



HAL
open science

The contribution of the computer to improving L2 oral production

Anthony Stenton

► **To cite this version:**

Anthony Stenton. The contribution of the computer to improving L2 oral production: An examination of the applied and theoretical research behind the SWANS authoring programme. *Etudes en didactique des langues*, 2012, *De l'anglais oral*, 19, pp.59-80. hal-04047775

HAL Id: hal-04047775

<https://ut3-toulouseinp.hal.science/hal-04047775>

Submitted on 27 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The contribution of the computer to improving L2 oral production. An examination of the applied and theoretical research behind the SWANS authoring programme

Anthony STENTON

LAIRDIL & Université Toulouse I Capitole

anthony.stenton@univ-tlse1.fr



© ND

Visual representations of language for students in Europe rely heavily on the imperialistic traditions of the Latin alphabet and the archaic complexity of the international phonetic alphabet. The computer can be used for graphic representations showing pitch height (fundamental frequency) and curves showing volume and length (oscillograms) but the accurate interpretation or decoding of such data generally limits its use to experts in phonetics and other acoustic scientists. The verdict after even a cursory examination of such representational systems is inescapable; foreign language learning is bedevilled by man's incapacity to visualise accurately the invisible: the vibrations of sound waves as they travel through the air. Students who have taken the trouble to acquire 10,000 words in a foreign language might reasonably hope to attain an "advanced" C1 level in oral production. The reality is that most European university students fail to advance beyond A2/B1 levels in spoken production in their chosen foreign language. Problems of articulation do exist but their importance is considerably smaller than that of problems related to listening perception. Studies of language perception in literate students thus try to analyse the activity of sound decoding rather like Plato's famous description of the prisoners in the cave: his prisoners do not see reality but only vague shadows against a wall. Similarly, L2 learners do not always perceive foreign sounds accurately but generate deviant deconstructions according to the nature of their mother tongue.

The name we give to this process is mother tongue interference (MTI) and it constitutes a major obstacle both to perception and oral production and hence to human communication. The interference often becomes fossilised, lifelong, and particularly resistant to change. Even gifted bilinguals make frequent errors of lexical stress when in emotional contexts or when levels of monitoring concentration begin to decline. The nature of the interference, however, varies locally according to the language pair in question. Whereas MTI is a universal phenomenon in foreign language learning, its particular manifestation is an index of syllabic structure, lexical stress patterns and other local features. Human beings may have been endowed with the same basic language learning capacities since the dawn of time but the playing field inherited from our more recent ancestors is not always level. In the case of languages with greater syllabic complexity, the acquisition of reading skills may take longer. In fact every language has a unique combination of letters and sounds, or graphemes (written system units) and phonemes, as well as having a particular relationship between them. When the relationship between graphemes and phonemes is transparent, children learn to read and speak earlier (European project: Enhancing literacy development in European languages – ELDEL). Languages said to have a good grapheme-to-phoneme correspondence include Bulgarian, Basque, Finnish, Georgian, Hungarian, Korean, Sanskrit, Croatian and Serbian.

Studies into dyslexia have also suggested that English is harder than most languages. Dyslexia is diagnosed much more easily when children speak English than when they speak romance languages such as Spanish or Italian. A 2001 test of 600 children in primary schools in fifteen European countries showed that it took British children two and half years to master basic reading skills acquired after only one year in Greece, Finland, Spain and Italy (Seymour, 2003). English has a complex syllable structure with many different types of syllables, most of them closed, (*e.g.* CVC, CVCC, CCVC).

Figure 1 (from Adsett & Marchand, 2010) shows the percentage distribution of CV syllables (which are thought to be easier to learn than other syllables) among a variety of languages: Italian, European Portuguese, Spanish, French, Dutch, English.

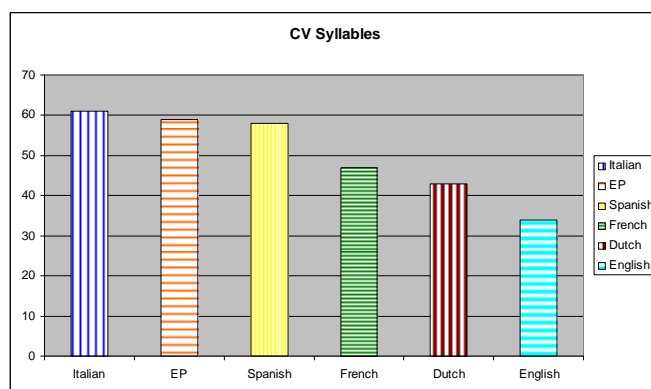


Figure 1: Percentage of CV syllables used in different European languages

The hypothesis is that a high percentage of CV syllables makes learning to read easier. If the hypothesis is valid, then learning to read the English language (which has 37% CV syllables) is more difficult than learning to read Italian (where CV syllables constitute 61% of all syllables).

A test which appears to confirm this position was carried out in 2003 by measuring success and failure in the detection of non-words. Children whose native languages were Finnish, French, Greek, Italian, Spanish, Portuguese (all believed to have simple syllabic structure) had significantly lower error rates in detecting non-words than children whose native languages were Austrian, Danish, Dutch, English, German, Icelandic, Norwegian, Swedish (all thought to be more syllabically complex), (Seymour *et al.*, 2003). It is estimated that the pronunciation of English vowels is only 51% consistent over different words (Trieman *et al.*, 1995). For example, *a* is pronounced differently in *cat*, *call*, *car*, *cake*, and *care*.

Mother-tongue interference in L2 perception and production is a natural and universal phenomenon described by Polivanov, Trubetsky and the Prague school in the 1930's and today known as "neural commitment" (Kuhl, 2000). The brain's commitment is a kind of highly specialised tuning capacity, partly acquired in the womb before birth, which is perfectly adapted to detecting one language and to excluding all others as parasitical "noise". Dulay *et al.* (1982) define interference as the automatic transfer, due to habit, of the surface structure of the first language onto the surface of the target language. Ellis (1997) describes interference as "transfer" which is "the influence that the learner's L1 exerts over the acquisition of an L2". The age of learning (before or after puberty) also intrudes upon the question (Critical Period hypothesis, Lenneberg, 1967). Sebastián-Gallés *et al.* (2005) stress the significance of phonology for L2 learning; they claim that learning an L2 once the L1 phonology is already internalised can reduce individuals' abilities to distinguish new sounds that appear in the L2. Peperkamp and Dupoux (2003) describe French L2 perceptual processing difficulties with English as "phonological deafness" which is prelexical. The name "Mr Watson" will be pronounced with a strong final syllable stress – "Mr What's on?" – because the French student has not perceived that the vowel in the final syllable is reduced. The French Minister for Foreign Affairs, recently deplored "the famine in Africa" leaving his Anglophone listeners completely mystified because he pronounced "famine" to rhyme with "machine" and placed a second syllable stress on "Africa": "the **famine** in Africa". Deafness occurs when non-native sounds are "assimilated" to the closest native phoneme category or "when two sounds are mapped on the same category" (Iverson *et al.*, 2003). Stress deafness is very persistent, and is still found in relatively proficient late learners of Spanish (Dupoux *et al.*, 2007).

