

---

# SWANS, un système auteur de synchronisation et d'annotation pour un apprentissage multimodal des phénomènes accentuels en langue vivante 2.

## Partie 2 : mise en œuvre informatique et évaluation empirique

**Aryel Beck<sup>?</sup>, Nicole Décuré<sup>?</sup>, Pascal Gaillard<sup>?</sup>, Nabil Kabbaj<sup>?</sup>, Anne Péchou<sup>?</sup>, Anthony Stenton<sup>?</sup>, Saïd Tazi<sup>?</sup>, Antoine Toma<sup>\*</sup>, André Tricot<sup>¶</sup> et Christine Vaillant-Sirdey<sup>?</sup>**

<sup>?</sup> Université Toulouse I, UFR Informatique ; Place Anatole France – 31 042 Toulouse Cedex [aryelbeck@yahoo.fr](mailto:aryelbeck@yahoo.fr) et [Kabbaj@insa-tlse.fr](mailto:Kabbaj@insa-tlse.fr)

<sup>?</sup> Laboratoire Interuniversitaire de Recherche en Didactique des Langues (LAIRDIL) Université Toulouse III, Département des Langues et Civilisations; Université Toulouse I Place Anatole France – 31 042 Toulouse Cedex [{anthony.stenton, anne.pechou}@univ-tlse1.fr](mailto:{anthony.stenton, anne.pechou}@univ-tlse1.fr)

<sup>?</sup> « Hypermédias et apprentissages » ERT 34 - IUFM de Midi-Pyrénées et Laboratoire Jacques Lordat, IFR 96 « Sciences du Cerveau » et Université de Toulouse le Mirail, 5 Allées Antonio Machado - 31 058 Toulouse cedex [pascal.gaillard@univ-tlse2.fr](mailto:pascal.gaillard@univ-tlse2.fr)

<sup>¶</sup> « Hypermédias et apprentissages » ERT et Laboratoire Travail et Cognition, UMR 5551 CNRS – EPHE – Université de Toulouse le Mirail [andre.tricot@toulouse.iufm.fr](mailto:andre.tricot@toulouse.iufm.fr)

<sup>?</sup> LAAS-CNRS Avenue Colonel Roche 31077 Toulouse Cedex 4, et Université Toulouse I. [tazi@laas.fr](mailto:tazi@laas.fr)

<sup>\*</sup> LAIRDIL, Université Toulouse III, Département du Monde Anglophone, Université de Toulouse le Mirail, [toma@univ-tlse2.fr](mailto:toma@univ-tlse2.fr)

<sup>?</sup> LAIRDIL, IUT Université Toulouse III, 115 Route de Narbonne -31077 Toulouse Cedex 4. [christine.vaillant-sirdey@iut-tlse3.fr](mailto:christine.vaillant-sirdey@iut-tlse3.fr)

---

**RÉSUMÉ :** Cet article fait suite à un premier article, qui présente notre problématique didactique et étaye notre approche. Les apprenants francophones ayant de grandes difficultés à encoder et à produire l'accent tonique en anglais, notre projet est de développer un système auteur qui permette à des enseignants de générer semi-automatiquement des documents multimédias où la syllabe accentuée est codée visuellement et auditivement. Nous présentons ici l'architecture

*informatique de ce système auteur SWANS. Nous présentons aussi l'évaluation empirique d'une maquette, ce qui nous permet de confirmer qu'il y a bien un effet positif du double codage de la syllabe accentuée sur sa mémorisation. Cependant, d'autres résultats sont nécessaires pour que notre approche soit complètement validée.*

*MOTS-CLÉS : accent tonique, prosodie, apprentissage de l'anglais, didactique des langues, double-codage, multimédia, multimodalité, système auteur*

---

## **1. Introduction**

Dans un premier article, nous avons montré qu'apprendre la place de l'accent tonique était un enjeu didactique important. Les apprenants francophones n'encoderaient pas l'accent tonique en anglais. Ils ont par conséquent du mal à mémoriser la place de l'accent tonique et à le produire ou à le reproduire, ce qui a un effet négatif sur la compréhension de leur discours par des interlocuteurs anglophones.

