

SWANS : Synchronised Web Authoring Notation System

Peut-on inhiber l'interférence de la langue maternelle dans la perception et la production d'une L2 ? L'utilisation de la synchronisation et de l'annotation typographique adaptée à l'apprentissage dans les centres de langues plurilingues.

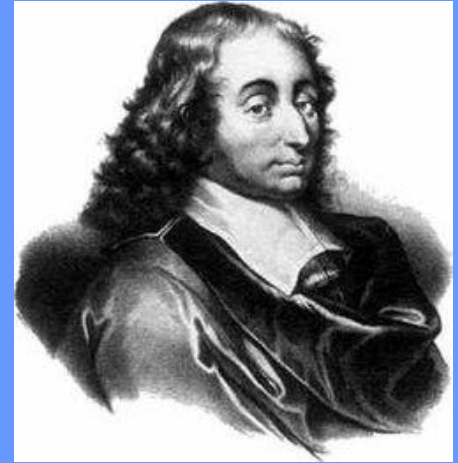
L'alphabet, la musique, la mathématique

- Qui a inventé le système sténographique[1] employé pour écrire rapidement les discours de Cicéron et toujours pratiqué mille ans après ?
- Qui a inventé les techniques de transcription nécessaires pour les partitions musicales ?[2]
- Qui se rappelle du nom du génial inventeur du zéro[3] ?

Des pionniers inconnus ?

- [1] Marcus Tullius Tiro, 63 avant J.C., (source : Cambridge Encyclopedia of Language p. 206)
- [2] On attribue l'invention d'une portée de quatre lignes à Guido d'Arezzo ou Guido Aretinus. Ce moine bénédictin s'est fait connaître à la cathédrale d'Arezzo, entre Sienne et Florence, comme professeur de musique. c.990–1050.
- [3] Son usage moderne, à la fois comme chiffre et comme nombre, est héritée de l'invention indienne des chiffres [nagari](#) vers le [Ve siècle](#). Le mot indien désignant le zéro était [ṣūnya](#) (sūnya), qui signifie «vide» «espace» ou «vacant». Le mathématicien et astronome indien [Brahmagupta](#) est le premier à définir le zéro dans son ouvrage Brâhma Siddhânta. Ce mot, traduit par les Arabes en « sifr », ce qui signifie «vide» et «grain», est la racine des mots chiffre et zéro. La graphie du zéro, d'abord un cercle, est inspirée de la représentation de la voute céleste.

Inventeurs de représentations symboliques



Intuitions

- L'accentuation lexicale en anglais représente un problème important pour les français et en particulier pour les étudiants non spécialistes de l'enseignement supérieur
- Ce problème est plus grave pour les français que le même problème chez les espagnols, les italiens ou les allemands.
- Les conséquences d'une mauvaise accentuation lexicale sont graves car l'interlocuteur peut 'décrocher' après quelques fautes rapprochées.

Intuitions

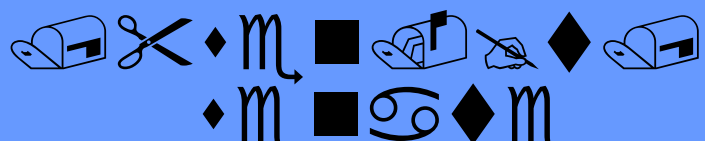
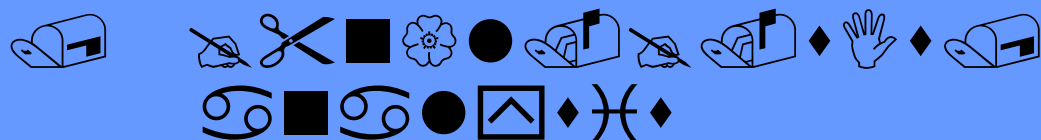
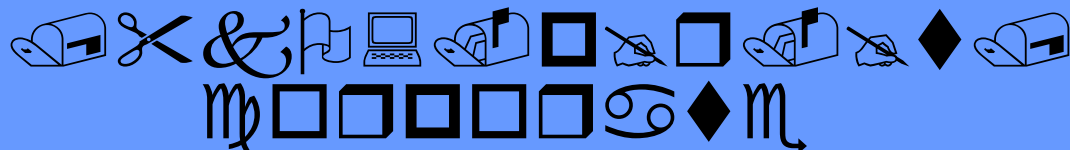
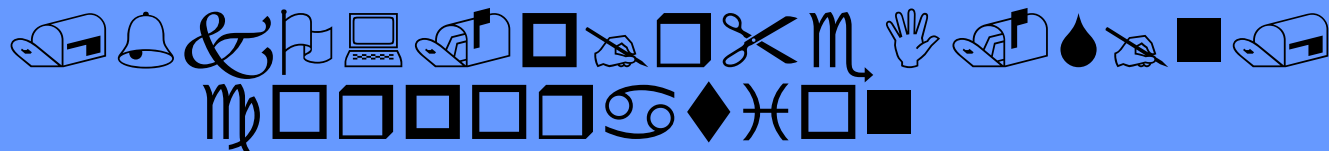
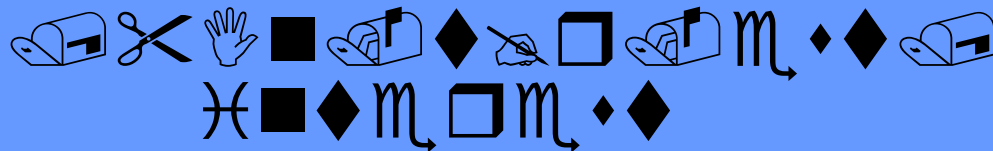
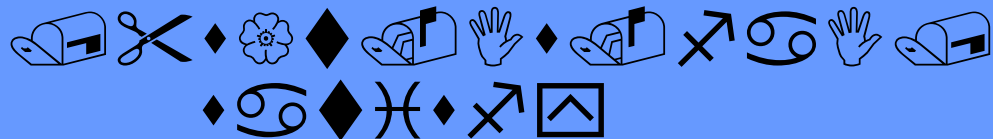
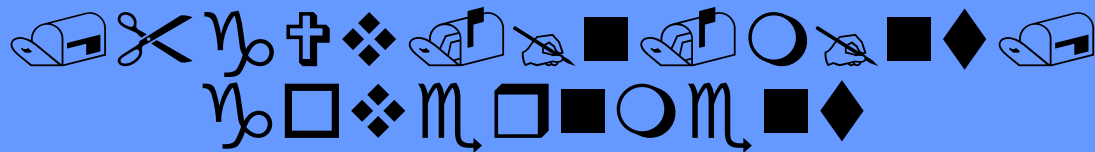
- Le problème de mauvaise accentuation est sous-estimé par les enseignants d'anglais francophones car ils arrivent à communiquer en anglais avec grande facilité malgré leur propre formation superficielle dans ce domaine.
- Le problème est sous-estimé par les enseignants natifs en France car ils sont habitués à entendre certaines déformations
- Le problème est sous-estimé par les étudiants français car assez souvent par politesse les interlocuteurs anglophones ne signalent pas les difficultés de compréhension.

