MODELING LEARNING PROCESSES IN LEXICAL CALL

Robin Goodfellow, Aston University, and
Diana Laurillard, Open University, United Kingdom

ABSTRACT

To devise ways of tracking learning processes is one of the goals of current CALL research. However, the data collected must be able to account for the learner's approach as well as his/her performance. A study of a novice learner of Spanish using a CALL system designed for vocabulary enlargement, shows that introspective evidence may be used to validate performance data within a theoretical framework which characterizes the learning approach as "surface" or "deep," depending on the degree of the learner's engagement with principles of structure underlying the target language lexis. It is proposed that this framework offers a basis on which effective computer based vocabulary teaching strategies can be developed.

KEYWORDS

CALL, vocabulary, lexical structure, performance data, introspective data, learning approach, teaching strategy.

INTRODUCTION

The idea that data generated in the interaction between learners and CALL programs could provide us with information about language learning processes has created a lot of interest (e.g.: Garrett 1993, Sussex 1991, Chapelle and Jamieson 1989, Jung 1987 etc.). The attraction is that the computer's ability to record complex processes accurately and unobtrusively means that we can use it to tell us exactly what learners do. However, whilst the general principle is clear, precisely what we should do with this information is not. One suggestion is, for example, that the computer can “pile up” language acquisition data relating to the frequency and quality of student errors (Jung). Another is that keystroke data can be used in the description of students' monitoring strategies.
(Chapelle and Jamieson). A third is that CALL environments could provide a “cognitive workbench” for research (Sussex). A fourth proposal is for "detailed and sophisticated assessment instruments" to be used in the evaluation of our theories of language learning (Garrett 1993).

But the problem is that, however thorough and detailed the computer’s record may be, it doesn’t describe what the learner thinks s/he is doing. This can lead to misconceptions, and, as Chapelle has demonstrated, misconceptions concerning what subjects do during instructional activities lead to inaccurate interpretation of results (Chapelle 1990). For this reason, the promise of rich research pickings held out by the computer-as-tracker has yet to be fulfilled.

In this paper I will argue that we also need to take into consideration the subjective dimension of the learner’s approach to the task, if we wish to interpret the data we collect in terms of learning processes. The learner's approach describes the relation between the student and the learning s/he is doing (Ramsden 1992). In CALL it is based on their perception of the task as represented by the interaction design. It is likely to involve alternative strategies and learning theories to those the program design presupposes, and may even come into conflict with them. Thus, even where we have a principled cognitive model underlying the design of the interaction, a further understanding of what the learner thinks about what s/he is trying to do is essential if we are to describe his/her learning processes.

To illustrate, I shall describe a CALL system which has been designed on the basis of a cognitive model of word production, to support self access vocabulary learning. Some results from a study of the use of the system by a learner of Spanish are discussed, showing that while his approach, revealed in introspective data, appears to confound the pedagogical intentions of the program, analysis of it in terms of a "deep/surface" dichotomy (Ramsden, 42) offers a principled way to relate it to the performance data. The conclusion derived from both sets of data is that the learner's interpretation of the task stems from a 'surface” approach which is consistent with his level and the difficulty of the task, and that the program has failed, in his case, to promote the “deep” approach which the design intended should result in a better quality of learning outcome. It is
proposed that the concept of deep/surface approach could be applied to the design of CALL research programs, which then could support both the elucidation of the learning processes involved in learning languages with computers, and the development of more effective teaching strategies.

LEXICAL CALL

Lexical Knowledge

Systematic vocabulary learning makes sense if you are in a self-access language learning environment (i.e.: one that isn't in a target-language-speaking country or classroom), because incidental learning of words (i.e.: by exposure alone) requires the sort of linguistic input and repetition that is only available through natural communication, and this is, by definition, absent. However, a systematic approach is only half the story; for effective learning we also need to tailor the content. Lexical knowledge can be viewed qualitatively — the learner's ability to relate specific items to specific contexts (i.e.: to distinguish between senses and connotations, to appropriately collocate, to correctly inflect etc.), or quantitatively — the size of the learner's vocabulary in terms of their ability to recognize and produce items irrespective of context and parameters of quality described above.

Proficiency clearly requires development along both these dimensions, and teaching approaches can emphasize either, but currently qualitative knowledge is very much in favor, and writers such as McCarthy (1984), Clarke (1992), etc. have stressed the importance of discourse to the way we access meaning, even to the extent of dismissing the traditional definitional dictionary, with its alphabetical access, as an “obstruction” to a learner's understanding (Clarke, 144). There are, however, limitations in the capability of CALL to explicate this kind of knowledge. Many systems which address the problem of lexical knowledge rely on a hypertext approach to generate “links” between vocabulary items and the multiplicity of senses, contexts and discourses in which they are deployed. Such systems require a highly specified and labor intensive organization of the learning material before they can be of use to learners (e.g.: see Lyman-Hager 1993), and are inevitably limited in the amount of input they can provide for the learner.
Vocabulary Size

