Interlanguage speech recognition by computer: implications for SLA and computational machines

Larry Selinker and Rita Mascia, Centre for Interlanguage Studies, Birkbeck College, University of London

During the last decade, there has been a rapid growth in research into speech recognition by computer (SRC). Computerised voice recognition systems have been developed which are being used for a variety of applications. However there remain a whole range of issues which have to be elucidated and investigated before SRC can be broadly useful including for language learning purposes. It is well documented that speaker variability caused by accent is one of these issues and one of the major hurdles in accurate speech recognition. Foreign speaker recognition is particularly problematic to program for reasons that our work is beginning to suggest. In this paper we describe and compare the SRC of an interlanguage speaker of Italian/English versus a native speaker of English, both with repetitive strain disorder (RSD) and thus highly motivated, using the same software, DragonDictate, from Dragon Systems. Cognitive processes such as language transfer, fossilization and communication strategies are examined in light of the research. We illustrate the possibility of using SRC in second language research with particular emphasis on phonology. In this paper we not only explain our views of the potentials of this new technology in facilitating second language acquisition research but go to a more general applied linguistics issue where we briefly discuss some implications for the design of speech recognition systems for interlanguage speakers. This focus, we believe, can help make applied linguistics a main stream discipline, thereby increasing the job space for applied linguistics graduates.

Introduction

As far as we know, this is the first paper[1] on the topic of interlanguage speech recognition by computer.

In this paper we explore the potential of one new class of technology in facilitating second language acquisition research: speech recognition by computer. There are a range of issues in second language research that we feel can be elucidated by this technology and we discuss these. Computerised voice recognition systems have been developed which are being used for a variety of applications. In second language acquisition terms, speaker variability caused

by accent is one of the major hurdles in accurate speech recognition and will be referred to throughout this paper, passim. Non-native interlanguage speaker recognition is particularly problematic to program for reasons that our work is beginning to suggest; there are computer science questions concerning hardware and software requirements to cope with interlanguage speech recognition by computer which we will discuss elsewhere (Mascia & Selinker, in preparation). In this paper we focus on the role of automated speech recognition in second language acquisition research. Specifically, we describe and compare speech recognition by computer of an interlanguage speaker of Italian-English versus a native speaker of English, both with repetitive strain disorder who are thus both highly motivated. Both use the same software and hardware. Here, we illustrate the possibility of using speech recognition by computer in second language research with particular emphasis on phonology.

Our plan of the paper is as follows: we begin by discussing in an introductory manner three interlanguage cognitive processes in ways that fit most neatly into our empirical study; then we describe for the non-initiated, some principles underlying speech recognition technology. Next we present an empirical study which attempts to move speech recognition by computer into the interlanguage area and end the paper by discussing how this technology can be related to second language acquisition research and, importantly, possible implications for the design of speech recognition systems for interlanguage speakers.

Three interlanguage cognitive processes

Language transfer

Language transfer is one of the first and defining concepts of applied linguistics and second language acquisition. One only has to look at the early work of Weinreich (1953) and Lado (1957) to see it as a robust concept even in the 1950's. In fact, *Rediscovering Interlanguage* by Selinker in 1992 (referred to as *RI* in what follows) was in essence an attempt to show the historical foundation of that concept and others in interlanguage studies. To put things in perspective one of the earliest references discussed there is over 100 (!) years old (Whitney, 1881). To say that language transfer can be discussed in many ways is an understatement, to say the least. One popular definition of language transfer is from *Random House Dictionary*:

> "Ling. the application of native language rules in attempted performance in a second language, in some cases resulting in deviations from target-language norms and in other cases facilitating second-language acquisition." (2009)

But this definition leaves out a host of effects that there is evidence in the literature for: transfer from a second to a third language called interlanguage transfer (see e.g. De Angelis, forthcoming), transfer back into the native language from an interlanguage, avoidance of a target language (TL) structure due to native language constraints and so on. (see Gass & Selinker, 1994, chapters 3 & 4; and forthcoming). Thus we think that a more useful way to conceive of language transfer is:

> "Language transfer is best thought of as a cover term for a whole class of behaviors, processes and constraints, each of which has do with CLI (cross-linguistic influence), i.e. the influence and use of prior linguistic knowledge, usually but not exclusively native language (NL) knowledge." (Selinker, 1992: 208).

In that light, we think that for the purposes of this paper, we think it is useful to provide ten principles of language transfer that we think have justification from the research literature:

1. Transfer as a 'selection process' (Weinreich, 1953; chapters 2, 7 in *RI*).
2. Interlingual identifications as the basic learning strategy, where you 'make the same what cannot be the same' (ibid.)

> "For example, /p/ in Russian (R/p/) is defined among others, by its' distinctive feature of non-palatality (in opposition to R /p'/, while the definition of English /p/ (E/p/) involves no such restriction. From the point of view of the languages, therefore, R/p/ and E/p/ cannot be 'the same." (Weinreich, 1953: 7)

In the next language, you look for what you have in your native language (Corder, 1983; chapter 6 in *RI*).
4. Perception and production have to be looked at differentially, at least in some transfer cases (Nemser, 1961; Briere, 1966; chapter 7 in *RI*).
5. 'Blends' and 'autonomous material' are to be expected in interlanguage. (ibid).
6. Structural models, such as Lado (1957) emphasize 'holes' in the interlanguage, the learner using translation then as a learning strategy, seen clearly in interlanguage morphology (Harris, 1954).
7. 'Transfer to somewhere' (Andersen, 1983, 1989; chapter 7 in *RI*) is often a necessary principle in language transfer (though 'transfer to nowhere' may also exist (Kellerman, 1996), especially with typologically distant L1's (Han, pc). An example would be Wode's (1976) claim that the earliest German-English negator in English is 'no' (i.e. not transfer), but that 'German post-verbal position to the negator *nicht*transfers to English only after the learner has developed a set of English auxiliaries' (cf. Andersen, 1989).

Andersen relates this principle to the 'one-to-many' principle in interlanguage, where the learner uses one interlanguage form for two or more language functions (cf. also, Rutherford, 1987)
8. Prediction of individual (on-line?) transfer effects may not be possible, so the principle of 'transferability', i.e. probability of transfer, was created by Weinreich (1954: 36) (which he attributes to Haugen) and recreated independently by Kellerman (1977) in the area of polysemous idioms. Cf. Kellerman's 'break' example where

> "[...] learners tend to avoid one-to-one correspondence between L1 and L2, where meanings were perceived as far from prototypical, even though this led them unwittingly to gratuitous error."

Kellerman relates this principle to another principle he calls "psychotypology", viz: typological closeness.
9. Underlying linguistic principles may be transferred, e.g. apparently, the underlying phonological principle of 'tonality' in Thai is transferred to Thai-

English and is one key reason why Thai speakers are often unintelligible in English (Rudaravanija, 1985; chapter 4 in *RI*).

10. To create equivalence in the next language, use key linguistic variables, especially structural and translation correspondences (ibid, all of the above).

But, even if complex, language transfer can not be seen as an isolated variable in determining the shape of interlanguage in second language acquisition. We must get beyond the tendency to describe isolated variables in isolated studies and make the difficult and serious attempt to interrelate important and crucial second language acquisition variables, no matter how hard the effort. Transfer has been seen to be a central factor in whether or not an item fossilizes in interlanguage and the relationship between the two has been termed the *Multiple Effects Principle* (Selinker & Lakshmanan, 1992), which we will return to at the end of this paper.

Fossilization

As with language transfer, there is a large literature which discusses fossilization and there are many and often contradictory points of view. Again, *Random House Dictionary* helps us begin by defining 'fossilize' in the following way:

> "Ling. (of a linguistic form, feature, rule, etc) to become permanently established in the interlanguage of a second-language learner in a form that is deviant from the target-language norm and that continues to appear in performance regardless of further exposure to the target language." (755)

One complication is that over the past decades, as further delineated in Selinker & Han (in press), fossilization has been discussed within two distinct traditions: developmental and ultimate attainment (cf. Rutherford, 1984). The former tradition encompasses a wide range of perspectives, but often emphasizes the sociolinguistic (Preston, 1989; Tarone, 1994); the latter, interestingly, analyses the problem almost exclusively in terms of one or another form of universal grammar. In the developmental tradition, we find the nub of the fossilization question to be: How do we as observers know that interlanguage development has ceased? In the ultimate attainment tradition the question is put slightly differently: How do we as observers know that the attainment to date is in fact ultimate and that final steady state grammar, if such a thing exists, *and this point is crucial to interlanguage speech recognition by computer,* has been reached? When the focus is on near-natives or those who seem to know the language very very well (Coppieters, 1987, Birdsong, 1992, White and Genesee, 1996), answering this question is particularly crucial.

In retrospect, it is clear that the earliest definitions of fossilization delineated five basic properties: first, fossilization is equivalent to cessation of development; second, fossilizable features pertain to each and every aspect of interlanguage, including phonetic, phonological, morphological, syntactic, semantic, lexical, discoursal and pragmatic features; third, fossilizable features are persistent and resistant; fourth, fossilization hits both adult L2 learners and child L2 learners; fifth, fossilizable features usually manifest themselves as backslidings in performance, with re-emergence of forms being a key indicative marker, again an important point in interlanguage speech recognition by computer.

We need to relate fossilization to input, the most appropriate perspective for our study below. According to Gass and Selinker (1994), fossilization results when new (correct) input fails to have an impact on the learner's grammar, in other words when, the correct input is not apperceived or it is not comprehended. So what happens when the machine fails to perceive the interlanguage speaker's attempt to interact with it? Given that language transfer and fossilization appear to be linked, it is important to emphasize that all speech recognition programs, as far as we know, are intended for native speakers with their basic lexicon native speaker based. Thus, we must also look to other variables that have been shown to have an influence on the formation of interlanguage and here we turn to communication strategies.