Un problème de négligence auditive

- En anglais dites-vous :
- « de-**vel**-oped » ou « de-vel-**oped** » ?
- « ne-**cess**-ar-y » ou « **ne**-cess-ar-y » ?
- 50% des étudiants français hésitent même après 10 ans de cours d'anglais, 10 % savent où se place l'accent dans « pre-si-den-cy »



70% to 90% of 250 Toulouse
students got these accents wrong



Lié au problème auditif : le problème des exposés en anglais des étudiants français au niveau avancé

- Fluide
- Bon débit
- Langage assez approprié
- Mais.....

Accentuation aléatoire

- La plupart des étudiants francophones ne stockent pas d'informations concernant des schémas accentuels de façon consciente. Lors des exposés, l'accent sera placé de façon aléatoire sur la deuxième ou la troisième syllabe du mot anglais
“ de**ve**lopment ” indiquant non pas une ‘surdité’ ou un problème de production mais une *négligence* au niveau du stockage.

Un problème français mais pas toujours francophone ?

- Le babillage des tout-petits de France et de Québec présente deux différences fondamentales associées au français québécois et au français de France.
- *"Chez les Français, les syllabes d'un même mot ont toutes la même durée, sauf la dernière, qui est allongée. Cette isochronie est observable dans le babillage vers l'âge de 14 mois. Toutefois, ceci n'est pas une caractéristique du français québécois et on ne l'observe pas dans le babillage des bébés au Canada."* (Maneva, 2001)

Are the problems with English stress patterns the same for all nationalities?

- Of course not
- In Czech, Latvian, Hungarian, Swiss German (Bernese dialect), Finnish, and Swahili, stress is always placed on the first syllable.
- In French, Spanish, Portuguese, Turkish, and Polish stress is rarely placed on the first syllable.

Adaptive annotations

- Texts should be annotated in several different versions to suit the needs of students of different nationalities. Students whose mother tongue does not rely on a high incidence of alternate strong and weak forms need particular help from visual annotation for improving perception and oral production of English.

L'effet sur la communication

- **Le seuil de tolérance est bas.**
- Cinq fautes d'accentuation rapprochées peuvent bloquer la communication.
- L'interlocuteur « décroche »
- Les Français ne font pas d'effort pour mémoriser les schémas accentuels car l'accentuation est prévisible en français

La visualisation de la parole.

Quelle méthode pour améliorer la perception auditive ?

- Courbes de Fréquence fondamentale (f_0)
- Courbes d'intensité ou volume
- Représentations spectrales
- ou Annotation textuelle ?

L'oreille perçoit globalement

- Les visualisations séparées de l'intensité, F0 et de la durée ne donnent pas une bonne image de la perception auditive
- Les apprenants ont du mal à lire ces informations car elles sont complexes et difficiles à interpréter.
- L'alignement du texte et les courbes pose problème

Annotation textuelle : méthode Robertson circa 1850

8

THE VICAR OF WAKEFIELD.

with the world, or each other.

We had an elegant house, situated

in a fine country, and a good

neighbourhood. The year was

spent in moral or rural amuse-

ments; in visiting our rich neigh-

bours, and relieving such as were

poor. We had no revolutions to

fear, nor fatigues to undergo; all

our adventures were by the fire-

La Méthode Robertson

TABLEAU DES SIGNES EMPLOYÉS POUR FIGURER LA PRONONCIATION.

SIGNES PRINCIPAUX.

1	égale A	dans ma nière.
2	= E	» de stin.
3	= I	» mi sère.
4	= O	» mo ment.
5	= OU	» mo ulin.
6	= EU	» seu lement.
7	= CH	» ch ou.
8	= TCH	» caou tch ouc.
9	= K	» k ermès.

SIGNES ACCESSOIRES.

— son grave et long.

◌ son aigu et bref.

◌ son très-ouvert :

Ces trois signes indiquent ••
outre la place de l'accent.

⌒ diphthongue.

• consonne douce ou relative

⌒ réunion de voyelles.

Brazil 1994

5.1

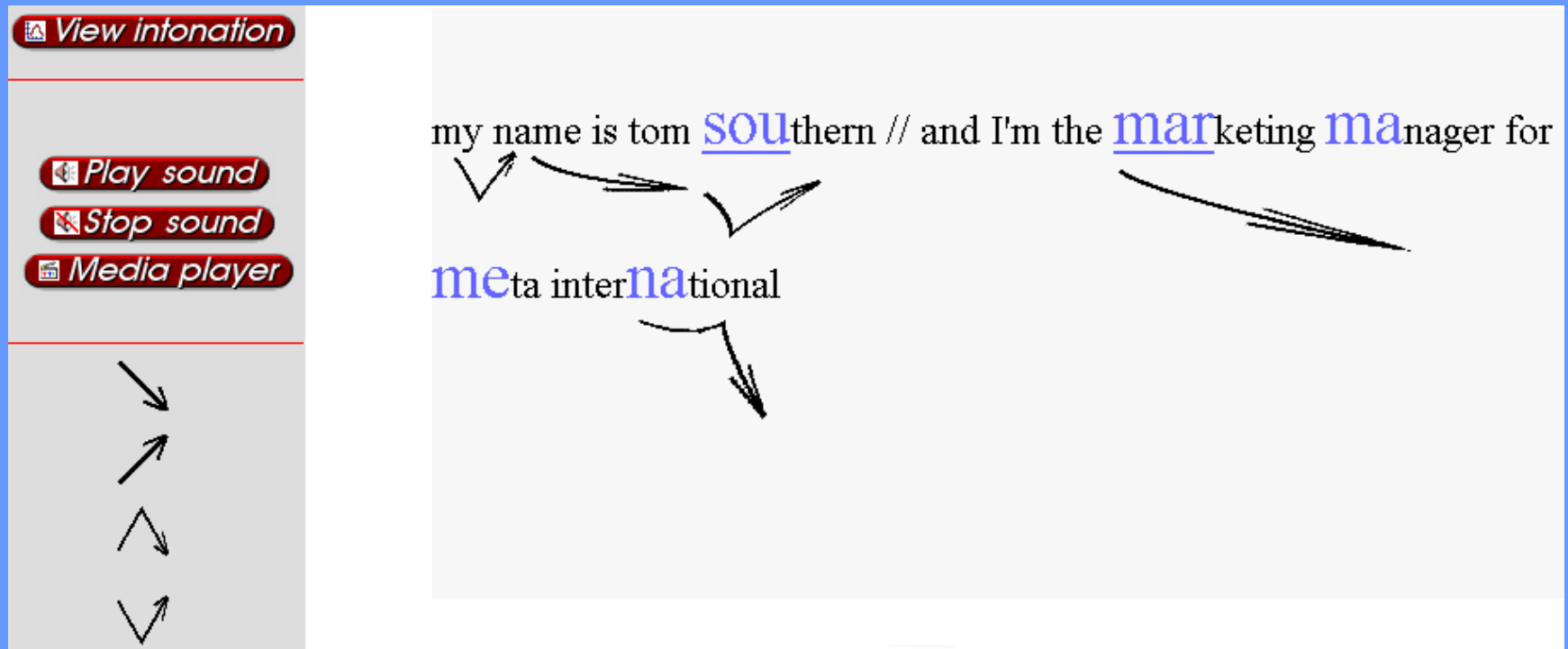
A rising tone is used in both versions of // GOOd EVening // etc.

