ASp
23-26 (1999)
Varia

Alain Cazade

De l’expérimentation multimédia à l’analyse

Quelques pistes pour tirer parti d’un parcours apprenant en langues

Avertissement
Le contenu de ce site relève de la législation française sur la propriété intellectuelle et est la propriété exclusive de l'éditeur.
Les œuvres figurant sur ce site peuvent être consultées et reproduites sur un support papier ou numérique sous réserve qu'elles soient strictement réservées à un usage soit personnel, soit scientifique ou pédagogique excluant toute exploitation commerciale. La reproduction devra obligatoirement mentionner l'éditeur, le nom de la revue, l'auteur et la référence du document.
Toute autre reproduction est interdite sauf accord préalable de l'éditeur, en dehors des cas prévus par la législation en vigueur en France.

revues.org

Revues.org est un portail de revues en sciences humaines et sociales développé par le Cléo, Centre pour l'édition électronique ouverte (CNRS, EHESS, UP, UAPV).

Référence électronique
Alain Cazade, « De l’expérimentation multimédia à l’analyse », *ASp* [En ligne], 23-26 | 1999, mis en ligne le 18 janvier 2012, consulté le 06 mai 2013. URL : http://asp.revues.org/2620 ; DOI : 10.4000/asp.2620

Éditeur : Groupe d’étude et de recherche en anglais de spécialité
http://asp.revues.org
http://www.revues.org

Document accessible en ligne sur :
http://asp.revues.org/2620
Document généré automatiquement le 06 mai 2013. La pagination ne correspond pas à la pagination de l'édition papier.
Tous droits réservés

Alain Cazade

De l'expérimentation multimédia à l'analyse

Quelques pistes pour tirer parti d'un parcours apprenant en langues

Pagination de l'édition papier : p. 441-462

Introduction - buts à atteindre

1 Ceux qui sont intéressés par l'étude des parcours apprenants sont bien obligés de constater qu'ils sont aujourd'hui encore peu nombreux, et la France ne fait pas exception en ce domaine[1]. Il est vrai que la logistique qu'il faut mettre en place pour soutenir un travail de recherche dans cette direction est assez lourde, tant au niveau du protocole de l'expérimentation proprement dit, que pour ce qui est des procédures de récupération des données et, en fin de parcours, de leur analyse.

2 Il est vrai, aussi, qu'une étude de linguiste trop ouvertement ancrée dans la didactique risque de susciter un enthousiasme mesuré à l'échelon national universitaire – notamment en anglistique – où on préférera souvent voir le dossier confié à des didacticiens déclarés. Reconnaissons que la recherche en didactique des langues opère sur un terrain délicat, où objet et sujet d'étude sont très intimement liés et où la modélisation est des plus incertaine, devant tenir compte d'un nombre infini (apparemment du moins) de paramètres et étant vite démentie par de multiples exceptions. Un travail appuyé sur l'analyse d'un corpus de productions d'apprenants pourra donner lieu à un travail plus légitimement convaincant, du fait qu'il reposera sur des bases palpables, plus facilement objectives, qui obligeront l'analyse à se limiter à des éléments clairement discernables.

3 Or, les ressources informatiques deviennent de jour en jour plus accessibles et nombreuses. Elles offrent au chercheur dans ce domaine des possibilités naguère encore inespérées, soit en renouvelant ou en simplifiant les méthodes qui ont déjà fait leurs preuves et qu'il n'est pas question d'abandonner, soit en révélant véritablement de nouvelles pistes de travail. Cette plus grande accessibilité peut désormais ouvrir la voie à une expérimentation plus large, à des échanges plus nombreux de choix de méthodes et de résultats.

4 Il est temps de susciter un travail concerté dans cette direction et, en mutualisant les ressources matérielles et les compétences des uns et des autres, de lui donner enfin les lettres de noblesse qu'il mérite. Pour ce qui ressort de l'anglistique, qui me concerne plus particulièrement, le DEA « anglais de spécialité » multipôle pourrait, et devrait même, être le lieu de réflexion et l'organe de transmission qui manque encore.

5 Sans prétendre être expert ou spécialiste en la matière, sans vouloir concurrencer les didacticiens ou les métacogniticiens, à force de traiter des « fichiers traces » j'ai fini par acquérir quelques idées de ce que l'on peut en faire. Je ne prétends pas ici dépeindre l'état de l'art, mais plutôt discerner de façon concrète quelques pistes de recherches et quelques procédures pratiques simples.

6 Je ne pourrai éviter de faire référence dans ce qui suit à certains produits du marché, les uns inévitables, qui permettent à la bureautique sur PC ou sur Mac d'exister, les autres plus discrets. Je mentionnerai même un logiciel que j'ai moi-même conçu afin, précisément, de pouvoir récupérer pour mon travail de recherche les données dont j'avais besoin. Mais bien d'autres produits pourraient être cités : en informatique, tout ou presque peut être adapté ou fabriqué, si l'on est inventif ou/et si l'on sait programmer. Les produits que je mentionne permettent de travailler en ayant le moins possible recours à un véritable langage de programmation. Le didacticiel de recherche que j'utilise pour récupérer mes données apprenantes, *Help Yourself* (programmé sous 'Logiix', *Infoaccess*, et non commercialisé), serait lui-même aussi, voire plus efficace s'il avait été d'emblée programmé en utilisant l'environnement 'Toolbook', ou encore 'Director'[2].

1. Évolution des procédures d'observation

7 Les moyens que le chercheur peut utiliser pour analyser le travail de l'apprenant sont le plus souvent lourds, peu pratiques à mettre en place et les données délicates à analyser. Pour savoir ce que l'apprenant fait, pendant qu'il le fait, on a souvent recours à ce qu'on appelle des *think aloud procedures*[3] durant lesquelles, par exemple, l'apprenant testé enregistre sur un magnétophone ou encore sur un cahier de notes, un certain nombre de commentaires sur l'action qu'il entreprend, ses choix de stratégies, etc. Il peut aussi lui être demandé de répondre à un questionnaire durant ou après son travail. Le chercheur peut, parallèlement ou à l'exclusion de toute autre procédure, préférer prendre lui-même des notes qu'il estimera peut-être plus objectives car moins sujettes à la déformation que l'apprenant ne peut manquer d'inclure dans ses remarques et déductions. Il n'est pas inintéressant, en fait, de pouvoir bénéficier des deux sources d'observation pour faire apparaître des convergences et divergences toujours porteuses de sens. Néanmoins, on peut difficilement être en même temps juge et partie et cette restriction s'applique aux deux procédures.