What the durable notions of syllabic complexity and stress deafness both imply is a failure of pedagogical engineering to find suitable solutions to pressing problems of perception, oral production and ultimately communication.

By relying on the relatively primitive printing technology of our ancestors, we are clumsily advancing into an increasingly wired-up, connected, multilingual world with ill-adapted, basically monolingual tools which leave children and adults as victims of the poverty of our engineering imagination just as the blind suffered for millennia until tactile methods for learning to read could be invented. In theory, the IPA should have brought multilingual clarity and the light of unambiguous scientific rigour as academics agreed on normalised international definitions, in practise most non-specialist students have voted with their feet. Even the Cambridge Examining Board now gives feedback on the oral examinations without employing phonetic transcriptions. The IPA, like the Latin alphabet, has great prestige, a quasi monopoly situation and no valid competition. Unfortunately, this is still no recipe for success.

Questioning the symbols we use for language representation involves challenging considerable vested interests, whether those of the publishing industry or academia. In the field of pedagogical engineering, however, the arrival of the computer age has inevitably led to renewed typographical experimentation. Contrary to static typographical representations on paper, the electronic screen of the computer offers *animated movement, colour, more flexible size, and sound synchronisation*. Given the plasticity and ubiquity of the computer/smart phone environment, typographical design has recently acquired a new and legitimate impetus which the problems and ever-growing needs for foreign language learning in a multilingual world have made manifest.

Hypothesis: the alphabet amplifies L1 interference in the perception and oral production of an L2

It is the central hypothesis of this article that, in the European context the Latin alphabet intensifies mother tongue interference in the perception of an L2. Using the same symbols for two different sound systems is a dubious practice because confusion becomes rife. Speaking in a foreign language requires a continuous and high level of concentration to monitor oral production and eliminate interference. The English word “development” is mispronounced (given a third syllable stress) by most French citizens either because initial perception was deviant and so deviant pronunciation springs to mind or because concentration levels are insufficient to eliminate L1 interference. Problems of interference are certainly well-known and well-analysed but are underestimated pragmatically to the extent that such problems vary greatly according to the language pair in question. In fact, interference appears proportional to morphological proximity. The closer together two languages are in terms of morphology, the higher the likelihood of serious interference problems in perception and production. In the European context, one could argue that the French-English language pair are unique in that English is made up of 30% French and 30% Latin and yet the lexical stress patterns of the two languages are radically different. The result is that the English and French have intense, long-

lasting problems in acquiring spoken intelligibility; a finding confirmed for the French by recent Danish research (Rikke, 2005).

Anecdotal evidence suggests we can go further. Not only is the L2 speech production of English and French students mutually deviant but high intolerance levels of their deviant production are particularly widespread. The emergence of English as a lingua franca has reduced the negative effect of such intolerance in the case of the English simply because fewer and fewer English people make the effort to speak French. The poor intelligibility of their spoken French is generally invisible or unheard. In the case of the French, the effect is often the opposite. As French schools and universities have intensified efforts to teach English, the numbers of fluent French speakers of English have increased dramatically. Their oral production in English is stigmatized all over Europe and the stigma is arguably unjustified because the Francophone difficulties in acquiring intelligibility in English are manifestly considerable. In fact, both the English and the French suffer from opaque orthographies. The writing systems of Serbo-Croatian, Finnish, Welsh, Spanish, Dutch, Turkish and German are on the whole much more regular in sound-symbol correspondences than those of English or French.

A recent TOEFL test of some 120 nations placed French students 69th in the world for their skills in English (Baumard, 2009) (see also the Bonnet report where French students finished 7th out of 7, Bonnet *et al.* 2004). We would argue that such a result not only underlines an often undeniable truth about the intelligibility of deviant Gallic oral production, it also suggests excessive intolerance of the specific qualities of Gallic mother tongue interference in English language oral production. Whether or not such an attitude is based upon an appreciation of the stakes of a “linguistic 100-years war” still being waged – perhaps the French are subconsciously striving to bring back the English language into its “rightful” acoustic orbit? – it often appears suspiciously close to an expression of prejudice – an expression of scorn for a collective failing of Gallic DNA – and as such deserves condemnation by the academic community. One might argue that, compared to other nations, the French have an advantage in acquiring English lexis, much of which is Old French, but language learning is often a fragile affair. Once the notion of being “weak” in one skill area is acquired, demoralisation can rapidly spread and destroy motivation in other areas.

Why fluent L1 readers still require specific training in L2 sublexical reading techniques. The “cat may still have to sit on the mat” for L2 readers of all ages.

What lies behind the paradoxical problem of needing a visible form of language to acquire vocabulary quickly and yet suffering from amplified interference when trying to memorise L2 sounds in texts, is of course the practice of reading. Reading in the mother tongue is generally acquired at an early age when the brain “hard-wires” (*i.e.* renders efficient and accurate)

associations between orthographic form and sound. Students' brains naturally resist learning different sound/symbol associations which use the same symbols and which might destabilize one of the most useful skills of a life-time: L1 reading. Foreign language learning in Europe is a discipline apart for this very reason. It is not just difficult for some students, this specific kind of new learning can be spontaneously resisted by the brain. Recent research suggests that as the L1 reading brain becomes more efficient and automatic and improves its organisation in the left hemisphere, it partially displaces the brain's face-recognition functions which are shipped out to the right hemisphere (Dehaene, 2010).

In the context of such unpredictable cerebral reorganisation (how does the brain decide on priorities in establishing internal organisation?) we propose to examine syllabic textual annotation as a way of nudging the brain's perception capacities towards an improved level of accurate L2 performance. To improve L2 speech production, we propose to improve reading accuracy or perhaps we should more simply admit that we propose to address a curiously neglected problem in a highly literate European higher education student population: the art of identifying and pronouncing segments (syllables and/or phonemes, the controversy continues) in a foreign language.