2a // ♫ beFORE // ↗ i introduce tonight's SPEAker // → there's ER // ♫ ONE //
↗ important reMINder //

3a // ♫ NEXT month's // ↗ MEEting // → will BE // → OUR // ♫ ANNual GENeral
meeting //

4a // → AND er // ↗ on that oCCAsion // → we're HQping for // → a GOOD // → and
SPIrited // ↗ aTTENdance //

Une innovation : l'intégration de textes et courbes.
glisser/déposer de flèches extensibles de *Sounds Right*
(Stenton& Péchou 2002)



Synchronisation avec SMIL

- L'arrivée du langage SMIL (Synchronised Multimedia Integration Language) en 1999 et les travaux du groupe de travail du W3 Consortium - 'Timed-Text Initiative' (2003) - constituent une percée importante. Le langage SMIL permet la **synchronisation fine** du son et du texte avec annotation à l'écran.

Effet de la synchronisation

- La synchronisation ligne par ligne (et non pas mot par mot) signale un nouveau confort oculaire
- Les yeux balayent la ligne pendant l'écoute du son

double codage avec SMIL.

La mémoire visuelle au secours de la mémoire auditive défaillante

TCAN - Microsoft Internet Explorer

Fichier Edition Affichage Favoris Outils ?

Précédente Recherche Favoris OK Liens icq

Adresse C:\Documents and Settings\All Users\Documents\BBC Temp\BBC2.HTML

BBC2.HTML

Start End Play Paused 13.0

Text Accent



Politicians may be worried and businesses around the world unhappy but foreign tourists in New York can barely contain their delight at the weak dollar. It means they get to stay in posh hotels, eat in the best restaurants and shop to their heart's content, all without huge credit card bills when they get back home. The BBC's Emma Simpson has been hitting the tourist trail. New York at Christmas there's nothing quite like it. This is high season from Manhattan. So this month a night at the Waldorf will cost you around \$600. Pricey, but here they're always busy. This year, though, foreign visitors are up by a third.

Permet de visualiser les accents ou de voir le texte "nu"

Permet de suivre le son

Terminé Poste de travail

Techniques d'annotation testées (1)

- 'Nue' : butterfly

- 'Simple' 

- 'Complexe' = arrière plan noir



Techniques d'annotation testés (2)

- texte ordinaire : couleur noire , taille 18
ex : «university »
- accent primaire : couleur bleue, taille 22
ex : 'ver' dans « uni**ver**sity »
- accent secondaire : couleur rouge, taille 20
ex : 'u' dans « **u**niversity »
- forme faible (le 'schwa') : couleur orange, taille 14
ex : 'a' dans « a**g**ree**a**ble »

Techniques d'annotation testées (3)

Hem, there is a **C**ertain,
definitely
a **C**ertain
re**a**lity to that.
There is a very ...
di**St**inct **d**ifference
be**tween** what
American **b**usinesses,
and ... **A**mericans in general,
ex**p**ect from their
government,
and what French **b**usinesses

Annotation automatique et manuelle

- **L'automatisation de l'annotation** : l'ajout automatique d'annotations au texte pour indiquer accents primaires et secondaires
- **L'annotation manuelle** : l'ajout par l'enseignant d'annotations au texte pour indiquer :
 - les accents primaires et secondaires
 - les voyelles neutralisées, voyelles « faibles »



Cimabue 1300

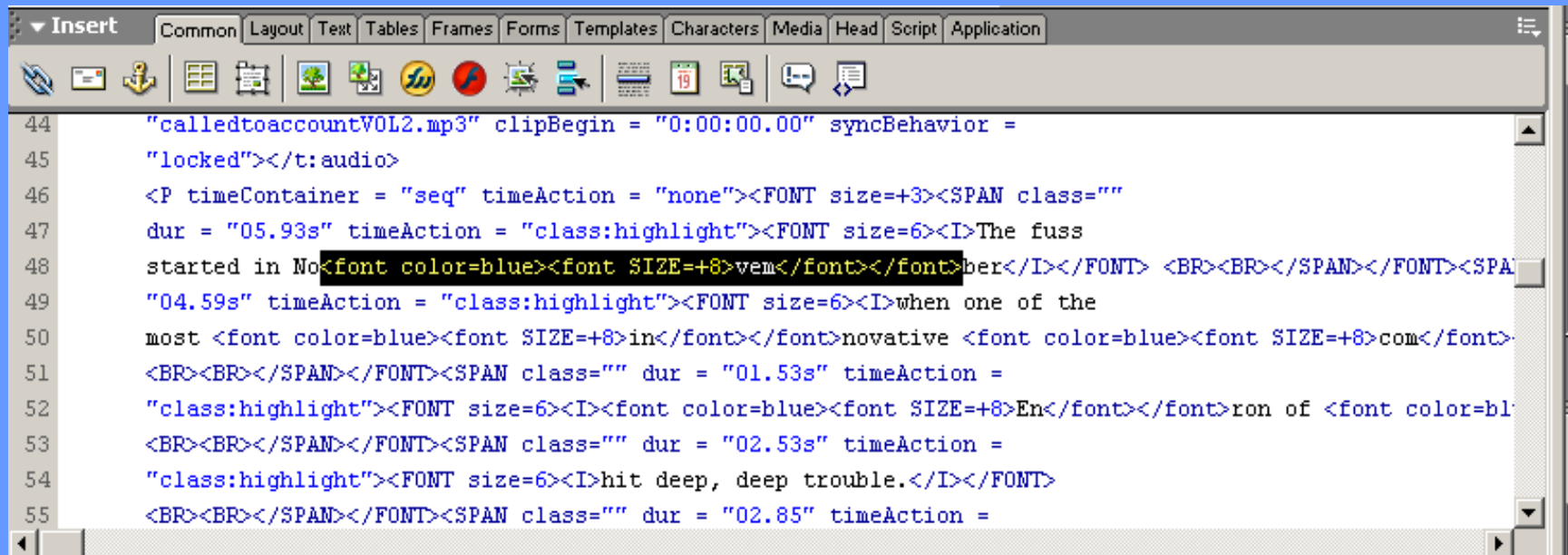


Rogier van der Weyden 1450

Projet: un système auteur pour accélérer l'annotation des textes

- La synchronisation et l'annotation d'une page XHTML avec Dreamweaver et Magpie est laborieuse.
- 'A la main' il faut compter 2 heures pour annoter et synchroniser un texte court (30 lignes) avec un film ou un fichier son et les transférer dans une page web.