Another strand in vocabulary acquisition research has concerned itself with issues of vocabulary size and has taken up the challenge of the scale of the learning task and the apparent slowness with which L2 acquisition proceeds compared to L1 (e.g.: Meara 1980). This area of interest is related to work which has been done in psycholinguistics on 'lexical decision" tasks in the native language, in which the simple recognition of the fact that a string of letters is a word constitutes knowledge of that word (see Forster 1989 for a discussion of the theories involved). The idea that the biggest problem in vocabulary learning is the size of the task has long been recognized in teaching approaches which stress initial enlargement and productive practice of a learner's vocabulary rather than refinement of it. For example, Nattinger (1988) considers that it is more important for students to use the newly stored language as effortlessly and quickly as possible than it is for them to wait for control of precise vocabulary or perfect grammar; Judd (1978) argues that it is necessary to tolerate some vagueness in the initial teaching of some lexical items and that more precise meanings will develop as the vocabulary is repeated. This approach has proved far more amenable to CALL, and a large percentage of the vocabulary programs that Jung observes in his survey of the field (Jung 1988) take a restricted view of what word knowledge should consist of and focus on the learner's ability to enlarge on the number of words they can produce given an appropriate prompt. But there is a persistent problem with this “quantitative” approach, and that is the essential dullness of the activities (largely to do with memorization) with which it is associated. A commonly expressed view among the learners I have spoken to in my own research is that they don't do vocabulary learning exercises either in books or on computers because they are 'boring.'

CALL DESIGN

Lexical CALL design, therefore, has to overcome two problems: it has to avoid a huge materials preparation effort, which means that it is reliant on existing teaching resources such as dictionaries, thesauruses, etc., and it has to be interesting to the learner, which means that it has to offer something more than these traditional resources do. One solution, providing a 'halfway house" between quantitative and qualitative approaches to vocabulary learning, and offering scope in content and opportunity for more interesting practice activity, is to promote a discovery approach to
words, based on the idea of grouping. Grouping makes use of structural principles of organization of items in the L2 lexicon. These are not generative in the sense that syntactic structure is (e.g.: once you know the rule for forming past tense endings for regular verbs you can generate the past tense of any regular verb you meet), and they do not extend to all areas of vocabulary, but they are sufficiently marked in many areas (e.g.: hyponomy — the covering of specific terms by more general terms) to be used by teachers and learners as a way of organizing new items so as to render them more learnable. This idea is the explicit basis for CALL programs such as LEXNET (Swartz 1991), which employs associative networks to generate practice of semantically linked words in French.

The issue of what constitutes lexical structure is too complex for full coverage here (Carter and McCarthy 1988, 18-38 presents a good overview), but it can be summarized as the tendency of many words to enter into relations of meaning and to share features of spelling and pronunciation, syntactic role, collocational combination etc. This is a 'halfway house' in the sense that a learner's target words can be related to each other even if they are not given their whole significance within a particular discourse. The qualitative aspects which CALL is good at can therefore continue to be exploited, but the learner can also aspire to a level of processing which involves more than simple memorization.

A design for lexical CALL which aims at the enlargement of vocabulary and at an increased awareness of lexical structure, but which does not need extensive materials preparation, can thus be proposed. It is based on the assumption that the learner can be given on line access to fairly large lexical resources, such as text, dictionary, concordancer, etc. and that a pedagogical strategy which directs their use of these resources towards an awareness of how words can be related, and uses this awareness to promote productive practice can be incorporated. To implement such a design necessarily involves modelling learning and cognitive processes.

THEORIES AND STRATEGIES

The learning model adopted for this design is a representation of a theory about conscious word learning procedures. The cognitive model is a representation of lexical storage and retrieval at the level of a theoretical description of non-conscious cognitive processes. The pedagogical strategy is the organizational principle which links the two
representations to the objective of learning words via their structural associations.

Learning Model

The learning model proposes three logical stages in conscious word learning: reception of new words, their integration into the mental lexicon, and their retrieval for comprehension or for production. In the first stage the learner identifies unfamiliar items and decides whether to select them as targets for study. In the second stage they discover something about the meanings and behavior of the new items and organize the information in a way that will help them to retrieve them by association. In the third stage they practice recalling the items. (For the purposes of computer based self access learning the target language is assumed to be in written form; the vocabulary items are derived from written text, looked up in dictionaries and thesauruses, etc. and recalled by being typed in response to a prompt). The stages described by this model are what a CALL design for vocabulary learning must be able to support, and the learning processes involved are those that require elucidation from the data which the computer collects and those which the learner’s introspections provide.

Cognitive Model

The cognitive model represents the unconscious information processing which is assumed to underlie the learner’s behavior in the interaction. It is based on the processes which are described in the psycholinguistic literature on L1 speech processing (e.g.: Aitchison 1987; Dell 1986 etc.). The L2 learner’s mental lexicon is thought of as an active network of meanings and phonological forms. Items in the network have associative links of varying strengths according to the similarities of sense, sound, syntax or context that they share. These links are responsible for passing ‘activation’ from one item to another, so that a strong link will pass a lot of activation and a weak link will pass little or none. The activation level of an item is what is responsible for its being produced, for its ‘popping out’ of memory. Activation spreads simultaneously to all the neighbors of a highly activated item, so that there is a kind of competition for production among several items with similar levels of readiness. Depending on the time available for production the speaker either selects the item which best fits the semantic and syntactic frame which has been constructed, or is "bounced" into production of whichever item arrives first (the theory is used to explain L1 speech errors where a
wrong word or syllable creeps into an utterance without the speaker being aware of it, e.g.: “...Liszt's second Hungarian restaurant...,” Dell, 285). These processes are what a CALL design must be able to operationalize, in order to track, theoretically, a learner's parallel non-conscious processing as s/he uses the program.

