

## THE USE OF COMPUTER TECHNOLOGY IN EXPERIMENTAL STUDIES OF SECOND LANGUAGE ACQUISITION: A SURVEY OF SOME TECHNIQUES AND SOME ONGOING STUDIES

**Jan H. Hulstijn**

University of Amsterdam

### ABSTRACT

This paper first gives a brief characterization of the ways in which second-language acquisition researchers use the computer to elicit L2 production data or to record how L2 learners process L2 input. Eight tasks and/or techniques are described; most of them borrowed from the experimental toolbox of psychologists. The paper then describes the use of computer technology in some ongoing investigations in which the author participates. These investigations pertain to the acquisition of automaticity in L2 reading, writing, and listening, and to the use of electronic bilingual dictionaries.

---

### PART ONE: SURVEY

#### Characterization of the Domain Surveyed

The first part of this paper provides a survey of the ways in which the computer has been, and is used in the study of second language acquisition (SLA). The field of SLA research encompasses a broad variety of orientations. It is mainly in the cognitive and linguistic orientations that the computer is used as a data elicitation device. Thus, the computer does not figure prominently as a data elicitation device in research conducted from a functional/pragmatic, interactional, sociocultural, or sociolinguistic perspective. Some phonetic research is concerned with SLA (e.g., Flege, 1988) and most of the work conducted in the laboratories of phoneticians involves the use of computers. This area, however, will not be considered in this paper.

The paper will be confined to the use of the computer as a tool to elicit L2 production data or manifestations of the ways in which L2 learners process L2 input. Thus, the paper will not be concerned with the use of the computer for writing research reports (word processing), statistical analysis (e.g., SPSS), corpus analysis with concordancing software, communication (e-mail), or information gathering (internet).<sup>1</sup> Two software packages have been developed especially for the analysis of interlanguage: COMOLA (Jagtman & Bongaerts, 1994) and COALA (Pienemann, 1992). Some L2 researchers have used the Childes package (MacWhinney, 1995; Sokolov & Snow, 1994) for the uniform transcription and storage of learner utterances into a large database, making the data accessible to colleagues worldwide.

#### Literature of Computer-Aided Research Methods

Overall, one could say that SLA data are seldom elicited with the use of computer-aided techniques. Handbooks of research on SLA or applied linguistics (Brown, 1988; Larsen-Freeman & Long, 1991, chap. 2; Nunan, 1992; Seliger & Shohamy, 1989; Tarone, Gass, & Cohen, 1994), or journal specials devoted to research methods (Kasper & Grotjahn, 1991) do not mention the use of computer-assisted data elicitation techniques, except, of course, the use of software packages for statistical analyses (e.g., Hatch & Lazaraton, 1991). A perusal of papers published between 1990 and 1999 in the international SLA journals *Second Language Research*, *Studies in Second Language Acquisition*, *Language Learning*, and *Applied Linguistics*, and of papers published in the more "applied" journals, *System*, *The Modern*

*Language Journal*, and *Language Teaching Research*, revealed the use of only a small number of computer-assisted elicitation techniques in only a few areas within the SLA domain.

If SLA researchers use the computer at all, they almost always borrow the tools developed in psychology departments for psycholinguistic research. Most of these techniques are described in the monumental *Handbook of Psycholinguistics* (1994), edited by Gernsbacher. This excellent handbook has 34 chapters on spoken and written word recognition, speech perception and production, text and discourse comprehension and production, and many other topics. Psycholinguists employ a wide variety of elicitation tasks and techniques, most of which apply the computer in one way or another (e.g., the measurement of eye movements and event-related brain potentials). Psychology departments of most universities have developed their own software tools. Some tools are commercially available, such as,

- PsyScope, a product of Carnegie Mellon University (<http://psyscope.psy.cmu.edu>; see Cohen, MacWhinney, Flatt, & Provost [1993]);
- the Micro Experimental Laboratory (MEL) of Psychology Software Tool Inc. (<http://www.pstnet.com/melpro/melpro.htm>);
- E-Prime, the successor of MEL, also produced by Psychology Software Tools Inc. (<http://www.pstnet.com/e-prime/e-prime.htm>);
- NESU (Nijmegen Experimental SetUp) developed by the Max Planck Institute of Psycholinguistics (<http://www.mpi.nl/world/tg/experiments/nesu.html>);
- tools produced by Life Sciences Associates (<http://lifesciassoc.home.pipeline.com/>); and
- more tools described at <http://www.york.ac.uk/inst/ctipsych/dir/experiment.html>

These programs offer an environment for developing, prototyping, and executing experiments with a graphical, interactive design tool (selection of icons of experimental functions, which are dropped on time lines); they do not assume prior computer programming experience.

### Survey of Learner Tasks and Computer-Aided Presentation and Elicitation Techniques

**Grammaticality Judgement Task.** The most widely used task in linguistically oriented SLA research is the grammaticality judgement task. Normally, this task is administered with paper and pencil, but sometimes researchers are interested in the speed with which L2 learners pass their judgments. In such cases they use software developed in psychology departments for psycholinguistic research. L2 learners are presented with sentences (one at a time) on the computer screen and press one of two buttons or keys to indicate whether the sentence is grammatical or ungrammatical; the software allows the registration of participants' responses (correct or incorrect) as well as the speed (in milliseconds) with which the responses are given.