Récupération de commentaires ou prises de notes par logiciel de bureautique

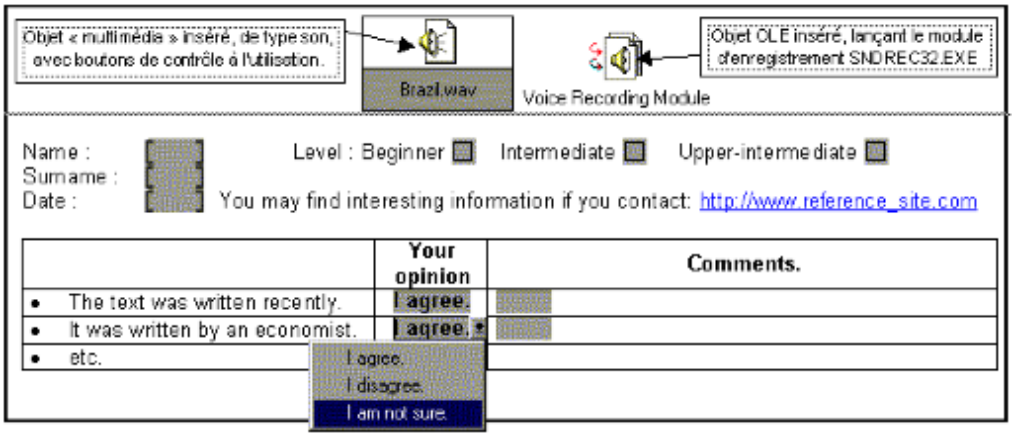
8 Si ces données écrites sont récupérées avec des outils informatiques dès le départ, une étape de transcription sera évitée, les données seront peut-être moins soumises à des erreurs de recopie et seront en tout cas plus rapidement exploitables. Il est aujourd'hui relativement aisé de constituer autour de l'apprenant un environnement d'expérimentation qui inclue des outils permettant facilement la prise de notes ou l'acquisition de réponses, à l'écrit ou même à l'oral. Un simple traitement de texte comme le 'Wordpad', livré d'office avec l'environnement *Windows 3x* et *9x*, chargé en parallèle d'un didacticiel éventuellement testé, sera amplement suffisant pour cela.

9 S'il est nécessaire de récupérer des réponses précises à des questions ciblées, on pourra valablement avoir recours aux 'champs de formulaires' que met à notre disposition un autre traitement de textes plus sophistiqué et devenu quasiment inévitable : 'Word' (*Microsoft*). Les champs réponses (voir figure 1 infra) peuvent gérer des cases à cocher ou bien du texte libre (réponses ouvertes) qui ne s'inscrira, si on a activé la protection du formulaire, que dans les champs prévus (et pas ailleurs : détail très utile pour pouvoir dépouiller et classer les réponses de tous les apprenants testés ultérieurement). Les formulaires peuvent encore présenter des menus déroulants comprenant des réponses types déjà préparées que l'apprenant pourra rapidement sélectionner sans perdre trop de temps à écrire et que l'on pourra analyser ensuite sans récupérer des réponses non prévues. Les champs disponibles pour les formulaires en ligne sont (dans la version française du logiciel) : { FORMTEXTE } = zone de texte, { FORMCASECOCHER } = case à cocher et { FORMLISTDÉROUL } = liste déroulante. Lorsqu'on a lancé le module « Formulaires » en cliquant, par exemple, sur le bouton droit de la souris dans la barre de menu, ils peuvent être insérés très simplement, déplacés, agrandis, colorés, etc. Leur utilisation sera de bout en bout gérée automatiquement par le logiciel. Un bouton 'Options' de la barre d'outils 'Formulaires' permet d'ajouter des noms ou signets à chaque case (cf. identification récupérable dans d'éventuelles macro-procédures, pour ceux qui aiment programmer plus finement, mais ce n'est pas indispensable), de préparer leur état par défaut (case cochée ou non, ou texte d'accroche, etc.), de tester contenu et mise en forme (majuscules ou non, réponses numériques, nombre de caractères maximum ou illimité, etc.) et de leur appliquer des macro-procédures écrites par ailleurs pour tester, en direct, avant ou après remplissage du champs, la présence de tel ou tel élément et déclencher une réaction adéquate (analyse de réponse, par exemple). En fin de travail, les contenus des champs peuvent être calculés, sommés, classés, etc. Pour ne récupérer que les données inscrites dans les champs de réponse préparés, il suffira de configurer auparavant le traitement de textes sur tous les postes apprenants avec l'option « Formulaires : enregistrer uniquement les données », dans le menu : Outils/Options/Enregistrement. Il est possible de faire de même si on désire récupérer les données sur imprimante (avec le menu : Outils/Options/Impression). Les données sauvegardées seront facilement importables dans un tableur ou une base de

données quelconque. Ces options ne fonctionneront qu'en cas d'utilisation d'un formulaire et n'empêcheront aucunement le fonctionnement normal du logiciel par ailleurs. Pour que le texte préparé par l'enseignant ne risque pas d'être sauvegardé une fois modifié par un étudiant, il sera sage de le sauvegarder au départ avec l'option « *read only* » ou « *lecture seule* »[4]. L'étudiant pourra ouvrir ce document, y insérer ses réponses et le sauvegarder avec ses propres références dans le répertoire prévu ou sur disquette[5].

10 La figure 1 ci-dessous montre, au passage, qu'il est possible d'insérer dans une simple page de traitement de textes des boutons son ou vidéo, éventuellement avec menus de contrôle (pause, retour, bouclage, etc.), des accès à des logiciels spécifiques (ici un module d'enregistrement de la voix de l'étudiant livré avec Windows), ou même des hyper-liens tournés vers des adresses Internet[6]. Le simple 'Wordpad' donne accès à tout ce qui précède, à l'exception des champs de réponse.

Figure 1



11 Il est donc possible, avec peu de moyens logistiques, logiciels et matériels, de tester la compréhension écrite et orale des apprenants à partir de fichiers sonores ou de séquences vidéos, de récupérer leurs réactions par écrit et même via un enregistrement oral (cf. en cas de *think aloud procedures* en mode oral, voir *infra*).