The Synchronised Web Authoring Notation System (SWANS) was developed in the context of widespread perception tests which tried to define more precisely the nature of MTI in the English language productions of university students in Toulouse. Student mother tongues included French (80%), Spanish, German, Italian, Portuguese, Bulgarian, Romanian, Russian, Norwegian, Arabic, Wolof, Chinese, Japanese.

In 2005, between 70% and 90% of students tested in Toulouse (levels A2/B1), mispronounced the following high frequency words by displacing the primary stress or neglecting to use reduced vowel sounds and hence giving an *impression* of displaced primary stress (e.g. "senate" and "corporate" rhyming with "sen-eight" and "corpor-eight")

"government", "satisfy", "interest", "corporation", "corporate",
"analysis", "senate", "presidency", "military", "apathy"

It will be noticed that all the words except "analysis" have first syllable stress thereby demonstrating a particular area of weakness for Francophone students.

The high level of stress pattern deviance and the associated intelligibility problems suggested that a solution might be found by a campaign of awareness-raising, specifically related to the role of stressed syllables. Literate students who were particularly advanced rapid readers in their own L1 were clearly failing to recognise the L2 stress patterns and failing also to understand that this "negligence" might have disastrous effects on their own oral communication.

Modern perception studies have revealed that syllables play a greater role in speech perception than previously assumed (Mattys & Melhor, 2005). Chait *et al.* (2005) argue that the short input sound streams (~30 ms) and longer input streams (~300 ms) are processed simultaneously and separately before being bound together in a stable representation called a syllable. The populations tested by these researchers are generally highly literate students profoundly conditioned by their L1 experience of learning to read and the use of the Latin alphabet but S. Dehaene and his team (Dehaene, 2010) have taken the process further. In their 2010 tests they analyse the functional neuroimaging of speech perception and reading in adult literate and illiterate students and demonstrate the plasticity of the brain which structures for vision and language during the process of learning to read. Most changes occur even when literacy is acquired in adulthood, emphasizing that both childhood and adult education can profoundly refine cortical organization.

The SWANS authoring system

The Swans project united 12 researchers from the fields of linguistics, psycholinguistics, computer science, cognitive science, acoustics and human-computer interface design from four Toulouse laboratories (LAIRDIL, LTC, LAAS-CNRS, LORDAT) financed in the context of the CNRS-TCAN programme 2004-2006. The working hypothesis was that students must be helped to see the lexical stress patterns which they patently did not always hear through L1 interference. A number of potential visual models were examined and rejected. Graphic models of sound waves were rejected on the grounds of excessive complexity and debatable accuracy.

Our own highly simplified in-house authoring system “Sounds Right” (Péchou & Stenton, 2002) allowed teachers to drag and drop a variety of ready made pitch arrows (rise-fall, fall-rise, etc.) on to the screen, just under the sound synchronised text and modify their appearance manually by dragging the corners of the arrows to obtain different lengths or angles.

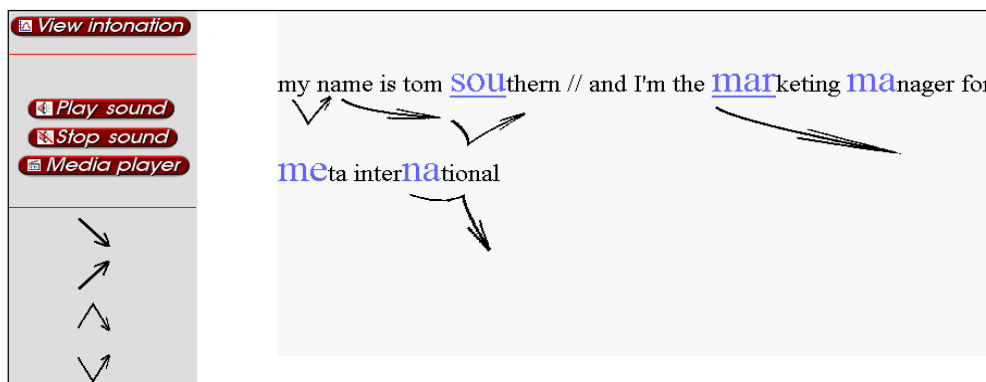


Figure 2. The programme Sounds Right (2002) which allows teachers (after first viewing the pitch curve of the phrase in “Speech Analyser”) to drag and drop appropriate pitch arrows onto the screen

This procedure, when synchronised with the sound, was useful for teaching intonation at the sentence level. The final conclusion of the Swans team, however, was that sentence level intonation was much less of a fundamental barrier to intelligibility than the interference from L1 stress patterns. Concentrating on lexical stress alone, it was hoped, would provide significant economies of scale as real communication progress might be more easily measured. Swans-type textual annotations, which are relatively modest and discreet, aimed above all to respect the essential requirements of a low learning curve for teachers and students and the possibility of rapid retention of stress patterns via dual coding to accelerate learning. According to the theories of Paivio (1986 & 1991), Sweller (1999) and Mayer (2001), such dual coding based on sound and typography should lead to better learning for novice learners and have no effect or even a negative effect (called “expert reversal effect”) on learners who already know the place of these accents.

WILLARD ROWLAND: Well, I think it's always the **problem** of that in a **profit-oriented environment**. There is a great need for the **producers** to **produce** not **necessarily** high **quality programming** but large numbers of **audience**. So in the **newspaper** it's a question of the **circulation size** and in the **broadcasting station** it's a question of the **ratings**. And to the **extent** that the **companies** are more and more **concentrated**, the **pressures** for that sort of **production of profit** are higher and higher. It's **exacerbated**

Figure 3. Textual annotation in SWANS using blue for primary stress (size 24), orange for reduced vowels (size 12) and purple for secondary stress (size 22). Ordinary black text is in (18).

The text is a 2009 UT1 video interview with Professor W. Rowland from Denver P.B.S. on the public media in the U.S.A. and available on a Moodle platform. With Swans, the above audio script is placed in a web page with the video film and synchronised line-by-line with the sound.

