Adapted typographical annotations for language learning: user-profiled design solutions to problems of auditory perception and speech production

Anthony Stenton[1], Saïd Tazi[2], André Tricot[3]

[1]LAIRDIL and Université Toulouse I, Département des Langues,Toulouse, France,
anthony.stenton@univ-tlse1.fr
[2]LAAS-CNRS and Université Toulouse I,Toulouse, France, said.tazi@laas.fr
[3]ERT 34 and Laboratoire Travail et Cognition, UMR 5551 CNRS- EPHE,
Université de Toulouse le Mirail,Toulouse, France, andre.tricot@toulouse.iufm.fr

Abstract: The study of English stress patterns is the royal road to improved communication for learners of English. Other pronunciation problems cannot be minimized but many such problems disappear once the lexical stress framework is in place. Learners of English encounter difficulties in perceiving the stressed syllables. As the source of the problem seems related to attention, the dual coding of information on the word using sound and visual annotations is a potential solution. SWANS is an authoring system which enables the semi-automatic generation of multimedia documents in which accents are marked visually and sound is synchronized. This article presents SWANS, the hypotheses that underlie its development and considers the implications for explicit and implicit learning in an adaptive context for multilingual classes.

Keywords : Language learning, perception, typography, annotation, synchronization, authoring systems

1.0 Introduction

The failure of adaptive hypermedia in the field of language learning is partly due to ambitious hopes that expert systems and automatic annotation techniques could handle some of the myriad varieties of individual learning needs. It is also linked to the absence of appropriate authoring systems (Bickerton et al. 1997). This paper presents new reading techniques and new authoring software (SWANS) which address both questions. Neurolinguistic and cognitive psychology research has shown that adult learning paths are not linear but convoluted, eccentric and marked by all kinds of perceptual biases, regressions, errors and unexpected twists. (Houdé 2002). Effective user-profiling and indeed a new pedagogy in this context are desirable but there is still strikingly little evidence that the computer has an important role to play in the process. Recent research into document design for improving

listening perception and speech stress, however, suggests that nationality-based user profiling may constitute a legitimate compromise for genuinely useful adaptation allowing learners to profit from broad categorizations about L2 (second or foreign language) stress patterns which are still relatively unexplored by practising teachers and even less by language learners.

French speakers of English, even at advanced levels, make many highly predictable errors of stress. "Deve**lop**ed" is still stressed on the third syllable, and only sometimes on the second, after 12 years of study. The effect on communication if the number of errors exceeds tolerance thresholds can be destructive. The Anglophone listener 'switches off' and communication breaks down. A new proposed solution is based on experiments with texts specifically annotated for problems of French students. Students of other nationalities are invited to read differently annotated texts according to the L1 (mother tongue). For example, in Czech, Latvian, Hungarian, Swiss German (Bernese dialect), Finnish, and Swahili, stress is always placed on the first syllable. It is regular and predictable. In contrast, for students from France, Turkey, Poland, Spain, or Portugal, it is often crucial to indicate words with first syllable stress as such words are relatively rare in the L1. Weak forms are also annotated as vowel reductions often present difficulties of production and play a key role in perception.

The suggestion that reading could have a negative effect on some aspects of learning may appear paradoxical. In the field of language learning, however, and more particularly in the field of English language learning by the French, it is clear that students are faced with a relatively unusual problem. Most foreign languages are just that: 'foreign'. For the French, the English language is far from foreign. The two languages share between 60 and 70% of the same lexis. The central problem for French learners is that the stress patterns of these words are radically different. English stress is relatively irregular and more often on the left. French stress is regular and more often on the right. For French students of English, reading and subvocalizing an English text is fraught with pitfalls. The eye reads an English word but, more often than not, the inner voice uses a French lexical stress pattern or places stress randomly. Reading may thus become an activity which reinforces errors and inhibits perception.