**Pedagogical Strategy**

I have described the pedagogical implications for L2 of such a cognitive model, in more detail elsewhere (Goodfellow 1993). Briefly: weakly activated items (i.e.: words that are barely known) need to be strengthened, and strongly activated items can be used to activate them, provided there is a link available. This means that if the computer has a model of the learner's mental network then this can be used to organize practice according to the status of individual target items. This is what I shall call the Learner Model. The way to build it is to get the learner to select, process and practice vocabulary items as if they were overtly related and structured, rather than distinct and idiosyncratic, and this is achieved by promoting the tactic of word grouping (e.g., see Oxford and Crookall 1990), whereby the learner organizes L2 target items into sets according to common features and then those common features are used as clues in practicing recall. The teaching strategy enables the program to identify those items which the learner has done more processing on (e.g.: put into more groups), but has not yet practiced. These items can be presented first, on the assumption that they have more associations and that the learner will be able to recall something of them even if not of their exact forms. The act of recall itself generates associations, so when unprocessed (i.e.: weakly associated), unpracticed items are presented later, it is assumed that the learner’s network will be in a more highly activated state.

**Summary — Learning Outcomes**

The conscious activity of word grouping, therefore, is intended to engage the learner in creating an implicit description of aspects of the lexical structure of the target language, and the learner model created whilst this description is being built, is used by the program to help the learner retrieve the individual items on which the description has been based. The required learning outcome at the information processing level is a set of target words organized into groups and successful retrieval of all the words in the set. At the learning processes level it is the learner's confidence that s/he is able to generate the words for use in a communicative context. The functionality of a system designed to implement this approach is described below, and a study of a learner's
using it follows. This provides the basis for a discussion of the issues surrounding the interpretation of data produced by interaction with the program. As well as the Learner Model and the Performance Trace (the sequential record of the learner's actions on mouse and keyboard), data is derived from the learner's own comments on the interaction and from a self-assessment exercise which was carried out at the end of the study.

PROGRAM FUNCTIONALITY AND RESEARCH METHODOLOGY

The three stages of the learning process described above (reception, integration and retrieval) are implemented as three separate processing modules, called Selection, Grouping and Practice. These modules are supported by a corpus (database of texts), lexical tools (bilingual dictionary, L2 thesaurus, keyword-in-context concordancer) and the Learner Model.

Selection Module

Learners select target vocabulary by highlighting the items they want in a scrollable window full of text. The learner selects items, which may be single words or phrases, by putting them, on the basis of a subjective judgement, into one or all of four category boxes entitled: SENSE (for items where it is the meaning which is of interest), SOUND (for items with some interesting pronunciation or spelling feature), SYNTAX (for items with some interesting grammatical feature), and CONTEXT (for items with some interesting contextual feature). An element of structuring of the target word list is thus present from the start, provided that the learner is consistent in his/her criteria for matching items with categories.

Grouping Module

Once items have been selected, the learner then uses the lexical tools (dictionary, concordancer, thesaurus) to get information about meaning, synonyms and contexts of occurrence for each item. The learner notes whatever information s/he wants, and the program saves the notes. The learner then puts the items together into groups according to their semantic, formal and contextual features and identifies each group by a title of their own choosing. The program saves these groupings.
Practice Module

A target item is selected by the system, and a practice prompt is given which consists of the truncated (60 character) context for every occurrence of that item in the corpus, with the item itself gapped out (Figure 1). The learner has to recall the missing item and enter it. There are a number of clues available, i.e.: a Groups clue, which retrieves the groups that the target item has been included in (the program displays the group titles plus all the other items which belong to those groups), a Notes clue which displays the notes that the learner has written for the target item, a wordshape clue which identifies the vowels and consonants in the target item, and a Grammar clue which identifies affixes, endings etc. in the target item where appropriate. The program gives feedback on the learner’s response, identifying correct strings of letters in the answer. The selection of items, structuring of the list, and the generation and utilization of Note clue information is thus the learner’s own responsibility. The program directly controls the sequence of presentation of target items for recall practice, the content of the two form clues (Wordshape and Grammar) and the assessment of correctness of the learner’s responses. The sequence of presentation is based on the representation held by the Learner Model.

Figure 1. Practice module showing prompt for target word “esfuerzos”
Learner Model

The Learner Model is a 'snapshot' of the types and amounts of information processing done on each item in each of the three modules. When the learner first selects an item in the Selection module the Learner Model sets the value for Selection-module-processing for that item to 1. Whenever that item is subsequently highlighted in this module the value is incremented by 1. Similarly in the Grouping module, using the concordancer and saving notes each result in the incrementing of the item's score for that type of processing. In addition, when an item is assigned to a group, the code for that group is added to the item's record. The Learner Model thus records how many and which groups each item belongs to. In the Practice module, the learner's attempts to recall items also result in changes to a value for Practice-module-processing for the item, incrementing it if the attempt is successful or decrementing it if the attempt is an incorrect one on a previously successful item. A representation is thus maintained, measured by the number and type of events involving a particular item in each module, of the presumed level of activation of that item in each of three general categories of processing:

**Context Processing** — involving awareness of the item in context. The events counted are highlighting the item in its textual context in the Selection module and doing KWIC concordances on it in the Grouping module.

