solutions have been suggested such as simply forcing a structure on the domain [Simon, 1973], or forgoing a structure altogether and focusing on non-model-based learning tasks [Davis & Tessier, 1996]. We believe, however, that by adapting current cognitive tutor modeling techniques, it is possible to develop a cognitive model for an ill-defined domain that provides structure while not over simplifying the complexity of the domain.

In this paper, we focus on the ill-defined domain of determining aspect in French second language learning. We discuss the difficulties of applying traditional modeling techniques in ill-defined domains as well as our solutions for adapting these techniques. Finally, we examine the model created using this procedure and outline future plans for developing a tutor. Our results suggest that while traditional modeling techniques are inadequate for ill-defined domains, adaptations in knowledge representation, problem presentation, and experimental design lead to effective solutions.

## ASPECT IN FRENCH LANGUAGE LEARNING

Problems in language learning fall at all points on the continuum of well-defined to ill-defined tasks. Successful cognitive tutors based on student models of observable behaviors have been implemented in well-defined language areas, for example the Capit system [Mayo & Mitrovic, 1997] teaches students capitalization and punctuation rules [see Gamper & Knapp, 2002, for review]. However, not as much work has been done in more ill-defined areas (e.g. the acquisition of grammatical gender) even though formal instruction in these areas is a necessary component of second language education [Norris & Ortega, 2000]. In particular, the distinction between the passé composé and the imparfait tenses in French is a prototypical ill-defined language learning problem that has clear start and goal states, but ill-structured rules and conditions for applying them [DeKeyser, 2005]. Mastering the distinction between these is a difficult task for both beginning and advanced French students and is reviewed often throughout programs of French instruction. This distinction is acquired by learning and understanding the concept of aspect.

Aspect is the relation between a situation and its associated interval of time [Comrie, 1976]. Students must know the role aspect plays in a sentence in order to both understand temporal qualities of actions and be able to accurately produce novel utterances. When speaking in the past in French, aspect is conveyed through the use of two tenses: the passé composé and the imparfait. The passé composé involves a completed action, as if viewed from an external perspective, while the imparfait involves an ongoing action, as if viewed from an internal perspective [Salaberry, 1998]. Examples of uses of the passé composé, translated into English, are "I went to the store on Tuesday" and "I stopped at the store," while uses of the imparfait include "I went to the store every Tuesday" and "I was at the store". The phrases "on Tuesday" and "stopped at" indicate that the action is finished, while "every Tuesday" and "was at" indicate ongoing actions. Students must use features of the sentence, like lexical semantics and calendric expressions [von Stutterheim, 1991], to infer properties of the action such as its duration that are relevant to making this aspectual decision [Ishida, 2004]. Because the past tense in English is 'completely ambivalent', this task is novel and particularly difficult for native English speakers learning French [Salaberry, 1998].

Aspect is difficult to learn, because it belongs to a class whose problems "express highly abstract notions that are extremely hard to infer, implicitly or explicitly, from the input" [DeKeyser, 2005]. Aspect has a relatively clear goal state: Experts might disagree on the aspect of a sentence, but only in certain ambiguous cases where the intention of the speaker is not clear. Difficulties in identifying aspect tend to arise because apsect is a problem with ill-structured operators: rules are hard to formalize, the conditions under which they apply are too specific and numerous to be described, and rules may be applied in parallel. It is difficult to find a formal description of the rules for identifying aspect, and descriptions often do not correspond to one another. Additionally, instructional texts often confuse rules, which will always return a correct answer when applied correctly (e.g., "If the action occurs a single time the tense is passé composé"), with heuristics, which may be easier to apply but are not always correct (e.g., "If the sentence contains a word like 'once' the tense is passé composé"). Instructional texts also generally do not cover the whole problem space, leaving students with cases that are difficult to classify. In fact, since rules that do cover the space are abstract, describing all the conditions under which they apply is important but impractical. To understand what is meant by a completed action, the student must be aware that it has a start, an end, or a specific duration. "Has a start" is then broken down into many sub-cases, such as verbs that imply a beginning action, and as these conditions become more specific, enumerating them amounts to identifying particular examples. To complicate the situation, there are sentences where multiple rules apply, and the student must perform conflict resolution. For example, the sentence, "All of a sudden, the sky was blue," appears to be both an ongoing description of circumstances and a completed change of state. The student has to know that this sentence is in fact not a description, but an event (as signified by *all of a sudden*) and the passé composé should be used. Because aspect is an ill-defined domain which is challenging and important for students to master, it is an ideal candidate for our attempts to model ill-structured problems.