Les solutions visuelles informatiques semblent constituer une bonne solution. Le repérage visuel aide à la discrimination des parties du discours où se posent des problèmes de perception et de compréhension. Le son prend alors une forme visuelle observable dans le temps à l'inverse de sa matérialité première constituée par d'éphémères vibrations de l'air. Voir ce que l'on entend est loin d'être un gadget [CAZADE 99] comme il était permis de le supposer de prime abord. Mais il faut parvenir à ce que cette visualisation soit synchrone, peu coûteuse à générer par des utilisateurs comme les enseignants, mais aussi non interférente et peu coûteuse pour les apprenants. Notre projet est donc de créer un système auteur qui permette à tout enseignant de générer les documents où l'accent tonique est marqué visuellement valider expérimentalement cette approche.

Les marqueurs paralinguistiques permettent de mettre en exergue des aspects du texte sur lesquels on veut attirer l'attention du lecteur [CELLIER et al. 01, 02]. Si un étudiant lit et entend simultanément le mot *de*velopment, où la syllabe accentuée est mise en exergue visuellement, traiter ce mot consiste pour lui en un encodage auditif d'informations linguistiques, un encodage visuel d'informations linguistiques et un encodage visuel d'informations paralinguistiques, ici le changement de couleur et un pas plus grand. Il est donc possible pour le concepteur de dispositifs d'enseignement de coder deux fois (auditivement et visuellement) un aspect sur lequel il veut attirer l'attention de l'étudiant. Selon la théorie du double codage de [PAIVIO 86], celle de l'apprentissage multimédia de [MAYER 01] ou celle de la charge cognitive de [SWELLER 99], un tel double codage devrait améliorer l'apprentissage chez des apprenants novices.

On peut en revanche s'attendre à un effet nul voire négatif chez des sujets connaissant déjà la place de l'accent. Cet effet a été mis en évidence par [KALYUGA et al. 03] et nommé *expertise reversal effect* : ce qui facilite l'apprentissage chez les novices peut gêner le traitement chez les experts. Ce renversement de l'effet positif

du double codage s'expliquerait par la relation entre mémoire à long terme et mémoire de travail. Le double codage, en enrichissant l'information à traiter, permet un codage par deux canaux au lieu d'un seul, et par conséquent un renforcement de l'encodage chez un sujet peu habitué à ce type d'encodage, ne disposant pas en mémoire à long terme des connaissances correspondantes. A l'inverse, chez l'individu qui possède déjà ces connaissances en mémoire à long terme, c'est-à-dire un expert, le double codage constitue une modalité inhabituelle de présentation de l'information, qui vient interférer avec la mobilisation des connaissances en mémoire à long terme.

## **2. Développement du système**

La démarche originale de ce projet est d'associer une modalité de représentation graphique et/ou typographique à la modalité orale de l'accentuation. L'association de style typographique à la qualité du discours parlé a été peu explorée. Notre objectif dans un premier temps est de comprendre ces associations et de trouver des conventions satisfaisantes dans le domaine de l'enseignement des langues. Pour cela, l'objectif à long terme du système auteur SWANS est de servir d'atelier d'expérimentation pour trouver les meilleures représentations typographiques (ou graphiques) possibles adaptées aux différents contextes pédagogiques.

Nous allons faire le point sur les techniques informatiques d'association d'une représentation graphique (ou typographique) à la modalité orale du texte, sachant qu'à part le système *Prosodic Font* [ROSENBERG & MACNEIL 99] qui associe une présentation dynamique de polices de caractères au signal de sa prononciation, très peu ont fait des associations d'une représentation typographique à une mélodie sonore. Il existe deux types de solution pour associer une représentation graphique à la prononciation orale d'un discours : celle qu'on qualifiera d'automatique et celle qu'on qualifiera d'interactive. Des solutions intermédiaires combinent l'interactivité et l'automatisation, comme celle qu'on propose avec SWANS.

### **2.1. Les solutions automatiques**

Une solution automatique consiste à mettre en entrée un discours oral (fichier audio, ou vidéo) et à générer une représentation graphique des sonorités, traduites sous formes de graphiques, et/ou de texte, avec des variations typographiques correspondant à l'intonation et la prosodie. Cette solution si elle présente l'avantage de décharger complètement l'auteur d'intervenir sur le texte ou sur la partie sonore, présente l'inconvénient de ne pas satisfaire complètement ce qu'on souhaite faire c'est-à-dire pouvoir choisir parmi plusieurs représentations possibles. Elle ne donne pas complètement satisfaction car la génération d'un graphique partir du signal n'est pas bien admise comme meilleure solution didactique pour l'enseignement de la prononciation [CAZADE 99].