The SWANS authoring system uses typographical techniques, changes of letter size and colour, and line-by-line synchronisation, to help students notice and memorise the target lexical stress patterns. We refer the reader to Stenton (2008) for a basic introduction to the SWANS authoring system and, in particular for an account of crucial design choices in terms of text size and colour, and to (Stenton *et al.*, 2005a) for an account of technical considerations in the creation of sound synchronised web pages with the authoring system. Throughout its development stages SWANS has been tested by language teachers across Europe in CERCLES language centres and we are grateful for feedback particularly on multilingual interference questions and the notion of

training teachers from different countries through shared networked annotation of the same basic documents. Recent pedagogical feedback highlights the practical use of SWANS in multimedia laboratories in the presence of teachers, in self-study facilities in Language Centres and in other contexts including cyber-café's, and working from home.

Synchronisation

The practice of synchronisation of audio and text dates back to the origins of the cinema and television. Al Jolson's performance in *The Jazz Singer* (1927) was first subtitled in Danish in the early 1930's and the BBC broadcast German opera with subtitles on television in the 1940's. Subtitles are often a limited technique for synchronisation as the text is rarely the full focus of attention. The spectator is invited to watch a film, or look at pictures, or listen to sound and the text is discreetly added at the bottom of the screen as an extra, as a bonus, rather than as an equal partner in an audiovisual reading experience. In the context of the modern cinema, subtitles have evolved little since early experiments. They remain fleeting (and therefore visually distracting), are rather small and are still not always perfectly in time with the sound track. In the context of CALL, however, high definition, on-screen synchronisation is becoming a more flexible tool which modifies the reading experience and holds out promising, unexplored educational potential. One important breakthrough came with the invention of SMIL language (W3C 1999) SMIL (or "Synchronised Multimedia Integration Language") which permits fine-tuned synchronisation measured for the first time in hundredths of a second. The SWANS programme uses SMIL language to generate its web pages from within Internet Explorer.

On the computer or LCD screen different techniques of synchronisation have emerged in software in the last 12 years.

- subtitles, particularly for films, where a short line of text appears and disappears in time with the soundtrack;
- karaoke, which is usually based on word by word synchronisation and which is visually distracting through the jerking movement of the animation;
- highlighting, where a whole page of text is visible for much longer on the screen and is highlighted with a band of colour behind the particular lines of text which the listener is listening to. The technique of highlighting, which is used exclusively in SWANS, allows the student to read and re-read elements of automatically scrolling text. The fact that several lines of text remain visible on screen for a longer period of time constitutes an improvement in the reading experience compared to the reading of subtitles or karaoke.

For the Swans team, synchronisation through line-by-line highlighting of annotated text offered other potential advantages which few language teachers have considered. When silently reading an L2 text, all language students who have acquired a basic grounding subvocalise to a certain degree, that is to say they "hear" much of the text in their head. As this process engenders L1

interference, what students hear undoubtedly suffers from perceptual distortion. When L2 scripts and L2 soundtracks are synchronised and annotated, however, the reading brain can no longer subvocalise in the same uninhibited way. The L2 soundtrack and the textual annotation interfere with the L1 interference which obliges the brain to choose between perceptually distorted subvocalisation and the L2 sounds the ear is busy decoding. Simultaneous listening and reading is not just multitasking, it is creating a destabilising dilemma for the brain which is receiving mixed messages from the eyes and ears. Whether or not this destabilising experience can become a positive reinforcement of the authentic L2 speech patterns, accelerate the elimination of L1 interference fossils such as “development” and thereby enhance intelligibility in oral production, is precisely the challenge our research set out to measure. In this context, the possibility within SWANS web pages of switching between an annotated web page and an unannotated web page, while listening to the sound, plays an indispensable role in helping students notice what is at stake in this eye/ear dilemma in terms of the strong and weak forms of the L2 lexical stress pattern.

Line by line synchronisation of text through highlighting also offers a degree of ocular comfort which should not be underestimated. As the essential semantic content of the text is already “known” to the student through prior discussion, the value of the activity hinges around the degree of cerebral concentration applied to the exploration of language form through the reading activity. Synchronisation, we suggest, represents a lightening of the cerebral burden as the eyes are guided and can focalise more easily on essential linguistic clues. Feedback from students specifically refers to reduced eye fatigue. Here is a student talking to himself while listening to and reading to an unannotated script: “the word I am hearing starts with a prefix ‘in’. The prefix ‘in’, I remember, is not usually stressed (*e.g.* ‘intelligent’ = 2nd syllable stress, ‘intellectual’ = 3rd syllable stress) but this word is ‘interesting’ which, if I’m not mistaken, has a stress on the first syllable. Let me check in the annotated page... Yes I’m right...‘interesting’ does have a first syllable stress. Now that’s interesting!”). The preceding inner dialogue requires a degree of concentration on linguistic form which is relatively rare in multimedia language laboratories or classrooms. Facilitating such thought processes through highlighted “metaguiding”, which is exactly what we do when placing our finger under a word to help the eyes find the right place on the page, represents one of the most important contributions to reducing the physical effort of reading that pedagogical engineers can make.

The use of audio synchronised text to promote language learning is clearly on the increase. Not only is such reading being experimented on smart telephones and E-readers but vast numbers of public domain readings of books (world literature before 1923) are now available online as synchronised audiobooks where the reader only has to click on the page to hear it read out loud with highlighted synchronised text (LibriVox, whose ambition is to record

every book in the public domain and the open platform Dinglabs, allow anyone to freely publish an audio and a transcript in synchronised form).

Testing synchronisation and annotation techniques

Testing the dual coding effect of synchronisation and SWANS annotation techniques, included formal testing of perception and oral production before and after training, and questionnaires to obtain student feedback.

Early testing of Swans took place with over 250 students in four Toulouse institutions (Université des Sciences Sociales, Université Paul Sabatier, IUT Ponsan, Collège Château de l’Hers) between 2004-2006 (test results of 2004-2005 are analysed in (Stenton *et al.*, 2005a and Stenton *et al.*, 2005b).

More recent testing in the University of Toulouse (2007-2010) concerns over 300 students in postgraduate Master courses: Statistics M2, Finance M2 (levels B2/C1) Human Resources M1 courses (level B1/B2) and post-doctoral students from Toulouse research laboratories preparing “English for Communication in International Conferences” in the fields of computing, linguistics, sociology, law, economics, archaeology, geology, mathematics, and meteorological studies (levels B1 to C2).

Student L1s were approximately 80% French. 20 % of students were of some 30 different nationalities: European, North African, African, Chinese, Japanese.