** Meaning Processing** – focusing on semantic features. The events are looking up and noting translations or definitions for the item and putting it into groups in the Grouping module. Although it is possible that items will be grouped on syntactic or phonological/orthographic criteria, it has been observed in pilot studies that a) most groupings are semantic, and b) where they are not semantic, the meanings tend to be looked up anyway as part of a general focusing strategy. For this reason the assumption is made that grouping implies processing of meaning.

**Production Processing** — involving producing an item which has already been processed for meaning. The key event is the retrieving of an item in the Practice Module.

This representation enables items to be compared in terms of their relative levels of assumed activation, according to the cognitive model which underlies the design of the teaching strategy. This comparison is used, by the teaching strategy, to sequence the
presentation of items for recall so as to maximize the effects of activation spreading from highly processed items downwards. The order is determined according to the following rationale: items combining a high level of meaning processing and a low level of production processing are practiced first, items with high levels of context processing are practiced next, and items with high levels of production processing are practiced last. This rationale, embedded in the program’s pedagogical strategy, is equivalent to an assumption that there are stages in the learning of each item. Providing that it can be validated, the Learner Model, at any given point in time, stands as a representation of a particular state of the learner’s knowledge of the target list as a whole.

Performance Trace

The Performance Trace consists of the translations and other information which has been recorded during the integration stage and also a sequential record of every action executed via keyboard or mouse. There are 39 possible actions (including entering and leaving modules), so the procedural part of the trace is likely to be quite large, and any analysis is dependent on the "chunking" of groups of actions together to be understood as procedures related to the achievement of some sub goal underlying the main learning objectives. The procedures which are of the most relevance to a description of learning processes are those related to the implicit objective of the pedagogical strategy, which is to build a structural description of the target lexis. Procedures relating to this objective might include searching through texts for items which can be related to other items already selected, reassigning a selected item to another category in order to add it to an existing group, creating new groups, assigning "extra" items to groups (i.e.: words which are not in the target texts but which are known and expected to generate useful associations), using group clues in the practice module, etc. Provided that it can be correctly interpreted, the Performance Trace stands as a representation of the learner’s approach to the task as a whole.

Learner Strategy and Research Methodology

Strategic structuring of the target list is encouraged by the pedagogical design of the interaction, i.e.: the use of the four category boxes at the time of initial selection and the facility to group target words together and then to use the groups as clues for retrieval. One other feature also attempts to promote word grouping, which is that there is a limit to the number of items (25) which can be contained by the four category boxes. When
the limit is reached, some items have to be 'finished with" (in the terminology of the program) before new ones can be selected. When an item has been finished with, it continues to be selected for testing in the Practice module, and prompted in the same way (i.e.: the practice prompts, see Fig 1) but the only clue that is available is the Groups clue (i.e.: no Notes, Wordshape or Grammar information). If the learner has not included the item in any groups then s/he gets no clues. The theory is that this will foster grouping because in a very large target-word set (say 250 words), 90@o of the items will be finished with and the gapped contexts of the practice prompts are often not sufficiently informative to guarantee recall. Extensive grouping as a strategy ensures that there is always supplementary clue information available and that this information relates the target item to other L2 items in the learner's lexicon.

But word grouping is not the only strategy logically available to the learner in this interaction. They could, for example, ignore the grouping functions altogether and focus on the production of extensive notes for individual items. Or they could simply associate each item with its L1 translation. Or they could adopt some other unpredictable strategy involving off line activities that the program is unable to recognize. These eventualities separate the meaning of the Learner Model from that of the Performance Trace, the learning theory from the cognitive model, and make them unreliable on their own for the description of learning processes. This is the reason why a separate source of data from the learner is required.

Introspective data can shed light on the nature of unpredicted strategies and also serve as an interpretative link between the cognitive and learning theories. If the outcome represented by the Learner Model is in broad correspondence with the learner's assessment of his knowledge, then we can say that the processes recorded in the Performance Trace (because the Learner Model is a product of them) include processes representative of the learner's approach (because the assessment is a product of it). These processes can then be identified in one of two ways: either they were intended by the pedagogical strategy, in which case they can be described directly by the learning theory together with the learning outcome, or they were not, in which case they must first be described in terms of the learner's approach (revealed by the introspective data) and the learning outcome, before the learning theory is broadened to take account of them.
In the case study described below, the learner does not adopt the strategy which is implicit in the program design. But a comparison of the program's data with his assessment of the learning outcome can be made, because the different parts of the Learner Model representation can be made to correspond to the Learner's subjective judgements: 'I remember the context of this item," 'I know the meaning of this item' and 'I could use this item.' The Use judgement corresponds to a high level of production processing (use implies knowledge of meaning). The Meaning and Context judgements correspond to high levels of meaning and context processing. From this comparison, and with the aid of the deep/surface dichotomy, it can be shown that both sources of information support the same conclusion, which is that the learner's approach arises from a misconception about lexical structure and is not an optimal one for the task. In this way it can be shown that the Performance Trace, at least in this case, is capable of giving us information about learning as well as about information processing.

CASE STUDY

A single case will be described here and selected results considered. The objective is to establish the extent to which the data in the Performance Trace/Learner Model, a) corresponds to the Subject's self assessment and introspective reports, b) reflects the pedagogical intentions of the program, and c) matches the quality of the learning outcome. A deep/surface framework for learner approaches is used to explain the agreements and discrepancies between the different data sources and to unify the conclusions which can be drawn from them.