Communication Strategies[2]

A third set of cognitive processes are usually termed *communication strategies* (CS). As Kasper and Kellerman (1997) put it:

> "Identification of CS depends to a great extent on what one considers a CS to be, and in this respect, it matters very much whether one conceives of CS as intraindividual or interindividual events." (8)

Again, as can be seen, there are many ways to approach this area as well. We have found the Kasper and Kellerman book to be particularly helpful in both giving us prime sources that are often not easily accessible but also in intelligently pulling the various issues together.

According to Faerch and Kasper (1983), cited as a central source in Kasper and Kellerman (1997) communication strategies are "potentially conscious plans for solving what to an individual presents itself as a problem in reaching a particular communicative goal". For example, interlanguage speakers trying to communicate with an interlocutor may use a communication strategy such as paraphrasing if they do not know or cannot access a particular lexical item. Kasper and Kellerman (1997) argue that this definition fits within what they call "the intraindividual view", a view widely held by early researchers in the field, which saw communication strategies as underlying processes occurring individual mind and importantly which did not have to engage the interlocutor for resolution.

The opposing view has been termed "the interindividual view" with Tarone (1983) as one of its main proponents, again cited as a key source by Kasper and Kellerman (1997, 3). Tarone (1983) sees communication strategies as used by both the interlanguage speaker and the interlocutor in attempts to "[...] bridge the gap between the linguistic knowledge of the second-language learner, and the linguistic knowledge of the target language interlocutor in real communication situations". Requests for clarification and comprehension checks are two examples of interactional communication strategies, which "operate on input which is too far ahead of the learner's current interlanguage competence and size it down to what the learner can manage" (Kasper & Kellerman, 1997, p.5). According to Larsen Freeman and Long (1991), again cited as a key source in Kasper and Kellerman (1997) "[...] all CS are helpful for acquisition because they enable learners to keep the conversation going and thereby provide more possibilities for input". We think this is not the whole story.

Kinahan and Selinker (1997) in an online paper, http://alt.venus.co.uk/VL/AppLingBBK/DB/kinahan/, argue that researchers may have overlooked the possibility that "communication strategies could be used as a learning tool to reveal the gaps between a learner's interlanguage and the target language". The conclusion from their data is that the particular communication strategies used by the learner studied during a taped conversation with a native speaker of the target language which are then analysed afterwards by both the learner and the native speaker, that this combination of procedures could help identify some of the gaps to the learner in that particular learner's interlanguage with relation to a particular target. This understanding, we believe, could give the learner the opportunity to receive "target-like input tailored specifically to that particular learner's needs" (Kinahan & Selinker, 1997). In that paper, it is argued that this type of *joint analysis* conducted by the learner with the native-speaker interlocutor is likely to reveal *gaps* in the learner's interlanguage which the learner may not specifically know about and data is presented there to argue for that claim. According to Perkins (1985 as cited Kaplan, 1997), "ability to find problems can be as important an ability as ability to solve problems" in second language learning. Kinahan and Selinker (1997) also argue that "it may be for some learners that the bringing to their consciousness of mistakes or errors could be a learning strategy in itself". However, in other cases learners may need to apply learning strategies in order to integrate into their interlanguage target-like grammatical structures or lexical items identified through the analysis of their use of communication strategies.

To continue the argument, Kinahan (1999) further hypothesises that target-like grammatical structures and lexical items brought to a learner's consciousness in this way are more likely to be noticed by the learner and integrated into their interlanguage system than target-like grammatical structures or lexical items corresponding to errors or mistakes the learner may have made during production but which caused little or no difficulties during communication (whether for the learner or the interlocutor). Thus communication strategies are well-integrated into the second language acquisition research literature and, it can be hypothesized that some sort of strategy use will occur when interaction with voice recognition software occurs. We have to ask what types of communication and thus communication strategies are involved between the human and the computational machine in this application.

One interesting possibility, which we believe we have evidence for, is that the human interlanguage user can find themselves treating the computer as another human interlocutor. In this case, the communication strategies used can become a set of intraindividual and interindividual strategies while the human interlanguage user is trying to dictate a text. It is our hypothesis that the only communication strategy likely to lead to the software 'understanding' a statement, is to ensure that in any interactions with the software, each and every word is pronounced exactly the same way each time in order to promote voice recognition. Communication strategies in real life are based around face-to-face communication. Thus, the reader should note that these strategies when applied to human-computer interaction may be counterproductive. Communication strategies will thus most likely be of no help at all to the interlanguage user.

Before we can describe our study so that it makes sense to those new to this area of technology, we present some background underlying speech recognition technology.

Background to speech recognition technology

During the last decade, there has been a rapid growth in research into speech recognition by computer. Although speech recognition systems have been developed in different application areas, there are still some key aspects that need researching before these applications can truly become effective. First of all, why has been so difficult to program speech recognition by computer? It has taken many decades, after numerous false starts, to produce software that is commercially viable and this has only happened in the past few years. A second, related question is how do language engineers tackle this difficulty? To tackle questions such as these, we need to have some basic understanding of how speech recognition programs work.

Current speech recognition programs operate by converting signals into a sequence of short spectrum representations which are then analysed to identify the phonemes of the corresponding words. In brief, the speech recognizer necessitates three elements in order to perform its task: an acoustic model, a language model and a pronunciation archive. The majority of current systems are based on *Hidden Markov Models*, acoustic statistical models of speech data used to represent these signals. They are called hidden because the sequential model is unknown and therefore 'hidden'. These statistical models are created by training an extensive corpora of samples. Based on the three previous elements, an algorithm is then created to perform the decoding of the speech units[3]. Doing this type of search/decoding is computationally expensive and this we believe is part of the problem for the inadequacy of current system in recognizing foreign speakers accurately.

We will now briefly illustrate current research performance in removing some of the limits affecting speech recognition which we will describe below . The speech recognition problem is how to produce a machine readable transcription of spoken input in a world of great linguistic variation. In effect, the essence of speech is given by a combination of elements from different sources (syntax, phonology/phonetics, semantics, discourse analysis and the lexicon). The main challenge faced by programmers who design automated speech recognition is the variability of these sources in producing the spoken signal. We can summarise relevant variables according to:

- speaker variation: cultural-geographical, accent/dialect; gender; physiological, where the speaker can be sad, nervous or sick, for example.
- style variation: formal versus casual; reading versus conversing; clear versus unintelligible; fast versus slow speech and the accompanying reductions and deletions.
- environmental conditions: for example, voice carried through a telephone channel as opposed to face to face speech; background noise versus quiet conditions.

In order to cope with these array of variables and many more, speech recognition by computer has to limit them in some manageable way with the

goal of current studies being in fact to *remove* ambiguity caused by these factors entirely, if at all possible! Current applications attempt to do this:

- by limiting vocabulary size: from simple applications like speech recognition for English Second Language having a very limited active vocabulary of about 600-1000 words to more sophisticated applications like DragonDictate and ViaVoice whose vocabulary ranges between 30,000 and 120,000+ words when customised by the user.
- by limiting mode of spoken input:
  a. discrete speech recognition, requiring a slight pause in between words which makes the dictation a bit artificial and less resembling natural speech, but requiring less co-articulation.
  b. continuous speech recognition, allowing text to be inputted at more or less natural speed. This latter is important as continuous speech recognition from unscripted sources has been particularly difficult to program.
- by limiting the loudness factor, resorting to the use of microphones and head mounted microphones
- by limiting the speaker variable:
  a. from speaker-dependent applications where the system needs training in order to recognise the characteristics of a particular speaker; to speaker-independent applications where the system can be used by anyone.
  b. customising the application according to regional accent and dialect, e.g. American English versus British English.
  c. fluent versus non fluent, e.g. software of English Second Language customised for different levels of learners
  d. native versus non-native, the variable which we attempt to describe in this paper.

The good news is that some of these limitations are being removed little by little. However, given the way speech recognition works, and considering that interlanguages have much variation, and in this sense are 'unstable', by definition, speech recognition of foreign speakers is still far from reaching the same level of accuracy than that of native speakers. In the section below we will describe our study, trying to highlight the problems experienced by our interlanguage subject and the way the use of this technology has helped us to refine our understanding of the user behavior. We will also relate the findings of our study to current concerns in second language acquisition research.

The study

Subjects

The two subjects we compared are both chronic RSD sufferers, strongly motivated to learn how to use the dictation software. The first user is a native speaker of Italian and Sardinian, 37 year old female, a very advanced interlanguage speaker of Italian-English, who has lived in an English-speaking countries for many years and who works and lives daily in English. The second user is a native speaker of standard American English, a 60 year old male. Both are well-versed in computers and both hope to replace the mouse and the keyboard by using as much speech recognition software as they possibly can.

Materials

Although continuous speech recognition is a recent technological innovation, for our particular purpose (an application to be used by RSD sufferers) we chose discrete speech recognition as the best solution because it allows the speaker to control the PC and the software programs by voice commands virtually 'hands-free'. After a review of both discrete and continuous speech recognition programs we selected DragonDictate Classic Edition, from Dragon Systems, Inc. (hereafter referred to as *DD*).

> "DragonDictate represents pronunciations as hidden Markov models that are built from the speech of a reference speaker according to phonemic spellingsand (this allows DragonDictate) to adapt quickly to the user's own speech." (Mandel, 1992, 246).

The first element for creating a speech recognizer is what is technically called a 'vocabulary' (i.e. a list of words which have to be recognized). The vocabulary contains the most frequently used words and from this corpus a language model is assembled. In theory, if someone wanted to train a speech recognizer, they would have to pronounce each word until the whole vocabulary was complete. This method is obviously impractical for large size vocabularies and it is with this hurdle in mind that the model of *DD* was engineered.

> "DragonDictate's [...] acoustic processing is based on a three-level analysis of each word: the phoneme, the phoneme in context, and the phonetic element. Theoretically a single model of each phoneme, together with rules for coarticulation and contextual allophony, would allow recognition of any word conforming to the phonology of the language. In actuality we need many models; some we extract from speech, and from these we generate others. For each language, the speech is that of a reference speaker; once the product is built, the models adapt to the user's speech in the process of being used." (Mandel, 1992, 238).