Edition avec Dreamweaver



```
44 "calledtoaccountVOL2.mp3" clipBegin = "0:00:00.00" syncBehavior =
45 "locked"></t:audio>
46 <P timeContainer = "seq" timeAction = "none"><FONT size=+3><SPAN class=""
47 dur = "05.93s" timeAction = "class:highlight"><FONT size=6><I>The fuss
48 started in No<font color=blue><font SIZE=+8>vem</font></font>ber</I></FONT> <BR><BR></SPAN></FONT><SPA
49 "04.59s" timeAction = "class:highlight"><FONT size=6><I>when one of the
50 most <font color=blue><font SIZE=+8>in</font></font>novative <font color=blue><font SIZE=+8>com</font>
51 <BR><BR></SPAN></FONT><SPAN class="" dur = "01.53s" timeAction =
52 "class:highlight"><FONT size=6><I><font color=blue><font SIZE=+8>En</font></font>ron of <font color=bl
53 <BR><BR></SPAN></FONT><SPAN class="" dur = "02.53s" timeAction =
54 "class:highlight"><FONT size=6><I>hit deep, deep trouble.</I></FONT>
55 <BR><BR></SPAN></FONT><SPAN class="" dur = "02.85" timeAction =
```

Called to Account : the Enron Scandal BBC February 2002

```
<?import namespace = vid implementation = "file:///C:/enroncollected/audiovideoSCRIPTS/cp.
vid:cp id=cp target="main">
```

*The fuss started in Nov***em***ber*

*when one of the most***in***novative* **com***panies in* **Ame***rica*

Projet CNRS : TCAN

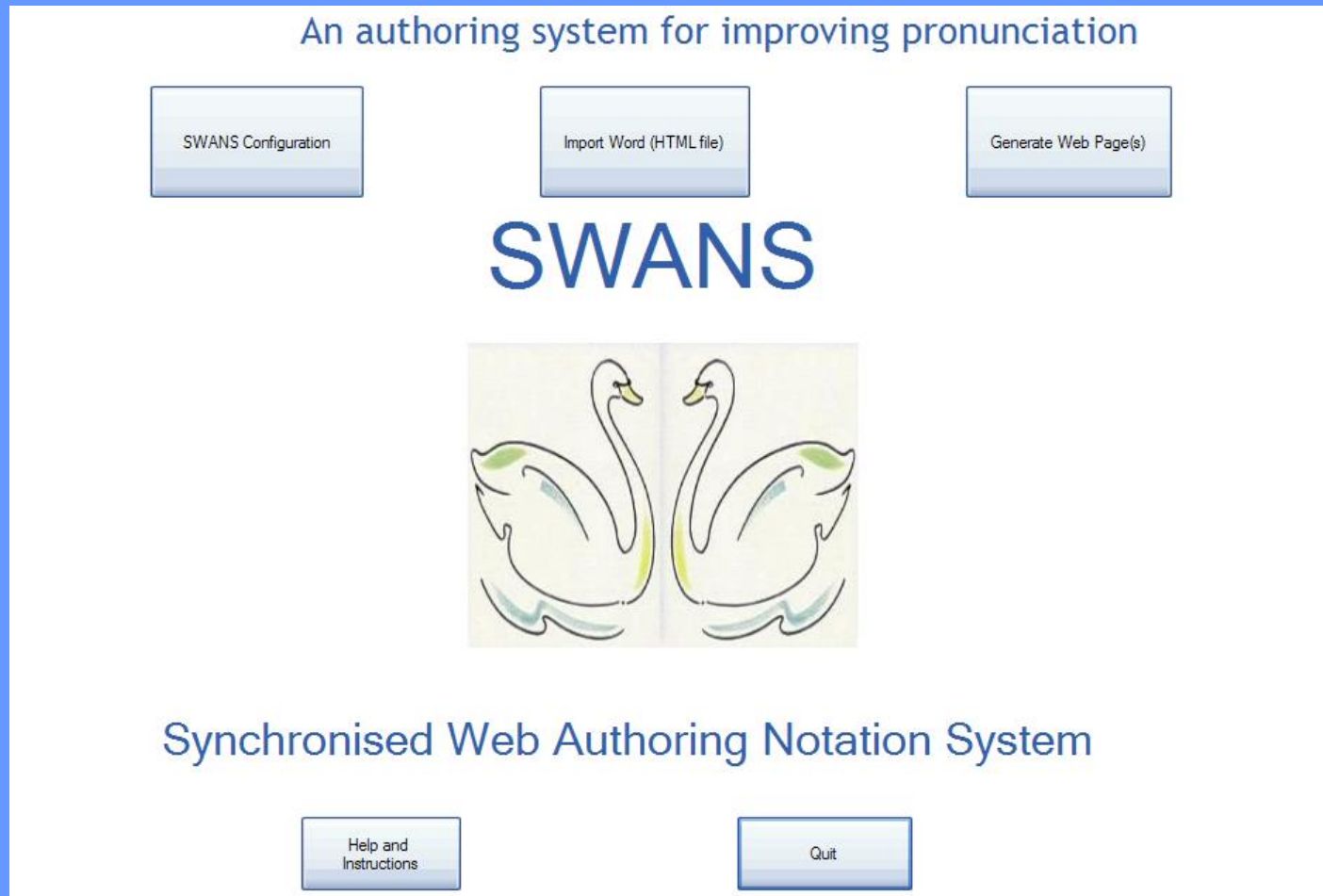
Traitement Cognitif Apprentissage et NTIC

- Le développement d'un prototype pour améliorer la perception auditive des schémas accentuels de l'anglais