## TRADITIONAL MODEL-BUILDING AS APPLIED TO ASPECT

A cognitive model is a formal description of a problem-solving process. It includes both expert and novice behavior, and correct and incorrect actions. It is required when building a cognitive tutor so that the tutor can provide contextual feedback on problem-solving. Cognitive models are generally developed using a combination of theory-driven and data-driven approaches. Although a theory-driven approach can identify what is relevant about a task and pinpoint thought processes that are not necessarily visible through behavioral observation, it may not ultimately reflect the steps novices take to solve a problem. A data-driven approach can highlight problem-solving strategies and misconceptions in actual users, refining the initial formal model.

We modeled the specific task of identifying the aspect of a verb. Students were presented with a French sentence containing a verb in its infinitive form and asked to indicate whether the sentence should use the passé composé or imparfait. For example, a student was given the French translation of "While I was doing my homework, the telephone _________ (to ring)" and asked to select the appropriate tense of the sentence. We first performed a rational task analysis and then enhanced our model through think-aloud protocols from experts and students.

### Rational Task Analysis

In a rational task analysis, a formal specification of the task is combined with consultation with experts to produce a model of ideal performance and identify places where novices may make errors. The result is generally a set of production rules (see Anderson et. al., 2004) arranged in a sequential and deterministic structure. For example, this approach was used by Siegler (1976), who proposed a decision tree representing children's ideal performance on balance scale problems, identified subsets of the tree representing novice performance, and validated his model empirically. A properly developed theoretical model forms a basis for the design of an effective cognitive tutor.

A model of performance on determining aspect in French was developed. Logical analysis of the model suggested that the task involves production rules of the following structure: "IF the sentence contains feature X, AND feature X is a member of class Y, AND class Y indicates tense Z, THEN the tense of the sentence is Z." An example of one of these rules is, "If the sentence contains an expression of time, and the expression of time indicates a one-time action, and a one-time action indicates the passé composé tense, then the tense of the sentence is the passé composé." Because the instructional texts and experts we consulted tended to disagree on the formalization of the rules, we chose a set of five that were agreed upon by the majority of sources and covered the full problem space. We then attempted to identify how these rules might be applied in sequence to efficiently reach a correct result. The full decision tree is shown in Figure 2. In our model, students ask themselves a sequence of increasingly abstract and more difficult questions. Notice that the default question is not a yes or no question but a catch-all which covers the full problem space. It requires students to make a high-level judgment about the action in the sentence without resorting to the heuristics targeted in previous questions that are easy to answer but not always accurate. The model predicts that experts (and novices) go through a sequential process of querying the sentence for features and use the first positive response to arrive at a decision. We intended to deal further with the problem's ill-structuredness by providing scaffolding (described in later sections) for the identification of features in the resulting tutor.