Mandel explains the three-level analysis in detail. He mentions that phonemes in English constitute a total of 24 consonant, 3 resonants, and 17 vowels and that for each vowel there are 78 phonemes (due to different stress levels). Although phonemes can also be marked for appearing in syllable final, Mandel explains that the cost of computing that feature is too high and it is thus disregarded. The second analysis is based on the phoneme in context (PICs) which is "[...] an augmented triphone, comprising:

- the phoneme before the one being modelled, as context
- the phoneme being modelled
- the phoneme after the one being modelled, as context
- the degree of prepausal lengthening of each phoneme due to its position in the word." (Mandel, 1992, 239).

It is at this point that the reference speaker records a subset of the vocabulary (what needs to be recorded are all the PICs, not all the words). The third element for the analysis is the phonetic element (PELs), a steady state which phones may share. For example, the /t/s of the triphones /eɪtɪ/, /ɪtɪ/ and /ɪta/ have very

similar onsets (for a more detailed phonetic description of the triphones, see Mandel, 1992, 239).

The importance of PELs derives from the fact that it is impractical to code all the PICs for the 25,000 words that DragonDictate allows. To obviate to this inconvenient, acoustically similar onsets are implemented using PELs. The exemplar of pronunciation is then regarded as a *Markov model* of an acoustic event[4]: a stochastic mechanism that in recurrent periods would originate acoustic events. More precisely Mandel talks of a *hidden Markov model (HMM)* because, as he puts it,

> "We know the output of the model, but not the model itself (therefore 'hidden'), and the speech recognizer's task is to reconstruct it. Thus , from our language model together with the prior context, and our set of acoustic word models together with the acoustic data of the word just spoken, we have derived an ordered list of words that the speaker could have meant, arranged from most to least probable. Building on this ordered list, the DragonDictate interface lets the user select the correct word with minimum effort. In the best case, where it is at the head of the list, he or she needs only dictate the next word in the text to confirm it." (Mandel, 1992, 240).

The key feature of SRC is the process of *adaptation*. The first time people use the system, they have to create their own voice-file, which initially is based on the reference speaker's model. At each successive use of the program, *DD* adapts to the pronunciation of the user. Bamberg and Mandel (1991) carried out some tests to check the effects of adaptation of *DD* used by different users and in different conditions. Their tests showed that, initially, recognition performance is not very accurate when the user is of a different sex from that of the reference speaker used to model the recognizer. However, the recognition improves if the software is initially used by someone of the same sex as the user, confirming that adaptation does take place.

DragonDictate Classic Edition *(DD)*, like the majority of the other speech recognition packages, features an enrolment session that lasts for a couple of minutes. The purpose of the enrolment is to create an individual user file which records the main characteristics of the user's voice. The enrolment prompts the speaker to indicate the type of their voice (high pitch or low pitch), test the microphone, and adjust the microphone volume if necessary. Although the enrolment session is sufficient to enable one to proceed to use the application, there is an optional tutorial which lasts approximately 20 minutes and it's called *Quick Training*. This training enables *DD* to recognise the users' particular speech pattern and adapt to their pronunciation. The *Quick Training* is made of 4 modules: *Correction Words* (up to 65 words), *Common Commands* (up to 110 words), *Dictation Words* (up to 230 words), *Additional Words* (up to 365 words). In each module the words appear in a dialogue box and the user is supposed to pronounce them following the machine's instructions (see next picture below). If a pronunciation is not correctly recognised, the machine invites the user to repeat the word several times until it is recognized. It achieves this by showing some dots underneath the misrecognized word, which is highlighted until the pronunciation is satisfactory.

DD continuously adapts to the user's pronunciation. This is particularly important when using the system in *Dictation Mode*. Since many types of word pronunciation are accepted, it is important to correct misrecognized words as they take place, as the software can only correct the previous 16 words.

Accordingly, when dictating text, it is not advisable to leave misrecognized words until the end and then do a spell check (which would probably be the norm in a conventional use of a word-processing package). By doing so the program tends to make more errors in successive uses, because of the process of adaptation. Finally at the end of a particular task, the program confirms with the user if they want to save their user file, and if the speaker decides to save the file, the program enriches the user's speech archive.

Once the users have mastered the basic use of *DD*, there are a number of options which can be explored to improve the system's voice adaptation. For instance, in the feature *Vocabulary Manager* there is an archive of dictation words which can be re-trained to achieve better accuracy.

Procedures

We carried out the following tasks:

1. Videorecording of training sessions

This task consisted of an enrolment session during which the users had to complete the Create New User Wizard (as detailed in the previous section), run the tutorial by repeating 15 basic command words, follow the demonstrations of WordPad use, Calculator use and again WordPad use. The entire task took approximately 20 minutes. The tutorial showed the speaker how to run those applications by using the control commands (which control the main drop-down menus) and the dictate commands.

2. Videorecording of training session and Dictation Task 1

For this task, the users had to complete four *Quick Training Modules* (as detailed above). The training lasted approximately 25 minutes in total and was followed by a dictation test. The test consisted of reading the summary of a book chapter[5], summarise it and dictate the gist of the synopsis using Microsoft WordPad. During the test the users made full use of correction commands such as Word History and Spell Mode.

3. Audiorecording of Dictation Task 2a

This task required the subjects to dictate three paragraphs from a different book chapter[6], but this time without correcting any unrecognised words, without using any formatting or any control commands, and without checking any spelling. The word-processing package used was Microsoft WordPad as before. The idea behind this test was to have an approximate indication of the degree of accuracy of word recognition for the two speakers. The text was compared to the text of the other user, and to the original text.

4. Audiorecording of Dictation Task 2b

This task was a replica of the previous one after about four months had elapsed between the two dictations. Prior to this, the two subjects used *DD* on average

15 hours per week, working in their own environment carrying out their normal type of work. The ultimate goal of this recording was to discover how quickly the system adapted to the two speakers' pronounciation and to compare the two texts in order to discover if there were any significant systematic differences between the two users.

Design and administration

This study was conceived of as a comparative case study of two people, an interlanguage speaker and a native speaker, interacting in a certain way with a computer. The speech corpus was collected as spontaneous speech, although some of the tasks administered required samples of elicited speech and the sessions were recorded[7]. But in general, we followed as naturalistic an approach as possible. To get the feel of the aims of the program, i.e. the more the program was used, the better it was supposed to adapt to the user, as much like ordinary work conditions were maintained. The two subjects of the study then used the program on their own and in their own time. The only intervention was represented by technical assistants who limited their interventions to setting up the equipment and giving basic training instructions. The subjects therefore experienced very similar situations as having to deal with difficulties entirely on their own. During the administration of the first three tasks, the two subjects were unaware of each other's progress. In order to ensure correct operational measures, the validity of the study was constructed using videorecording, audiorecording and recorded interviews.

Results

Native speaker speech recognition

Table 1 shows the statistics from the *Quick Training Modules:*

Table 1. Native speaker recognition accuracy in Quick Training Modules.

Quick Training Modules	words dictated	words misrecognized	recognition accuracy
correction commands	46	6	86.9%
common commands	124	14	94.7%
dictation words	94	12	87.2%
other common command words	35	0	100%
additional words	43	8	81.3%
total number of words	342	40	88.3%
total % of recognition accuracy			88.3%

The results from Dictation Task 1 are shown in Table 2:

Table 2. Native speaker recognition accuracy in Dictation Task 1.

CHART NUMBER 1

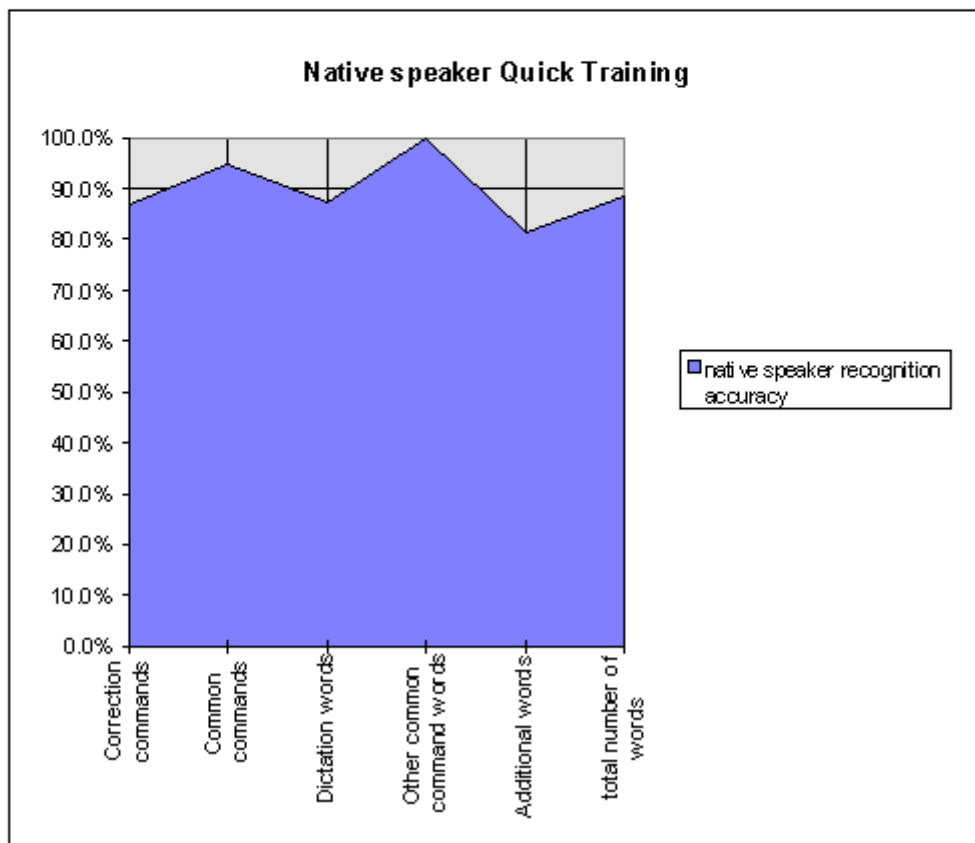|	native speaker recognition accuracy
Correction commands	86.9%
Common commands	94.7%
Dictation words	87.2%
Other common command	100.0%
Additional words	81.3%
total number of words	88.3%

Table 3. Native speaker recognition accuracy in Dictation Tasks 2a and 2b.