**Sentence matching task.** Only rarely do SLA researchers use the more complicated technique of sentence comparison or sentence matching (e.g., Beck, 1998; Clahsen & Hong, 1995; Eubank, 1993). In the sentence matching (SM) task, which was originally developed by Freedman & Forster (1985), two stimulus sentences are presented, one after another and for a very brief time, on a computer monitor. The first sentence appears left-justified in the uppermost portion of the screen, and the second appears well offset to the right in the lowermost portion. The first member of a stimulus pair remains on screen for a predetermined period of time (the delay; e.g., 1700 ms), and, when the delay elapses, the second member of a stimulus pair appears while an internal timer starts to measure response latencies (RLs). The subjects' task is to determine, as quickly as possible, whether the first and the second stimuli are identical or not. When they make a decision they press one of two buttons. At that point, the timer is stopped and the RL is registered (Beck, 1998, pp. 322-323). Bley-Vroman & Masterson (1989) discuss the pros and cons of the measurement of reaction times in grammaticality judgement and sentence matching or comparison tasks.

**Oral Production.** Psycholinguistic software also allows the measurement of the speed of oral responses. Participants speak their responses to stimuli into a microphone which is connected to the computer's internal clock via a voice activator. Linguistically oriented SLA researchers have seldom used this technique. A rare example is the study conducted by Beck (1997), in which participants (native and nonnative users of English) were shown the stems of regular and irregular verbs and had to produce, as quickly as possible, the form of the simple past tense of these verbs.

**Word Recognition.** The most widely used technique in cognitively oriented SLA research is the measurement of word recognition via the so-called lexical decision task. De Bot, Cox, Ralston, Schaufeli, and Weltens (1995) describe this task as follows:

In a lexical decision task subjects are presented with letter sequences (in the visual version of the task) or phoneme sequences (in the auditory version) that may or may not be words in a given language. The subjects' task is to indicate whether the sequence is a word or not by pushing a 'yes' or 'no' button as quickly as possible. Stimuli typically consist of words ('desk'), pseudo-words ('derk') and sometimes nonwords ('dker'). In a lexical decision task with repetition priming, the same stimulus is presented twice in the course of the experiment. This repetition leads to a shortening of the latency on the second presentation because the level of activation of that particular word is still high as a result of the first presentation. In another variant the target word is preceded by a word that is a semantically and/or associatively related word, the 'prime', as in the pair 'nurse' < 'doctor'. Semantic priming also results in a shorter latency for the target word as compared to the same target word preceded by an unrelated prime." (p. 3)

De Bot et al. used this technique in their research on the organization of the bilingual lexicon, that is the bilingual's mental representation of words known in L1 and in L2, to test the interdependency hypothesis against the independence hypothesis. The interdependency hypothesis predicts that there should be repetition and semantic priming between languages, while the independence hypothesis predicts that this should be only so within languages. An overview of the vast empirical literature on the organization of the bilingual lexicon, most of which based on lexical decision data, is given by Kroll & De Groot (1997). Lexical decision also forms the central task in research by Segalowitz (summarized in Segalowitz, 1997; see also Segalowitz, Segalowitz, & Wood, 1998) on the attainment of automaticity and fluency in L2 reading. An aural lexical decision is used in research conducted by Poelmans, Van Heuven, and Hulstijn, described in [part two](#) of this paper.

**Sentence and Paragraph Reading.** In the empirical literature on cognitive processing of L2 reading processes we come across the occasional measurement of sentence or paragraph reading times. For instance, Horiba (1993, 1996), presented L1 and L2 readers of both Japanese and English with stories for a story recall task. Readers were shown the texts on the computer screen one sentence at a time. Participants read at their own pace, advancing through the text by pressing the arrow key. Reading times for each sentence and text were recorded by a timer in the computer. In this type of research, sentence and/or text reading times have been analyzed in comparison with a number of other variables, such as quantity and quality of the verbal recall, level of L2 proficiency, and text variables (e.g., degree of coherence). Furthermore, reading times in L2 have been compared with reading times in L1; also, the L2 reading task has been repeated two or three times in order to measure any decreases in reading times when the text was reread two or three times, in relation to the quality of the recalls, which had to be given after each reading. In section 2.1, a longitudinal project is described which uses laptop computers for the measurement of word recognition as well as for sentence reading in L1 and L2, in three annual sessions, in order to assess whether the role of lower-order processes in L2 reading (and L2 writing) changes over time, with increasing L2 proficiency.