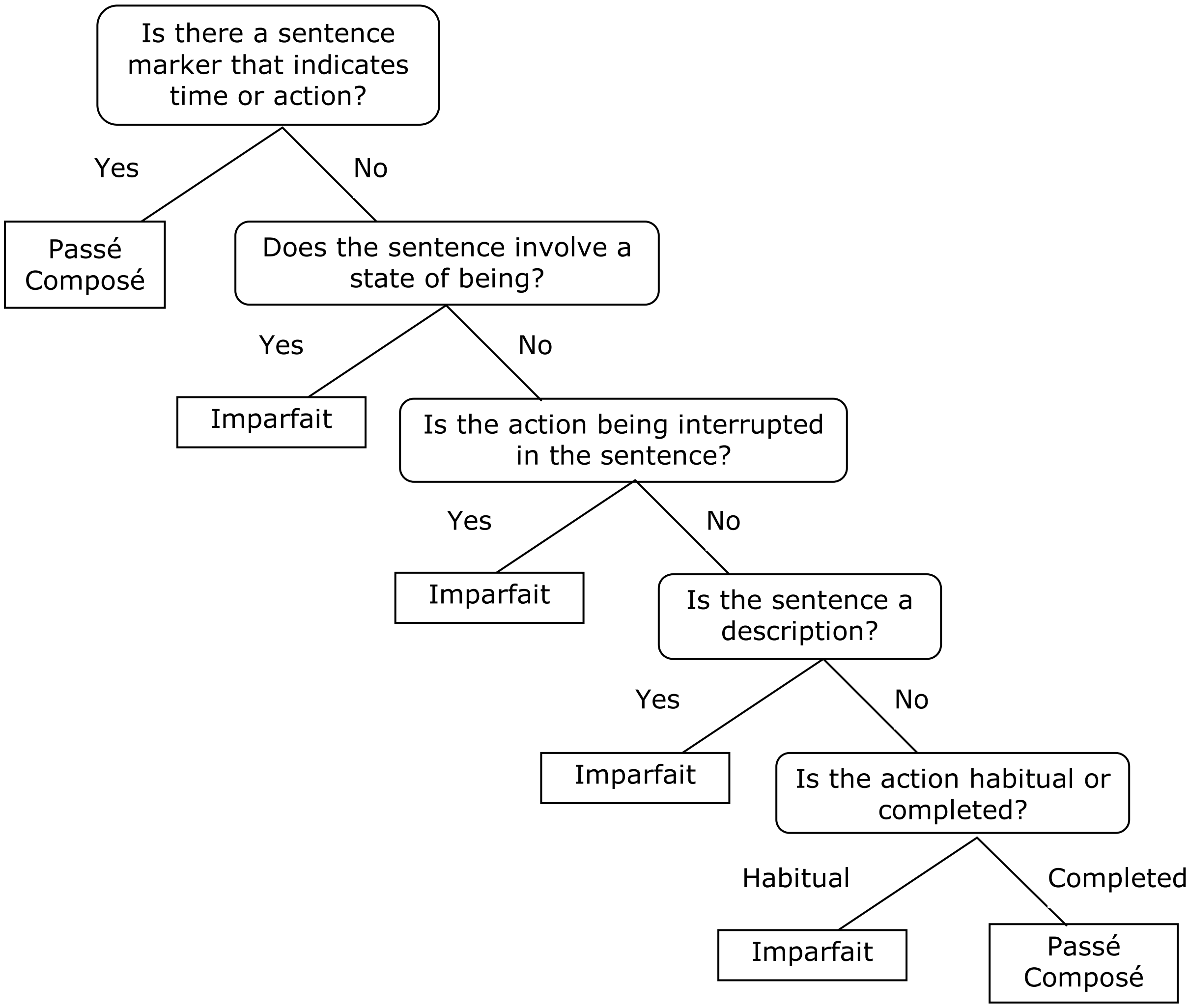


Figure 1: Decision Tree used in original model

### Think-alouds

Our next step was to collect data using a think-aloud procedure (Ericsson & Simon, 1984). In a think-aloud, individuals are asked to verbalize all thoughts and actions while solving a problem. Participants are not asked to explain or justify what they are doing nor are questions posed by the experimenter during the process. The result is a stream of conscious record of the actions being performed and is thus valuable for collecting data to build or refine a model. Think-alouds have been shown to not significantly interfere with the problem-solving processes (Schooler et. al., 1993). We performed think-alouds on both expert and novice subjects.

We attempted to refine our model of expert performance on the aspect identification task by conducting a think-aloud with four French professors. We generated twenty initial problem sentences, which were simple and without context, and prompted experts to think out loud as they solved the problems. The results did not support our model. Experts did not appear to go through a sequential process of decision making or base their choice on the rules in our model. There was never a response in the explanation that indicated they were taking a 'no' branch of the decision tree. Secondly, the first three rules of the tree were never mentioned. Finally, their explanations were holistic, often using phrases such as "This sentence is ..." rather than referring to individual parts of the sentence.