*Prosodic Font* est un système développé au Média Lab du MIT [ROSENBERG & MACNEIL 99] dans le but de générer des polices de caractères dynamiques qui varient avec le temps et les variations de sonorité d'un discours parlé. Le but est de générer des textes animés en fonction de l'intonation et la prosodie du discours. L'avantage d'une telle solution est qu'on est capable de voir le texte avec des polices variées se générer au fur et à mesure que le discours parlé évolue. L'inconvénient est que l'auteur n'a pas la maîtrise de ce qui est restitué.

## **2.2. Les solutions interactives**

Les solutions interactives des systèmes auteurs consistent à permettre à l'auteur de choisir sa propre présentation d'une part, et de décider où il souhaite mettre l'accent d'autre part. L'auteur peut également décider d'agir sur certaines parties du discours afin de varier l'accentuation ou la durée des pauses pour des raisons pédagogiques. Le plus grand avantage des solutions interactives est de permettre à l'auteur de faire varier les présentations en fonction du type de matériel utilisé, ou en fonction du niveau des apprenants. Ces variations se font par l'intermédiaire d'annotations insérées manuellement sur des morceaux de textes sélectionnés. L'inconvénient majeur de telles solutions est le temps mis par l'auteur pour annoter le texte restitué. Toutefois l'efficacité de tels systèmes ne devraient pas être mesurée seulement en fonction du temps passé à créer le matériel pédagogique. Il faut mesurer l'efficacité en termes de la fonction pour laquelle le système est créé : améliorer l'apprentissage. L'efficacité concernant la communication d'un document annoté pour améliorer sa lecture et son accessibilité a été démontré [TAZI 05].

## **3. SWANS**

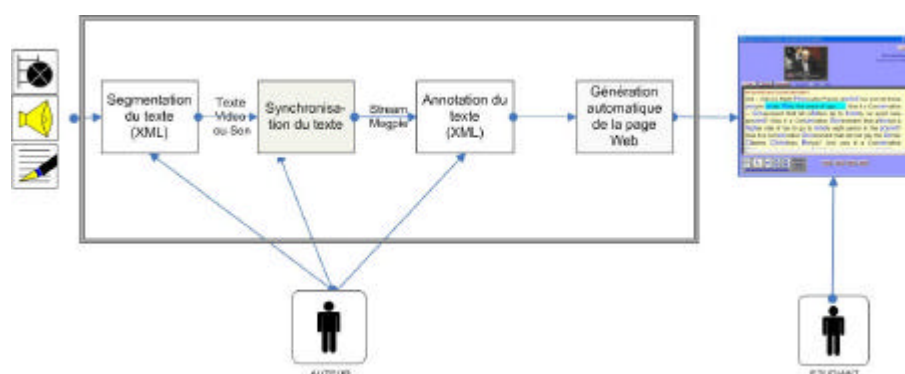
SWANS (*Synchronised Web Authoring Notation System*) est un système auteur développé dans le but d'automatiser le processus de création de documents multimédia pour aider à accélérer les processus de synchronisation et d'annotation par des enseignants non informaticiens. Le but de ces documents est l'apprentissage multimodal de l'anglais comme seconde langue. A long terme, permettra aux professeurs de choisir de nouveaux styles d'annotation afin d'adapter et d'améliorer l'apprentissage. Dans l'état actuel du système, il est considéré comme un système auteur décrit dans la section suivante.

### **3.1. Architecture**

Le système SWANS est composé de cinq modules tels qu'ils sont présentés dans la Figure 1. Ces modules permettent de faire les opérations suivantes :

- Importation de texte et média dans l'environnement de travail. Cette fonctionnalité permet d'importer différents formats de média.