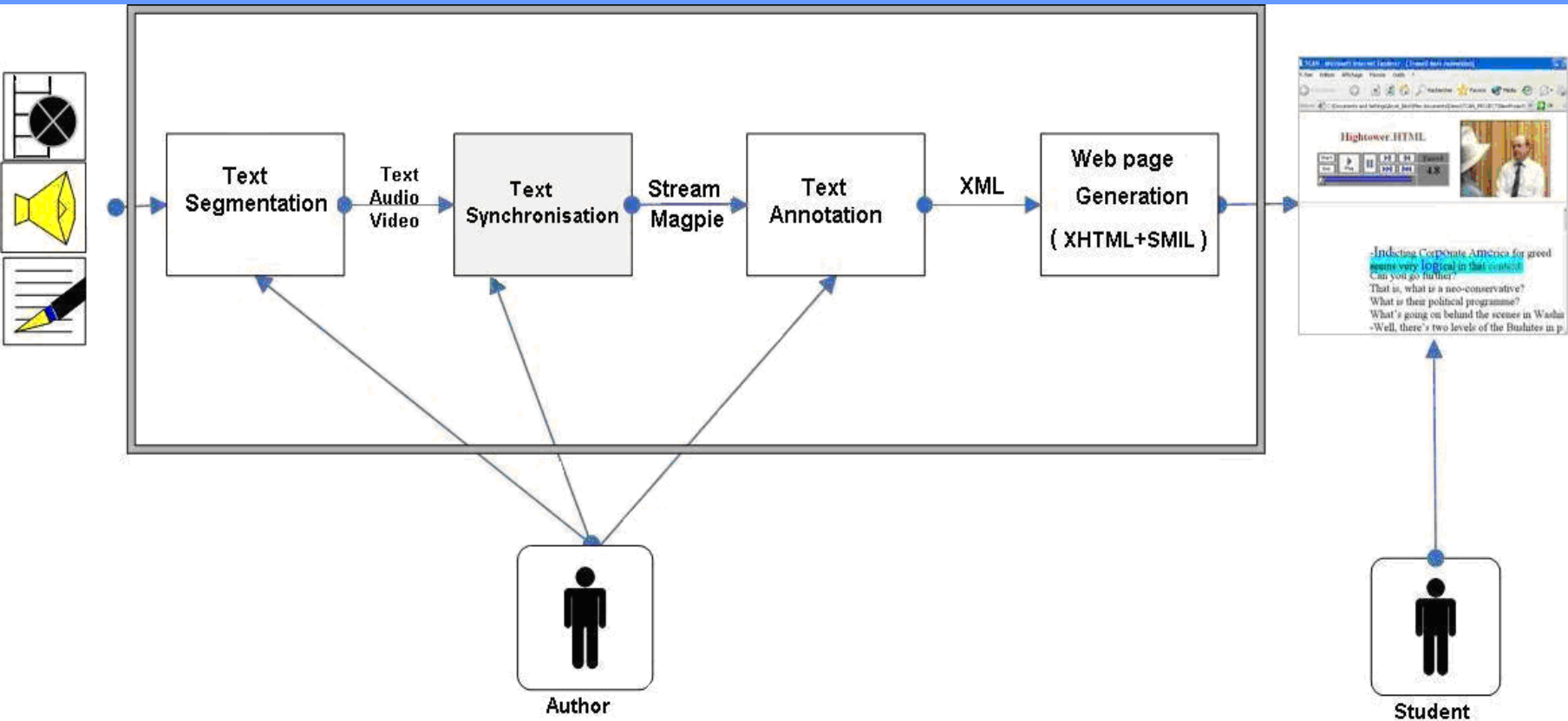
Disciplines des équipes partenaires

- Linguistique
- Perception auditive
- Psychologique cognitive des apprentissages
- Sciences et Technologies de l'Information et de la Communication
- Interface Home-Machine

La génération de textes annotés et synchronisés

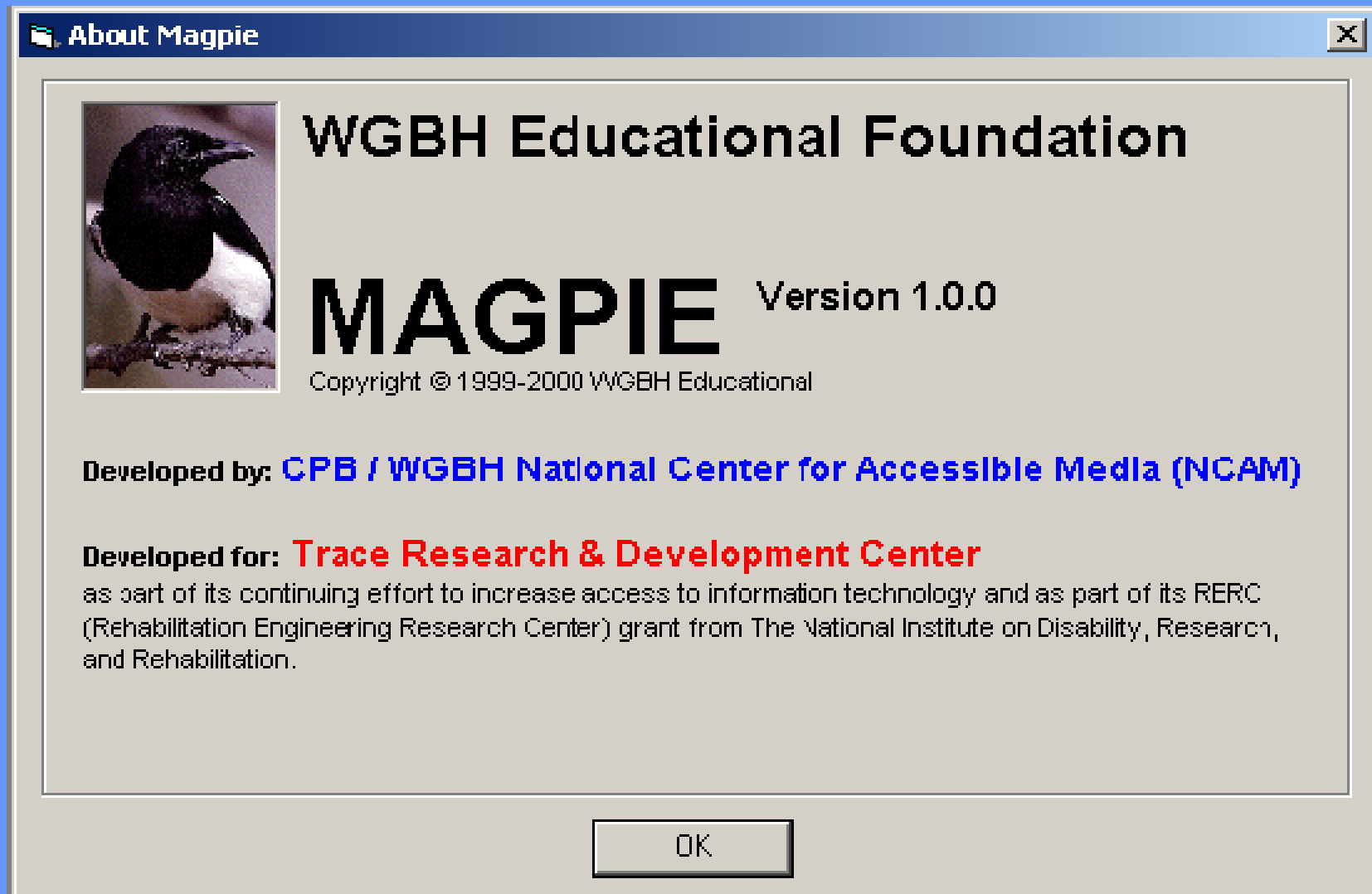


Les étapes de création avec SWANS



Student

Magpie synchronise et génère le code SMIL (Synchronised Multimedia Integration Language)