We used three subjects for the novice think-aloud. All had completed elementary French instruction so in theory should have been able to complete the task. However, their level of expertise made a great difference in performance. Our first subject had one semester of French and didn't remember conjugations other than the present tense. The second subject had the most French experience, and completed the task with a high level of automaticity. He could not explicitly articulate the process he followed when making the aspectual decisions. The third subject had sufficient French experience, but since time had elapsed since her last instruction she had difficulty completing some of the sub-steps (e.g. understanding the sentence, identifying the tense, conjugating the verb) of the task. Due to the variation in student performance, we were limited in the amount of process data we were able to collect.

### Discussion

Our expert and novice think-aloud protocols revealed three main issues with our model and techniques. First, there were problems with the data-collection methodology. Our task was structured so that our novice think-aloud participants did not provide verbal protocols that could decisively confirm or disconfirm our model. In fact, there seemed to be a narrow window of participant expertise where the process of solving this type of task can be verbalized; two of our participants were too inexperienced and thus did not have the necessary knowledge to complete the task let alone explicitly state the rules being applied, and one of our participants was too expert and relied on implicit knowledge to complete the task. One solution is to include additional scaffolding that would elicit more information about participant problem-solving strategies at all levels. Despite the limitations of the problem design, we were able to collect some data from our expert and novice think-alouds. However, the data did not appear to confirm our model. In our design of the decision tree, we did not differentiate between rules, which are correct all the time but can be difficult to apply, and heuristics, which are not always correct but can be easier to apply. More seriously, it appears that the decision tree representation is not appropriate for this ill-defined domain. Based on the expert performance, a task analysis which incorporates a representation that includes the application of rules in any order and allows the holistic processing of sentences would be more appropriate.

## ADAPTING MODEL-BUILDING TECHNIQUES

### Rational Task Analysis

Our first step in adapting traditional modeling methodology was to develop an alternative representation that is more suitable for use in this class of ill-defined domains. Instead of a decision tree, we adopted a model loosely based on scientific experimentation that better explains the data from our initial exploration and better fits the parameters of this ill-structured task. In this model, as individuals begin to solve the problem, they keep in mind two competing hypotheses: *H1* – The sentence should use the passé composé, and *H2* – The sentence should use the imparfait. The individual applies methods (or rules) for gathering evidence that support one of the hypotheses with a certain degree of weight. The weight depends on the difficulty of the problem and the experience of the student. After students have gathered enough evidence in favor of a hypothesis (i.e., the level of supporting evidence crosses a certain threshold), they make a decision. Notice that this representation can be translated into English production rules such as: 1) IF method *A* yields result *B*, THEN increment evidence for *H1* by *C*, and 2) IF *D* is the amount of evidence for *H1*, *E* is the threshold for accepting *H1*, and $D > E$, THEN accept *H1*.

Using this representation, it is easier to formalize the rules and heuristics relevant to identifying the aspect of a sentence. Experts read a sentence from beginning to end and apply the three methods in Table 2 simultaneously. The evidence provided by these methods leads them to select the tense of the sentence. For an expert, at least one of the methods always provides evidence strong enough to confirm one of the hypotheses. It is important to note that although Table 2 represents a complete mapping of the problem space, individual expert knowledge ranges from this level of specificity to a much more general representation (e.g., one containing only the complete/incomplete distinction, but covering all cases).

| Method | Evidence for Passé Composé | Evidence for Imparfait |
|---|---|---|
| Determine the manner of action | Action is one time<br>Action is repeated (specific number of times) | Action is habitual<br>Action is repeated (generalized) |
| Determine the duration of action | Action has a start<br>Action has an end<br>Action has a specific duration<br>Action is completed | Action is incomplete<br>Action is in progress |
| Determine the role of action in sentence | Action is an event<br>Action is a change | Action is a description<br>Action is a context |

Table 2: Available expert problem-solving methods

In this model, an expert might read the sentence, "It was raining" and apply the three methods simultaneously. The manner and the duration of the action are not entirely clear, and would produce results that weakly confirm the hypothesis that the action is in the imparfait. However, the role of the action is clearly a description, and strongly confirms the hypothesis. The expert would therefore choose the imparfait tense. It is important to note that we are treating these evidence-gathering methods as black boxes, and avoid describing exactly how one determines that a sentence is a description.