12 Il va sans dire que les réponses de type questionnaire seront tout aussi facilement récupérables directement dans une base de données ou dans un tableur (avec tous les avantages que peuvent présenter de tels outils pour trier, analyser et présenter les données en fin de procédure d'acquisition) mais il semble que leur utilisation ne soit pas aussi banalisée que celle d'un traitement de textes. Qui plus est, soulignons que, par le recours aux champs de réponses mentionnés, il est possible de présenter à l'apprenant sans aucune programmation et avec une facilité déconcertante un environnement de travail assez sophistiqué, avec textes mis en forme, enrichissements multimédias, liens hypertextes avec d'autres textes ou même avec des adresses sur l'Internet, URL de sites particulièrement intéressants, consultation d'encyclopédie, de dictionnaire, de bibliographie 'en ligne' etc. Cette richesse de possibilités multipliera d'autant les directions de recherche qu'elle pourra stimuler la réponse apprenante.

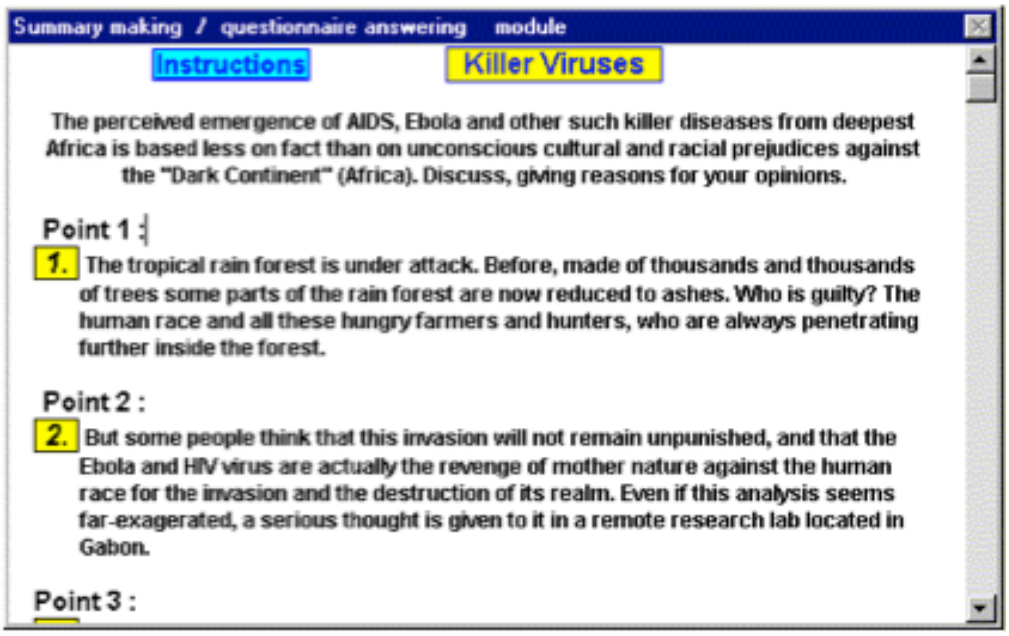
Récupération de commentaires ou prises de notes par didacticiel spécialisé

13 Tout environnement « auteur » offrira des facilités différentes pour ce faire. Avec *Help Yourself* (auquel il sera fait référence ultérieurement par le sigle : HYS), j'ai prévu plusieurs possibilités de récupérer automatiquement les réponses de mes apprenants. Une procédure a été spécialement mise au point parmi d'autres pour répondre aux besoins évoqués ici, concernant la prise de notes ou le commentaire écrit en direct à propos d'un point particulier (visualisation d'une séquence vidéo, lecture d'un texte, écoute d'un fichier son, etc.). La procédure a été mise au point pour récupérer ce que ne pouvait produire un traitement de texte. Un bouton spécial, très simplement inséré[7] dans le corps d'un écran de travail (*frame*), au milieu d'un texte que l'on peut faire défiler sur l'écran (« scrollage », sic !) s'il est un peu long, déclenche, lorsqu'il est activé, une boîte « combo »[8] permettant de récupérer une réponse d'une longueur

de 500 caractères maximum. La limitation à 500 caractères me permet d'obliger l'apprenant à sérier sa réponse s'il a besoin d'écrire davantage, lorsque je demande un résumé ou un commentaire organisé à partir d'un document quelconque. L'intérêt principal de la procédure est que, lorsque l'apprenant veut ultérieurement revenir sur le contenu de sa réponse pour la modifier, il lui faut réactiver le bouton de saisie de réponse, ce qui permet d'inscrire, dans le fichier 'trace' automatiquement créé et alimenté par le logiciel, la modification apportée à la réponse, séparément du précédent texte entré pour la même question. On comprend que cette procédure donne, au moment d'analyser le fichier trace, la possibilité de déterminer, par comparaison, ce qui aura suscité un ajout, une amélioration, une rectification d'erreur, etc. Il peut être éclairant de noter, par exemple, dans quel ordre de priorité les idées sont posées ou quelles erreurs sont vues et quelles autres ne sont pas vues. Le fichier trace révélera ainsi la progression de l'acte d'écriture, avec la notation exacte du moment où les modifications sont apportées, les retours éventuels après d'autres tâches, divers types d'interruptions, etc.

14 La figure 2 montre un exemple d'un texte[9] en train d'être rédigé par un étudiant de magistère (Bac + 5). Les fichiers traces montreront que certains points ont été ou sont sur le point d'être rectifiés : dans le point 1 : ajout d'une virgule après *before* et de majuscules à *Tropical Rain Forest*, dans le point 2 : ajout d'un trait d'union après *far* et rectification du mot précédemment mal orthographié : *thougt*. Mais d'autres détails pourront encore être modifiés.

Figure 2



Récupération de données orales

15 La prise de notes orale peut éventuellement se faire via un utilitaire comme le 'magnétophone' virtuel livré avec tout ordinateur multimédia. Sur un PC tournant sous Windows, l'utilitaire de base s'appelle '*Sound Recorder*'[10]. Il peut rester chargé inactif en parallèle avec toute application et être activé à volonté tandis que l'apprenant pourra continuer à écrire quelque chose sur son traitement de texte. Toutefois il sera un peu rudimentaire et ne sera pas très maniable. On lui préférera certainement d'autres outils plus performants, comme ceux livrés par exemple avec certaines cartes vocales de bonne qualité[11]. Précisons que des logiciels sont actuellement développés[12] qui doivent dépasser les possibilités actuellement offertes et permettre de gérer les accès sons/images/vidéos de manière beaucoup plus riche, avec, indexés sur un marquage numérique des fichiers multimédias, des affichages de textes (sous-titres, commentaires, traductions, consignes etc. Le développement de la compression des fichiers sons (format MP3, couche sonore du format MPEG) rend même possible désormais une fonctionnalité très intéressante pour la compréhension orale : la diffusion d'un fichier son recalculé de manière à ce que son débit soit ralenti sans que la hauteur des sons ne change[13].