**Form-Function Mapping.** According to the Competition Model of Bates and MacWhinney (1989), an essential part of language learning consists of discovering the functional meanings of the grammatical forms of the language to be learned. As there seldom exists a one-to-one relationship between forms and their meaning(s), and as the form-function mappings of different forms and functions may stand in competition with each other, language learners have to find out what the strength is of this relationship for each form separately, and to what extent there is competition between these form-function relationships. An impressive number of empirical studies have been based on the Competition Model. In most of these studies, which were conducted with the aid of the computer, participants were presented with a series of three words, two nouns and one verb in any order (VNN, NVN, NNV). Participants had to indicate which of the two nouns they interpreted as the subject or agent of the action expressed by the verb. In some experiments they did so by pressing one of two buttons, in other experiments they called the noun out loud in a computer-connected microphone. The computer registered both the choices and the reaction times. Most of these studies were conducted with native speakers, but some studies used this technique with L2 learners (e.g., Issidorides & Hulstijn, 1992; Kilborn & Cooreman, 1987; Kilborn & Ito, 1989).

**Connectionist Simulations.** A relatively new and increasingly important type of SLA research aims to simulate learners' language production over time as correctly as possible, thereby trying to describe and explain language acquisition. Such simulations are based on neural networks, often of the so-called connectionist type (Broeder & Plunkett, 1994). Connectionist simulations have been shown to be fairly successful in reproducing the acquisition of morphology by L1 learners, and recently also by L2 learners of Danish (Jensen & Ulbaek, 1994), French (Sokolik & Smith, 1992), and Russian and German (Kempe & MacWhinney, 1998).

**L2 Learning Experiments.** L2 learning experiments conducted under tightly controlled conditions form another relatively new type of SLA research. Hulstijn (1997) reviews twenty of these studies published since 1988. In many of these laboratory studies the computer was used for input presentation, learning instructions, feedback, and the elicitation and registration of responses, with or without reaction times. One of these studies, reported in Yang and Givón (1997) and Yang (1997), will be briefly described here, as it illustrates nicely (in this author's opinion) how computers can be used in L2 learning research in a number of perspectives: linguistically and cognitively, as well as educationally. In this study, which was called the Keck (also called Keki) Second Language Learning Project, a group of monolingual English speakers spent two hours a day, five days a week, during a period of five weeks, learning Keki, an artificial language specially constructed by the researchers, under one of two learning conditions. One group (the pidgin group) received grammatically simplified pidgin input for 20 hours of instruction and then fully grammatical input for the remaining 30 hours. The other group (the grammar group) was introduced to the grammar via fully grammatical input right from the start. Besides the single manipulation of the input, all other aspects of instruction and testing were identical (Yang & Givón, 1997, p. 176). The aim of the study was to test Givón's competition hypothesis that posits that, in early L2 acquisition, vocabulary and grammar compete for memory, attention, and processing. Because one can communicate with vocabulary in absence of grammar but not vice versa, Yang and Givón proposed that the pidgin group would acquire vocabulary more efficiently than the grammar group, challenged with the dual task of acquiring vocabulary and grammar simultaneously (Yang & Givón, 1997, pp. 173-174). The computer was used for almost all learning and testing tasks. Testing tasks consisted of word recognition (Keki words, English words, and nonwords), semantic priming in English and Keki, word translation from Keki into English and vice versa, sentence elicitation (immediate and delayed recall memory for both English and Keki sentences), grammaticality judgments of Keki grammatical and ungrammatical sentences, picture description (not conducted with the computer), and narrative translation and comprehension (participants were presented with a short Keki narrative, one sentence at a time and were asked to translate each sentence as accurately as possible; they then answered four comprehension questions). Yang's study, as well as the laboratory language learning experiments of DeKeyser (1997) and

De Graaff (1997), demonstrate that it is possible to motivate individuals to devote themselves to the learning of an artificial language during several weeks. This is an important methodological feat and should encourage scholars, whether interested in SLA from a linguistic, a cognitive or an educational perspective, to devise experiments adopting a similar method.

### **Research on the Use of Multimedia Software**

In more "applied" areas, related to SLA, the computer is used as well. For instance, in the area of language teaching, electronic multimedia hardware and software has replaced the traditional audio and video systems in the language laboratory. Hence, educational research comparing the relative benefits of various teaching techniques and methods is bound to investigate the use of these multimedia systems. Unfortunately, there is far too little research of this type. The most likely reason for this dearth of research is that, on the one hand, the companies that develop and produce these systems are hesitant to finance such investigations, and that, on the other hand, foundations for the funding of academic research, often find such "applied" investigations less important than research considered more "fundamental." Examples are to be found in the journal *Computer Assisted Language Learning (CALL)*, published by Swets and Zeitlinger, and the present journal *Language Learning and Technology* (<http://lt.msu.edu>).

## **PART TWO: ONGOING RESEARCH IN WHICH THE AUTHOR PARTICIPATES**

This part gives a description of some studies in which the author has been, or is, involved. The first group of studies is concerned with the acquisition of automaticity in L2 reading, writing and listening. In these studies, the computer is mainly used for the measurement of reaction times in word and sentence recognition. The second group of studies is concerned with the use of electronic dictionaries. At the Department of SLA of the University of Amsterdam, we are currently planning a series of computer-controlled experiments of L2 input processing; however, the design of these experiments is presently too rudimentary to be rendered here.)