This evidence-gathering approach to task analysis also gives some insight into where novices may make errors. There are four areas where novice performance might differ from expert performance. First, novices may be unaware of all the methods they can use to determine the aspect of the sentence. For example, they may not know that the duration of the action relates to the aspect. Second, novices may know that they should be using a method, but may be unable to do so. For example, they may be unable to identify the duration of an action. Third, novices may understand that a given method provides evidence, but be unsure of what the evidence means. For example, they may know that the duration of the action is important, but forget whether it provides evidence for the passé composé or imparfait tense. Finally, novices may be aware of a method and able to apply it correctly to support a particular hypothesis, but they may be unclear to what extent the evidence supports the hypothesis.

In addition, novices may compensate for their lack of expertise by using imperfect methods for gathering evidence. Three such methods are listed in Table 3. Although these heuristics can supplement expert methods for determining the tense, they sometimes yield erroneous results. If the two pieces of evidence gathered are conflicting, the novice uses the weighting of each piece of evidence to resolve conflicts and determine whether the tense is passé composé or imparfait.

| Method | Evidence for Passé Composé | Evidence for Imparfait |
|---|---|---|
| Identify a temporal keyword | Word signifies a completed action or an action which occurs a specific number of times | Word signifies a habitual or incomplete action |
| Identify verb type | There is an action verb in the sentence | There is a state verb in the sentence |
| Identify an interruption | There is an interrupting clause | There is a interrupted clause |

Table 3: Supplementary novice problem-solving heuristics

We believe that this model addresses the concerns that arose from the decision tree-based model. Namely, unlike the previous model which made no mention of implicit knowledge, now implicit knowledge is accounted for by the existence of evidence gathering methods at different levels of abstraction. Expert knowledge can be represented as a single evidence gathering method ("Identify the completeness of the action"), rather than the more specific methods we have described. Moreover, rules can be applied in any order, represented by specifying a list of evidence gathering methods than can be used, rather than a sequence of actions that must be performed to solve the problem. Finally, the approach handles uncertainty by specifying a threshold for accepting a given hypothesis, and allowing different evidence gathering methods to be weighted based on characteristics of the problem and the skill of the problem-solver. This uncertainty also allows heuristics, or rules that do not necessarily work all the time, to be counted as valid evidence gathering methods that can be used by novices. Following traditional modeling procedure, this model was then refined using data from additional think-aloud protocols.

### Think-aloud Protocols

As evidenced by the first think-aloud study, simply observing student output fails to capture the underlying processes by which the student arrived at his or her final decision. As such, we supplemented our primary task with secondary scaffolding tasks that elicited richer responses from the student.

The first task we added was requiring students to provide an explanation for their answer in addition to the answer itself. When students are required to explain their decisions, the tutor learns the rule the student is trying to apply. We experimented both with rule-based and freeform explanations. Further, self-explanation, both free-form and rule-based, has been shown to increase student learning in intelligent tutoring systems [Aleven, 2002].

Unfortunately, simply providing an explanation does not necessarily yield information regarding the features of the sentence that lead students to initially apply the rule. For this information, we asked students to identify a sentence comparable to the problem sentence with respect to the features for choosing aspect. This process highlights the features students use to make their decision. With the added scaffolding, not only is the student required to actively process the text but insight into the process being used to derive the answer is now available to the tutor; they are no longer simply making a binary decision. Forcing students to make comparisons between examples is particularly helpful for teaching feature discrimination in ill-defined domains [DeKeyser, 2005].

Finally, we added some instruction at the beginning of the task, based on our model of expert performance, in order to remind students of the uses of the aspects and insure that all participants shared a common foundation for the domain. Additionally, based on interviews with experts, we determined that context is critical for removing ambiguity in the selection of aspect. For that reason, we opted to situate the individual problem sentences within a paragraph. See Figure 2 for a screenshot detailing the types of scaffolding provided.