Pedagogical methods

Ten-week courses using SWANS documents aimed to increase awareness of English lexical stress patterns as part of the wider course objectives in improving oral communication skills. Video film of fluent English-speaking French politicians and managers who ignored English stress patterns were used to demonstrate potential communication problems. Students were first introduced to Swans annotations methods, (blue for primary stress, orange for reduced vowels, purple for secondary stress) and invited over a period of six weeks to simultaneously listen to and read short synchronised documents (three minutes each) in the multimedia laboratory, in the self-access centre or at home. The language content of the films had previously been discussed in class so that students approached the activity predisposed to concentrate on form rather than content. Documents used were short video films or audio interviews talking about science at B1, B2 or C1 levels. The documents were available for consultation/downloading from a server. Students were then invited to speculate on the rules governing stress (*e.g.* stress on the penultimate syllable before “-tion” ending) and compare their speculations with a four-page simplified document giving basic rules. As the PC platforms and networks used for this exercise were of varying specifications, the quality of the experience (headphone quality, processing speeds, bandwidths) was also variable. A student questionnaire revealed that students found the use of Swans-generated web

pages easy to access and manage and that the sound quality for the listening activity varied between “good” to “acceptable”.

Laboratory pedagogical exploitation was organised around five activities:

- reading out loud from the unannotated script in pairs, with mutual correction after consulting the on-screen annotated text;
- annotating scripts on paper after listening to the sound, with self-correction via the annotated texts;
- carousel activities: repeated three-minute oral summaries on research topics in constantly changing pairs. The same presentation is made 4 times to different partners each time. The idea of repeated summaries is that after the first presentation the student stops worrying about content during subsequent presentations which frees brain resources for concentrating on spoken form;
- distance teacher correction of student annotated keywords, followed up by “Powerpoint” oral presentations in class using the same keywords;
- EXPLICS case studies where oral performance was continually monitored.

By moving from formal CALL-assisted academic analysis of stress patterns to purposeful, voluntary student organised presentations and case studies, we attempted to construct a context where students would become aware of a need for ear training, of the need to filter out the “fossils” in their oral production whose credibility could easily be undermined by an accumulation of misplaced lexical stress. In other words, we sought to help the students become more autonomous and reflect on their needs. One possible evidence of success in this field was the fact that online consultations of some 60 SWANS-annotated documents available increased towards the end of the course rather than at the beginning as students began to appreciate the reality of the problems and the need to find solutions. When informed that only 10% of 250 fellow students at B1/B2 level knew where to place the primary stress in the word “presidency” some students simply shrugged: “Where’s the problem?”. Convincing students that working on lexical stress should be a priority for many of them is not easy partly because during eight to ten years of previous English courses the subject has rarely been explicitly mentioned.

Testing procedures and student profiles

After this initial training, the following profiles emerged. Tests of perception and oral production revealed striking parallels. If scores were high in perception they were almost always high in oral production too. By contrast, annotation tests, where students were requested to manually annotate a text by underlining stressed syllables and reduced vowels, were far less predictable. Annotation tests included items which were unsegmented (“fantastic”) and pre-segmented (“fan-tas-tic”). Tests of listening perception were based on multiple choice tests in the multimedia laboratory where the students listened to both isolated words and to contextualised items in dialogues before deciding where the lexical stress was placed. Tests of spoken English used computer laboratory student recordings of individual items and of phrases and three-minute

classroom presentations on themes chosen by the students and prepared in advance by the annotation of selected keywords chosen by the students and corrected via e-mail by the teacher. Juries were composed of experienced native speaker (NS) and non-native speaker (NNS) teachers.

Four student profiles

Student profile 1. *New to annotation, perception & production weak*

| | High score | Low score |
|----------------------|------------|-----------|
| Textual annotation | | x |
| Listening perception | | x |
| Spoken production | | x |

Student profile 2. *Has learnt stress pattern rules but has received no ear training*

| | High score | Low score |
|----------------------|------------|-----------|
| Textual annotation | x | |
| Listening perception | | x |
| Spoken production | | x |

Student profile 3. *No experience of annotation but good listening and speaking skills*

| | High score | Low score |
|----------------------|------------|-----------|
| Textual annotation | | x |
| Listening perception | x | |
| Spoken production | x | |

Student profile 4. *Clear evidence of progress or original high level*

| | High score | Low score |
|----------------------|------------|-----------|
| Textual annotation | x | |
| Listening perception | x | |
| Spoken production | x | |

The four student profiles above help understand the distribution of learning problems in this field. What was striking, in France, during initial pre-course testing, was the omnipresent nature of L1 interference. The term “stress deafness” is by no means excessive even if it does not refer to a real physical impediment. Student profile 1 was common among students across the spectrum of previously tested levels, that is to say even relatively fluent students were sometimes judged unintelligible owing to an accumulation of lexical stress errors and their problems were confirmed by low scores on auditory perception.

Student profile 2 was particularly common among students with excellent visual memories, for example the Chinese students, who relied heavily on successful memory of annotated text but who failed to transfer this memory to auditory perception or oral production.

Student profile 3 was also common suggesting that learning the Swans annotation system requires more time for some students. Clearly, previous experience of working with lexical stress and syllable definitions played an important role here. For some, detecting primary stress was simple and intuitive, for others, even syllable segmentation presented problems and a third category recognised the stressed vowel (the nucleus) in the primary stress but had problems with syllabic frontiers (e.g. “significant” where the “g” of the first syllable has been added to the second stressed syllable). Among annotation errors in keywords chosen by the students themselves for their presentations (level B2) we noticed (underlining shows what the student *thought* was a primary stress):

offence, infringement, analysis, linguistics (from a doctoral student studying “linguistics”) economics (from an economist), enigmatic, distressed, paranormal, understanding, incomprehensible, accident, sequence, realisation, development

If the reader reads the above words while putting on a French accent, it’s surprising how “natural” the French stress pattern can sound.

Notice that many of the above annotation errors were not present in the listening perception tests or in the oral production of the students in question. It thus becomes particularly difficult to identify the source of the error. We might add, in this context, that the linguistic behaviour which characterises bilinguals is often of an extreme subtlety in disguising their ignorance or doubts concerning L2 lexical stress rules. They deliberately reduce pitch range and volume of the “difficult” words in question and by removing these crucial acoustic contours hope to escape unnoticed. The technique is no doubt to be found in all espionage manuals. The fact that they are regarded as bilingual is often proof of success in this matter. Their behaviour is not imitated by native speakers who have no qualms (or relatively few) about using a wrong or controversial lexical stress pattern with little known five- or six-syllable words and do so often.