– Segmentation du texte selon les unités de souffle, et balisage avec des codes Xml. La segmentation permet aussi bien de faire un découpage du texte en fonction de la ponctuation qu'en fonction de la longueur de la phrase. Cette étape permet à l'utilisateur de gagner du temps sur l'étape de synchronisation en automatisant une grande partie du découpage du texte. Il faut noter que cette étape est semi-automatique et ce volontairement afin que les utilisateurs puissent intervenir sur ce découpage.



**Figure 1.** Architecture de SWANS

– Synchronisation du texte et des médias par l'insertion de la durée de chaque unité de souffle. Pour cela on utilise dans l'état actuel du système les fonctionnalités du système MagPie [MAGPIE 04], qui permet de synchroniser le texte avec le média en insérant des balises SMIL dans le texte afin de distinguer les différentes unités de souffle. Le texte est préparé de manière à ce qu'il soit lu sous forme de Karaoké. Après avoir synchronisé le texte, l'application effectue un contrôle de cohérence et met le fichier synchronisé au format SWANS c'est-à-dire une unité de souffle suivie de deux temps : le temps de début d'élocution de la phrase et le temps de fin qui est nécessaire pour la précision du Karaoké. Le passage par un format SWANS qui est en fait un ensemble de balises Xml, permet à l'utilisateur de rester libre quant au logiciel qu'il va utiliser pour synchroniser le texte, et nous permettra d'ajouter notre propre système de synchronisation.

– Annotation des textes en fonction des accents et de l'intention des auteurs concernant l'objectif pédagogique. Dans l'état actuel du système, cette fonctionnalité est exécutée manuellement de manière interactive. Toutefois, nous envisageons de l'automatiser en partie afin de réduire la charge de l'auteur. La modélisation des règles de prononciation de la langue anglaise à partir de différentes recherches effectuées dans le domaine de la phonologie et de la graphématique faciliterait cette semi-automatisation. Cette partie permet aussi optionnellement d'annoter l'intonation de la phrase et donc s'occuper de la prosodie. Une automatisation totale de l'annotation réduirait la liberté des choix des stratégies pédagogiques.

– Génération de pages XHTML+SMIL [W3C 02] prêtes à l'emploi. Cette fonctionnalité permet de visualiser le document tel qu'on le voit dans la Figure 2. D'autres possibilités d'interface sont envisageables. Cette partie commence dès que l'utilisateur dispose d'un texte au format SWANS, elle va servir à saisir les paramètres nécessaires à la génération de la page Web, à savoir le chemin d'accès au media, le chemin d'accès au script synchronisé (cette saisie est automatique si on a réalisé l'étape précédente) le temps de début de lecture et le temps de fin de lecture (des valeurs par défaut sont proposées) enfin il reste à choisir la taille des caractères dans une liste et donner un titre à la page web. Un bouton génère la page Web, lance le traitement et une fois que la page est créée elle est affichée avec le navigateur par défaut, en l'occurrence Internet explorer si l'on travaille sous Windows.

### 3.2. Les fonctionnalités de SWANS

Le Logiciel SWANS génère une suite de pages Web qui assurent les fonctionnalités suivantes :



**Figure 2.** Interface générée par SWANS

- Ecouter un son et/ou voir une séquence vidéo, par le professeur ou l'apprenant ;
- Lire le texte correspondant à une séquence audio ou vidéo ;
- Lire et écouter d'une manière synchronisée avec la possibilité d'afficher le texte annoté ;

– Effectuer des tests de mémorisation sur l'accentuation. Cette fonctionnalité est destinée aux apprenants.

L'application génère quatre types de pages : une page avec le texte seul non annoté permettant l'écoute et la lecture, une avec le texte synchronisé avec le son ou la vidéo permettant aussi de visualiser les annotations phonétiques, une page éventuelle si l'utilisateur a annoté l'intonation permettant de voir le texte avec l'intonation annotée et enfin une page avec un exercice de prononciation avec six phrases du texte dans laquelle l'étudiant pourra s'exercer et vérifier qu'il a appris à placer l'accentuation.

La Figure 2 présente une page avec synchronisation et annotation. Les boutons « *Text* » et « *Accent* » permettent de basculer entre texte nu et texte annoté.