The hypothesis of L1 interference with perception and production of L2 has been widely studied. In 1931 Polivanov claimed that the phonemic representations of a second language are perceived according to the system of the first language. Although at times correct production can precede correct perception, most analysts still follow Polivanov's lead. Kuhl (2000) suggests that perceptual mapping of L1 speech sounds creates a "complex network or filter, through which language is perceived".

1.1 SWANS
The programme SWANS (Synchronised Web Authoring Notation System) developed by a group of 12 researchers working in four research laboratories in Toulouse, attempts to use synchronisation and enhanced typography to transform the experience of reading and listening. New exercises developed with SWANS tap into the brain's

adaptive capacities. The use of visual stimuli as a potential remedy for negligent auditory perception is possible because all the brain's intelligences are connected. Consequently, the use of dynamic synchronised audiovisual events, more simply, but at times misleadingly, described as "karaoke for language learning", may have implications for memorisation, comprehension and oral production. The implementation of synchronisation and annotation technologies within an authoring system is the basis of the SWANS. SWANS generates web page documents for the student and integrates audio and video materials synchronized with XML-based SMIL tags.

1.2 Synchronizing Solutions

The use of synchronized sound and text has been particularly timid in the field of language learning. This timidity must be linked principally to the absence of suitable tools for synchronization. It is unlikely that synchronization will seriously interest teachers if authoring tools are not simple to learn and rapid to use as the 'Rapido' analysis of authoring systems suggests (Bickerton et al 1997). The captioning of video film is open to the fundamental objection of insufficient reading time. When the dialogue accelerates in the film the script is readable on screen for a few fleeting milliseconds and must be simplified, not to say distorted, to allow readers to keep up. For language learning, the physical separation of script and video film is indispensable to allow the reader to read all the script verbatim and to adjust reading techniques to individual preferences.

1.3 SMIL and SWANS

An important breakthrough for permitting accurate fine-tuned synchronisation (measured in milliseconds) arrived in 1999 with the Synchronised Multimedia Integration Language (SMIL) as a recommendation from the W3C consortium. If the eye can recognise a word in L1 in an eighth of a second (Pinker) and the ear can recognise a word in a fifth of a second then our tools for exploiting such breathtaking accuracy must offer similar performances. Speed reading techniques rely on reducing eye fatigue by flashing text in the middle of the screen thus avoiding the complex and time-consuming calculations needed to focus the eyes on the middle of the next words about to be read. By showing a block of blue colour behind synchronised text, SWANS permits a similar reduction in fatigue (prototype presentation Europe-SMIL, 2003). The eyes are free to scan a highlighted line of text (not usually a sentence but tone units separated by pauses for breath) but are guided to the next line automatically in time with the playback of the sound. Contrary to traditional karaoke methods, this approach avoids the distracting hop from word to word which is intrusive as it troubles the field of vision as the eye scans the line backwards and forwards. During field tests, the central hypothesis tested was one of dual coding. The focal point of explicit learning, the place of the primary and secondary accents, was encoded twice: visually and aurally. According to the theories of Paivio (1986), Meyer (2001) and Sweller (1999) such dual coding should lead to better learning for novice learners and have no effect or even a negative effect on learners who already know the place of these accents.

Figure 2. *Showing text and annotations synchronized in an IE web page using the language SMIL. The blue annotations can be shown or hidden as the user wishes.*

1.4 Authoring synchronised and annotated texts with SWANS

In SWANS, annotation is a set of typographical features added to syllables to show modifications of pitch, amplitude and length while listening to the text. The act of annotation can be divided to three subtasks:

a) The more difficult sub-task in annotating a text in SWANS is to decide where to put on the annotation. For example, dictionaries often indicate where stressed syllables start but leave the user to deduce where they end. Syllables are like mountains - we can all 'see' the summit but identifying the frontiers of the foothills is not always easy. Development underway of an expert system (based on Deschamps and Guierre) and a dictionary data base should accelerate the process by offering semi-automatic annotations which the user can validate or modify.

b) The second subtask consists of choosing the best typographical feature (size, kerning and colour), which is adapted to the contextual substring, according to the pedagogical goal.

c) The last subtask consists of applying the annotation to the substring.