Dictation Task 1	words dictated	words recognized	recognition accuracy
command words	267	186	69.6%
command words + spelling words	295	214	72.5%
dictation words	256	71	27.70%
dictation words - spelling words	228	43	18.8%
command + dictation words	523	257	49.10%
total percentage of recognition accuracy			49.10%

Interlanguage speaker speech recognition

The results from the *Quick Training Modules* are shown in Table 4 below:

Table 4. Interlanguage speaker recognition accuracy in Quick Training Modules.

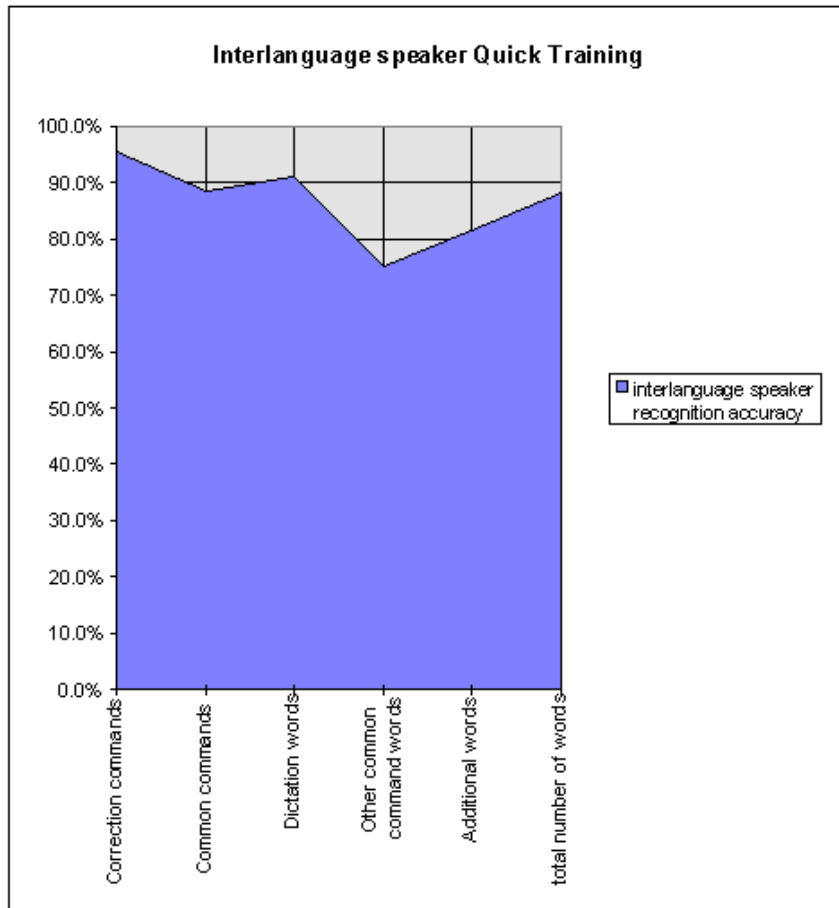| DICTATION TASK 2 a AND b	native speaker accuracy
dictation a	85.10%
dictation b	84.70%

Table 5 shows the results of Dictation Task 1 and Table 6 the results of Dictation tasks 2a and 2b.

Table 5. Interlanguage speaker recognition accuracy in Dictation Task 1.

Quick Training Modules	words dictated	words misrecognized	recognition percentage
correction commands	45	2	95.5%
common commands	122	14	88.5%
dictation words	134	12	91.0%
other common command words	24	6	75.0%
additional words	65	12	81.5%
total number of words	390	46	88.2%
total percentage of recognition accuracy			88.2%

Table 6. Interlanguage speaker recognition accuracy in Dictation Tasks 2a and 2b.

CHART NUMBER 3

	interlanguage speaker recognition accuracy
Correction commands	95.5%
Common commands	88.5%
Dictation words	91.0%
Other common command words	75.0%
Additional words	81.5%
total number of words	88.2%

Interlanguage speaker Quick Training

We now discuss the data, first focussing on a series of patterns we identified in this speaker's interlanguage (word numbers refer to the words listed in Appendix 2):

The first pattern: the interlanguage speaker sometimes had to 'guess' the target pronunciation. There were certain words which were unknown to the speaker for which the pronunciation was not obvious. This resulted in the interlanguage speaker relying on the Italian phonological system for filling in that gap (e.g. 1, the command word 'oops' which was pronounced as it had been an Italian word, or , 'begin document' /begin/ /dokument/ for the word 'begin' the user was unaware of the target language pronunciation for the vowel 'e' in the first syllable.

The second pattern: a wide-spread language transfer from the L1. For instance, *the presence of cognates* in the L1 resulted in transfer of phonological features (e.g. in #51, 'Diana' was pronounced once as /diana/ and once as /daiana/, the first word pronounced as if it was the Italian cognate word 'Diana'), including features such as stress (e.g. in 15, /'bravo/ /bra'vo/ where the user changed the syllable stress since the vowel sounds are the same as those in the Italian cognate word 'bravo'). A *visual interference* of the target word reminded the speaker of the L1 pronunciation (e.g. in 19, 'equal sign'/i'kwol//saɪʲ/, the presence of /ʲ/ in 'sign' is another hypothesised instance of phonological transfer from the L1 where the two letters 'g' and 'n' when together are pronounced as above (/ʲ/).

The third pattern: several instances of apparently fossilized forms like the use of [d] instead of [ð] in 20 (greater than/greɪter/ /den/) and the lack of aspiration in words containing [h] as in 26 ('home key' /o'm//ki/. There were also some instances of *backsliding* to earlier interlanguage forms, for instance in 20, where the speaker backslided to earlier Italian/American English interlanguage forms where [r] would be pronounced at the end of a word.

The fourth pattern: a large amount phonological variation of a type often unexpected. Here we can only give a hint of this. One pattern obvious from the data is that the interlanguage seemed to alternate between production of target-like and non target-like phonological features. An example of both target-like and non target-like features is 4 ('DragonDictate' /dragondɪkteɪt/, the 'a' in 'Dragon' is pronounced as the Italian /a/ instead of the target English /æ/, while the 'a' in 'dictate' is pronounced as in target English [eɪ]). Other examples found were the variation of occurrence for schwa [ə]which sometimes appeared especially in final position as in 5, /kompju'tə/ or 10, /kalkleɪtə/, and sometimes was replaced by a final [r] as in 18, /oskar/. Likewise, we found free variation between the use of dark [ɫ] and clear [l] (e.g. 16, /deɫta/ and /delta/, and the use of [h], which sometimes was pronounced and sometimes was not (e.g. 26, home key /o'm/ /ki/ /ho'm/ /ki/).

The fifth pattern: the consistent use of various communication and learning strategies[8] which characterised this speaker's interlanguage interaction with the machine and which resulted in different effects. For instance in 13 ('back space', /bak/ /speɪs/,/bak/ /speɪs/,/bek//speɪs/) the interlanguage user reported that she assumed that her pronunciation in the two first dictations was incorrect and tried to make it more target-like by changing the vowel sound in 'back' from /a/ to /e/. This looks like a type of *hypothesis testing* possibly due to the fact that the interlanguage speaker finds it difficult to distinguish between the different vowel sounds of the English phonological system[9]. Based on the same pattern, we also found instances of *self-correction* (e.g. in 21, 'alt key', /alt/ /ki//olt/ /ki/, where the interlanguage speaker assumed that the pronunciation was incorrect and as a strategy, she corrected herself). *Hyper-correction* was also

reported (e.g. in 34, 'among' əmong/ and /əmon/, the first pronunciation had a [g] instead of [ŋ] and was not recognized, so in the second pronunciation the interlanguage speaker hypercorrected herself by eliminating the [g] all together). *Overgeneralization* took place in the form of the occurrence of phonological forms like in 16, 'delta' /delta//deɫta/ where in the second pronunciation of 'delta' there was an approximation of a velar /ɫ/ and the interlanguage speaker overgeneralized the use of the dark [l] to this word, where there are only clear [l]s in Italian.

As a general observation, and not consistent enough to be categorised as patterns, the speaker tended to simplify the target language phonology (e.g. 24, 'enlarge 50%' /enlardz/ /fɪftɪ/ /per/ /sent/), lack discrimination between short and long vowels (e.g. 44, 'team'/tɪm/) and fail to produce diphthongs and triphthongs throughout.

The rate of accuracy for the native speaker was quite high compared to the interlanguage speaker as shown in Table 7 below. The interlanguage speaker showed a very low score on accuracy after the first and the second test (Dictation Tasks 2a and 2b, in Table 6 above), but there was a considerable improvement in performance. We tend to interpret this result as a consequence of the variability of the interlanguage system. Whilst the native speaker performance is relatively stable (hence the minimal variation after 4 months of use of the dictation software), the interlanguage speaker's performance is more subject to variation due to the instability of the interlanguage system. As a consequence of this variation, the accuracy of recognition tends to stay quite low on the whole, but improves considerably as a result of the interlanguage speaker's continuous learning.

The interlanguage speaker's metalinguistic awareness results in a pronunciation progressively closer to the target language as she interacts with the machine and becomes more careful especially in the pronunciation of those words which she perceives more difficult to articulate or those words for which she has had positive or negative feedback from native speakers. This might also explain why originally the results obtained for the first tasks (Quick Training Modules and Dictation Task 1) showed a better performance of the interlanguage speaker versus the native speaker. During the initial training, the interlanguage speaker was particularly conscious of her interlanguage phonology and therefore made the effort of pronouncing words carefully. However, during everyday use, she backslid to her normal pronunciation. This in turn resulted in the system adapting to her pronunciation in a much slower way than for the native speaker whose pronunciation did not change considerably from the original pronunciation of the training, except perhaps sometimes, due to performance effects such as casual articulation.

Table 7. Native vs. interlanguage speaker accuracy in Dictation Tasks 2a and 2b.