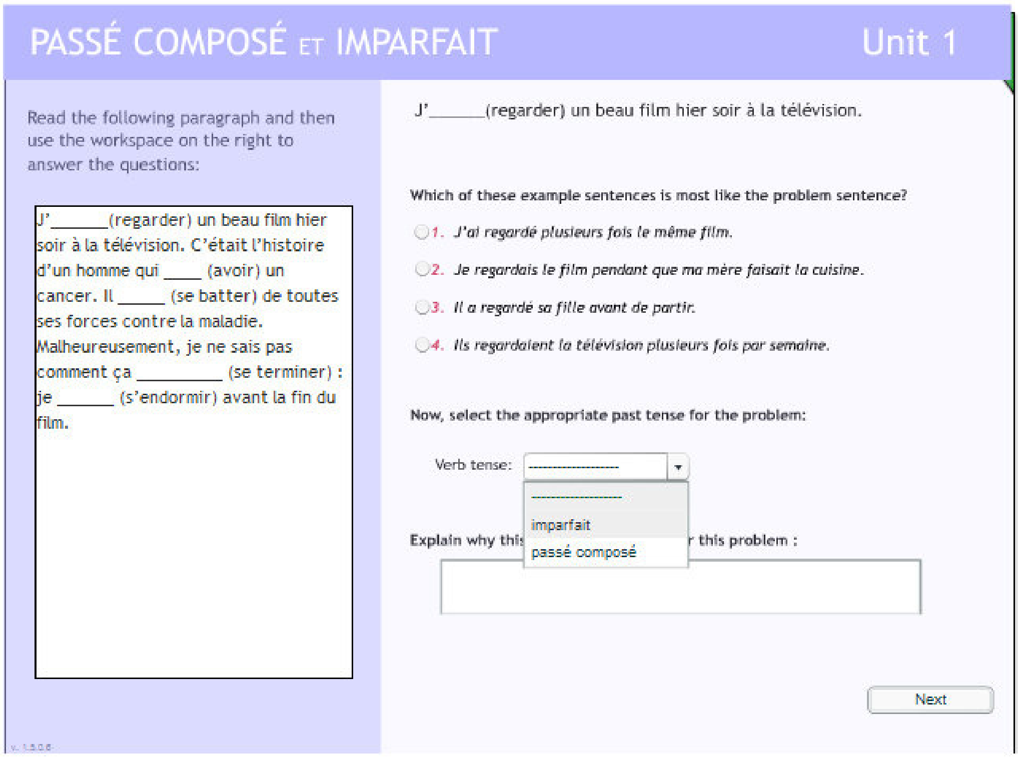


Figure 2: Screenshot of tutor interface with context and scaffolding

We collected think aloud data from six novice French speakers between the ages of 18-29. All participants had recently completed at least the first semester of college French (or equivalent) but had not continued beyond the second semester. Participants had differing levels of exposure to the passé compose and the imparfait. Students were given a pretest, instruction regarding the use of aspect, and then asked to think out loud as they solved the task. Students completed this activity for four different paragraphs with four different types of scaffolding (no scaffolding, comparison example, self-explanation, combined comparison example and self-explanation). During this phase, students were given immediate feedback on their performance. Each session concluded with students completing a post-test and transfer assessment.

## Results

Much of the behavior we saw supported our original model. Since students were primed with the rules and heuristics of the model during the instruction phase, it is not surprising that the language they used when explaining their reasons was similar to that which was presented. However, the fact that they were able to acquire and successfully use the model given only brief exposure (average time reading instruction = 4.5 minutes) might suggest that our model doesn't deviate much from students' existing models. We also found some direct evidence for the idea that students were in fact conducting evidence gathering when making their decisions. When presented with conflicting evidence, students would verbalize the conflict and look for other features to support one aspect or the other. Even when conflicts were absent, some students were reluctant to make a decision based on limited information. For example, P1's behavior and statement "[the sentence] doesn't say exactly when it happened, no specific time, but it's an event" suggest that she was trying to gather more evidence that the sentence was in the passé compose before making a final decision.