Dans l'environnement de travail que j'ai préparé avec HYS, plusieurs utilitaires sont prévus à cet effet, pour présenter à l'apprenant du texte, des graphiques, une séquence vidéo etc. et récupérer juste après sa réaction à l'oral[14].

Récupération de 'données discrètes' par procédures automatiques

16 Un autre type de procédure, spécifique, celle-ci, d'un environnement informatisé, peut contribuer à fournir une base de travail pour une analyse de parcours apprenant. Cette voie permet d'aller au-delà de ce qui vient d'être mentionné en espionnant, pour ainsi dire, l'apprenant sans qu'il en ait conscience —bien qu'il doive en être averti, pour satisfaire aux exigences légitimes de la CNIL[15]. Dans le cadre d'exercices de types variés (trous, QCM, classements, traductions, questions à substitutions multiples automatisées, etc.), toutes les opérations réalisées par l'apprenant peuvent être 'discrètement' accompagnées d'une écriture dans un fichier trace, comportant des indications laissées au choix du programmeur.

17 L'avantage de telles procédures est qu'elles ne perturbent pas le travail de l'apprenant en l'obligeant à s'arrêter pour réfléchir à ce qu'il fait. On est tenté de dire que les données récupérées sont également plus proches de l'acte apprenant, plus authentiques, d'une certaine manière, puisque non modifiées par l'apprenant qui s'auto-analyse ou par l'expérimentateur scrutateur, les deux devant inévitablement subir, dans leurs prises de notes, le filtre de leur subjectivité.

18 Je mentionnerai infra, en point 2, les choix que j'ai privilégiés avec HYS et les types de procédures automatiques que j'utilise mais il est clair que tout autre environnement auteur existant sur le marché peut faire de même —et infiniment mieux— pour peu que le programmeur s'en donne les moyens ou que son cahier des charges le lui demande, ce qui n'est hélas pas le cas avec la quasi totalité des produits existant actuellement. Des protocoles de programmation et des procédures pré-écrites sont disponibles dans les environnements déjà cités : 'Toolbook', 'Director' et 'Authorware'[16]. Le premier, certainement le moins inaccessible à l'enseignant non spécialiste mais curieux, et volontaire, comporte toute une série de fonctions et d'outils particulièrement destinés à un travail en langues vivantes.

19 La difficulté qu'un tel système opératoire pose est que, en fin de parcours, il faut savoir interpréter les contenus des fichiers traces obtenus, et donc il faut savoir choisir, au préalable, quelles données discrètes y faire figurer. Un autre élément à prendre en compte est que l'analyse qu'on développera risque d'être très fortement marquée par le quantitatif, même s'il est, d'expérience, toujours possible de renforcer l'aspect qualitatif. Il importe d'en tenir compte au moment où on conçoit le protocole d'une expérimentation.

2. Quels éléments récupérer dans des fichiers traces

20 C'est probablement le problème majeur à affronter. De ce choix dépend la qualité et la pertinence des analyses ultérieures, bien plus que de savoir manipuler tel ou tel outil pour traiter les données. Néanmoins, il faut bien admettre que, pour effectuer un choix correct, il faut déjà avoir une certaine habitude de ce que l'on peut faire avec un tableur, une base de données ou un concordancier. Pour rassurer ceux qui en ont une idée insuffisante, ajoutons que si un outil peut être véritablement plus efficace qu'un autre pour traiter un aspect et mettre en évidence une tendance, tous les logiciels d'analyse travaillent, à la base, sur un ensemble d'opérations qui sont toutes plus ou moins apparentées et qui tournent autour de quelques idées simples : classer, trier, rapprocher, compter, déduire, formuler une hypothèse et donc projeter, puis re-classer, re-trier, re-rapprocher et vérifier l'hypothèse.

21 Pourtant, il faut toujours se rappeler que les données peuvent être constamment réorganisées pour s'adapter à un autre logiciel d'analyse que celui auquel on avait pensé en premier. Pour nous aider dans ce sens, il est toujours possible d'avoir recours à un procédé que les utilisateurs de concordanciers connaissent bien. Une étape de repérage (*tagging*) du corpus constitué aidera ultérieurement pour cadrer l'analyse de multiples manières et sera même, dans de nombreux cas, indispensable. Ce repérage peut être automatiquement fabriqué par le logiciel fabriquant les fichiers traces ou peut être ajouté par la suite. Des utilitaires de recherche/remplacement existent pour faire cela assez facilement mais il est préférable de

prévoir dès la récupération des éléments traces un minimum de séparateurs clairs pour que les transformations puissent se faire de façon automatisée par la suite. N'importe quel traitement de texte fera cela, mais pour traiter des fichiers traces un peu longs ou nombreux, cela risque de poser problème. 'Wordsmith Tools' (OUP[17], écrit par Mike Scott), l'outil qui a succédé à l'OCP, est livré avec un TEXT CONVERTER, lequel peut traiter 16.000 fichiers différents par une seule demande de traitement, l'ensemble des commandes de transformations dans tous les textes traités pouvant atteindre 150 opérations différentes et successives.

22 Un marquage minimum insérera dès le départ des marques de tabulation pour séparer des éléments de même nature et des fins de paragraphe pour séparer deux enregistrements différents (voir en point 4, infra).

23 Ceux qui ont déjà vu les fichiers sources de pages écrites en HTML (*Hypertext Markup Language*) ou en SGML (*Standard Generalized Markup Language*) ont déjà une idée de ce qui n'est qu'une évolution des anciens marquages de type COCOA[18]. De tels marquages permettent de baliser le texte étudié, en faisant figurer, entre chevrons le plus souvent (<… >, certains utilisent les accolades : {…}) des étiquetages incluant le type d'étiquette traité (ou de variable, si ce terme est plus satisfaisant pour certains) et la valeur locale de l'étiquette. Un marquage possible pour désigner une réponse d'un apprenant (*learner*) appelé John, à la question n° 23 d'un QCM, pourrait être : <L=JOHN EX=MCQ QN=23>. Le début d'une réponse ouverte (*open question*) dans un fichier à analyser pourrait être marqué ainsi : <OQ>, puis viendraient le corps même de sa réponse et enfin la séquence </OQ>, pour marquer la fin de cette réponse.