Dictation Task 1	words dictated	words recognized	recognition accuracy
command words	340	208	61.1%
command words + spelling words	394	262	66.4%
dictation words	206	109	52.9%
dictation words - spelling words	152	55	36.1%
command + dictation words	546	317	58.0%
total percentage of recognition accuracy			58%

Summarizing, the results of the two speakers' performances in the use of the speech recognition program were not significantly different during the accomplishment of the training and during the very first dictation (Dictation Task 1). In fact, at first, the results from the interlanguage speaker seemed to outrun those of the native speaker suggesting that there were no major differences. Subsequently, after the dictation of the first Trask's text (Dictation Task 2a), it became clear that there was a substantial contrast between the two, and that the adaptation for the interlanguage speaker was definitely slower, after four months, than the one for the native speaker. The accuracy of recognition seemed to particularly affect uncommon words (like 'glasnost' or 'rock'n'roll') and justifies the choice of using the Trask's chapter which included unusual words and many proper names. Finally the improved recognition of the interlanguage speaker could be a positive sign of the interlanguage system developing to incorporate more target language features, and this is an important aspect of language learning taking place. By contrast, the accuracy rate of the native speaker in the second dictation task (Dictation Task 2b) was more or less the same. The recognition percentage of the native speaker's dictation is high enough to justify the claim that the system adapts better, or perhaps faster, to the native speaker than to the interlanguage speaker, whose pronunciation is more unstable and requires more training.

Conclusions

In this paper we primarily explored the potential of one new class of technology – speech recognition by computer – in facilitating second language acquisition research. We would not be the first to claim the enormity of that potential, but what is new here is, as we said in the first paragraph of the paper is that as far as we know, this is the first paper on the topic of interlanguage speech recognition by computer,and, thus, importantly, there is no research literature to compare these exploratory results to. In the back of our minds, and we kept coming back to it in discussions with software people, was a secondary issue, a wider applied linguistics issue, the fit of speech recognition by computer, in

principle, with interlanguage speakers, in bothe software and hardware terms, which we briefly discuss at the end of the paper.

What we hope is that we have shown some important detail leading to an extension of our usual research paradigms in SLA and interlanguage studies. We determined that there are a range of issues in second language research that can be elucidated by this technology and focussed on three central processes: language transfer, fossilization and communication strategies. We also referred in detail (passim) to the problem of speaker variability caused by accent, in both native and interlanguage speakers as being one of the major hurdles in accurate speech recognition. In the interlanguage area, as might be expected, variation is clearly greater than in the native speaker area. This is particularly important due to the fact – and its depth only became clear to us as a result of doing this research – that such recognition systems *must have stability of pronunciation as input to the machine system.* A consistent result for us in this case study is that the native speaker subject achieved this stabilization easier and faster than the interlanguage speaker, which has led us to the new concept of *purposefully directed positive fossilization*, where what we have here is a type of fossilization which could become a benefit, a topic which should be pursued in second language acquisition research.

It is generally accepted that language transfer is best thought of as a cover term for a whole class of behaviors, processes and constraints having to do with cross-linguistic influence (Selinker, 1992), in this case from several possible sources, the subject's native Sardinian, learned standard Italian and classroom French. In terms of the ten principles of language transfer listed above, in the one of the pronunciations produced, it is very clear that the principle 'transfer is a selection process' was shown here to be the case in that not all potential predictable instances of transfer occurred. The transfer principle of interlingual identifications where, the learner 'makes the same what cannot be the same' (see Weinreich, 1953), is also operative in our data in that equivalences were created across linguistic systems that were unexpected. Also, the transfer principle where 'blends and autonomous material' are to be expected in interlanguage, in the same manner held. Another thing that was clear is that the principle 'to create equivalence in the next language, use key linguistic variables' is operative in these data in that 'phonetic similarity' at times played a key role.

In the case of *DD* the interlanguage speaker is left with the option of testing new, and often untried, types of pronunciations until a satisfactory level is reached for the machine, if it is reached at all; as is shown in Appendix 2, in some cases, it is absolutely not reached at all. The implication of this testing, and we think it may be positive in terms of language learning, is that the interlanguage user may be forced to check her pronunciation against the phonological transcription of a dictionary or of a native speaker. This can become a successful strategy of 'communicating' with the computer if a pronunciation stability can be arrived at.

Also as a result of this work, we became aware that other ways to research second language acquisition begin to appear. Speech recognition by computer could be investigated as a device for investigating the improvement of pronunciation. What seems to be clear from our research is that the interlanguage user can become *metalinguistically aware* of a 'correct' pronunciation and, research could show if, in time, she be able to train the machine with a pronunciation very close to the target language, but care should be taken because backsliding can occur, as we have seen in our results. What is

important is that in subsequent uses of the machine, the interlanguage speaker has to maintain the stability of that type of pronunciation and with the help of the native speaker/teacher check that she does indeed match the pronunciation originally recorded for the training.

We believe that the entire system could be made more suitable for pedagogical and research purposes if it were equipped with a speech synthesiser for then instances of misrecognition could be pointed out against a target language pronunciation model and then consulted by the learner at their request. These types of programs are already available[10] although their purposefulness often comes from feedback given on grammatical structure and not on phonology. One could research applications of this type which could be particularly useful for learning tonal languages like Chinese for instance. In this case the hypothesis would be that speech recognition by computer would enable the learner to check which tonally distinguished word-form has been recognized by the system.

One of the main theoretical implications of this study resides in the distinctive use of the recognition system by the interlanguage speaker. In this study, language transfer takes place extensively, but we think not necessarily in the same ways as it would occur in ordinary conversation. Language transfer, of some special, and perhaps, unknown kind, seems to be one of the major causes of poor machine recognition, though it appears to be mainly of the type that has been called 'negative transfer'. The reason for this is interesting: the interlanguage speaker, according to comments gained as retrospective data, *consciously suppressed* phonological transfer by trying to pronounce words in a target-like fashion, although in a normal conversation she would not have pronounced the same words that way. This is an example of a domain result in that we hypothesize that it would only occur in the software training domain.

But once the training task was over, transfer occurred mainly for those words which had cognates in the mother tongue. An implication of this phenomenon, which is peculiar to interlanguage speakers, is our conclusion that speech recognition by computer cannot function satisfactorily if the reference speaker model adopted is the same for native speakers and non-native speakers. If language transfer is a major interference to the speech recognition system, then we think that separate programs for interlanguage speakers may have to be developed which, among other things, will have to take into consideration issues like the presence of cognate words and their special linguistic and cognitive status. The serious computational problem is whether currently conceived hardware will suffice or whether new sorts and expensive sorts, of hardware will have to be envisioned. This is one place we feel sure that applied linguistics can help computer scientists and programmers understand issues of import to them[11].

One thing is clear to us from this research: dictation packages such as *DD,* in order to be successfully used by interlanguage speakers cannot afford to have a native English speaker as its reference model, the native speaker *cannot* have a priviledged place. Theoretically it would have to have an interlanguage speaker as the reference speaker, in this particular case, we think, an Italian speaker. Although objections can be made about the issue of which accent of Italian should be chosen for the modelling, the same can be said for the choice of the English accent (i.e. from which dialect of English should the reference speaker be chosen). In fact the choice of accent does not prevent the system from working fairly efficiently (albeit not wonderfully) for native speakers of English

and given that the model employed is a fuzzy model, that should allow interdialect variations.

In this scenario, the issue of fossilization takes on a new dimension. One effect of fossilization has to do with the user's need to interact efficiently with the machine and this appears to result in the interlanguage speaker actually becoming fossilized in the use of some forms for the purpose of being recognized. The interlanguage speaker commented that sometimes forms which she would not normally use, for instance /ð/, would appear in her interlanguage but as it was not recognized she pronounced it as she 'normally' would (i.e. /d/). Hence for instance the pronunciation of 'the' as [de] (as it was recorded initially) was becoming a candidate for fossilization because when the target language pronunciation [ðe] also appeared (as a result of linguistic variation), the machine would not recognize it, and the interlanguage speaker would revert to [de] in order to achieve better recognition. This is the case in which fossilization, in this context, can be seen as 'positive' with regard to interacting with the computer.

The paradoxical aspect of the use of the speech recognition by computer is that it becomes an instrument for gaining greater linguistic awareness of target language pronunciations and noticing the gaps between interlanguage and target language, as discussed above. As a consequence apparently fossilized forms might become destabilized within the specific domain of human computer interaction. The interlanguage speaker might find herself constantly reminded of the interlanguage pronunciation and, unlike in normal human/human communication where politeness plays a part in not pointing out a pronunciation deviant from the target language, the electronic medium which is oblivious of such pragmatic issues, draws attention to the error and functions as a very pedantic sort of teacher. The positive implication of this type of unsolicited feedback is that perhaps in the long term, this destabilization of apparently fossilized forms *might* extend to other domains outside the virtual environment.

As a consequence, there is a possibility that on the one hand, speech recognition by computer might reinforce interlanguage forms, stabilising or maybe even fossilizing interlanguage forms for reasons of efficiency of interaction with the adaptive recognition algorithm. On the other hand, speech recognition by computer might result in the interlanguage speaker becoming more aware of their interlanguage system, and as a positive consequence of that, the speaker might develop an interlanguage system closer to the target language. We need serious second language acquisition research of a new type related to our current technological age.

In this study we describe an aspect of human-computer interaction specifically comparing interlanguage speakers versus native speakers interacting with an electronic medium, in this case speech recognition by computer, to our knowledge a comparison not undertaken before and there is thus no literature available on it. This seems strange to us as there must be millions of non-native interlanguage users in a world where computers seem to have been built for California teen-agers[12].

The results here suggest that, although the ability of the machine to recognize interlanguage speakers, after 4 months was 65.6%, the degree of accuracy or the user friendliness of the program was illusory and far from being satisfactory. During the initial use of *DD*, for a successful recognition the

speaker had to use as many command words (the majority of which were used for spelling and editing) as dictation words. This resulted in a tedious and frustrating experience. It could be argued that perhaps these disappointments are attributable to the design of the software and not, as suggested, by the non-nativeness of the user. However the results of the dictation tasks refute this conclusion given the considerable discrepancy in recognition between the two speakers (i.e. the native speaker recognition accuracy was 85% while the interlanguage speaker accuracy only 65 %, up from 49%).