Student profile 4 was relatively rare, suggesting that dealing with fossilised problems of pronunciation after eight to ten years of acoustic reinforcement probably requires more than the ten weeks or thirty hours available. On a more optimistic note, student feedback did suggest a certain placebo effect. By annotating the keywords and studying the corrections, many students really did improve the intelligibility of their presentations and this in turn improved self-confidence. Making a presentation while knowing that some of the worst pronunciation pitfalls have been eliminated apparently boosts morale.

Results

Globally, dual coding improves performance. 75% of students exposed to dual coding over ten weeks scored at least 10% higher than in initial lexical stress recognition tests. 25% showed no change. Student feedback showed the synchronised multimodal experience in SWANS documents was appreciated

and, given the choice between monomodal techniques (text and sound separated) or multimodal techniques (SWANS), students preferred to use SWANS. This was a potentially important finding as the multimodal synchronised experience is theoretically more demanding. Annotations appear to improve short-term memorisation in most cases. Perception of reduced vowels in listening tests and in written tests was clearly more difficult for students than the perception of stressed syllables, suggesting that some secondary school teachers have perhaps neglected this area.

Oral testing based on a three-minute presentation showed significant improvement in controlled conditions for certain individual students and a small global improvement after preparation via keyword annotations. On the other hand, the results of oral testing in the context of subsequent spontaneous conversation did not confirm the idea of improved oral production. “Fossils” returned to the L2 oral performance perhaps because concentrated filtering efforts were reduced and the language content moved away from the controlled area of the presentation. It is noticeable that correct use of L2 stress patterns is closely connected with concentration levels and that any loss of concentration results in a rapid downward spiral in performance quality.

Feedback from teachers using the authoring system

Feedback from teachers using the SWANS authoring system raised fundamental questions about teacher education and how and when it should be provided. Much of the most useful information relating to lexical stress patterns is scattered among academic papers and rarely assembled in practical fashion in teaching manuals. For example –how many words in American English have a different stress pattern from British English (*e.g.* “laboratory” or “adult”) and what kind of tolerance levels apply to those who mix elements of American and British? One recent comparative study gave 960 examples of difference but of course gave no indication on the more subjective question of tolerance.

Annotating texts to suit local needs of student L1 interference often requires intimate knowledge of lexical stress variety. In Czech, Latvian, Hungarian, Swiss German (Bernese dialect), Finnish, and Swahili, stress is always placed on the first syllable. In French, Portuguese, Turkish, and Polish stress is never, or rarely, placed on the first syllable. Many teachers expressed the desire to be warned automatically in advance, before reading a text, about which words might be a cause of deviant perception or production in their students according to their L1.

The generation of an annotated synchronised SWANS web page of some thirty lines with a two- or three- minute video file or audio file, requires no programming knowledge but takes at least ten minutes for experienced language teachers and even more for teachers discovering the nature of lexical stress patterns. The need to make lexical stress pattern annotation automatic rather than manual and to improve networked sharing of finished documents between

European language centres was apparent from the start. Textual annotation, however, raises far wider questions for the fields of teacher education and language policy. Language teachers cannot be expected to speak all the languages of their students but when the rules of two sound systems are mutually exclusive, communicating key elements of such knowledge to students is essential. What cross language knowledge about lexical stress patterns can be considered indispensable? Should trainee teachers be studying “pitch accent” – as in Scandinavian languages, “stress accents” – as in English or Spanish, or “tone variations” - as in Chinese, and what exactly is the difference? Should teachers study deeply the phonotactic¹² constraints of the language they teach or is the current somewhat “superficial” level of knowledge acceptable? Which consonant clusters can be used before a vowel and should teachers be able to answer this question just for their own language or for all the languages used by their students? It is striking that all computer programmers who have developed syllabifiers (automatic syllabification programmes) can answer such questions whereas most teachers cannot. Understanding the origins of interference problems such as epenthesis (the addition of one or two more sounds to a word, *e.g.* “street” pronounced /suto – i – to/ by Japanese students), means raising tolerance levels and thereby improving teaching. Many would argue that today’s language teachers are not properly trained to teach in a multilingual world. A hybrid solution might lie in the development of automatic syllabification annotation systems for all European languages (putting more expert knowledge into machines) while increasing networked, language centre-based programmes for sharing out tailor-made annotated documents among flesh and blood human teachers.

Automatic syllable annotation

The syllable is a useful tool in computational linguistics and prototypes of syllabification programmes exist for many European languages even if no accepted standard algorithm for automatic syllabification exists. If these programmes have not found their way into mainstream language teaching it is largely because of the surrounding complications rather than commercial factors related to the highly competitive automatic speech processing market. According to the circumstances (such as variable rate of enunciation), a given phoneme sequence may be syllabified differently. Rule-based and statistical approaches are common (see Hammond's and Fisher's implementation of Kahn's original research in his well-known 1976 thesis, Hammond 1997) as are data-driven approaches based on machine readable dictionaries or databases of

¹² Phonotactics deals with the permissible combinations of phonemes or syllables in a given language. It defines permissible syllable structures, consonant clusters and vowel sequences and is still relatively ignored in much language teacher training throughout Europe.

known syllabifications. A common rule, little-known to many teachers, that is often cited, is that more consonants are generally present in the onset of a syllable than in its coda; this is known as the maximal onset principle. Such knowledge can be built into a syllabifier.

To help students decide the number of syllables in a word the following steps can be employed:

- count the vowels in the word;
- subtract any silent vowels;
- subtract one vowel from every diphthong;
- the number of vowel sounds left is the same as the number of syllables.

In the case of machines, highly successful syllabification can now be performed either on the orthographic representation of words or on the phonetic transcription of words, according to the particular task required. Schmid presents a statistical approach to syllabification in German and claims an accuracy level by word of 99.85% (Schmid *et al.*, 2007). In the United States, the ToBI framework (Ohio State University) has been used for transcribing the intonation and prosodic structure of spoken utterances in several languages notably English and German. For the French language, Avanzi *et al.* (2007) have used the Dutch programme PRAAT to achieve an effective degree of syllable recognition accuracy. Marchand *et al.* (2009) have developed an algorithm for determining syllable boundaries in the orthographic form of unknown words that works by analogical reasoning from a corpus of known syllabifications. According to the last authors, data-driven techniques outperform the rule-based systems in word and juncture accuracies by a very significant margin in the case of English but require training data and are slower. Adsett & Marchand (2007) argue that this is equally true in the case of Italian which, like Finnish and Spanish, is considered to exhibit simple syllable structure. Their syllabification by analogy procedure achieved 97.70 % accuracy for Italian whereas the best rule-based methods attained 89.77%.