24 Différents types de marquage pourront être choisis, suivant les données traitées et la direction de recherche privilégiée. On ne sera pas étonné que les termes habituellement utilisés pour les différencier nous viennent du monde anglo-saxon, où ce travail est nettement plus avancé qu'ailleurs. On s'intéressera plus particulièrement aux parties du discours, qu'on pourrait appeler PDD pour faire pendant à l'expression anglaise *POS tagging* (*part of speech*) ; au *phrasicon tagging*[19] (le terme cherchant à se distinguer de : *lexicon*), c'est-à-dire au repérage des enchaînements types que privilégient les apprenants ; au marquage syntaxique, au marquage sémantique, ou au *discoursal tagging*. Il est possible d'avoir recours à un traitement automatisé satisfaisant des deux premières catégories citées, mais il n'en va pas encore de même pour les autres[20].

25 Fort de cette malléabilité potentielle des résultats, et puisqu'il est possible de ne pas être obnubilé par les contraintes de l'analyse logicielle, il reste nécessaire de se pencher sérieusement sur la véritable problématique de la fabrication du corpus apprenant : quels axes de lecture va-t-on chercher à suivre en priorité ? Va-t-on se concentrer sur la façon d'apprendre des *non native speakers* (NNS) ou sur la façon d'enseigner ; sur la réponse de l'apprenant à un stimulus, sur le parcours qu'il choisit d'emprunter pour aller d'un point à un autre, d'une requête à son aboutissement, plus ou moins direct ou au contraire sinueux, organisé de façon géométriquement systématique, heuristique ou plus – apparemment, du moins – aléatoire[21] ; ou va-t-on se pencher plus attentivement sur l'accompagnement pédagogique qu'il recherche, sur l'impact d'un renforcement visuel à l'intégration d'un point abordé, ou encore sur le temps qu'il met pour faire tout cela ? En fait, tout est envisageable et doit, à un moment ou un autre, être envisagé, ne serait-ce que pour formuler, puis infirmer ou confirmer une hypothèse.

26 Afin de ne pas alourdir inutilement les fichiers traces, il importe donc, pour commencer, de se déterminer sur un certain nombre de points :

- les catégories de '**didactèmes**'[22] auxquelles on veut s'intéresser en priorité, le terme pouvant désigner les données minimales à enseigner ;
- de quels '**actèmes**' on veut suivre la trace, le terme voulant représenter les actions apprenantes minimales : écriture, effacement, validation, activation de bouton son, vidéo, demande d'aide, mouvement intra- ou/et extra-modulaires (changement d'écran de travail, retour à un menu, etc.), accès à un dictionnaire, recherches d'autres occurrences de termes dans l'environnement de travail, avec ou sans tentative de lemmatisation, ouvertures/fermetures de modules etc. ;

- quels '**productèmes**' seront ciblés, le terme indiquant les productions apprenantes minimales : les réponses finales aux questions posées ou bien celles-ci ainsi que toutes les tentatives successives qui auront précédé, les productions écrites seules ou/et orales, etc.

27 Une catégorie d'éléments étudiés n'est parlante qu'au regard de plusieurs autres. Et donc, s'il faut garder à l'esprit l'idée que : plus on a de données, plus elles sont délicates et lourdes à manipuler, ce qui justifie un travail approfondi de sélection des pistes à suivre mentionnées ci-dessus, l'expérience montre malgré tout qu'il faut aussi laisser des portes ouvertes pour l'inattendu, préserver des approches différentes ultérieures d'un même corpus de données après constitution.

28 La base type d'une requête que l'on formulera pour analyser un fichier trace pourra comporter des questions telles que :

- qu'est-ce qui pose le plus de problème, qu'est-ce qui en pose le moins, (données, vecteurs de données, conditions d'utilisation, à long/court terme etc.) ?
- quel type d'activité attire le plus, le moins ?
- qu'est-ce qui valorise, qu'est-ce qui a un rôle neutre ?
- pour qui les tendances décelées sont elles probantes, pour qui ne le sont-elles pas ?
- est-il possible de déterminer un ou plusieurs types de paramètres qui influent sur ces résultats ?

29 Suivant les réponses trouvées, une hypothèse sera alors formulée, qui précédera une nouvelle expérimentation jouant sur les paramètres qu'on estimera déterminants.

3. Mes choix avec HYS

30 Chacun des items définis supra entraînera l'écriture dans un fichier trace d'une série d'enregistrements mélangée dans ses constituants mais chronologique, chaque enregistrement correspondant à une action ou décision minimale donnée et devant être clairement identifié. Dans ce but, en plus de l'identité de l'apprenant, j'ai choisi de récupérer : le groupe pédagogique auquel il appartient (et donc le niveau possible de ses études, son profil général, ses attentes et centres d'intérêts possibles), les références de l'exercice, le numéro de l'écran de travail (chaque écran portant sur des points précis), le numéro du bouton-question ou de tout autre bouton activé, un horodatage à la seconde près[23]. Chaque détail identifiant tout enregistrement permet par la suite de suivre la progression du travail de l'apprenant, d'exercice en exercice, de question en question.

31 Les didactèmes peuvent être choisis à l'infini et ne sauraient être listés dans une simple phrase. Parmi bien d'autres pistes possibles, mentionnons seulement qu'ils peuvent être d'ordre lexicologique, stylistique, qu'on peut viser la maîtrise ou non des formes du discours, le traçage de points de langue acquis ou/et les erreurs commises dans leur emploi, la capacité d'organiser un commentaire ou de reconnaître l'organisation interne d'une série d'informations, etc., tout cela en comparaison avec d'autres corpus NNS ou d'un ou plusieurs corpus NS.

32 Les exercices possibles sont, pour beaucoup, fabriqués à base de procédures de masquages/démasquages, sur lesquelles beaucoup a déjà été écrit mais qui peuvent fournir, pour peu qu'on réfléchisse solidement à tous les possibles, des outils remarquables pour tester les connaissances apprenantes et même stimuler leur travail. Des questions que j'appelle « *colour questions* » permettent de colorier des éléments de texte suivant un codage de couleur qu'on peut redéfinir pour chaque exercice. Cela permet de faire apparaître des syllabes accentuées d'une couleur différente de celle réservée aux autres syllabes, de même pour les voyelles diphtonguées ou non, les différents phonèmes entre eux, la transformation en *schwa* de la sonorité centrale de certaines particules, du *to* devant un mot commençant par une consonne ou non, etc. Le « coloriage » du texte permet également de distinguer ses différents éléments constitutifs, autant d'un point de vue sémantique que structurel. L'intérêt est d'abord de pouvoir rassembler plusieurs éléments dans une même classification (traités séparément dans d'autres types de questionnement), ensuite de profiter d'effets synesthétiques pour faire jouer l'intellect dans un contexte conceptuel légèrement décalé. J'ai également prévu un système de substitutions multiples (*Multiple Substitution Questions*,ou MSQ) pour renouveler légèrement

le principe de la question à trou afin de tester, là encore, la cohérence de tout un ensemble de points. Le questionnement par MSQ permet aux contenus manquants d'être remplis par différentes propositions préparées au fur et à mesure qu'on clique sur les éléments testés. La validation d'une question ne se fait que lorsqu'on pense que tous les éléments sont correctement placés.