This in turn provided a potential answer to considerations of comparative accuracy of interlanguage and native speakers: The accuracy rate was a great deal less than that for the native speaker (about 35% less after the first dictation task (Dictation task 2a), reduced to around 20% less at the end of the second task (Dictation task 2b)). A major result is that the adaptation for the interlanguage speaker was definitely slower (after four months) than the one for the native speaker and the accuracy of recognition seemed to particularly affect uncommon words. The improved recognition of the interlanguage speaker was interpreted as a positive sign of the phonological interlanguage system developing to incorporate more target language features, although apparently fossilized forms were found to co-exist with non-fossilized forms. Finally, the recognition percentage of the native speaker's dictation was considered to be high enough to justify the claim that the system adapts better, or perhaps faster, to the native speaker than to the interlanguage speaker, whose pronunciation is more unstable and requires more training.

We also wanted to know if the use of this new technology affects the development of interlanguage systems: does it improve or impede them? According to retrospective comments by the interlanguage speaker, it often appeared that words initially not recognized, were recognized the second time because the speaker often remembered the pronunciation that she used for the training. In light of this situation, it appears that speech recognition by computer forces the speaker to pronounce words in a consistent way and if the pronunciation is target-like then, at least in this domain of computer mediated interaction, thanks to this new technology, the interlanguage phonology could be improved. We could argue here that this is a type of 'forced stabilisation'. In this domain of speech recognition by computer, the interlanguage speaker's phonological system might become stable. Whether this phenomenon would extend to other domains of interlanguage is a question that can be pursued for possible future research together with the question of whether the interlanguage would change as a result of the use of speech recognition by computer. This question requires a study which takes into consideration other domains of interlanguage (see e.g. Selinker & Douglas, 1989).

This research also provided we think some insights into the area of cognition in human-computer interaction. It appears that the interlanguage speaker treats the machine as an expert native speaker: if it does not recognise her, she thinks her pronunciation is wrong and she varies it. This phenomenon is deeply cognitive. Although at some level, the interlanguage speaker knows that the machine is ready to accept any articulation she gives it and it is ready to be trained to recognise that, she actually sits in front of the screen expecting the machine to respond best to RP English or standard American English. This causes her to doubt her own pronunciation because she is aware it is not RP or standard American English and that leaves her to having to 'guess' the target pronunciation often relying on the L1 phonological system for filling in that gap.

Another aspect of this particular type of technology is the visual influence of the program which activates a different type of cognition. The interlanguage speaker reported in her retrospective comments that sometimes the visualisation of a certain word during the training would trigger certain pronunciations clearly influenced by transfer of phonological features from the L1 and she seemed to alternate between production of *some* target-like and non target-like phonological forms.

Fossilization plays an important part as well since it becomes an instrument for achieving successful interaction by the interlanguage user with the machine, thus the new concept of 'positive fossilization'. At the same time we can see that certain apparently already stabilised features of the interlanguage speaker (e.g. the lack of aspiration of word with [h] in initial position) will be difficult to eradicate. Yet there was evidence that in certain conditions fossilized forms were being destabilized (for instance the [r] at the end of words was progressively being substituted by a [ə]] although the two forms both appeared in the system, showing a linguistic variability). However, given the essence of fossilization, the real test can only come from longitudinal studies on interlanguage SRC.

Finally, in terms of second language acquisition research, one of the results of the study, we should not forget is that the interlanguage speaker gained a better awareness of her phonological interlanguage system, as well as the gaps with the target language system, as a result of the use of the software and in some cases she was able to improve her pronunciation making it more target-like. The transfer effect of the user's native language on the interlanguage, which often seemed to result in misrecognition by the computer, and inadequacy of the interlanguage communication strategies, especially where treating the machine as a human created problems, where communication strategies that worked with humans like repeated attempts at target pronunciation that worked with humans but not with computers. Given these factors, in virtue of the Multiple Effects Principle described above, it appeared that those words misrecognized more often as a result of transfer effect were the most likely candidates for fossilization. Thus, if the fossilization proved to be positive, as we think it did in some cases, though longitudinal studies would have to be carried out to be sure, the interlanguage speaker achieved a certain degree of stability in her interlanguage for reasons, perhaps, of efficiency with the computational machine.

Finally, we believe that it is incumbent upon applied linguists to become, where it is reasonable, part of the larger social and economic fabric, looking for ways to make applied linguistics a main stream discipline. This is not seen by us as an entirely altruistic activity since we feel that we should be consistently seeking ways to increase the job space for applied linguistics graduates. In pursuing the work of researching interlanguage speech recognition by computer, we have spoken to many software and hardware colleagues struggling to be the first to perfect SRC; many of these are indeed linguistically sophisticated and were easily able to integrate the interlanguage notion into their perspective. The topic of greatest interest was whether with more and more continuous recognition systems, to accommodate the sorts of interlanguage recognition problems we discovered here, would only new software solutions be adequate or would very expensive new sorts of hardware have to be developed. Not one of our computer informants was totally willing to bet on the only software option totally. Most of them could see a place, if they weren't so busy

surviving and creating product, for joint research ventures between interlanguage applied linguists and computer programmers and engineers. As was anticipated in the introduction, the task faced by computers in understanding human speech is extremely complex. A computer can recognize human speech by using model matching but it needs to overcome the problem of variability of different speakers, and that is the reason why programs like DragonDictate are speaker dependent. It is our conclusion that:

The speech recognition by computer solution of model matching works well for native speakers but less well for non-native speakers where interlanguage is often changing and where interlanguage speakers are often unsure of what the target should be.

It is an accepted fact by now that interlanguage speakers achieve variable success in second language acquisition and it is our conclusion that this results in speech models whose characteristics are not as fixed as native speakers' speech models. This has implications for speech recognition devices that applied linguists might wish to pay serious attention to for the reason listed above. Applied linguistics deserves to become a main stream discipline: after all, it is the only discipline that has the solution of practical language problems as one of its main foci. This is of concern also to other colleagues in applied linguistics who wish to work on language problems related to computer development, in general.

As we have seen above [ebov][13], the interlanguage speaker's phonological variation seems to be one of the causes for the slower process of the software adaptation. An important factor for successful use of the software by non-native speakers is the relationship between the cognition of the interlanguage speaker and the social environment in which they interact. Note that in this case, the "normal" human social environment has been replaced by an artificial or virtual social environment, the electronic medium. In the speech recognition by computer systems we have seen, and of course one can *never* be totally up-to-date in these matters – the speaker only receives tacit feedback while using the system and the only explicit feedback she receives is negative in essence, because it appears only when there is word mis-recognition. This is opposed to "natural" human interlanguage communication, where human interlocutors are unlikely to interrupt a non-native speaker to correct their pronunciation (unless successful communication is at stake) and the native speaker's body language or linguistic behaviour may well function as an indirect form of feedback to the interlanguage speaker. Interlanguage speakers seem to rely on very sophisticated pragmatic systems to meet their communicative needs. It is our view that such natural systems may *not* be very effective in the area of virtual life where the virtual medium uses brute language engineering whilst providing not always useful feedback to the user, which means to us that new sorts of communication strategies may have to be evolved by interlanguage speakers and by designers of speech recognition systems. If applied linguists work hard trying to make applied linguistics a main stream discipline in this way and begin to work with computer colleagues on these sorts of problems, we believe that there is a good chance that the job space for applied linguistics graduates may begin to approach the unfillable job space that now exists in Silicon Valley for computer people who may well need our skills, if we are inventive enough.

Endnotes

1   We acknowledge with gratitude the help and inspiration of our 'tech guru', Jim Tyson, who not only pushed us technologically, but also inspired us to explore and rethink our applied linguistics 'content' in terms of new and developing technological dimensions.   For us, Jim's constant intelligent help puts us on the side of those who argue in discussion groups that for content to progress in a challenging technological world, those who do the technology cannot be outside the content but *must* know and care about the very essence of the content, as Jim does.

2   We thank Carol Kinahan for help with this section.

3   We will describe speech recognition by computer in a more detailed way in the section on materials below.

4   "The technique of (hidden) Markov modelling [...] extracts probabilistic information about speech units, but requires a vast amount of training data in order for the system to 'learn' how to recognize speech segments" (Murray, 1995, 149).

5   The summary section of chapter 6 in Selinker (1992) of the quintessential interlanguage concept.

6   Trask, R.L., (1996) 'Lexical and semantic change' from *Historical Linguistics*. London: Edward Arnold.

7   The results of one task, a reading of a portion of a book on historical linguistics is presented in Appendix 1 at two points in time.

8   We will here not get into the extensive debate of the differentiation between learning and communication strategies.

9   One obvious reason that this probably happens because Italian has between five and seven vowels (depending on the dialect spoken), as opposed to the many vowels sounds found in English.

10 Programs like 'TriplePlayPlus' use speech recognition (licensed from Dragon Systems, Inc.) to give feedback on pronunciation. While in principle these programs constitute an innovative multimedia language learning aid, in practice at present they are still far from being useful. As part of this study the second author tested the Italian version of this program and discovered that even recognition of native speakers is really poor, let alone that of non-native speakers. The program is very sensitive to inflection and stress. The second author tried to use it speaking naturally and there were words or phrases which were not recognized even after several repetitions. Since the reference native speaker models used to program the software were representative of different accents of Italian, she tried to mimic their accents and noticed that recognition improved. This accent variation would be very hard to achieve for non-native speakers who at beginner level, although they might be able to perceive different accents, cannot mimic them in production. Another factor which seemed to affect recognition was the stress pattern. She tried to alter the syllable stress and noticed that the system responded to this variable. Here again, a non-native speaker would not be aware of the target language stress pattern at a beginner stage, since transfer of prosodic features often pervades the interlanguage phonology. On the whole she found the program definitely entertaining from a learning point of view but not effective to be used with speech recognition.