### ***3.3. XHTML+SMIL au service de l'enseignement des langues***

SMIL [W3C 01] est un langage qui permet d'élaborer des présentations multimédias et de synchroniser leurs éléments. C'est un langage qui utilise une grammaire simple et rigoureuse basée sur une DTD XML, il est normalisé par le W3C (World Wide Web Consortium). Le XHTML+SMIL intègre une partie des spécifications de SMIL 2.0 dans XHTML. Il inclut les modules de SMIL 2.0 qui supportent les animations, les objets multimédia, le contrôle, la temporisation et la synchronisation du contenu ainsi que les effets de transitions. Les caractéristiques de SMIL 2.0 sont intégrées directement dans XHTML et CSS et peuvent être utilisées pour manipuler leurs caractéristiques. Ce langage est conçu pour des clients Web qui supportent le balisage XHTML+SMIL tel que Internet Explorer 5.5 ou plus.

```

<STYLE>
t\:* {BEHAVIOR: url(#default#time2)}
.time { behavior: url(#default#time2)}
.highlight {BACKGROUND-COLOR: cyan;}
</STYLE>
<t:par id=main begin = "indefinite" fill = "freeze">
<t:video src = "parl.mpg" syncBehavior = "locked" clipBegin="19.0s"
clipEnd="115.0s"> </t:video>
<p id="regtext" class="time" timeAction="none" timeContainer="seq">
<span class="time" dur = "3.4s" timeAction = "class:highlight"> Will...Will my
Right <a id="acc1">Hon</a>orable friend </span>
<span class="time" dur = "3.4s" timeAction = "class:highlight">re
<a id="acc2">mind</a>
me and all those <a id="acc3">peo</a>ple</span>
<span class="time" dur = "5.55s" timeAction = "class:highlight">
<a id="acc4">un</a>der <a id="acc5">thir</a>ty five years of age .....</span>
.....
</p>
</t:par>

```

**Figure 3.** Code XHTML+SMIL généré par SWANS

Nous montrons dans la Figure 3 le code généré qui permet de synchroniser le son et le texte en présentant le texte sous forme de Karaoké. On remarque qu'il y a différents types de balises, celles qui servent à la synchronisation et celles qui servent à la présentation des différentes annotations.

#### 4. Tests et évaluations

Les évaluations ont été conduites pendant le développement de SWANS, sur du matériel simple (des listes de mots) généré « à la main » sur l'environnement informatique d'apprentissage des langues LAVAC [TOMA 93]. Il s'agit d'évaluer l'efficacité du double codage de la syllabe accentuée sur sa mémorisation en fonction du niveau d'expertise des apprenants.

##### 4.1. Hypothèses

Les résultats empiriques que nous avons rapportés dans les parties 1 et 2 ainsi que la pratique d'enseignement de l'anglais de la part de plusieurs d'entre nous, nous permettent de défendre le point de vue suivant : le double codage de l'accent tonique devrait avoir un effet positif sur sa restitution par des élèves et des étudiants peu performants. Cet effet devrait être observé quelle que soit la modalité de restitution du mot : écrite et orale. Ce point de vue est à l'origine du système auteur SWANS. Il nous conduit à formuler des hypothèses : le double codage de la syllabe accentuée améliore les performances des novices ; il n'améliore pas ou il détériore les

performances des experts ; ce double effet inverse s'observe avec les modalités de restitution écrites et orales ; les performances seront meilleures pour une tâche de rappel écrit que pour une tâche de rappel orale.

Nous contrôlons la présence d'un indice (fort versus faible) pour les modalités de restitution écrite : la présence d'un indice fort dans une tâche de rappel améliore normalement les performances. Ces hypothèses ont été testées empiriquement dans une étude que nous rapportons ici.

#### **4.2. Participants**

Les participants sont trois groupes de niveau scolaire différent. Les deux premiers groupes sont des étudiants en deuxième année de l'IUT GEA de Toulouse. Le premier groupe (les experts ci-après) est composé de 64 étudiants formés pendant plusieurs mois à percevoir, à placer et à produire l'accent tonique. Le deuxième groupe (les intermédiaires) est composé de 52 étudiants « non formés ». Ces deux groupes sont simplement distincts par le fait qu'ils ont réalisé l'expérience en début d'année (avant formation à la perception de l'accent tonique) ou en fin d'année (après formation). Le troisième groupe (les novices) est composé de 50 élèves du collège Château de l'Hers de Toulouse (classes de 4<sup>ème</sup> et 3<sup>ème</sup>). Les élèves de collège travaillent pas ou très peu sur l'accent tonique.