In the case of SWANS, similar prototype automatic syllabification development in English and French is currently based on multiple lexical databases of pre-syllabified words assembled by partner institutions. Progress has been relatively slow in the absence of appropriate software development budgets but the emergence of the syllabification by analogy procedure now holds out hope of an improved rate of progress.

Conclusion

The practice of annotation has a history almost as old as that of writing itself. Modern day computer-based annotation, however, can now go far beyond the supplementary information provided by Talmudic commentary or the modern-day wiki. By animating text, changing its size and colour and adding sound synchronisation, we intervene intimately in the process of decoding and manipulate more thoroughly the thought processes themselves. SWANS uses

typographical annotation in colour and sound to raise awareness of lexical stress patterns and promote noticing skills in the belief that “more noticing leads to more learning” (Schmidt, 1995). The on-screen textual annotation of syllables attempts to achieve, via simultaneous heightened stimulation of eyes and ears, what A. Tomatis (1971) claimed to achieve through more cumbersome and relatively complex auditory manipulations with an “Electronic Ear”: the triggering of a mental control mechanism permitting the readjustment of voice quality (in particular with regards to stressed syllables and weak vowels) and thus a clearer, more comprehensible, L2 vocal production.

Although initial testing of the SWANS authoring system permits a modest “yes” to the question “Can dual coded texts improve perception and production?”, we are convinced that the battle for the eyes and ears of 21st century students will remain an arduous one. Modifying L1 reading practice when tackling an L2 text means fighting deeply engrained habits that have taken years to put in place. The brain’s resistance to novelty is not a sign of conservatism but simply a sign of natural mental health. Choosing the right priority between reading on the one hand and face recognition on the other is not a decision for the brain to take lightly. As Ong puts it, “Freeing ourselves from typographic bias is probably more difficult than any of us can imagine” (Ong, 1982). Only extreme motivation, such as the blind learning Braille with their fingers, can provide quick solutions. Fortunately, new technology holds out increasingly powerful and increasingly nomadic devices to face up to this very real 21st century challenge. As the publishing world prepares to swamp educational markets in the next decade with e-Readers (22 million in 2012), 19% of the US adult population in the case of tablet computers and e-readers (pewinternet.org), it is time for the academic community to lend a hand to transform such devices into genuinely useful tools for reading foreign language texts. Electronic readers for European citizens should offer automatic annotation of stressed syllables and weak vowels for texts in all European languages just as word processing programmes offer spelling correctors. The technology and know-how exist, the finances deserve to be found. The limited nature of traditional European project financing, whereby nine million euros are divided up annually into relatively small and pedagogically unambitious projects, needs to be revised. In the late 20th century the educational world stood by passively while the IBM-led private sector took decades to put spelling correctors in place. The quality of such learning tools can always be called into question but today their fundamental utility is self-evident. Much as teachers are loath to admit, spelling correctors have made the greatest ever contribution to correct written form. We believe that syllable annotation represents an equivalent tool for the acquisition of correct spoken forms. Paradoxically, by dint of massive investment and the profit motive, we have learnt how to make machines more intelligent than our teachers, just as we taught computers to beat Garry Kasparov at chess, but there is nothing inevitable or “natural” about such a process. The

intelligence of text-to-speech synthesis or the intelligence of automatic syllable annotation can be re-engineered to make it understandable and memorable for people, to make our students more intelligent and communicative rather than machines. From the point of view of the human race, there is little doubt that students deserve investing in more than machines. Such a process might not generate the short term profits so eagerly sought by manufacturers but the educational benefits would most certainly be a worthwhile, permanent and very human compensation.

References

- ADSETT, C.R. & Y. MARCHAND. 2010. Syllabic complexity: a computational evaluation of nine European languages. *Journal of Quantitative Linguistics* 17 : 4, 269-290. <<http://dx.doi.org/10.1080/09296174.2010.512161>>.
- ADSETT, C.R. & Y. MARCHAND. 2007. Are rule-based syllabification methods adequate for languages with low syllabic complexity? The case of Italian. 6th International Speech Communication Association ISCA Workshop on Speech Synthesis Bonn Germany pp. 58-63. <http://ebookbrowse.com/adsettmarchand-syllabificationitalian-2007-pdf-d210619860>>.
- AVANZI, M., J.P. GOLDMAN, A. LACHERET-DUJOUR, A. SIMON & A. AUCHLIN. 2007. Méthodologie et algorithmes pour la détection automatique des syllabes proéminentes dans les corpus de français parlé. *Cahiers of French Language Studies* 13 : 2, 2-30.
- BAUMARD, M. 2009. Les étudiants français toujours aussi nuls en anglais. *Le Monde*, 25 août 2009. <http://www.lemonde.fr/imprimer/article/2009/08/25/1231684.html>>.
- BONNET, G. & J. LEVASSEUR. 2004. Évaluation des compétences en anglais des élèves de 15 ans à 16 ans dans sept pays européens in: Note évaluation, 1, s. 1-6. Direction de l'évaluation et de la prospective. <<http://www.education.gouv.fr/stateval>>.
- CHAIT, M., S. GREENBERG, T. ARAI, J. SIMON & D. POEPEL. 2005. Two Time Scales in Speech Segmentation. *Proceedings of the ISCA Workshop on Plasticity in Speech Perception*, University College, London.
- DEHAENE, S. *et al.* 2010. How learning to read changes the cortical networks for vision and language. *Science* (online). <<http://dec.risc.cnrs.fr/Pdf/col/Dehaene.pdf>>.
- DULAY, H., M. BURT & S. KRASHEN. 1982. *Language Two*. New York: Oxford University Press.