### **Automaticity Studies**

The first study is called "Transfer of higher-order processes and skills in reading and writing in Dutch and English." This is a four-year research project, conducted by a group of seven researchers: Amos van Gelderen, Rob Schoonen, Annegien Simis, Patrick Snellings, Marie Stevenson, Kees de Glopper, and Jan Hulstijn, SCO-Kohnstamm Institute for Educational Research and Faculty of Humanities, Universiteit van Amsterdam. The project, which is funded by the Netherlands Organization for Scientific Research (NWO, grant 575-36-001), consists of four interconnected studies. The first one is a longitudinal study of higher-order and lower-order reading and writing proficiency (in terms of product and process) in English L2 and Dutch L1 among approximately 300 students from grade 8 to 10. Students are tested in three consecutive years (1999-2001) on a large battery of tests in Dutch and English, measuring four dimensions of proficiency:

- Functional reading and writing
- Knowledge of vocabulary and grammar
- Automaticity of word and sentence recognition
- Metacognition.

The main research question is, *To what extent is the correlation between higher-order reading skills in Dutch L1 and in English L2, and the correlation between higher-order writing skills in L1 and L2, dependent on the degree of automatization of L2 knowledge?* In order to measure the dimension of automaticity with which L2 knowledge can be used, students perform speed tests in both L1 and L2, measuring word and sentence recognition. Word recognition is measured in the traditional form described in section "Word Recognition." The measurement of sentence recognition is, as far as we know, new in

SLA research. Students are presented with propositions that are true or false in terms of normal, every-day knowledge of the world, such as, "John likes sports, so he plays in a soccer team" (T), or "John likes sports, so he never plays soccer" (F). The sentences are made up of vocabulary and grammatical constructions that have been taught earlier. Whereas the lexical decision task may allow students to react only on the basis of word *form* recognition, without accessing word meaning, the sentence test forces students to process linguistic forms to the level of *meaning*. What we aim to investigate is whether reaction times in the word and sentence recognition tasks in L2 go down over the three-year course of the study and to what extent such a decrease might affect performance on L2 reading comprehension tests, thus allowing the correlation between L1 and L2 reading comprehension to increase.

Study 2 consists of an in-depth longitudinal investigation of the processes of L1 and L2 reading and writing. A sample of 16 students is followed during three consecutive years, from grade 8 to grade 10. These students perform all of the automaticity tests of study 1, and, in addition, they perform reading comprehension and writing tasks under think-aloud or stimulated-recall conditions. Students perform the writing tasks at the computer. The computer registers all writing actions, composing so-called log files, which allow an unobtrusive observation of the writing process. What students say aloud is recorded on a conventional audio tape recorder.

Study 3 is an intervention study, in which grade 8 students are given a training aimed at increasing the automaticity of their L2 word knowledge. The aim of the study is to provide evidence for the hypothesis that automaticity of lower-order skills (processing at the word and sentence level) is a necessary condition for the successful deployment of higher-order reading comprehension skills. In this experiment, which adopts a pretest-treatment-posttest design, computers (again laptops) are used both for pre and posttesting (as in study 1) and for training. During the training sessions, students do various tasks, under increasing time-pressure, requiring the processing of L2 words, all of which have been taught previously in the school curriculum. These tasks are presented as a kind of computer games. In a pilot study, conducted in June 1999, two exercise types were implemented. In the first one, students are shown L2 target words, one at a time, along with two L1 words, one being the correct L1 translation. Students have to press one of two keys to indicate which of the two L1 words offers the correct translation. Students go through the entire set of target words in three rounds. In the first round, they always receive feedback, regardless of whether their response has been correct or incorrect. The feedback consists of either a green tick-off sign or a red cross accompanied by the correct translation. In the first and second rounds, stimuli remain on the screen until students respond but never longer than six seconds. In the third round, however, stimuli are only briefly flashed on the screen. In all three rounds, students are encouraged to respond as fast as possible. After each set of 10 items, students are given their mean correct scores and their mean response times so far. In the second exercise type, students are shown L2 sentences (one by one) containing a gap. Underneath, two L2 words are shown, one of which does while the other does not fit the gap. Students press one of two keys to indicate which of the two words fits the gap. In all other respects, this exercise type is identical to the first one.

In this pilot study, the investigator Annegien Simis pretested word recognition (lexical decision) on 100 L2 words. Then, 40 of the words were trained in two 50-minute sessions within a one-week period with the two exercise types described. Finally, the word recognition test was administered again. The results showed that the RTs of all 100 words decreased from pre- to post-test, but this was significantly more so for the trained words than for the non-trained words. Also, the coefficient of variability (Segalowitz & Segalowitz, 1993) in the post test was smaller for the trained words than for the non-trained words.

Study 4 is also an intervention study, designed to run in parallel with study 3. The aim of study 4 is to investigate the role of automaticity of productive word knowledge and grammar skills in the attainment of fluent L2 writing. The treatment of this study is still in the planning phase.