#### **4.3. Matériel**

Trois séries de 10 mots anglais sont présentées aux participants sur LAVAC. Chaque liste est présentée sur un écran d'ordinateur, dans une colonne, à gauche de l'écran. En cliquant sur l'un des dix mots, l'étudiant entend ce mot en même temps que les mots s'affichent au centre de l'écran. Le mot reste affiché jusqu'à ce que l'étudiant clique sur un autre. Chaque mot peut être cliqué autant de fois que l'étudiant le souhaite. Deux variables sont contrôlées : le nombre de syllabes et la transparence lexicale. Les mots sont tri ou quadrisyllabiques, transparents (leur quasi-homophone en français désigne le même mot : *national* = national) ou non-transparents. La répartition de ces deux caractéristiques (nombre de syllabes, transparence) est équivalente dans les 3 listes. Chaque liste de mots est présentée de trois façons différentes :

– modalité 1 dite «codage audio seul » (de la syllabe accentuée) : le mot est affiché au centre de l'écran et entendu ;

– modalité 2 dite «double codage » : identique à la modalité 1 mais l'accent tonique est marqué visuellement (sur la syllabe accentuée le pas est plus grand et la couleur est différente) ;

– modalité 3 dite « complexe » : identique à la modalité 2 mais d'autres phénomènes acoustiques linguistiques sont marqués visuellement (avec une 3<sup>ème</sup>

couleur), et la couleur de fond change (blanche pour les modalités 1 et 2, elle est noire avec la modalité 3). Cette modalité ne sera pas traitée ici.

#### **4.4. Protocole et mesures**

Les trois groupes de participants sont divisés en trois sous-groupes. Chaque sous-groupe travaille sur une seule modalité de présentation des 30 mots. La restitution est évaluée, pour chaque participant selon trois modalités successives (pour les étudiants de l'IUT) ou selon deux modalités successives (pour les élèves de collège). Chaque modalité de restitution correspond à une liste de 10 mots.

Les trois modalités de restitution sont trois épreuves de rappel. Les deux premières sont écrites tandis que la troisième est orale. Parmi les deux épreuves écrites, la première offre un indice « faible », tandis que la seconde offre un indice « fort ». L'épreuve orale ne comporte aucun indice :

– modalité A : l'étudiant entend un mot. Il doit cocher sur une feuille de réponse reproduisant la liste des dix mots, en face du mot correspondant, la syllabe accentuée. L'indice fourni est le nombre de syllabes du mot.

– modalité B : l'étudiant entend un mot. Il doit cocher sur une feuille de réponse reproduisant la liste des dix mots, en face du mot correspondant, la syllabe accentuée. L'indice fourni est un découpage syllabique du mot.

– modalité C, l'étudiant entend un mot. Il doit prononcer correctement chacun des dix mots de la liste. Ses réponses sont enregistrées sur l'ordinateur.

La succession des trois modalités est contrôlée, à l'intérieur de chaque groupe. Les réponses sont codées de façon binaire : la réponse est correcte si le ou les accents toniques sont correctement placés ; la réponse est incorrecte si un accent n'est pas correctement placé. La performance maximale est donc de 10 points, la minimale de 0 point. Pour l'évaluation de la modalité C, les cas les plus évidents sont traités « à l'oreille » par deux linguistes de l'équipe (l'une française, l'autre anglais). Pour l'analyse des cas ambigus, l'équipe a employé le logiciel *Speech Analyzer* pour afficher la courbe de la fréquence fondamentale où les chutes sur les syllabes sont visibles et mesurables.

#### **4.5. Résultats**

Les résultats (Tableau 1) sont globalement conformes à nos hypothèses. Nous devons cependant regretter une perte importante de traces des performances des participants pour la modalité de restitution orale (perte due à un problème matériel). Les résultats devront être confirmés avec des enregistrements complémentaires.

Globalement, le double codage de la syllabe accentuée améliore les performances des novices et détériore les performances des experts pour les restitutions à l'écrit. En revanche, pour la restitution orale, le double codage semble détériorer ou n'avoir pas d'effet sur les performances. L'effet sur les performances du groupe des

intermédiaires est moins net. Nous procédons maintenant à une analyse détaillée des résultats.