33 Toute une logistique, qu'il serait trop long de décrire entièrement, entoure le fonctionnement des exercices : accompagnements sonores pendant la recherche, commentaires écrits et sonores après réponse, différents niveaux d'aides, les unes prévues automatiquement par le logiciel, les autres par l'enseignant-auteur. Parmi les facilités automatisées, un système de bonification graduée permet de régler le niveau d'exigence ou de tolérance des procédures d'analyse de réponse. L'idée générale est de permettre à l'apprenant de pouvoir constamment adapter son environnement de travail « à sa mesure », et donc de se responsabiliser. On n'apprend que ce que l'on est prêt à apprendre, à la fois par ce qu'on sait déjà mais aussi par la décision qu'on a prise d'intégrer des éléments nouveaux.

34 Il faut donc prévoir un espace d'appropriation, un certain nombre d'éléments clés qui donneront le sentiment à l'apprenant qu'il « peut » se prendre en charge, même si certains préfèrent qu'on les aide fortement à prendre une décision, ce qui, en soi, est déjà une décision apprenante.

35 Il faut également préserver un « espace de séduction », dans lequel les aspects ludiques jouent un rôle privilégié, sinon essentiel. Les accompagnements sonores, pendant la recherche d'une réponse et au moment où le logiciel analyse et commente les réponses données, en font partie. Alors qu'une procédure permet de supprimer tout ou partie de ces accompagnements (une trentaine de thèmes de films célèbres, sous forme de fichiers midi[24], toute une série de bruits animaliers ou domestiques plus ou moins suggestifs en cas de réponse bonne, inattendue ou insatisfaisante, etc.), tous les collègues qui utilisent HYS dans notre centre de ressources constatent avec étonnement que, les neuf dixièmes du temps, les étudiants préfèrent les conserver. Je dois même préciser que c'est sur leur demande insistante que je les ai ajoutés, et en nombre si élevé. Il faut parvenir à garder le contact, ne pas lasser trop vite les blasés, ne pas décourager les faibles mais au contraire les faire avancer même quand il ne savent pas comment, et proposer dans le même temps différents niveaux de défis possibles aux plus forts.

36 Toutes les décisions concernant le paramétrage de l'environnement de travail par l'apprenant – le degré de difficulté, de bonification ou d'accompagnement ludique choisis, les choix privilégiant tel type d'activité plus qu'un autre, l'ordre d'enchaînement de toutes les actions –, font partie des actèmes que l'on peut faire intervenir comme variables afin de suivre l'influence de l'environnement sur les résultats obtenus. Le problème est de nouveau de choisir quels actèmes analyser, puisque tout peut, en théorie, être pris en compte, des déplacements de curseur aux mouvements de souris, des accès aux points d'aide (internes, ou dictionnaires extérieurs, etc.) au temps d'attente entre les différentes actions tentées, qu'il faudra savoir parfois identifier comme temps de réflexion ou d'hésitation.

37 Les productèmes que je privilégie sont essentiellement des réponses écrites aux différents types de questions posées. Un des aspects essentiels est la possibilité de pouvoir analyser, non seulement les réponses finales à des questions données, mais surtout les différentes étapes pour parvenir à une réponse finale, tentatives complètes ou partielles, fructueuses ou non. Il sera même possible d'analyser les cas d'« abandons », qui aboutissent à ce que le logiciel montre la réponse, laquelle pourra soit être entrée par l'apprenant (option par défaut qui vise à valoriser l'acte de production de l'apprenant et à favoriser la fixation de l'item non reconnu) ou entrée par le logiciel en cas de blocage sévère ou de sentiment de perdition ! La récupération de toutes les étapes de production, qui me semble être une nécessité, n'est hélas pratiquement jamais permise par d'autres produits existant sur le marché. On imagine pourtant sans peine l'intérêt d'essayer de suivre les processus mentaux, les démarches déductives que chaque tentative de réponse non ou partiellement couronnée de succès permet de déceler chez l'apprenant. Pour repérer chaque productème (i.e., *tagging*), je suis amené à faire s'ajouter aux éléments d'identification mentionnés plus haut, et pour chaque enregistrement, la réponse attendue, la réponse donnée, l'état du texte travaillé après réponse (en cas de masquage, par exemple),

un signe qui précise la manière dont la réponse a été accueillie[25], le numéro de la tentative de réponse par rapport au nombre total d'essais autorisé, le score obtenu après réponse. Les références d'une requête d'aide seront répertoriées de la même manière. D'autres données pertinentes peuvent être facilement calculées ultérieurement et ne seront pas inscrites pour ne pas charger le fichier trace inutilement. Il en va ainsi, par exemple, des écarts de temps entre les réponses ou des pourcentages de réussite relatif et absolu atteints par l'apprenant au moment d'une réponse (lesquels ne seront récupérés globalement qu'à la clôture d'un module). Il va sans dire que je suis amené à changer les éléments qui paraissent pertinents suivant les besoins d'une expérience donnée.

4. Traitement des réponses

38 Les suggestions de traitement de fichiers traces présentées ici ne sont que des exemples parmi d'autres possibles. Le standard d'écriture de fichier que j'ai choisi est le plus commun : il s'agit de fichiers « tabulés », c'est-à-dire dans lesquels les différentes rubriques pour chaque enregistrement sont séparées par des tabulations (code ASCII : 09), chaque enregistrement étant séparé du suivant par un « retour chariot » (code ASCII 13 + 10). D'autres types de fichiers : SYLK, de type tableur ; DBF, Data Base File ; BAS, langage Basic classique… sont tout autant utilisables. Les différences tiennent au choix des séparateurs entre rubriques, entre les enregistrements, dans la présence nécessaire ou non en premier enregistrement de noms de variables pour identifier la structure du fichier et les différentes rubriques, etc.