In order to be independent of the widely different hardware and software facilities of the participating schools, the computer tests in the four studies of this project are administered on laptop PCs. The use of computers in this project was an obvious choice because all four studies pertain to the measurement of processing speed at the word and sentence levels, whereas the second and fourth study also pertain to the measurement of writing processes. However, the use of computer technology, of course, comes at a considerable price, both in terms of financial costs and in terms of logistics. The purchase and maintenance of 30 high quality laptops required a considerable amount of money, which will have to be written off in a relatively short period of four project years. The laptop PCs are expected to be outdated after the four years of this project and hence not to be suitable anymore for further research purposes. It was therefore decided that the laptops will be disposed off by lottery among the participating students after the last testing round, as one of the measures to keep student motivation high. Furthermore, as no commercially available software was deemed to suit the requirements for all tasks in all four project studies, it was decided to have the software programmed exclusively for these studies (including protection devices against unwanted student use), producing a heavy financial burden on the project budget.

Finally, the decision to conduct research in some 20 schools over a three-year period of data collection using laptops has important logistical consequences. When not used, the 30 laptops must remain locked in a special room, which is provided with 30 power outlets, allowing simultaneous recharging of batteries. No two schools can be visited at the same day, nor can more than two classes be tested at the same time. As the batteries last for only 2.5 hours, testing sessions cannot be planned for longer periods. Between school visits, backup files must be made of test data of all 30 laptops. For transportation, a minivan is needed as well as containers on wheels. A team of research assistants travels from school to school carrying 30 laptops. So far, no laptops have been stolen or damaged!

**L2 listening.** The second automaticity project to be mentioned here is called "Automatization of word recognition in Dutch as a second language." This study is conducted by Poelmans, Van Heuven and Hulstijn, and also funded by the Netherlands Organization for Scientific Research (NWO, grant 575-21-009). In this study on the attainment of fluency in the skill of listening in Dutch as a second language, an aural lexical decision task is used. Participants make lexical decisions on two types of aurally-presented word and pseudoword stimuli: half of the stimuli are words or pseudowords pronounced in isolation (with the intonation and sentence accent of a self-contained, one-word utterance), whereas the other half of the word and pseudoword stimuli were originally pronounced unaccented, in a sentence context, and subsequently excised (through digital tape editing) from their spoken context. Preliminary results suggest that responses to the pseudowords pronounced in isolation offer the best criterion distinguishing native from nonnative listeners.

### **Use of Electronic Dictionaries in Investigations on L2 Reading and Writing Processes**

In one study (Hulstijn, 1993), we wanted to investigate when L2 readers would look up the meaning of unknown words occurring in a L2 text which had to be read in order to answer comprehension questions at the higher levels of discourse, that is, beyond the understanding of individual words or sentences. The English L2 text which participants in that study (82 Dutch grade 10 and 11 students) read, was available not only on paper but also on a computer monitor. If students wanted to know the meaning of a word, they were to move the cursor to the desired word and press Enter; the L1 translation of the word immediately appeared in a window. Pressing Enter again made the window with the translation disappear. The computer program compiled log files of all these word consultations, for each individual participant. This technique allowed us a truly unobtrusive observation of participants' look-up behavior. We were particularly interested in the look-up incidence of 16 target words, selected according to a 2 x 2 design, involving the factors *relevance* and *inferability*. As for *Relevance*, half of the words were "relevant" in terms of reaching the reading goal (answering a number of comprehension questions) and half of the words were "irrelevant" in this respect. As for *inferability*, the meaning of half of the words could easily

be inferred from context whereas the meaning of the other half of the words could not. As expected, the results showed that participants almost always looked up the meaning of the *Plus Relevant & Minus Inferable* words and least often the *Minus Relevant & Plus Inferable* words (mean scores of 3.5 and 1.8 respectively, out of a maximum of 4.0). Thus, far from using the consultation facility in a blind or random fashion, participants in this study approached their task in a strategic manner, taking into account the relevance (determined by the reading goal), and, to a lesser extent, the inferability of the words they encountered.

In another study (Hulstijn & Trompetter, 1998), we observed participants' look-up behavior using a similar technique as in the 1993 study, measuring also retention of the looked-up words. The aim of the study was to investigate incidental vocabulary learning after dictionary look-up under two conditions of text processing, namely reading and writing. We wanted to find out whether dictionary use for writing purposes leads to higher incidental vocabulary learning than dictionary use for purposes of reading comprehension. We hypothesized that this was likely to be the case because using new lexical information for writing purposes would entail more mental processing than would using new lexical information for reading purposes. In this study, 110 students (L1 = Dutch; L2 = French) either read or wrote a L2 weather report containing weather terms, many of which were until then unknown to them. If they wished, they could look up the meaning of unknown words with the aid of a computer (from L2 into L1 for students in the reading condition, and from L1 into L2 for students in the writing condition). Students did not know that the computer kept a record of their look-up actions, nor did they know that they would be tested the next day on their knowledge of the words they had looked up. After students had completed their reading or writing task, individual post-tests were composed on the basis of the computer log files containing students' individual look-ups. The study had an incidental-learning design since students were unexpectedly tested on their knowledge of the meaning of the words looked up. The results showed that the number of words looked up hardly differed between readers and writers, but that, not surprisingly, writers needed more time to complete their task than did readers. Retention appeared to be higher among the writers but firm conclusions cannot be drawn because not all participants were tested on the same words (the tests had been individualized on the basis of students' individual look ups during the completion of their reading or writing task). The use of the technique of unobtrusively recording students' look-up actions with the aid of the computer in this study proved to be only a rather superficial measure in that it only revealed *that*, *when*, and *in which order*, but not *how* students processed the lexical information. A replication study will be conducted in the fall of 1999.