Texte segmenté et annoté dans Word avant traitement dans SWANS

YouTube is one of the Internet's fastest success stories
and it's two OWners now stand to join the ranks of the world's super rich
after Google agreed to buy it for almost £900 million.
“We just want to say thank you.
Today we have some exciting news for you.
We have been acquired by Google !”
The social networking site was created 18 months ago
by a pair of ‘20-something’ Californians.
It has just 67 staff and is based above a pizza restaurant.
But it's become a worldwide phenomenon.
So what makes YouTube such a hit ?
Well, it's all about turning ordinary people into celebrities|

La génération dans SWANS de la page Web avec audio ou vidéo synchronisé

MAGpie stream Address

C:\Program Files\MAGPIE\Stream1.rt

Select a stream from the list or search your disk

C:\Program Files\MAGPIE\sagginstream1.rt
C:\Program Files\MAGPIE\Stream1.rt
C:\Program Files\MAGPIE\Stream1perception.rt

Open...

Video or Sound track address

C:\Documents and Settings\Tony\Bureau\chomskydoss\chomsky2.rm

Open...

Web page name

Chomsky on Manipulation

Adjust text scrolling

Standard Size = 180

180

Frequency of line jumps

Standard frequency = every 7 lines

7

☒ Tick this box if your text contains typographical annotations

Generate Web Page

Return to Menu

Performance

- Le prototype SWANS a été utilisé pour la génération de 100 documents annotés et synchronisés (anglais, espagnol et allemand) en 2005 et 2006. Le temps de développement a été réduit de 2 heures à 10 minutes. La création est aujourd'hui à la portée des enseignants non informaticiens.

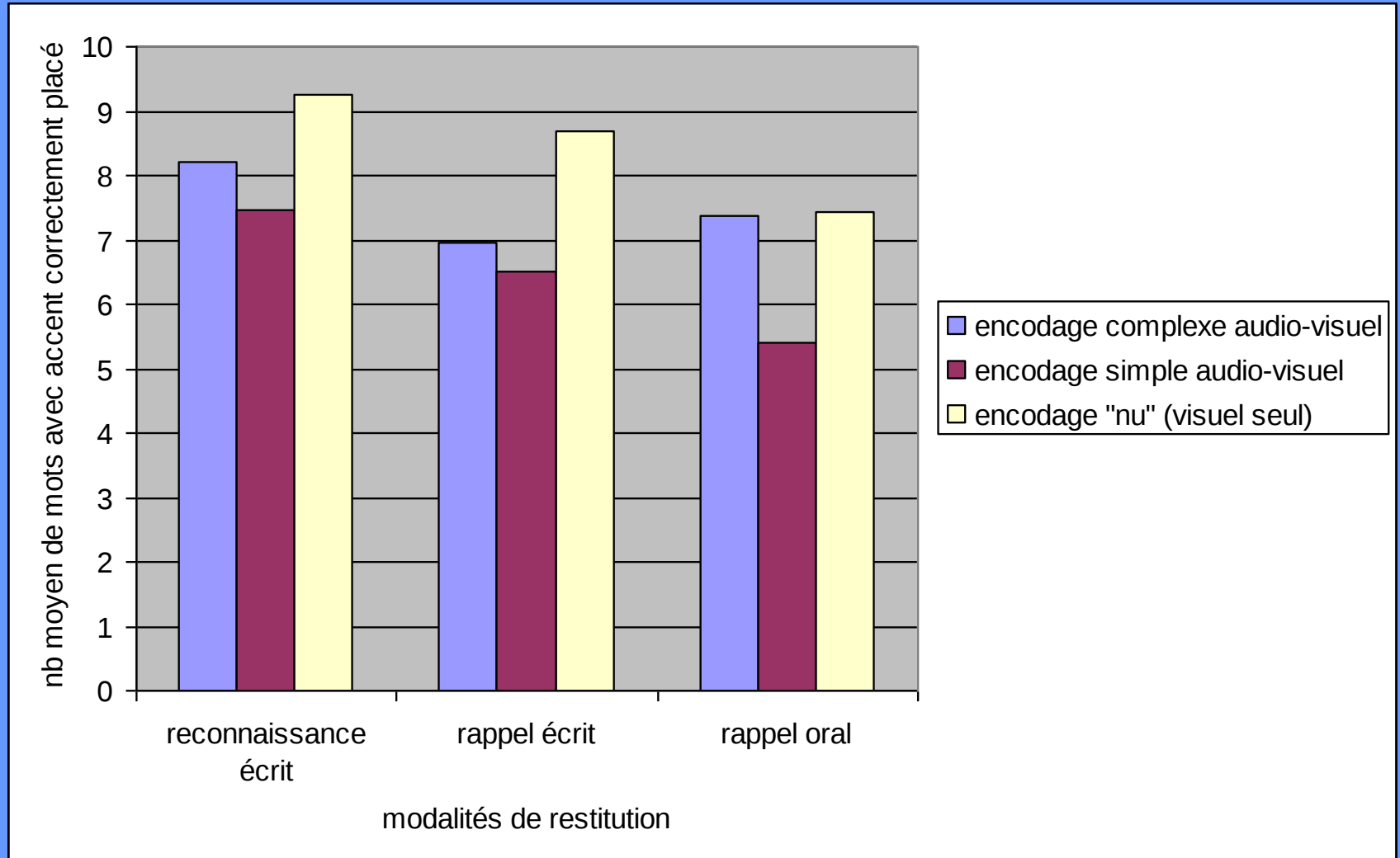
Hypothèses-I

- 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.
- SWANS devrait accélérer le processus de création de pages synchronisées et annotées.