39 Pour donner une idée de la simplicité d'utilisation de tels fichiers, j'indiquerai seulement qu'il me suffira de donner à ces fichiers l'extension XLS[26] pour que, une fois un tableur plus ou moins incontournable comme 'Excel' (*Microsoft*) installé, il suffise de double-cliquer sur son nom dans l'*Explorateur* pour automatiquement ouvrir ce fichier dans le tableur et voir tous les enregistrements proprement disposés à l'écran. Rappelons que ce logiciel n'est mentionné que pour illustrer plus concrètement ce que, en fait, on pourrait réaliser avec tout autre tableur. Dans l'extrait suivant (tableau 1), bien que j'aie supprimé un certain nombre de données pour qu'un enregistrement puisse figurer sur une seule ligne, il est possible de voir quelques progressions de réponses à différents masquages simples sur des points divers.

Tableau 1

Name:	File/screen/bt.nb	Date/time	Question item	Sign	Mask after:	Tries:	Pts:	Answer
Hubert	k_virus[3] -62582	May 26 1998 16:33:18	deadly viruses		**deadly	******	1|8	78|125
deadly								
Hubert	k_virus[3] -62582	May 26 1998 16:33:34	deadly viruses	+	deadly viruses	2|8	83|13D	viruses
Hubert	k_virus[3] -62658	May 26 1998 16:33:47	aggressive	-	ag********	1|6	83|13D	agressive
Hubert	k_virus[3] -62658	May 26 1998 16:33:59	aggressive	**	ag********	2|6	83|13D	agresive
Hubert	k_virus[3] -62658	May 26 1998 16:34:46	aggressive	*-	ag********	3|6	83|13D	agresive
Hubert	k_virus[3] -62278	May 26 1998 16:36:01	primeval	*-	prim****l	2|5	83|13D	primarly
Hubert	k_virus[3] -62278	May 26 1998 16:36:44	primeval	*-	prim****l	3|5	83|13D	primary
Hubert	k_virus[3] -61914	May 26 1998 16:36:52	Sound button					
Hubert	k_virus[3] -62278	May 26 1998 16:37:14	primeval	*-	prim****l	4|5	83|13D	primival
Hubert	k_virus[3] -62736	May 26 1998 16:37:27	process	*-	p******	1|4	83|13D	powers
Hubert	k_virus[3] -62736	May 26 1998 16:38:10	process	+	process	2|4	88|135	process
Hubert	k_virus[3] -63116	May 26 1998 16:38:25	mutation	+	mutation	1|3	93|14D	mutation
Hubert	k_virus[3] -62658	May 26 1998 16:38:50	aggressive	+	aggressive	4|6	98|145	aggressive

40 Alors que l'apprenant cherchait à répondre en s'aidant exclusivement de la lecture, d'une recherche de cohérence dans l'existant et de ses souvenirs du texte déjà étudié sous un autre angle, c'est l'activation d'un bouton son qui lui a finalement permis de se tirer d'une difficulté de compréhension sur le mot *primeval*, même s'il est intéressant de constater que la réalisation écrite de la pénultième sonorité du mot n'est pas encore maîtrisée. Un traitement de cette erreur pourra être inclus dans le module ultérieurement, sous forme d'analyse de réponse. On pourra de même chercher à tester si ce type de graphie interprétative de sonorité mal perçue peut être rapprochée d'autres cas plus ou moins apparentés, s'il concerne un apprenant de façon exceptionnelle ou si le trait est dominant chez d'autres également.

41 On remarquera qu'à propos d'une difficulté d'orthographe classique liée au mot anglais *aggressive*, la réponse négative du logiciel aura perturbé l'apprenant et l'aura poussé à envisager de façon insistante une solution là où on ne l'attendait pas (*agresive*). On notera également qu'un détour par d'autres démasquages victorieux aura convaincu l'apprenant de

revenir sur le point non résolu précédemment, après que le stress de l'erreur ou plutôt de l'incompréhension –tout relatif et minime qu'il puisse être – aura été ainsi calmé.

42 On notera encore chez l'apprenant une densité d'activité assez soutenue, un aspect dont il est intéressant de noter qu'il est assez constant chez tous les apprenants, mais dont la variation suivant les sujets et les objets d'étude est intéressante à suivre.

Tableau 2

Name:	File/screen/bt.nb	Date/time	Question item	Sig	Mk after:	Tries:	Pts:	Answer
Xavier el_nino[14]	-63627	Apr09 1998 18:09:25	marine life	•-	•••r••• ••••	1|7	102|135	rain
Xavier el_nino[0]	-63402	Apr 09 1998 18:09:41	rainfall	!	HELP	1|6	102|135	rai | fal
Xavier el_nino[14]	-63402	Apr 09 1998 18:09:54	rainfall	+	rainfall	2|6	107|140	rainfall
Xavier el_nino[14]	-63252	Apr 09 1998 18:10:07	predictions	••	•e•i••i••s	3|7	107|140	s
Xavier el_nino[14]	-63252	Apr 09 1998 18:10:12	predictions	••	••e•i••i•ns	4|7	107|140	n
Xavier el_nino[14]	-63252	Apr 09 1998 18:10:33	predictions	••	pre•i••i•ns	5|7	107|140	pr
Xavier el_nino[14]	-63252	Apr 09 1998 18:10:39	predictions	•-	pre•i••i•ns	6|7	107|140	previsions
Xavier el_nino[0]	-63252	Apr 09 1998 18:10:46	predictions	!	HELP	6|7	107|140	dic|pr|on
Xavier el_nino[14]	-63252	Apr 09 1998 18:10:58	predictions	+	predictions	7|7	112|145	predictions

43 Dans l'extrait ci-dessus (tableau 2), le masquage du mot *predictions* a dans un premier temps été identifié comme cachant les mots *prevision* ou *information* dans la phrase test suivante :

New •••••••••• from both government and private scientists say the latest El Niño could be a severe one.

44 L'apprenant a d'abord essayé la lettre « i », puis la lettre « e ». Ne pensant toujours pas au mot *predictions* et comprenant que le mot *information* n'était plus possible, il a proposé le mot : *previsions*. Sans succès. Le recours à l'aide et l'affichage de certains éléments constitutifs du mot masqué lui ont finalement ouvert les yeux et permis de retrouver le mot *predictions*.

45 On peut penser qu'un embryon de stratégie a été envisagé dans ces deux cas et il sera intéressant de vérifier que cette stratégie est soit passagère soit répétitive, qu'elle est incluse parmi d'autres dans un système plus général ou non.