We note that Eskenazi (1999) commenting on TriplePlayPlus in the context of 'error detection', comes to similar conclusions which are arrived at independently. We interpret this to mean that in order for interlanguage speech recognition to be effective, one needs an additional tool which provides a special type of feedback, one that will provide additional information to the user as to how close s/he is to the target.

11 Again the reader is referred to Mascia & Selinker (in preparation) for discussion of these computer science issues.

[12]   The only potentially relevant study we have found in the literature, involving two interlanguage speakers of Dutch/English, was the report carried out by Dirksen and Ruys (1998) using DNS in the British English version, a type of continuous speech recognition. One of the results they obtained was that training is necessary for the recognition accuracy to be satisfactory and some of the tests they administered aimed at verifying the accuracy of recognition in dictation. They concluded that, given the two subjects were non-native speakers, the 95% accuracy rate as advertised by the program manufacturers, could be obtainable if the speakers' mother tongue was English. They graded the speech recognition performance between 80% and 90% but as they did not have native speakers as control subjects, they could not report on the differences (if any) versus native speakers of English and, importantly qualitative interlanguage analysis was not performed.

[13]   See Appendix 2.

References

Andersen, R.W. 1983. Transfer to somewhere. In S. Gass & L. Selinker (Eds.) *Language Transfer in Language Learning.* Rowley, MA: Newbury House.

Bamberg, & Mandel, M. 1991. Adaptable phoneme-based models for large-vocabulary speech recognition. *Speech Communication, 10*, 437–452.

Birdsong, D. 1992. Ultimate attainment in second language acquisition. *Language, 68*, 706–755.

Briere, E. 1968. *A psycholinguistic study of phonological interference.* The Hague: Mouton.

Coppieters, R. 1987. Competence differences between native and near-native speakers. *Language, 63*, 544–573.

Corder, S. 1983. A role for the mother tongue. In S. Gass & L. Selinker (Eds.), *Language Transfer in Language Learning.* Rowley, MA: Newbury House.

De Angelis, G. & Selinker, L. 2000. Interlanguage transfer and competing linguistic systems in the multilingual mind. In Jessner, U., Hufeisen, B. & Cenoz, J. (Eds.), *Cross-linguistic aspects of L3 acquisition*. Clevedon: Multilingual Matters.

Eskinazi, M. 1999. Using automatic speech processing for foreign language pronunciation tutoring: some issues and a prototype. *Language Learning and Technology, 2(2)*, 62–76.

Faerch, C. & Kasper, G. (Eds.) 1983a. *Strategies in interlanguage communication.* London: Longman

Faerch, C. & Kasper, G. 1983b. Plans & strategies in foreign language communication. In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (p20–60). London: Longman.

Gass, S. & Selinker, L. 1994. *Second language acquisition: an introductory course.* Hillsdale, NJ: Erlbaum.

Harris, Z. 1954. Transfer grammar. *IJAL, 20,* 259–70.

Kaplan, T. 1997. General learning strategies and the process of L2 acquisition: a critical overview. *IRAL, 36(3),* 233 – 246.

Kasper, G. & Kellerman, E. (Eds.) 1997. *Communication Strategies: Psycholinguistic and Sociolinguistic Perspectives.* Harlow, Essex: Longman.

Kellerman, E. 1995. Crosslinguistic Influence: Transfer to nowhere? *Annual Review of Applied Linguistics, 15,* 125–150.

Kellerman, E. 1977. Toward a characterization of the strategy of transfer in second language learning. *Interlanguage Studies Bulletin 2*, 58– 145.

Kinahan, C. 1999. Communication Strategies: How they may help the learner notice the gaps in their interlanguage. Paper presented at the 1999 Summer Institute Symposium, Manoa, Hawai'i.

Kinahan, C. & Selinker, L. 1997. Learning/Teaching Strategies pilot database. Webbed paper. http://alt.venus.co.uk/VL/AppLingBBK/VLDB.html

Lado, R. 1957. *Linguistics across cultures.* Ann Arbor: University of Michigan Press.

Larsen-Freeman, D. & Long, M. 1991. *An introduction to second language acquisition research*. London: Longman.

Mandel, M. 1992. A commercial large-vocabulary discrete speech recognition system: DragonDictate. *Language and Speech, 35,* 237–246.

Mascia, R. & Selinker, L. in preparation. Speech recognition of interlanguage: is new hardware necessary?

Murray, D.E. 1995. *Knowledge Machines: Language and information in a technological society.* London: Longman.

Nemser, W. 1961. *An experimental phonological study in the English of Hungarians.* The Hague: Mouton.

Perkins, D.N. 1985. General cognitive skills: Why not? In S.F. Chipman, Segal & R. Glaser (Eds.), *Thinking and Learning Skills, Vol. 2.* Hillsdale, NJ: Erlbaum.

Preston, D. 1989. *Sociolinguistics and Second Language Acquisition*. Oxford: Basil Blackwell.

Radaravanija, 1966. *An analysis of the elements in Thai that correspond to the basic intonation patterns of English.* Unpublished Ph.D. dissertation. Columbia University.

Rutherford, W. 1984. Description and explanation in interlanguage syntax: state of the art. *Language Learning 34,* 127–55.

Rutherford, W. 1987. *Second language grammar: learning and teaching.* London: Longman.

Selinker, L. 1972. Interlanguage. *IRAL 10(3),* 209–31.

Selinker, L. 1992. *Rediscovering interlanguage.* London: Longman.

Selinker, L. & Douglas, D. 1989. Research methodology in contextually-based second language research. *Second language research, 5,* 93–126.

Selinker, L. & Han, Z. in press. Fossilization: moving the concept into empirical longitudinal study.

Selinker, L. & Lakshmanan, U. 1992. Language Transfer and fossilization: The multiple effects principle. In S. Gass & L. Selinker (Eds.), *Language Transfer in Language Learning.* Rowley, MA: Newbury.

Tarone, E. 1983. Some thoughts on the notion of 'communication strategy'. In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (p61–74). London: Longman.

Tarone, E. 1993. On the variability of interlanguage systems. *Applied Linguistics, 4,* 143–63.

Trask, R.L. 1996. *Historical Linguistics.* London: Edward Arnold.

Weinreich, U. 1953. *Linguistics across cultures.* Linguistic Circle of NY (reprinted in paper by Mouton).

White, L. & Genesee, F. 1996. How native is near-native? The issue of ultimate attainment in second language acquisition. *Second Language Research, 12,* .

Whitney, W. D. 1881. On mixing in language. *Transactions of the American Philological Association, 121,* 1–26.

Wode, H. 1977. On the systematicity of L1 transfer in L2 acquisition. In C. Henning (Ed.), *Proceedings of the Los Angeles Second Language Research Forum.* Los Angeles: University of California.

Appendix 1

The four readings of the original dictation text (from Trask's 'Historical Linguistics')

I. Original text
Speech recognition task
Trask 'Historical Linguistics'
Reference page 17
Chapter 2
Lexical and semantic change

Undoubtedly the most conspicuous type of language change is the appearance of new words. When a new word appears in the language, there will be an occasion on which you hear it for the first time, and you may very well notice that you have just heard a new word and remember the occasion. Depending on your age, you may perhaps remember the first time you heard somebody mention acid house, or chunnel, or glasnost, or floppy disc or laser; you may remember the first time president Lyndon Johnson spoke of the escalation of the war in Vietnam, or even the first time you heard the word television. I myself can clearly remember the first time I heard somebody use the word rock'n'roll to denote a new kind of music he was hoping to promote.
Apart from being conspicuous, the creation of new words is also exceedingly frequent. New words have been pouring into English through its history, and today the language is acquiring many hundreds, perhaps even thousands, of new words every year. One of the major tasks faced by lexicographers in preparing new editions of their dictionaries is to collect the thousands of new words which have appeared since their last edition, perhaps only three or four years earlier. Some dictionaries now come with cards tucked inside which invite readers to send in examples of new words they have come across. And one or two publishers even bring out annual volumes of new words. Where do all these new words come from?
There are, in fact, many different ways of acquiring new words, some of them exceedingly common, others rather unusual. In this chapter we will review these sources of new words, beginning with the simplest and most obvious source of all.

II. Native Speaker Dictation Text A

(The words in **bold** are misrecognized words.)
Speech recognition task
Trust, *'Historical Linguistics'*
Reference page 17
Chapter **to**
Lexicon and **suited** change

Undoubtedly the most conspicuous type of language change is the appearance of **Newham** words. **Where** a new word appears in the language, there will be an occasion on which you **here** it for the first time, and you may **bury** well notice that you have just heard a new word and remember the occasion. **20** on your age, you may **Bradford** remember the first time you heard somebody mention acid **concert**, or **child**, or **classless**, or **flight desk**, or **later**; you may remember the first time **Britain Lenin** Johnson spoke of the **estimation** of the war in Vietnam, or even the first time I heard somebody use the word **rival** to **develops out** new **concert** of **unique E. Wise** to promote.
Apart from being conspicuous, the creation of new words is also **truly driven. Newham** words have been pouring into English **throat** its history, and today the language is

acquiring many hundreds, **Bradford EM** thousands, of new words every year. One of the major **task case I lexicon** in preparing new editions of their **dictionary** is to collect the thousands of new words which have appeared since **there** last edition, **Bradford** only three or for years earlier. Some **dictionary** now come with cards typed inside which invite readers to send in examples of new words they have come across, and one or two publishers even bring out annual **blinds** of new words. **We're too** all these new words come from?

There are, in fact, many different ways of acquiring new words, some of them exceedingly common **common** others **rabbit** unusual. In this chapter we will review these sources of new words, beginning with the simplest and most **artists** source of all.

III. Native Speaker Dictation Text B

(Four months later; the words in **bold** are misrecognized words.)