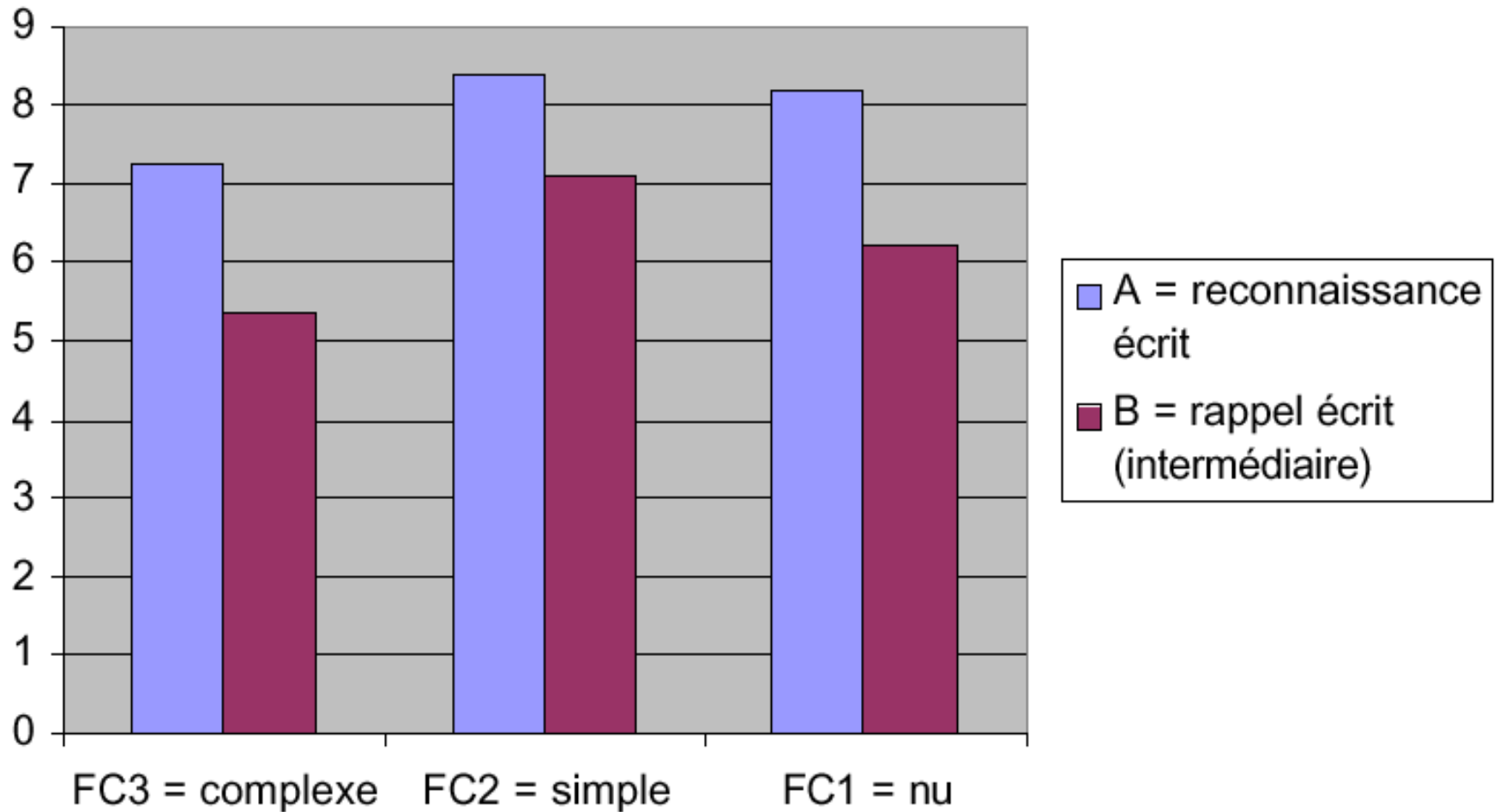
Hypothèses-II

- La lecture synchronisée, les modifications typographiques, l'utilisation des animations et des effets dynamiques exigent une concentration accrue de la part de l'étudiant
- L'effet bénéfique sera démontré si nous arrivons à prouver une mémorisation supérieure et une production orale plus compréhensible

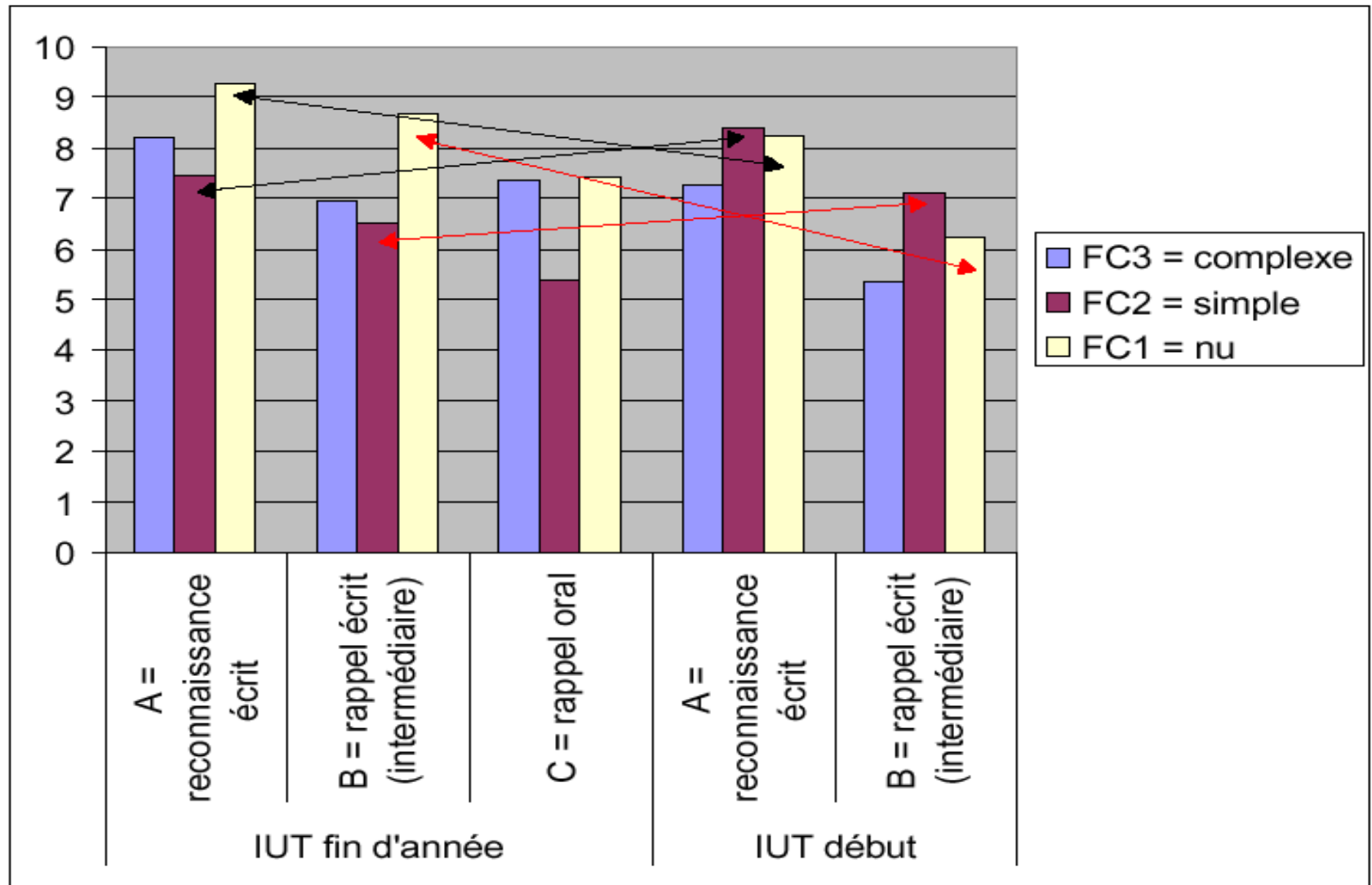
Première observation avril 2004 IUT fin d'année