Perhaps more interesting is our model also accurately predicted areas where students would have difficulty. We proposed four main areas where novice behavior may differ from expert behavior. Each of these are re-examined and supported by samples of student behavior:

1. *Unawareness of evidence gathering methods* -- It was common for participants to rely on a handful of well-known rules and attempt to classify the aspect based on this small subset only. For example, P1 never used description as a way of determining aspect, even when it would have been appropriate, suggesting that she was not aware of that method for gathering evidence.
2. *Inability to apply evidence gathering methods correctly* -- Students also failed to correctly apply the methods even when it was clear that they correctly understood the concepts behind them. For example, when determining the aspect of the following sentence: "Nous *avons dû* ranger les affaires en vitesse", the student failed to recognize that "en vitesse" is a time keyword. We know that the student was looking for a keyword because transcripts of the session show the student incorrectly acknowledging that there is "no specific mention of time".
3. *Inability to link the results of methods to particular hypotheses* -- Only one participant, P4, exhibited frequent incorrect mappings between the evidence in the sentence and the hypothesis being supported. In one exercise, she used the explanation, "a one time action" to explain both uses of the imparfait and passé composé. Later, the incorrect mappings between results and hypotheses became more evident when she commented that specific duration and completed action meant using the imparfait when in fact the passé compose should be used.
4. *Lack of understanding of how much evidence contributes to a given hypothesis* -- We also saw evidence that suggested novices placed too much weight on some heuristics, often failing to continue examining the sentence. For example, P2, heavily weighed the use of time keywords. Upon noticing one in the sentence, she automatically chose the passé composé even when other evidence in the sentence suggested otherwise.

Participants used several heuristics when solving the problem, only a handful of which were identified in the previous model. A table of the heuristics including the participants who used them follows:

| **Method** | **Passé Composé** | **Imparfait** |
|---|---|---|
| Identify the type of verb | Action verb (2, 4, 6) | State verb (1, 3, 7)<br>Not an action verb (2) |
| Identify the specificity of the action | Specific (1, 4, 6) | Not specific (1, 4, 6)<br>Vague statement (1) |
| Identify the suddenness of the activity | Happened all of a sudden (2)<br>Interruption word (4, 7)<br>Happened once (2, 3, 6, 7) | Something that occurs frequently (1)<br>Ongoing Activity (3)<br>Habit (3, 4, 6, 7) |
| Identify a mention of times | Time keyword (2, 3, 6) | No mention of exact time (1, 2)<br>Unspecified number of times (3, 4, 7) |
| Look for a comparable verb | The tense of the other verb is passé composé (2, 6) | The tense of the other verb is imparfait (2, 6) |

Table 4: Heuristics used by novices in second Think-Aloud

## Discussion

The preliminary results suggest that the model developed using a combined theory-driven and data-driven approach is a decent representation of student behavior, and an improvement over the previous model. The new model is based on a representation that is suitable for ill-defined domains: It incorporates holistic sentence processing, the application of rules in any order, and allows both rules and heuristics to be employed by students.

In addition, we improved our data-collection methodology so that student data could better inform the model. The added prompts for self-explanation and comparison to examples lead students to talk more about the rules and sentence features they were using to solve the problem, and exposed some misconceptions that would not otherwise have been clear. Ultimately, the data collected fits the model. Students show an evidence gathering approach to identifying aspect, and use the rules and heuristics that we have described to help them arrive at the correct answer. It is important to note that this model is preliminary, and serves as a plausible interpretation of the data. Further validation and specification of the model will be necessary.

In particular, there are still unanswered questions with respect to the completeness of the model. Because we provided students with the rules involved in the model when we gave them instruction on identifying aspect, we could not fully evaluate whether the rules map to actual novice performance. However, these rules are similar to instruction that students actually receive on the difference between the passé composé and the imparfait, so they do represent knowledge that novices should have already been exposed to, and represent knowledge that students are expected to learn. Therefore, the fact that students employed the rules after limited exposure to them might suggest that a student model for solving the problem is similar to the theory-driven model that we described. Additionally, we did not do a full analysis of the "helping" strategies that students used to solve problems such as translation of the sentence or looking for context in surrounding sentences. In the future, we intend to incorporate those types of heuristics into our model. Finally, we limited our analysis to a high level of abstraction, and did not look at how students parse the sentence to identify features that may be relevant for the conditions of the rules in our model. We believe that scaffolding and tutoring mechanisms can be built from our model which do not require such a fine grain of analysis to be effective.