Procedure

The program used was as described in the section above, with the exception that it did not contain a Spanish thesaurus. The Subject was a 25 year old English learner of Spanish in London. His level was just above elementary, as he had not studied Spanish formally and had just been "picking it up" for about 12 months in short visits to Spain and in occasional classes and conversations with Spanish speaking friends in London. For this study he set himself the task of using the program to learn 50 new Spanish words in 2 weeks. There were 5 observed sessions during this period: an introductory session, 3 observed practice sessions, and a recorded interview. He was permitted to use any other on- or off-line methods to achieve the 50 word goal as he saw fit.
program recorded all his actions with the mouse and keyboard during the observed sessions (the Performance Trace) and also built a quantitative model of his lexical knowledge in the form of a set of integer values attached to each selected item, representing the amount of processing of different kinds that he had done on that item (the Learner Model). His comments and replies to questions were noted by the observer. Questions usually took the form of 'why did you do that?' referring to some action or other, or 'what are you thinking about now?' when he lapsed into silence.

In the final interview he was asked to assess his success at learning his target words. The assessment took the form of a series of judgements about each of the L2 words in his target list. Three criteria for assessment were proposed to him, corresponding to the kinds of processing assumed for him by the Learner Model: i) whether he could remember anything about the context in which he originally encountered the word, ii) whether he understood the meaning of the word, iii) whether he could use the word. He was asked to consider each word in the list and mark it 'yes,' 'no,' or 'maybe' for each criterion. He was asked to think aloud whilst doing so, and his comments were recorded.

**Quantitative Results**

During the five hours, eight minutes of the Subject's use of the program (spread over two weeks), the Performance Trace recorded 552 mouse or keyboard events, distributed across the program modules (Table 1). The learning outcome can be represented initially as the total number of words selected, the number successfully retrieved, and the number of groups created (Table 2).

This falls far short of the 50 word target. The learning rate (23.7 minutes of study for every 1 successful retrieval) is also considerably slower than the average (16.5 minutes for every retrieval) for 6 similar pilot studies carried out with intermediate learners of Spanish. The small number of items selected (27) and the number of those retrieved (13) reflect the relatively short amount of time he spent in the Selection and Practice modules (41.7%), but the single group he created in no way accords with the time he spent (59.7%) or the extent of his activity (60.0% of events) in the Grouping module.
Qualitative Results

An explanation of what the Subject was doing during the 59% of the time he spent in the Grouping module was elicited during one of the observed sessions:

...Once I've selected the words I want, I go into the Grouping module. I click on each of the words and write it down on a piece of paper. Then I go into the dictionary and look up each of the words in turn. I write down next to it the main translations. Then I go back and go through clicking them again. I do a concordance on each one and also look it up in the big (off line) dictionary. I work out which of the meanings is right and enter that into the notes box...
His strategy was, apparently, to memorize the English translations for the items he selected. Where the program did not support this, e.g., dictionary and concordancer were inadequate or could not be used simultaneously, he resorted to off line solutions such as noting words to look up in a book dictionary, printing out the text in hard copy, etc. As a result, he was operating with a mental list of L1-L2 pairs and not with a network of L2-L2 associations. In the absence of context or production processing, many of the pairs subsequently became decoupled; this interpretation is supported by his comments on the words acontecen, comportamiento and rodeado, elicited during the assessment exercise (Table 3).

<table>
<thead>
<tr>
<th>Word</th>
<th>Subject's comments as he tried to assess his knowledge of it</th>
</tr>
</thead>
</table>
| acontecen    | (Eng. meaning = “they happen”)
  "...I've got a few meanings in my head...it could be another one...”
  "...it could have meant “confront” but I think enfrentando means that...”
  "...I was looking at comportamiento...” |
| comportamiento | (Eng. meaning = “behavior”)
  "...I think that meant “the behavior...” |
| rodeado      | (Eng. meaning = “surrounded”)
  "...what's come to me is the English translation of a word I used which was “surround”...”
  "...I don't know if rodeado means “surround” or if that is another word... ...or acontecen...” |
| enfrentando  | (Eng. meaning = “confronting”)
  "...I think that means “to confront” |

(re: Eng. word “surround”)
  "...“surround” is either acontecen or comportamiento...one means “behavior”...”
  "...if I was to guess I'd say acontecen means “surround”...”

**Table 3. Subject’s comments during assessment**
Self Assessment vs. Performance Data

Does the Subject’s assessment confirm the performance data? It is clear from the quantitative data that the Subject was not successful in either of the strategic objectives assumed by the program design, i.e., create groups/ maximize retrieval. The bulk of the information contained in the Performance data must therefore be representative of his alternative, i.e., L1 translation, strategy. A lot of this processing is therefore opaque to the Learner Model, e.g., the Trace records that in session 3, in the Grouping module, in 29 out of the 38 events where the Subject clicked on an item, the highlighting of the item was not followed by either a grouping event, dictionary or concordance event, or a notes saving event. In session 5, none of the 29 clicks on items in the Grouping module had any follow-up. In session 6, only 1 of 27 events had any follow-up, etc. The Subject’s comments on this strategy (he was clicking each word as he wrote it down on paper), elucidate this otherwise uninterpretable data, but his failure to conform to the program’s expectations means that much of his processing remains unrepresented in the Learner Model.