We have planned more carefully controlled studies on electronic dictionary use, in which we will also ask students to think aloud during the completion of their reading or writing task or will be stimulated to recall their look-up behavior. The aim of this new project, which is based on Hulstijn and Atkins' (1998) proposals for the systematic study of electronic dictionary use, is to investigate how beginning and advanced L2 learners (1) use, (2) learn to use, and (3) can be instructed to use bilingual electronic dictionaries in a flexible way. The study focuses on the question of how to bring the dictionary to the user and the user to the dictionary. Participants carry out a number of translation tasks (French-Dutch and Dutch-French) with the aid of an electronic dictionary designed for this research with various customization routes to help users deal with the great quantity and complexity of the lexical information in the dictionary. Participants' cognitive processes during electronic dictionary use are tapped through the registration of their actions conducted at the computer, by creating log files, as well as through the elicitation of think-aloud and stimulated-recall protocols. The study will compare the effect of two learning conditions in a between-subject design, consisting of approximately four learning sessions. Two customization conditions will be implemented: in one condition learners will begin under a pro-active, computer-guided user regime (the first two sessions) and switch to a reactive, free user regime in the third session. In the computer-guided regime, students do not obtain all information of a lexical entry simultaneously on their screens (information concerning spelling, derivational flexion, word class,

grammatical selection restrictions, etymology, literal and figurative meanings, expressions, L1 translations, etc.) as in a normal hard-copy dictionary. Instead, they will be led to their goal by a number of menus with options, from which they have to choose. Examples of menu questions are, "Do you want to know more about the word's form or meaning?" "Do you want examples of usage?" and "Do you want L1 translations?" In this regime, students must go through the steps laid down by the program; they cannot skip steps.

In the other condition, students will have a free regime in all four sessions. They can choose menu-driven routes, they can program preferred routes (by ticking off preferred options), or they can opt for obtaining all information of an entry simultaneously on their screen (as in a hard-copy dictionary) and scroll through it. It is hypothesized that a computer-guided regime in which, initially, not all possible options are open to the user yet will have a better learning outcome (in terms of search behavior) than a regime which leaves all options open to the learner from the very beginning. We expect this hypothesis to hold true (1) for students relatively weak in terms of metalinguistic knowledge and L2 proficiency, and (2) for certain word types (polysemous words, words that cannot be translated literally, and words whose corresponding dictionary entry can only be found when the words are morphologically changed to their entry form).

### CONCLUDING REMARKS

Computers, computer software, and computer-assisted data elicitation tasks are, and will remain, tools--nothing less and nothing more. The software and tasks described in this article have had an enormous impact on the study of language acquisition and use. Until some twenty years ago, empirical research on language acquisition and use was restricted to the observation and measurement of language input and output. With these computer-aided tools, however, researchers have the means to get closer to the *processes* of language acquisition and use. Of course, even the most sophisticated tools, such as pet scans and computer simulations of neural networks, remain tools, that is, they do not *explain* phenomena for us. Explanatory theories are developed and formulated by humans, not methods. Yet theories and methods influence each other. New theories may lead to the search for and invention of new methods and tools, whereas new tools and methods may give rise to new theoretical thinking. Thus, neural network thinking, which, for a number of years remained based only on the network metaphore, stimulated researchers to actually program the processes of excitation, inhibition and activation spreading in considerable detail, thus simulating language acquisition. In turn, these computer implementations influenced theories of language acquisition (e.g. Broeder & Plunkett, 1994; Ellis, forthcoming; Landauer & Dumais, 1997). Similarly, the impressive progress made during the past twenty years on theories concerning the construction and working of the mental lexicon, both the monolingual and bilingual lexicon (Kroll & De Groot, 1997), was made possible through an ever cumulative body of empirical research based on a computer-implemented tool such as the lexical decision task.