Synchronising text with audio allows simultaneous listening and reading. Synchronising is accomplished via 'Magpie', a freeware from NCAM that generates time codes and produces output in SMIL format.

2.0 The evaluation of SWANS

Informal testing of SWANS began in 2004 with the generation of over 40 synchronized documents mainly from a bank of three-minute video news items. A considerable increase in speed of web page production was observed. The time taken

to generate a web page of some 30 lines of annotated and synchronized text was cut from 2 hours (computer programmers using 'Dreamweaver' under guidance from language teachers) to 10 minutes (language teachers alone using Magpie and SWANS).

2.1 The evaluation of Dual Coding

Testing took place after presentations on screen using word lists which were annotated to show stress and linked to sound. Control groups were exposed to the same words with no annotation and the same sound. Students were requested to indicate the place of the tonic accent in two different written exercises and to record their own pronunciation of chosen words on the computer. Our initial hypothesis, based on teaching practice, was that dual coding of primary stress would have a positive effect on novice students whether the feedback demanded was written or oral. This point of view was at the origin of the SWANS authoring system. The corollary of the first hypothesis was that 'expert' students would show no improvement or deterioration in performance called 'expert reversal effect'.

2.1.2 Participants

Testing included 64 second year undergraduates who had experienced relatively intensive training in pronunciation (hence the label 'experts'). A second group of 52 undergraduate students tested at the start of the 1st year were called the "intermediate" group as they had received no specific tuition on stress. Finally we were able to compare these results with those of "novices" - 50 pupils from a local high school (14 to 15 years of age) who had never, or hardly ever, studied tonic accents.

2.1.3 Results.

Academic Level	Student production mode / Coding mode	A. written	B. written	C. oral
Experts (end of 2nd year students)	1. audio coding	9,26	8,69	7,55 *
	2. dual coding	7,45	6,50	5,19
Intermediate (start 1st year students)	1. audio coding	8,23	6,23	7,36
	2.. dual coding	8,37	7,11	7,17 *
Novices (High school)	1. audio coding	5,89	4,41	7,42
	2. dual coding	7,62	4,87	7,14

Table 1 : The effect of different coding methods on average performance out of 10 according to academic level and to student production modes .[1]

Globally the dual coding of the stressed syllable improves the performance of the novices and deteriorates that of the experts for the written exercise. By contrast, for the spoken exercise (where students recorded their own pronunciation on the

---

[1] The asterisk * indicates groups with less than10 individuals and thus statistically less significant.

computer), the dual coding has a negative or little effect on performance. Average performance in the written exercises is between 5 and 9 out of 10. The second year university students have scores which are 0.5 points higher than those of 1st year students, who are themselves 2 points higher than the high school pupils. The dual coding improves the performance of the novices (by 1.1 points) and that of the intermediate students (by 0.5 points) while it has a negative effect (down 2 points) on that of the experts.

3.0 Discussion

The results of our study are encouraging because they show a positive effect for dual coding. The fact that this effect is not present in the oral exercise is perhaps due to insufficient data or an insufficient time lapse between presentation and testing.

Our results need to be completed. An experiment underway is now attempting to replicate these results with more complex and more convincing material from a didactic point of view (from word lists, we pass to discourse at the sentence level).

In time, we should be able to offer teachers an innovative tool for the generation of video web pages with annotated, synchronized scripts where tonic accents will be highlighted visually and whose effectiveness for language learning in terms of recognizing and producing appropriate stress will be attested through studies on novices at university and high school level.