This lack of correspondence is a direct result of the failure of his approach to conform to the expectations of the pedagogical design and shows the complexity of the problem of modelling the learner on the basis of computer generated data. However, if we look at what the Model does show and compare it with the learner’s evaluation, it is possible to find common ground. The table below (Table 4) shows which of the items recorded by the Model as having an above average activation level for context processing (see section 3.4 above) was also marked positive by the Subject for his ability to remember the context. Similarly, items recorded by the Model as high for meaning processing, and production processing are confirmed by the Subject as positive for knowledge of meaning and ability to use. The comparison shows that for 32 out of the 40 item/category judgements made by the Model (80%), there is a confirming evaluation made by the Subject.

The performance data, as represented by the Learner Model, is therefore validated by the learner’s assessment, to the extent that the convergence of his approach to the program’s pedagogical intentions permits.
<table>
<thead>
<tr>
<th>Item</th>
<th>Context</th>
<th>Meaning</th>
<th>Production</th>
</tr>
</thead>
<tbody>
<tr>
<td>estuercos</td>
<td>X</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>profundas</td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>palmeras</td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>favorecer</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>desprende</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>perturbado</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>fracasado</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>de hoy en día</td>
<td></td>
<td>X</td>
<td>-</td>
</tr>
<tr>
<td>acontecen</td>
<td>X</td>
<td>X</td>
<td>-</td>
</tr>
<tr>
<td>comportamiento</td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>motivaciones</td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>sueño</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>ventanas</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>rodeado</td>
<td>X</td>
<td>X</td>
<td>-</td>
</tr>
<tr>
<td>arcada</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>estrecha</td>
<td>X</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>enlozado</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>húmedo</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>agrega el estudio</td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>aporta</td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>enfrentando</td>
<td></td>
<td></td>
<td>-</td>
</tr>
</tbody>
</table>

- = assessment confirmed
x = assessment not confirmed
blank = no assessment

Table 4. Confirmation of learner model’s assessment by subject’s assessment
Performance Data vs. Pedagogical Intentions

Does the Performance data actually reflect the program's pedagogical intentions? As we have observed, the Subject did not adopt the preferred strategy, and if we look in the Performance Trace for procedures related to grouping or lexical categorizing we find them few and far between, for example:

Events in the Selection Module related to the reassignment of items to categories (SENSE, SOUND etc.): 7 out of 65. The program design requires the reassignment of items in this module if the learner wants to group together items which were initially put into different categories. This is supposed to encourage reflection on the way the lexical categories are being used. The Trace also shows that the Subject's use of categories was more or less restricted to the SENSE and CO@TEXT boxes (only 4 items were assigned to SOUND, none to SYNTAX).

Events in the Grouping module related to the creation of groups and assignment of items to them: 11 out of 330 (only 1 group created, containing 2 items). I have already pointed out that the Subject's strategy did not involve grouping because he focused on L2-L1 pairing. The Trace reflects the fact that he recorded notes for most of the items but only grouped 2 of them together on the superficial basis that they were expressions and not single words.

Events in the Practice module related to the strategic use of the groups clue: 4 out of 157. The total number of words selected and the extent of the Subject's use of the Practice Module also failed to meet the program's expectations, with the consequence that the embedded strategy for encouraging grouping (the procedure for 'finishing with" items - see previous section) was irrelevant. The Subject's overall reluctance to practice his target words also meant that values for retrieval in the Learner Model were low and this contributed to the incompleteness of the Model's representations in the Use category which depends on retrieval values for its assessment. The Trace therefore fails to reflect the program's intention that retrieval should be supported with grouping information.
Performance Data vs. Learning Outcome

To what extent do the Performance Trace and the Learner Model data represent the quality of the learning outcome? In terms of the explicit objective of maximizing successful retrieval of selected words, the outcome is poor (see table 2 above). The same is true for the implicit objective of learning about lexical structure. The only group created (‘expressions’) specifies a relatively trivial relationship. The subjective data also confirms that the optimal network creating strategy supported by the pedagogical design was not taken up. Judged by the overall aims of the design, therefore, the quality of the outcome is low, demonstrating Pask's principle that if the style of the teaching does not match the approach of the learner then the outcome is poor (Pask 1976). But in order to compare this qualitative evaluation with the representation of the learner's use of the program which the program itself holds, i.e.: the Performance Trace and Learner Model, it is necessary to describe the data in similar, general, qualitative terms. This is where a conceptual framework for describing the learner's approach is required. The framework should relate what the learner thinks about the task (introspective data), to what they did (Performance data), and to what the overall learning outcome was (qualitative evaluation on the basis of pedagogical strategy).