Niveau scolaire	modalités de restitution modalité d'encodage	A. écrit indice faible	B. écrit indice fort	C. oral
Experts (IUT fin d'année)	1. codage audio seul	9,26	8,69	7,55*
	2. double codage	7,45	6,50	5,19
Intermédiaires (IUT début d'année)	1. codage audio seul	8,23	6,23	7,36
	2. double codage	8,37	7,11	7,17*
Novices (collège)	1. codage audio seul	5,89	4,41	7,42
	2. double codage	7,62	4,87	7,14

**Tableau 1.** Effet de la modalité d'encodage sur les performances moyennes en fonction du niveau scolaire et des modalités de restitution (la marque \* indique les groupes comportant moins de 10 individus, peu exploitables donc).

Pour la première modalité de restitution, l'écrit avec indice faible, on observe un effet simple du niveau scolaire ( $F(2;110) = 11,6$  ;  $p < 0,001$ ), pas d'effet simple de la modalité d'encodage ( $F(1;111) = 0,005$ ) et un effet d'interaction entre le niveau scolaire et la modalité d'encodage ( $F(2;110) = 12,3$  ;  $p < 0,001$ ). Le double codage améliore la performance des novices, détériore celle des experts et ne semble pas influencer celle des intermédiaires.

Pour la deuxième modalité de restitution, l'écrit avec indice fort, l'ensemble des résultats à la même allure : effet simple du niveau scolaire ( $F(2;109) = 29,5$  ;  $p < 0,001$ ), pas d'effet simple de la modalité d'encodage ( $F(1;110) = 0,7$ ), effet d'interaction entre le niveau scolaire et la modalité d'encodage ( $F(2;109) = 9,5$  ;  $p < 0,001$ ). Le double codage améliore (peu) la performance des novices et (nettement) celle des intermédiaires, tandis qu'il détériore (très nettement) celle des experts.

En calculant une performance moyenne des deux modalités de restitution écrite, on peut proposer un résumé simple des résultats. Les performances moyennes sont comprises entre 5 et 9 points sur 10. Les étudiants en fin de seconde année d'IUT ont de meilleures performances que ceux en début d'année (avec une différence moyenne de 0,5 points), qui ont eux-mêmes de meilleures performances que les élèves de collège, de 2 points en moyenne ( $F(2;126) = 33,4$  ;  $p < 0,001$ ). Le double codage améliore la performance des novices (de 1,1 points) et celle des intermédiaires (de 0,5 points), tandis qu'il détériore (de 2 points) celle des experts ( $F(2;126) = 14,1$  ;  $p < 0,001$ ).

Pour la troisième modalité de restitution, l'oral, les résultats sont différents et obtenus avec un nombre trop faible de participants. On n'observe pas d'effet simple du niveau scolaire ( $F(2;72) = 1,9$ ), mais un effet simple de la modalité d'encodage ( $F(1;73) = 4,1$  ;  $p < 0,05$ ), et un effet d'interaction entre le niveau scolaire et la modalité d'encodage non-significatif ( $F(2;72) = 2,5$  ;  $p = 0,08$ ). Le double codage

détériorer (peu ou pas) la performance des novices et celle des intermédiaires, tandis qu'il détériore (très nettement) celle des experts.

Les performances sont meilleures pour la tâche de rappel écrit avec indice faible (8,17 points sur 10 en moyenne) que pour celle avec indice fort (6,77 points),  $t(1;93) = 6,4$  ;  $p < 0,001$ . Il n'y a pas de différence entre la performance moyenne à l'écrit et la performance à l'oral ( $t(1;72) = 0,2$ ), ce qui est contraire à notre hypothèse.

#### **4.6. Discussion**

Nous avons formulé quatre hypothèses.

- Le double codage de la syllabe accentuée améliore les performances des novices. Cette hypothèse est vérifiée à l'écrit mais pas à l'oral et la conclusion peut être étendue au groupe intermédiaire, celui des étudiants en début de 2<sup>ème</sup> année, avant leur formation au traitement de l'accent tonique.