4.0 Conclusion

Testing recognition of stress at the word level constitutes a prudent first step towards a much larger exploration of new intensive, and perhaps also extensive, reading techniques. Computer-based document design has slavishly copied the Gutenberg tradition since its inception. The arrival of SMIL in 1999 has brought the real potential plasticity of our electronic environments to the attention of developers. Automatically generated textual annotations using colour, animation, flashing letters, changes in size and spacing offer a new way forward for multimodal research. Whether visual memory can really be enlisted to improve oral production or not, we are convinced that such experimentation will facilitate mental representations and raise awareness of perception problems for both teachers and learners. By adapting annotations to predicted errors of production in French, Spanish, Italian, or Arabic speaking students, teachers in Toulouse are currently sharing their intuitions and expertise with novice teachers. Raising awareness through annotation extends explicit knowledge of English stress patterns. The essential question becomes one of implicit learning and the inhibition of the L1 tendency which is also related to the psychological constraints of speaking a foreign language.

The problem of coping with stress in oral production of learners has an authoring parallel in the problem of persuading and training teachers to experiment with a typographical revolution which, in time, will no doubt link sound and appropriate animations automatically to every word of every electronic text. The challenge to that 5,400-year-old cultural artifact - the alphabet - (can it really be true that the letter 'A'

is an icon representing the inverted head of an ox?) has not arrived with a flash and a blare of trumpets but is part of a permanent quest to redefine the segmentation of, or the labels we give to, the experience of our senses. Like most evolutions, it will undoubtedly be slow and erratic and may well require several decades to defeat the resistance of ingrained habits. As Ong says, "freeing ourselves from typographical conditioning may be more difficult than we imagine" (Ong 82). SWANS should be seen not as method for manipulating the minds of students but simply as an authoring tool for offering more choice in the variety of new adaptive reading techniques we can now place at their disposal.

References

1. Bickerton D., Ginet , A.Stenton , A., Temmerman M., Vaskari T., (1997) (Final Report of the RAPIDO Project, 91 pp., February 1997, University of Plymouth, UK (SOCRATES Project TM-LD-1995-1-GB-58).
2. Deschamps A.(1994), De l'écrit à l'oral et de l'oral à l'écrit, Orphyrs.
3. Guierre L. (1987), Règles et exercices de prononciation anglaise, Paris A. Colin-Longman.
4. Houdé O. (1998) "Développement cognitif " in Vocabulaire de sciences cognitives, Paris, PUF., Houdé, Olivier, Mazoyer, Bernard, Tzourio-Mazoyer, Nathalie, PUF- Presses Universitaires de France.
5. Kuhl P., (2000).A new view of language acquisition, *Proceedings of the National Academy of Science*, 97, 11850-11857.
6. MAGPIE (2004) National Centre for Accessible Media, http://ncam.wgbh.org/webaccess/magpie/ (consulted march 2006).
7. Mayer, R. E., (1983) Thinking, problem solving, cognition, W. H. Freeman and company, New York.
8. Ong, W. (1982): Orality and Literacy: The Technologizing of the Word. London: Methuen.
9. Paivio, A. (1991b). Dual Coding Theory: Retrospect and current status. Canadian Journal of Psychology, 45(3), 255-287.
10. Pinker, S. (1984) Language Learnability and Language Development, Harvard University Press.
11. Polivanov, E. (1931) "La perception des sons d'une langue étrangère" *Travaux du Cercle Linguistique de Prague* 4; in *Le Cercle de Prague* (Change, 3) Paris, 1969.
12. SMIL 2.0 Synchronized Multimedia Integration Language (SMIL 2.0) Specification W3C Working Draft 21 September 2000". W3C SYMM Working Group. Available at: http://www.w3.org/TR/smil20/ (consulted may 2006).
13. Stenton A., N. Kabbaj, S.Tazi (2003) : Remedial work for English Pronunciation with a smil(e) Europe-SMIL Conference, ENSAM, Paris.
14. Sweller, J., (1999) Instructional design in technical areas. Camberwell: ACER Press.