Deep/Surface Approaches

The optimal strategy for vocabulary learning encouraged by the program design is based on the principle that mnemonic associations generated between lexical items in the target language will serve a) to make those items easier to retrieve, and b) to build an understanding of the kinds of general and language specific relations that can exist between words in the L2. The program encourages lexical categorization from the moment that new words are first selected, supports the creation and expansion of common feature groups, and promotes the use of these groups as cues in recall practice of the constituent words. A learner who consciously adopts this strategy and who deliberately organizes his/her selections and processing around the need to find new classifications and groupings for existing items is thought to be approaching the learning task in a "better" way than one who concentrates on decoding the selected items in their contexts and expressing the resulting understanding in L1. This dichotomy in approach is what has to underlie any evaluation of the performance data if it is to be related to the learning outcome.
One way to characterize the dichotomy is to use the distinction between 'deep' and "surface" approaches to the learning task, which has been proposed in the context of the learning processes of students in higher education (e.g.: Ramsden 1992, Marton 1986, Laurillard 1979, etc.). The characteristics of a deep approach are that the learner attends to the significance of the task and is able to see a relationship between the way that the knowledge is organized and/or described, and the eventual learning outcome. A surface approach is one which focuses on the component parts of the task and on procedures which guarantee outcomes at that level only (Ramsden, 43) The first approach applied to reading a text in L1 would focus on understanding the message of the text, the second on memorizing the information in it. This dichotomy provides a useful tool for evaluating a learner's approach to a learning task, but it does present some problems in the language learning case. If we were looking simply at the ability of the Subject in the case study to understand the L2 text (which concerns the problems of adolescents in post industrial society), then we could investigate whether he had grasped the relationship between, for example, the study identified in the expression "agrega el estudio" ("...the study adds...") and the young people, described as "fracasado" ("...failed...") , "perturbado" ("...troubled...") etc. This would be equivalent to finding out if the learner understood the structure of the text as well as what it referred to.

A deep approach would be one that tried to apprehend the point of the study's findings on the problems of adolescence while a surface approach would stop at being able to quote the study. But, as I have already discussed in Section 1, the discourse is not all that we are interested in. We are interested in the words and the way they are related linguistically. So the structure we are interested in is not just in the message of the text but in the explicit relations that can be created between the words themselves, e.g.: that "fracasado" and "perturbado" both express negative states, that they can be linked as cause and effect, that they both have near equivalents in English, that they both have "...ado" past participles, that both states can be consequential on "esfuerzos" (efforts), that they form a logical group with words like "decepcionado" (disappointed) and "desgraciado" (unfortunate) etc.

This is easy to mistake for a surface approach itself, focusing on the signs rather than what is signified, but if the point of learning is "...to improve one's understanding of some phenomenon in the world around by means of taking part in a symbolic
representation of that phenomenon..." (Marton, 2) and if the phenomenon to be understood is the structure of lexis in a target language, then the symbolic representation is not only the text(s) in which the words are embedded, but also the dictionary, thesaurus, and KWIC citation, and the taking part is not only the comprehending of the structure of the text but also the comprehending of the structure of the information contained in these resources.

Where textual structure can be either hierarchical or sequential (Marton, 20). Lexical structure can be paradigmatic or syntagmatic, where textual reference can focus on the component parts or on the principle that holds them together (ibid), the L2 lexicon can be referred to as a list of L1-L2 equivalents or as a network of L2 associations. The deep surface dichotomy can therefore be applied to the approach of the vocabulary learner with regard to the way that the task, finding links between words, is related to the desired learning outcome, using the words. This means that the evaluation of an actual learning outcome as rich or poor, as has been done in the study under discussion, can be related to the evaluation of the learner's approach as deep or surface, according to the extent to which his actions make explicit some underlying structure among the lexical items he has selected.

Applying these principles to the interpretation of the Subject's performance as reported above, we can classify many of the actions and events which are recorded in the Performance Trace, and evaluate them for the contribution that they make to the quality of the learning outcome. The surface deep dichotomy is used to structure the record of learner actions into those which are not related to an approach intending to establish common features among words, and those which are. The surface level actions are those which are intended to establish links between target words and their L1 translations, deep level actions are those which make use of the explications of meaning, form and structure offered by the lexical resources to create meaningful associations among the target words. It is clear that not all classifications of learner actions can be made simply on the basis of the actions themselves, but that some require the Subject's introspective comments to elucidate his intentions. These comments are derived from his remarks during the self assessment session and from protocols recorded during the observed sessions. Table 5 illustrates some of the classifications possible from the case study.
These examples are illustrative of the classification of the learner’s actions and comments as part of a surface or deep approach. A full classification of the entire Performance Trace would obviously cover many more action types and, although there might be a certain amount of ambiguity in some of the interpretations, it is clear that an assessment could be made of the extent to which the learner’s performance conformed to a surface or deep profile, and that this assessment could be related to quality in the learning process. In this Subject’s case, the examples of deep processing identified above are almost the only ones that occur in the whole interaction, so the quality of the overall learning process, as represented by the data, can be characterized as surface level and therefore likely to yield a poor outcome. This, as we have seen, would be in accordance with the quantitative evaluation of the learning outcome.