Observation octobre 2004 IUT début



Les résultats comparés



élèves du collège du Château de l'Hers

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

Discussion

- **le double codage de la syllabe accentuée améliore les performances** des étudiants quand ceux-ci ont un niveau d'expertise faible pour la perception, le traitement et la production de l'accent tonique ;
- L'effet s'observe plus nettement quand on demande aux étudiants de marquer l'accent tonique sur un mot écrit que quand on leur demande de prononcer ce mot (alors que les tests que nous utilisons donnent des résultats moyens équivalents à l'écrit et à l'oral : c'est donc bien un effet spécifique du marquage visuel de l'accent tonique que nous avons obtenu, qui ne peut absolument pas être imputé à une 'facilité' des épreuves écrites) ;
- L'effet s'observe de manière équivalente quand le mot cible est isolé ou dans une phrase,

Observations à Toulouse III

- Etudiant/es de 2ème à 4 ème année d'anglais en IUP + L3 .
- La constitution d'une base de données sonores en ligne pour l'analyse des schémas accentuels employés par les français
- Dans les mots isolés
- Dans les mots affichés à l'intérieur des phrases

Tests de reconnaissance au niveau de la phrase

- Not really. **Ba-si-cally** he's a good guy, but with quite a **com-plex** personality. He's **com-plete-ly** under his father's thumb and that man is the **em-bod-i-ment** of **re-lig-ious con-ser-va-ti-sm**.
- What sort of **e-du-ca-tion** did John have?
- His father sent him to a **mil-i-tary** school and that was **de-lib-er-ate**: he **be-lieved** in keeping strict **con-trol** over children.
- He was a bit of a **lu-na-tic**, wasn't he?

Moyenne inférieure à 55%

- La moyenne est inférieure à 55% de réussite pour l'ensemble des tests, ce qui indique que le choix des mots était judicieux. Ces mots se trouvent, assez souvent, dans la liste de “niveau seuil” du Conseil d'Europe (2800 mots de base en anglais. La mise à jour de Cambridge University Press date de 1990). Certains mots à haute fréquence (“mil-i-tary”, “pre-si-den-cy”) sont en dessous de 10% de réussite, tandis que d'autres (“ba-si-cally”, “po-li-tics”, “con-trol”, “sat-is-fy”, “some-times”, “e-co-nom-ic”, “re-sults”, “re-cent-ly”, “re-lig-ious”, “per-cent”) avoisinent à peine les 30%.

Performances

- **Globalement, le double codage de la syllabe accentuée améliore les performances des novices**
- L'évaluation de l'effet au niveau de la production orale exige un apprentissage plus long qui passera par lecture à haute voix (technique téléscripteur) et l'entraînement aux exposés en classe avec production spontanée

Perspectives 2007 : annotation automatique dans SWANS

- L'annotation automatique se repose sur trois techniques possibles et certainement une solution hybride :
 - 1) Algorithmique : la méthode AIX-Marsec
 - 2) Annotation obtenue de façon automatique à partir d'un dictionnaire indexé
 - 3) Système expert selon les règles (Guierre, Deschamps, Duchet)

Annotation automatique :

1) avec l'intégration d'AIX-Marsec

- Une solution récente à ce problème vient du Laboratoire Parole et Langage d'Aix-en-Provence et le projet AIX-Marsec. « MACHine Readable Spoken English Corpus » (Auran et Bouzon 2003). Ce projet repose sur un corpus d'enregistrements sonores de 55 000 mots anglais transcrits orthographiquement (version ponctuée et non ponctuée), étiqueté morphosyntactiquement grâce au système CLAWS (Garside, 1987) et annoté prosodiquement par G. Knowles et B. Williams à l'aide de quatorze marqueurs tonétiques (High rise, Low fall rise etc...). Le système de phonétisation utilisé pour le corpus est fondé sur la recherche d'entrées dans un lexique qui est ensuite complété par l'utilisation de règles phonotactiques notamment pour faire face aux problèmes de réduction vocalique, d'assimilation, d'élision etc...

Annotation automatique (2) à partir du corpus TEAC

- Le TEAC (Toulouse English Annotated Corpus) est un corpus développé à partir des bulletins d'information de la BBC (BBC World 2004-2005) dont les scripts sont annotés et synchronisés pour SWANS. Ce corpus sera agrandi à partir des documents annotés par les centres participant dans la validation de SWANS (CERCLES).

3 : Méthode d'annotation automatique

- Une base de règles selon Guierre, Deschamps et Duchet qui signale syllabes accentuées
- La plupart des terminaisons contraignantes induisent une régularité qui avoisine les 100% (règles de <_ic, itis, osis> ou règle de <ION>...), (Deschamps p.33, 1994)
- Les suffixes inaccentués qui ne changent pas normalement l'accent du radical peuvent être affichés temporairement pour aider l'annotation de l'accentuation aussi.
- Ex : *-able, -age, -ance, -ancy, ant, -cy, -ful, -hood, -ist, -ise, -ize, -less, 'ly, ment, -ness, -or/-er, -ous*

Perspectives

- La validation de SWANS est prévue dans plusieurs centres de langue Européens de l'association CERCLES :
- London School of Economics
- Cambridge University
- University College Dublin
- University of Maastricht
- University of Padua

Conclusion

L'inhibition de l'interférence de la langue maternelle nécessite une formation explicite de l'oreille....

Un locuteur naïf est un auditeur naïf

Même les enseignants natifs trouvent l'annotation parfois complexe. Leurs jugements de hauteur ou d'intensité peuvent être erronés.

L'annotation des scripts selon la méthode visuelle de SWANS (aidée par exemple par le Cambridge Advanced Learner's dictionary) constitue une formation utile pour enseignants et étudiants. Les perceptions sont éducatibles.

La lecture des textes annotés et synchronisés a un effet positif sur la perception et la production orale contrôlée. L'effet positif sur la production orale n'est pas mesurable dans la conversation spontanée.