Speech **resignation** task
Reference page 17
Trust 'historian linguistics'
Chapter 2
lexicon and **Sinton** change

Undoubtedly the most conspicuous type of language **changed** is the appearance of new words. When a new word appears in the language, there will be an occasion on which you **here** it for the first time, and you may very well notice that you **Selinker** just heard a new word and remember the occasion depending on your age, you may perhaps remember the first time you heard **Selinker** mention **Aston** House, or **child**, or **Selinker**, **War flagging** disc **Selinker** or **later**; you may remember the first time president Lyndon Johnson spoke of the escalation of the war **an Tiananmen**, **War** even the first time you heard the word television. I myself can clearly remember the first time I heard **Selinker else** use the word **rough** and roll **two Kinnock** a new **comment** of music he was hoping to promote.

Apart from being conspicuous, the creation of new words is also **Selinker** frequent. New words have an pouring into English **front** its history, and today the language is acquiring many hundreds, perhaps even thousands, of new words every year. One of the major tasks **phased** by **lexicon** in **competitive** new **conditions** of their **dictionary** is to collect the thousands of new words which have appeared since **there** last **petition**, perhaps only three or four years earlier. Some **dictionary** now come with cards tucked inside which **neither** readers to send in **itself** of new words they have come across. And one or two publishers even bring out **able violence** of new words. Where do all these new words come from?

There are, in **Selinker**, many different ways of **quiet** new words, some of them exceedingly common, others **runner** unusual. In this chapter we will review **peas** sources of **Nicholls** words, beginning with the **Selinker** and most **countries sauce** of all.

IV. Interlanguage Speaker Dictation Text A

(The words in **bold** are misrecognized words.)

Speech repeated task
Reference age 17
Task advocate Israel enlisted.
Tottenham to.
Liverpool end. demand change.

Undoubtedly the **mosque constituents tired off linkage** change **ease** the appearance of new **work without offer** when a new **work** appears **been there linkage**, there will **the end appealing** on which you **here** it for the first time, **N. New my** very well **knotted** that you **haired task burned** a new **work end remember** the **occasional**. Depending on **Europe been**, you **main caps** remember **there** first time you **first family mainstream** acid house, **Orrell Channel**, **Orrell** glasnost**, Orrell Sabine East, Orrell leader**; you **main** remember the first time president **Lincoln Jan stock** of the **destination** of the **Worrall been given**, **Orrell** even the first time you **hurt somebody you** the **work** rock **primrose soon been age** new kind of **New Zealand** he was **hopping** to **France**.
Jazz space stateAbout from these constituents the Koreans of new work ease also exceedingly frequent author Neil work has been boring interval England France beat history, end today the language ease ignorant many Andrew, perhaps even000,of new work every year. One of the major task faced by Liverpool been preparing new BBC of their insurance ears to collect the talent of new work which has appears since their last BBC, perhaps Ali Green Orrell forward years earlier. Some missionaries now down with house tapped inside which invites readers to send been exemplified of new work they haired hand across, end one Orrell to publishers even bring out angel values of new work. Where to all these new work down from Bazoft
can there are, been bucked , many different ways of **ignorant** new **work, Stan** of **them EC,, honours** rather **NUM**. In **piece** chapter **with** will review these sources of new **work**, beginning with the simplest **end half of stores** of all.

Interlanguage Speaker Dictation Text B

(Four months later; the words in **bold** are misrecognized words.)

Speech recognition **past**
reference page 17
task destroyed linguistics
Chapter **to**
lexicon and semantic change.

Unpopular the most conspicuous type of language change **easethese** appearance of new **wards**. **Wife** a **Newell** word appears in the **cleanliness**, there be an **occasionalhonour** which you **hereEast** for the first time, **endyou'llmain merry** well **knottedpets** you have just **hurt** a new **wards** and remember the **occasional**. Depending **honour Europe aid**, **you'll main** perhaps remember the first time **you'll hurt** somebody mention **Hussey** house, or **Channel, poor Gazdar if** or **blocking this**, or **Labour**; you **main** remember the first time President Lyndon **Jansen** spoke of the escalation of the war in **Interlanguage**, or **Egan** the first time you **hurt** the **wards policeman**. I **might** can clearly remember the first time I heard somebody use the **wards Roxburgh** to **default** a new kind of music **key** was **Opie** to promote.
Apart from **the selinker**, the creation of new **wards Heath Olsen** exceedingly **treatment newer** words have been pouring into **Sorace**, **pronounciation it** history and today the language **required** many hundreds, perhaps even thousands, **on** new **wards** every year. One of the major tasks faced by **difficulties selinker** preparing new editions

pouredthemissionariesease to collect the thousands of new **wards** which have **appears** since **the** last **editions**, perhaps only three or four years earlier. Some **missionaries** now come with **typewriter tax** inside which **invites** readers to send in examples of **Newell** words **main come because** and **once war to** publishers even **Rayner** out **angel bodied** of new **wards**. **Whereby** all these new words come from?
The car, in fact, many different **waste** of acquiring new **wards**, **from of men illegally,,** **other** rather **ideal**. In **these** Chapter we will review these sources of new **wards**, beginning with the simplest and most obvious **sources** of all.

Appendix 2

Sample interlanguage data

The following sample data were among those analysed for insights on interlanguage patterns; they only hint at the range of potential variation and the effects of the central processes discussed in the paper: language transfer, fossilization and communication strategies.
For full effect, the 'reader' should listen to the audio tapes and view the sample viewtape referred to in endnote 8 of the main text.
The titles represent the different sections of the training from which the data were collected.

15 basic command words

1. oops   /oˈps/
    oops   /oˈps/
    oops   /oˈps/

2. international   /ɪnterneæʃonal/
    international   /ɪnterneæʃonal/

Quick Enrolment

3. begin document   /begin/ /dokument/

4. DragonDictate   /dragondɪkteɪt/

5. computer   /kompjuˈtə/

6. letters   /letəz/

7. tango   /tango/

8. romeo /romeɪo/

9. begin document   /begin//dokument/

10. bring up calculator /bring/ /up/ /kalkuleɪtə/

11. open square brackets /open/ /skwer/ /braket/

12. bring up calculator   (Several attempts)
bring up ca...calculator
bring up the...

Training session 2

13. back space   /bak/ /speɪs/
back space   /bak/ /speɪs/
back space   /bek/ /speɪs/

14. escape   /ɪskeɪp/

15. bravo   /ˈbravo/
bravo   /braˈvo/

16. delta   /delta/
delta   /deɫta/

17. november   /novembə/

18. oscar   /oskar/

19. equal sign   /iˈkwol/ /saɪʲ/
equal   /iˈkwol/
equals   /iˈkwolz/

20. greater than   /greɪter/ /den/
greater than   /greɪter/ /den/

21. alt key   /alt/ /ki/
alt key   /olt/ /ki/

22. bottom of document   /botom/ /ov/ /dokument/

23. computer please   /kompjuter/ /pliˈz/

24. enlarge 50%   /enlardʒ/ /fɪftɪ/ /per/ /sent/
enlarge 50%   /enlardʒ/ /fɪftɪ/ /per/ /sent/

25. function 2 /fankʃon/ /tu/
function 2 /fankʃon/ /θu/

26. home key   /oˈm/ /ki/
home key /hoˈm/ /ki/

27. horizontal size /horizontal/ /saɪz/

28. move left   /muˈv/ /left/
    move left /muˈv/ /left/
    move left /muˈv/ /left/

29. save user   /selv/ /juzez/

30. tilde   /tɪɫde/
    tilde /tɪlde/
    tilde /tɪlde/

31. above   /ebouv/
    above /ebouv/
    above /ebov/

32. already   /olredɪ/
    already /olredɪ/
    already /olredɪ/

33. although   /olðoˈ/
    although /holðoˈ/
    although /holðoˈ/

34. among   /əmong/
    among /əmon/

35. as   /ez/
    as /ez/
    as /az/

36. determine   /determln/

37. economic   /ekonomik/

38. equipment   /ekwikment/

39. gas   /gaz/
    gas /gez/
    gas /gæz/

40. is   /iˈz/
    is /ɪz/

41. law   /lʊ/
    law /lʊ/
    law /loˈ/

42. parents　/parenz/
parents　/parenz/
parents　/pærenz/

43. rational　/reʃonal/
rational　/raʃonal/

44. team　/tɪm/

45. toward　/tauard/
toward　/touard/

46. back 4　/bak/ /for/
back 5　/bak/ /faiv/

47. function 8　/faŋkʃon/ /eɪt/
function 8　/faŋkʃon/ /eɪt/
function 8　/faŋkʃon/ /eɪt/

48. How do I　/au/ /du/ /aɪ/
How do I　/hau/ /du/ /aɪ/

49. mouse up　/maus/ /ap/
mouse up　/maus/ /ap/
mouse up　/maus/ /op/

50. believe　/bɪli:v/
believe　/bɪlɪv/

51. Diana　/diana/
Diana　/daiana/

52. example　/ɪgzempol/
example　/ɪgzempol/

53. group　/grøup/
group　/grøup/

54. having　/eviŋ/
having　/heviŋ/
having　/heviŋ/

55. language　/leŋgwidz/
language　/laŋwidz/

56. motion   /moʃon/
motion   /moʃon/
motion   /moʃon/

57. saying   /se'iŋ/
saying   /se'iŋ/

Dictation Task 1

58. begin document   /begin/ /dokument/
begin document   /begin/ /dokument/

59. the /de/

60. were /wer/

61. done /dan/

62. several studies were done especially...by many linguists... (Edited version: several studies were done by many linguists..... especially...)

63. linguists   /liŋgwɪstz/

64. selinker   /selɪnker/

65. nemser   /nemzer/ (Continually recognized as 'never')

66. briere   /brie'r/

67. who   /u'/
who   /hu/

68. oops   /o'ps/
oops   /o'ps/
oops   /o'ps/

69. previous   /privis/

70. shed   /ʃed/

71. phenomenon   /fenomenən/

72. transfer   /transfə/

73. phonology   /fonolodzi/ (Never recognized and was dictated six times.)

74. differentiated   /di'ferenʃeted/ (*not* recognized and had to be spelled)

75. echo /etʃo/
     echo /etʃo/
     echo /etʃo/

76. still /stiˈl/

77. obligatory /obligatoɪrɪ/