**SUMMARY**

In the program described above, the main pedagogical objective is explicit in the interaction design, in the sense that the facilities for separating words from context, for testing them, etc., clearly promote the learning of vocabulary. But there is a second objective, implicit in the design, which is to develop an awareness of lexical structure, in the sense that learners who make sensible decisions about lexical categories, word
<table>
<thead>
<tr>
<th>ACTION TYPE</th>
<th>Surface................................</th>
<th>APPROACH TYPE................................</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assign or reassign</td>
<td>Assignment of item is arbitrary, e.g.: Subject’s comment “…I just put everything into the CONTEXT box…”</td>
<td>Assignment of item is considered, e.g.: he reassigned <em>aporta</em> from SENSE to SOUND because he noticed that it was a near-cognate: “…I wanted to change the box because it means ‘bring’ as in ‘port’…”</td>
</tr>
<tr>
<td>target item to category boxes (Text Module)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Create group (Grouping Module)</td>
<td>Group title does not reflect significant lexical feature, e.g.: his creation of group “expressions” including <em>agrega el estudio and de hoy en dia</em></td>
<td>Group title reflects significant lexical feature (no examples)</td>
</tr>
<tr>
<td>Write notes on item (Grouping Module)</td>
<td>L1 translation copied verbatim from dictionary: e.g.: Subject’s notes for <em>aporta</em>: “…show up, arrive, port…” were irrelevant to the contextual meaning “bring”</td>
<td>Select appropriate L1 sense or modify grammar to fit form of target item or write L2 paraphrase for selected sense and form of target item. E.g.: his notes for <em>fracasado</em>: “…have been unsuccessful…” reflected its contextual meaning rather than the literal translation “failed”</td>
</tr>
<tr>
<td>Refer to KWIC (Grouping Module)</td>
<td>Do the concordance on the current form of the target item. e.g.: he did concordances on word forms as they appeared in the target word list and not on their root forms, e.g.: <em>acontecen</em> (3rd pers. pl.) produces 1 citation — <em>acontec</em> would have produced 3</td>
<td>Do the concordance on a root form for the target item (no examples)</td>
</tr>
<tr>
<td>Select clue type (Test Module)</td>
<td>Use translation clue e.g.: he used translation to prompt retrieval of <em>favorecer, entrentando, estuerzos</em></td>
<td>Use groups clue, e.g.: Subject used groups clue to prompt retrieval of <em>agrega el estudio and de hoy en dia</em></td>
</tr>
</tbody>
</table>

Table 5. Classification of learner actions by approach type
groupings etc. get more help from the program. Data from the case study shows that a poor learning outcome can be related to the learner's failure to adopt (or the program's failure to promote) the learning strategy which is appropriate to the implicit objective. This conclusion, deduced from the perspective of a conceptual framework which describes the learner's approach in terms of deep and surface learning processes, is supported by both performance and introspective data.

CONCLUSIONS AND FURTHER RESEARCH

I have argued that we must study the accounts that learners give of their interactions with CALL programs, as well as the performance traces that the interactions produce, if we are to describe the learning processes which the medium supports. I gave as an example some data generated by a novice learner of Spanish using a vocabulary learning program designed for the selection and integration of new words. That the learner in this study adopted a surface approach to the task is not to be interpreted as characteristic of anything like a learning "style" particular to him. Learners themselves cannot be characterized in terms of a dichotomized description of learning because learning is determined partly by the nature of the task environment (Laurillard, 408). In this case the task itself, and the task environment, contained features which might have predisposed the Subject, a novice in Spanish, to focus on the unstructured relations between L2 items and their translations rather than on meaning in the text or lexical similarity within the target word list. The texts were authentic, with high lexical content and many unfamiliar words. Text and dictionary windows could not be consulted at the same time, and there was no text translation available. Under these circumstances, with a low level learner, it may be that the adopted approach was the optimal one. If that is the case, and it can only be confirmed by further case studies, then the program's learning model needs to be broadened to account for the fact, and its pedagogical strategy to be enhanced accordingly to provide the support which is necessary to get the learner to the stage where a deeper approach is possible. Future research should therefore focus on the following:
• The validation of the deep/surface approach conceptual framework for vocabulary learning via more studies of learners at both ends of the spectrum, showing that the learning processes so described are consistently represented in the nature and extent of learner use of the facilities the program offers for categorization, grouping and cluing of lexical items.

• The development of the program design, giving it the capability to further adapt the teaching strategy to the individual learner (e.g., by giving help in the initial assignment of new items to lexical feature categories), on the basis of an evaluation of that learner's current approach. This would necessarily involve the formal specification of more types of lexical relation, perhaps on the basis of thesaurus and collocation resource banks, and also a 'discussion level' mode of discourse for user and system by which these relationships could be explicated. (Some work of this type has already been done by Cumming et al. 1993 etc.)

The complexity of programs like the one described here (which is in fact much less complex than some of the multimedia based systems described at CALICO 93, see Borchardt and Johnson 1993) demands a degree of sophistication in the user/learner which cannot yet be taken for granted. Research is thus concerned not only with prototypical systems but with prototypical learners too. These prototypical learners are breaking new ground in the matter of learning processes. In their progress towards mastery of systems like these there is much they can tell us, not only about the systems, but also about themselves.

REFERENCES


AUTHORS' BIODATA

Robin Goodfellow is a Research Associate in the Department of Modern Languages and European Studies at Aston University where he is developing software for French and German grammar learning as part of the Technology in Teaching and Learning Project (funded by the UK Higher Education Funding Council).

Diana Laurillard is a Senior Lecturer at the Open University, Institute of Educational Technology, and is head of a project on Learner Use of Media.

AUTHORS' ADDRESS

Robin Goodfellow
TLTP Astcovea Project
Technology
Dept. of Modern Languages and European Studies
Aston University
Birmingham B4 7ET
United Kingdom

Phone: +44 081 359-3611 ext. 4221
E-mail: r.goodfellow@aston.ac.uk

Diana Laurillard
Institute of Educational Technology
Open University
Walton Hall
Milton Keynes MK7 6AA
United Kingdom