## FUTURE DIRECTIONS

Developing an accurate student model marks the beginning of full tutor development. With our current model, we plan to create a full cognitive tutor to be deployed and evaluated in an online French course. The tutor will likely incorporate the scaffolds used during the think-aloud procedure, but future studies are also planned to identify the exact combination of scaffolds that lead to the greatest learning gains. During this process, we will continue to refine and evaluate our model for identifying aspect in the French past tense. We hope that our model, representation, and techniques can be generalized to other ill-defined domains where the rules are ill-structured but the start and goal states are easily formalized.

## ACKNOWLEDGMENTS

Thanks to Dr. Christopher Jones, Dr. Vincent Aleven, Dr. Ken Koedinger, Anne Catherine Delmelle, and Alida Skogsholm for their help with this project.

## REFERENCES

Aleven, V., & Koedinger, K. R. (2002). An Effective Meta-cognitive Strategy: Learning by Doing and Explaining with a Computer-Based Cognitive Tutor. *Cognitive Science, 26*(2), 147-179.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y . (2004). An integrated theory of the mind. *Psychological Review 111,* (4). 1036-1060.

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. W. (2002). Learning from examples: Instructional principles from the worked examples research. Review of Educational Research.

Cadierno, T. (1995) Formal Instruction from a Processing Perspective: An Introduction to the French Past Tense. Modern Language Journal. 79. 179-193.

Comrie, B. (1976). Aspect. Cambridge: Cambridge University Press.

Davis, Tessier, (1996). Authoring and Design for the WWW. Advisory Group on Computer Graphics.

DeKeyser, R. (2005) What Makes Learning Second Language Grammar Difficult? A Review of Issues. *Language Learning, 55*, 1-25

Ericsson & Simon (1984). Ericsson, K. A, & Simon, H. (1984). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

Gamper, J. & Knapp, J. (2002). A review of intelligent CALL systems.Computer Assisted Language Learning, 15(4):329–342.

Ishida, M. (2004) Effects of Recasts on the Acquisition of the Aspectual Form te i(ru) by Learners of Japanese as a Foreign Language. Language Learning 54:2, 311-394.

Jonassen, D.H., Tessmer, M., & Hannum, W.H. (1999). Task analysis methods for instructional design. Mahwah, NJ: L. Erlbaum Associates.

Koedinger, K. R.; Anderson, J. R.; Hadley, W. H.; and Mark, M. A. (1997). Intelligent Tutoring Goes to School in the Big City. Journal of Artificial Intelligence in Education 8(1): 30–43.

Mayo, M., & Mitrovic, A. (2001). Optimising ITS behavior with Bayesian networks and decision theory. *IJAIED*, *12*(3), 124-153. Project homepage at http://www.cosc.canterbury.ac.nz/~tanja/capit.html

Norris, J., and Ortega, L. (2000) Effectiveness of L2 Instruction: A Research Synthesis and Quantitative Meta-analysis. *Language Learning, 50*(3), 417-528.
Ormerod, T.C. (2006). Planning and ill-defined problems. Chapter in R. Morris and G. Ward (Eds.): The Cognitive Psychology of Planning. London: Psychology Press.
Robert M. DeKeyser. (2005) What Makes Learning Second Language Grammar Difficult? A Review of Issues. Language Learning 55:s1, 1-25.
Salaberry, R. (1998). The development of aspectual distinctions in L2 French classroom learning. The Canadian Modern Language Review, 54, 508-542.
Schooler, J.W., Ohlsson, S., and Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General 122*, 166-183.
Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology,* 8, 481-520.
Simon, H. (1973). The structure of ill-structured problems, Artificial Intelligence, 4:181-201.
Spiro, R. J., Feltovich, P. J., Jacobson, M. J., & Coulson, R. L. (1995). Cognitive Flexibility, constructivism, and hypertext: Random access instruction for advance knowledge acquisition in ill-structured domains. In P. Stele, & J. Gale, *Constructivism in education* (pp. 85-108). Hillsdale, NJ: Erlbaum.
von Stutterheim, C. (1991). Narrative and description: Temporal reference in second language acquisition. In Crosscurrents in second language acquisition and linguistic theories (pp. 385-403). Philadelphia: Benjamins.