SLA research has so far been mainly dependent on language production data and grammaticality judgements. Production data were analysed in search of so-called acquisition orders. The main landmarks in this area were the "morpheme studies" of the seventies, summarized in Hatch (1983, chap. 6), the ZISA project, reported in Clahsen, Meisel and Pienemann (1983), and the ESF project, reported in Klein and Perdue (1992), and Perdue (1993). Despite their sincere efforts, the scholars involved in these studies have not succeeded in explaining SLA. Current SLA research often obtains its empirical evidence of whether a target structure has or has not been acquired (in generative terms, whether and when a parameter of UG has been set or reset for L2) from grammaticality judgement data. There are serious problems in using grammaticality judgements with L2 learners at the beginning and intermediate stages of L2 acquisition, in contrast to using grammaticality judgements with adult native speakers (Bley-Vroman & Masterson, 1989; Sorace, 1996). It is time to match these forms of SLA research with experimental, laboratory-type of L2 acquisition research into implicit forms of learning, using receptive rather than

productive modes of language processing. An example is Ellis' (1993) study of the acquisition of the Welsh soft mutation rule, in which learners were bombarded with thousands of stimuli on their PC screens, without having to produce any language (Ellis, 1993). This study was limited in the sense that it provided written stimuli only. What is needed, are studies investigating the bottom-up processing of oral L2 stimuli. The multimedia computer is the ideal tool to present linguistic stimuli, both in spoken and written form, and to register all reactions of learners in terms of both accuracy and time.

On the editor's request, the emphasis in this article has been on the methods, not on the theories or on theory-method interrelationships. It is hoped that the survey given here will encourage people interested in SLA research, whether linguistic, cognitive or educational in orientation, to explore some of the surveyed methods and tools, which in turn may stimulate their theoretical understanding of the phenomena they seek to explain.

#### **NOTE**

1. Sometimes, the computer is used by SLA researchers for the analysis of L2 learners' utterances, rather than for the elicitation of these utterances as such.

---

#### **ABOUT THE AUTHOR**

Jan Hulstijn is chair professor of second language acquisition at the University of Amsterdam. His research topics are implicit and explicit learning of L2 grammar, automatization of word recognition in L2 listening and reading, incidental and intentional L2 vocabulary learning, and methodology of L2 learning research.

e-mail: [hulstijn@hum.uva.nl](mailto:hulstijn@hum.uva.nl)

#### **REFERENCES**

- Bates, E., & MacWhinney, B. (1989). Functionalism and the Competition Model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 3-73). Cambridge: Cambridge University Press.
- Beck, M.-L. (1997). Regular verbs, past tense and frequency: tracking down a potential source of NS/NNS competence differences. *Second Language Research*, 13, 93-115.
- Beck, M.-L. (1998). L2 acquisition and obligatory head movement: English-speaking learners of German and the Local Impairment Hypothesis. *Studies in Second Language Acquisition*, 20, 311-348.
- Bley-Vroman, R., & Masterson, D. (1989). Reaction time as a supplement to grammaticality judgments in the investigation of second language learners' competence. *University of Hawai'i Working Papers in ESL*, 8(2), 207-237. Also available on the World Wide Web at <http://www.lll.hawaii.edu/esl/bley-vroman>.
- Broeder, P., & Plunkett, K. (1994). Connectionism and second language acquisition. In N.C. Ellis, (Ed.), *Implicit and explicit learning of languages* (pp. 421-453). London: Academic Press.
- Brown, J.D. (1988). *Understanding research in second language learning*. Cambridge: Cambridge University Press.
- Clahsen, H., & Hong, U. (1995). Agreement and null subjects in German L2 development: New evidence from reaction-time experiments. *Second Language Research*, 11, 57-87.
- Clahsen, H., Meisel, J.M., & Pienemann, M. (1983). *Deutsch als Zweitsprache: der Spracherwerb ausländischer Arbeiter*. Tübingen: Narr.

- Cohen, J.D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments & Computers*, 25, 257-271.
- De Bot, K., Cox, A., Ralston, S., Schaufeli, A., & Weltens, B. (1995). Lexical processing in bilinguals. *Second Language Research*, 11, 1-19.
- De Graaff, R. (1997). The eXperanto experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition*, 19, 249-276.
- DeKeyser, R.M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition*, 19, 195-221.
- Ellis, N.C. (1993). Rules and instances in foreign language learning: Interactions of explicit and implicit knowledge. *European Journal of Cognitive Psychology*, 5, 289-318.
- Ellis, N.C. (forthcoming). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction*. Cambridge: Cambridge University Press.
- Eubank, L. (1993). Sentence matching and processing in L2 development. *Second Language Research*, 9, 253-280.
- Flege, J.E. (1988). Using visual information to train foreign-language vowel production. *Language Learning*, 38, 365-407.
- Freedman, S., & Forster, K. (1985). The psychological status of overgenerated sentences. *Cognition*, 19, 101-186.
- Gernsbacher, M.A. (Ed.). (1994). *Handbook of Psycholinguistics*. San Diego, CA: Academic Press.
- Hatch, E.M. (1983). *Psycholinguistics: A second language perspective*. Rowley, MA: Newbury House.
- Hatch, E.M., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. New York: Newbury House.
- Horiba, Y. (1993). The role of causal reasoning and language competence in narrative comprehension. *Studies in Second Language Acquisition*, 15, 49-81.
- Horiba, Y. (1996). Comprehension processes in L2 reading: Language competence, textual coherence, and inferences. *Studies in Second Language Acquisition*, 18, 433-473.
- Hulstijn, J.H. (1993). When do foreign-language readers look up the meaning of unfamiliar words? The influence of task and learner variables. *The Modern Language Journal*, 77, 139-147.
- Hulstijn, J.H. (1997). Second language acquisition research in the laboratory: Possibilities and limitations. *Studies in Second Language Acquisition*, 19, 131-143.
- Hulstijn, J.H., & Atkins, B.T.S. (1998). Empirical research on dictionary use in foreign-language learning: Survey and discussion. In: B.T.S. Atkins (Ed.), *Using L2 dictionaries: Studies of dictionary use by language learners and translators* (pp.7-19). Tübingen, Germany: Niemeyer Verlag, Series Maior of *Lexicographica*.
- Hulstijn, J.H., & Trompetter, P. (1998). Incidental learning of second language vocabulary in computer-assisted reading and writing tasks. In D. Albrechtsen, B. Henrikse, I.M. Mees, & E. Poulsen (Eds.), *Perspectives on foreign and second language pedagogy* (pp. 191-200). Odense, Denmark: Odense University Press.