- DUPOUX, E., N. SEBASTIAN-GALLES, E. NAVARETE & S. PEPERKAMP. 2007. Persistent stress “deafness”: the case of French learners of Spanish. *Cognition* 106 : 2, 682-706.
- ELLIS, R. 1997. *Second Language Acquisition*. Oxford: Oxford University Press.
- HAMMOND, M. 1997a. Vowel quantity and syllabification in English? *Language* 73, 1-17.
- IVERSON, P., P.K. KUHL, R. AKAHANE-YAMADA, E. DIESCH, Y. TOHKURA, A. KETTERMANN & C. SIEBERT. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87, B47-B57.
- KAHN, D. 1976. *Syllable-based generalizations in English phonology*. Ph.D. dissertation, Cambridge (USA): MIT.
- KUHL, P. K. 2000. A new view of language acquisition. *Proceedings of the National Academy of Sciences* 97 : 22, 11850-11857.
- KUHL, P.K., E. STEVENS, A. HAYASHI, T. DEGUCHI, S. KIRITANI & P. IVERSON. 2006. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science* 9 : 2, 13-21.
- LENNEBERG, E. H. 1967. *Biological Foundations of Language*. Oxford: Wiley.
- MARCHAND, Y., C. R. ADSETT & R. I. DAMPER. 2009. Automatic syllabification in English: a comparison of different algorithms. *Language and Speech* 52 (Pt 1), 1-27. <<http://eprints.ecs.soton.ac.uk/15646>>
- MATTYS, S.L. & J.F MELHORN. 2005. How do syllables contribute to the perception of spoken English? Evidence from the migration paradigm. *Language and Speech* 48, 223-253.
- MAYER, R. E. 2001. *Multimedia Learning*. Cambridge: Cambridge University Press.
- ONG, W. 1982. *Orality and Literacy: The Technologizing of the Word*. London: Methuen. p.72.
- PAIVIO, A. 1986. *Mental Representation: A Dual Coding Approach*. New York: Oxford University Press.
- PAIVIO, A. 1991. Dual Coding Theory: Retrospect and current status. *Canadian Journal of Psychology*, 45(3), 255-287.
- PECHOU, A. & STENTON A., (2002) : Encadrer la médiation – le cas de l’intonation, *Compréhension et Hypermédia*, approches cognitives, communicationnelles et sémiotiques, Albi.
- PEPERKAMP, S. & E. DUPOUX. 2002. Coping with phonological variation in early lexical acquisition. I. LASSER (ed.). *The Process of Language Acquisition*. Frankfurt am Main: Peter Lang Verlag.

- POLIVANOV, E. 1931. La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague, 4: Réunion phonologique internationale tenue à Prague (18-21 décembre 1930)*; Kraus Reprint, 1968, 79-96.
- RIKKE, T. F. 2005. Pourquoi les Français éprouvent-ils autant de difficultés à apprendre l'anglais? Université d'Aarhus, Danemark. <<http://ressources-cla.univ-fcomte.fr/gerflint/Paysscandinaves2/rikke.pdf>>.
- SEBASTIÁN-GALLÉS, N. 2005. Cross-language speech perception. D. B. PISONI & R. E. REMEZ (eds.). *The Handbook of Speech Perception*. Oxford, UK: Blackwell, 546–566.
- SEYMOUR, P.H.K., M. ARO & J.M. ERSKINE. 2003. Foundation literacy acquisition in European orthographies. *British Journal of Psychology* 94, 143-174.
- SCHMID, H., B. MÖBIUS & J. WEIDENKAFF. 2007. Tagging syllable boundaries with joint n-gram models. *Proceedings of Interspeech 2007*, Antwerp, 2857-2860.
- SCHMIDT, R. 1995. Consciousness and foreign language: a tutorial on the role of attention and awareness in learning. R. SCHMIDT (ed.). *Attention and Awareness in Foreign Language Teaching and Learning*. University of Hawaii at Manoa: Second Language Teaching and Curriculum Center Technical Report # 9, 12-55.
- STENTON, A., S. TAZI., & A. TRICOT. 2005a. Document Design for Learning English Pronunciation. The Swan-song of Gutenberg? ED-MEDIA 05, World Conference on Educational Multimedia, Hypermedia and Telecommunication. Montréal, Canada. <<http://www.aace.org/conf/edmedia/default.htm>>.
- STENTON A., A. PECHOU, C.VAILLANT & A.TRICOT. 2005b. Effet du double codage synchrone de l'accentuation en L2 selon des modalités de restitution du sujet. Didcog, Colloque International de didactique cognitive, Université Toulouse 2. <<http://acoustic31.univ-tlse2.fr/didcog/Intro.html>>.
- STENTON, A. 2009. The inhibition of mother tongue interference in foreign language speech perception and production – a proposed solution for European university students. MEDNEZ-VILAS *et al.* (ed.). *Research, Reflections and Innovation in Integrating ICT in Education*, vol 1, 297-303. <<http://www.formatex.org/micte2009>>.
- SWELLER, J. 1999 *Instructional Design in Technical areas*. Camberwell, Victoria, Australia: Australian Council for Educational Research.
- TOMATIS, A. 1971. *L'oreille et le langage*. Paris : Point-Seuil.

- TREIMAN, R., J. MULLENNIX, R. BIJELJAC-BABIC, & E.D. RICHMOND-WELTY. 1995. The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General* 124, 107-136.
- TRUBETZKOY, N.S. 1939. Grundzüge der Phonologie. Travaux du Cercle Linguistique de Prague 7. French translation by J. Cantineau: *Principes de phonologie*. Paris: Klincksieck, 1949.

Software and Internet references

Synchronized Multimedia Integration Language (SMIL 3.0). <<http://www.w3.org/TR/SMIL3>>.

Dinglabs. <<http://www.dinglabs.com>>.

Librivox. <<http://librivox.org>>.

EXPLICS. <<http://www.zess.uni-goettingen.de/explics>>. The aim of the Socrates-financed EXPLICS project is to improve language competence of students by preparing models of best-practice in how to exploit Internet case study and simulation templates and by familiarising language teachers with these models. Participating institutions include universities from 12 European countries. Specialisations include: task-oriented and problem-based learning and teaching; use of global simulations; use of case studies in language teaching; development of language level descriptors and methods of language testing, use of ICT for corpus analysis and concordancing and the use of ICT for language testing.

Acknowledgments

The SWANS project was financed in the context of the CNRS TCAN programme 2003-2005. Our thanks to Nabil Kabbaj, Aryel Beck, Anne Péchou, André Tricot, Saïd Tazi, Nicole Décuré, Pascal Gaillard, Gail Taillefer, Antoine Toma and Christine Vaillant-Sirdey.