- Le double codage n'améliore pas ou il détériore les performances des experts. Cette hypothèse est vérifiée et même précisée : le double codage détériore très nettement les performances des étudiants de fin de 2<sup>ème</sup> année, ayant bénéficié de la formation au traitement de l'accent tonique.

- Ce double effet inverse s'observe avec les modalités de restitution écrites et orales. Cette hypothèse n'est que partiellement vérifiée, puisque l'effet de renversement est observé uniquement à l'écrit ; à l'oral, seule la détérioration des performances des experts est observée.

- Les performances sont meilleures pour une tâche de rappel écrit que pour une tâche de rappel orale : cette hypothèse est infirmée, puisqu'il n'y a pas de différence observée.

Les résultats de notre étude sont encourageants parce qu'ils montrent un effet positif du double codage. Cet effet n'est malheureusement pas obtenu à l'oral, mais cette absence d'effet est peut-être simplement due aux données perdues. Si ce résultat décevant venait à être confirmé (effet sur la restitution écrite mais pas sur la restitution orale), nous devrions nous interroger un éventuel effet de non-congruence entre modalité d'encodage et de restitution : en codant visuellement l'accent tonique on obtient un effet positif sur sa restitution visuo-graphique, mais pas sur sa restitution orale.

Le fait que la présence d'un indice fort à l'écrit détériore la performance des étudiants est peut-être dû à un biais de construction de notre matériel et à la façon dont nous avons codé les performances. En effet, la liste de mots utilisée pour tester la modalité de restitution B (indice fort) comportait trois mots avec deux syllabes accentuées (comme la modalité de restitution C – orale) alors que la modalité de restitution A ne comportait aucun mot avec deux syllabes accentuées. Comme nous notions 0 point pour une réponse incorrecte (0 ou 1 syllabe accentuée correctement sur deux), la tâche à réaliser avec la modalité B (et C) est objectivement plus difficile que la tâche correspondant à la modalité A.

## 5. Discussion générale et perspectives

Nos résultats nécessitent d'être complétés. Une expérimentation en cours tente de répliquer les effets obtenus avec un matériel plus complexe et plus pertinent d'un point de vue didactique (de la liste de mots, nous passons au discours de quelques phrases).

A terme, nous devrions donc être capables de proposer aux enseignants un outil permettant de générer des documents multimédias audio-visuels avec textes annotés, où la syllabe accentuée serait mise en exergue visuellement, et dont l'efficacité sur l'apprentissage du repérage et de la production de l'accent tonique serait attestée auprès d'un public d'élèves et d'étudiants novices.

### Remerciements

Ce projet est financé par le programme TCAN du CNRS ; nous remercions Mme Sylvie Dualé et ses élèves du collège du Château de l'Hers pour leur participation.

## 6. Bibliographie

- [CELLIER et al. 01] Cellier, J.-M., Terrier, P. Le rôle de la mise en forme matérielle dans le traitement cognitif de consignes. *Langages*, n° 141, 2001, p.79-91.
- [CELLIER et al. 02] Cellier, J.-M., Lemerrier, C., Mojahid, M., Terrier, P. A propos de la fonction sémantique de la mise en forme dispositionnelle du texte, *Colloque Inscription Spatiale* du Langage : structures et processus, Toulouse, 29 – 30 Janvier 2002.
- [KALYUGA et al. 03] Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. The expertise reversal effect. *Educational Psychologist*, n°38, 2003, p. 23-31.
- [MAYER 01] Mayer, R.E., *Multimedia learning*, Cambridge University Press, New York, 2001.
- [PAIVIO 86] Paivio, A. *Mental representations*, Oxford University Press, New York, 1986.
- [SWELLER 99] Sweller, J., *Instructional design in technical areas*, ACER Press, Camberwell, 1999.
- [TAZI 05] Tazi S. Improving access to and reuse of educational documents, *International Journal of Interactive Technology and Smart Educations*, sous presse.
- [TOMA 93] Toma, T. *Lavac*, 1993 [application informatique, C puissance 3 Informatique, Morgan Multimedia]

### 6.1. Références sur le WEB.

[W3C 02] <http://www.w3.org/TR/2002/NOTE-XHTMLplusSMIL-20020131/>

[W3C 01] <http://www.w3.org/TR/SMIL2/>

[MAGPIE 04] <http://www.thirdwishsoftware.com/magpie.html>