- Issidorides, D.C., & Hulstijn, J.H. (1992). Comprehension of grammatically modified and nonmodified sentences by second-language learners. *Applied Psycholinguistics*, 13, 147-171.
- Jagtman, M., & Bongaerts, T. (1994). COMOLA: a computer system for the analysis of interlanguage data. *Second Language Research*, 10, 49-83.
- Jensen, K.A., & Ulbaek, I. (1994). The learning of the past tense of Danish verbs: language learning in neural networks. *Applied Linguistics*, 15, 15-35.
- Kasper, G., & Grotjahn, R. (Eds.). (1991). *Methods in second language research. Studies in Second Language Acquisition*, 13(2), thematic issue.
- Kempe, V., & MacWhinney, B. (1998). The acquisition of case marking by adult learners of Russian and German. *Studies in Second Language Acquisition*, 20, 543-587.
- Kilborn, K., & Cooreman, A. (1987). Sentence interpretation strategies in adult Dutch-English bilinguals. *Applied Psycholinguistics*, 8, 415-431.
- Kilborn, K., & Ito, T. (1989). Sentence processing strategies in adult bilinguals. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 257-291). Cambridge: Cambridge University Press.
- Klein, W., & Perdue, C. (Eds.). (1992). *Utterance structure: Developing grammars again*. Amsterdam: John Benjamins.
- Kroll, J.F., & de Groot, A.M.B. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In A.M.B. de Groot & J.F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 169-199). Mahwah, NJ: Erlbaum.
- Landauer, T.K., & Dumais, S.T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), pp. 11-240.
- Larsen-Freeman, D., & Long, M.H. (1991). *An introduction to second language acquisition research*. London: Longman.
- MacWhinney, B. (1995). Computational analysis of interactions. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language* (pp. 152-178). Oxford: Blackwell.
- Nunan, D. (1992). *Research methods in language learning*. Cambridge: Cambridge University Press.
- Perdue, C. (Ed.). (1993). *Adult language acquisition: Cross-linguistic perspectives* (Vols. 1-2). Cambridge: Cambridge University Press.
- Pienemann, M. (1992). COALA < a computational system for interlanguage analysis. *Second Language Research*, 8, 59-92.
- Segalowitz, N. (1997). Individual differences in second language acquisition. In A.M.B. de Groot & J.F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 85-112). Mahwah, NJ: Erlbaum.
- Segalowitz, N.S., & Segalowitz, S.J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, 14, 369-385.
- Segalowitz, S.J., Segalowitz, N.S., & Wood, A.G. (1998). Assessing the development of automaticity in second language word recognition. *Applied Psycholinguistics*, 19, 53-67.

Seliger, H.W., & Shohamy, E. (1989). *Second language research methods*. Oxford: Oxford University Press.

Sokolik, M., & Smith, M. (1992). Assignment of gender to French nouns in primary and secondary language: A connectionist model. *Second Language Research*, 8, 39-58.

Sokolov, J.L., & Snow, C.A. (Eds.). (1994). *Handbook of research in language development using CHILDES*. Hillsdale, NJ: Erlbaum.

Sorace, A. (1996). The use of acceptability judgments in second language acquisition research. In W.C. Ritchie & T.K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 375-409). San Diego: Academic Press.

Tarone, E.E., Gass, S.M., & Cohen, A.D. (Eds.). (1994). *Research methodology in second-language acquisition*. Hillsdale, NJ: Erlbaum.

Yang, L.R. (1997). Tracking the acquisition of L2 vocabulary: The Keki language experiment. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 125-156). Cambridge: Cambridge University Press.

Yang, L.R., & Givón, T. (1997). Benefits and drawbacks of controlled laboratory studies of second language acquisition: The Keck Second Language Learning Project. *Studies in Second Language Acquisition*, 19, 173-193.