46 Pour rechercher dans un fichier trace une information précise, répondant éventuellement à plusieurs critères croisés (ici, par exemple, les recours à l'aide : signe : !, les réponses non reconnues partielles : signe : •- ou complètes : signe : – , positives partielles : •• ou complètes : +) il suffira, avec 'Excel', de sélectionner tout le tableau (Ctrl A, comme *all*) et d'activer dans le menu 'Données/Filtre' l'option 'Filtre automatique'. La première ligne sera automatiquement transformée en une ligne de sélection de critères. Il suffira de cliquer dans la case contenant le critère voulu et de choisir un item parmi la liste déroulante automatiquement créée (voir figure 3), contenant tous les items accessibles (dont on appréciera qu'ils soient déjà classés par ordre alphabétique) pour automatiquement ne faire apparaître devant nos yeux que les lignes répondant à la demande ainsi formulée. Il est possible de personnaliser chaque critère par un double jeu de booléens classiques et d'autres plus parlants : [ne] commençant [pas] par / [ne] contenant [pas] / différent de/ égal à /, etc. (options entre crochets : […]). Un deuxième critère de sélection portant sur cette même rubrique pourra être construit de manière semblable, et sera lié au premier par un ET ou encore un OU (inclusif). Il suffira de répéter l'opération sur plusieurs têtes de critères et on aboutira à une sélection multicritère qui pourra être éventuellement très fine. Pour savoir qui, par exemple, a répondu correctement (+ dans la rubrique *sign*) à une question donnée dès le premier essai, il me suffit d'entrer comme premier critère le numéro de la question dans la rubrique *File+button*, le signe + dans la rubrique *sign* et, dans *tries* (construite comme suit : nombre d'essais|total autorisé), la chaîne : « 1| » avec la condition booléenne : 'commence par', (et non pas 'contient', car cela risquerait de sélectionner un 11e ou 21e essai sur *n* possibles). Je saurai de suite si la question était trop facile, trop difficile, pour qui, pour quel groupe, etc.

Figure 3

Lenoucec.xls

	A	B	C	D	E	F	G	H	I
1	Name	File+button:	Date/time	Question item	sign:	After Mas	trie	Pts:	Answer:
2	LENOUR	POP.GUI[2] -30310	Feb 05 1998 17:40:	[Tous]	+	population	1|5	5|5	population
3	LENOUR	POP.GUI[2] -30538	Feb 05 1998 17:40:	[10 premiers...]	+	90000	1|3	10|10	90000
4	LENOUR	POP.GUI[2] -30614	Feb 05 1998 17:41:	[Personnalisé...]	+	250000	1|4	15|15	250000
5	LENOUR	POP.GUI[2] -30462	Feb 05 1998 17:41:	90000	+	production	1|5	20|20	production
6	LENOUR	POP.GUI[2] -30692	Feb 05 1998 17:41:	250000	+	growth	1|3	25|25	growth
7	LENOUR	POP.GUI[2] -30769	Feb 05 1998 17:42:	$25 billion=qr	+	devaluation	1|5	30|30	devaluation
8	LENOUR	POP.GUI[2] -30846	Feb 05 1998 17:42:	1650=qr	+	increasing	1|5	35|35	increasing
9	LENOUR	POP.GUI[2] -30922	Feb 05 1998 17:42:	1934=nr	+	social	1|4	40|40	social
10	LENOUR	POP.GUI[2] -30998	Feb 05 1998 17:43:	1950=nr	+	political	1|5	45|45	political
11	LENOUR	POP.GUI[2] -31206	Feb 05 1998 17:43:	1983=qr	+	decrease	1|3	50|50	decrease
12	LENOUR	POP.GUI[2] -31282	Feb 05 1998 17:43:	1997=qr	+	increase	1|3	55|55	increase
13	LENOUR	POP.GUI[2] -31358	Feb 05 1998 17:44:	a	+	insecurity	1|5	60|60	insecurity
14	LENOUR	POP.GUI[2] -31434	Feb 05 1998 17:44:	a 15% decrease	+	morale	1|3	65|65	morale
15	LENOUR	POP.GUI[2] -31510	Feb 05 1998 17:45:	a repetition of hi	*-	un****ed	1|4	65|65	g
16	LENOUR	POP.GUI[2] -31586	Feb 05 1998 17:46:	pensions	+	pensions	1|3	70|70	pensions

47 Il est aussi possible de fabriquer un « filtre élaboré », en recopiant la première ligne de critères de choix dans une plage vide du logiciel, en écrivant sous les rubriques choisies des critères de sélection qui peuvent être bien plus complexes que ce qui a été mentionné ci-dessus, en désignant la zone de recherche (le tableau initial) et la zone de recopie des résultats (la zone sous les critères de recherche) et une copie partielle du tableau sera effectuée, sur laquelle on pourra encore pratiquer de nouvelles opérations de filtrage.

48 Ne voulant pas transformer cet article en une formation à l'utilisation d'un tableur, je ne présenterai pas ici la description des formules disponibles ; je me contenterai de préciser que de nombreuses fonctions statistiques, les unes simples, les autres plus délicates, sont utilisables pour comparer le nombre des enregistrements sélectionnés par une première requête au nombre d'enregistrements soit total, soit issu d'une précédente sélection. Les simples mesures de dispersions de résultats (calculs d'écarts types à partir d'un échantillon de données : ECARTYPE, ou de toutes les données : ECARTYPEP (où P = population), les tests de Fischer, Student, Pearson, les mesures de pente, courbes de régression etc. sont naturellement tous prévus.

49 Je ne donnerai qu'un exemple rapide pour montrer qu'avec un peu de pratique, les données peuvent assez facilement parler. Les temps de réponse, les attentes entre les différents actèmes sont intéressants à analyser car ils permettent d'évaluer les hésitations, les temps de recherche avant qu'un productème ne soit fourni, avant qu'une erreur ne soit rectifiée, etc. Les mesures obtenues pourront éventuellement différencier des données semblables (positives ou négatives) chez plusieurs apprenants et faire apparaître des tendances là où il ne semblait pas en exister. Elles peuvent être simplement calculées à partir des champs horodatés que j'ai prévus.

50 Chaque enregistrement contient une chaîne de type : mois jour année heure:minute:seconde (ex. : Jun 04 1998 10:09:26) dans laquelle la sous-chaîne commençant au 13ème caractère et se prolongeant sur les 8 caractères suivants correspond à l'heure (à la seconde près seulement, ici) de l'enregistrement. Il suffit d'isoler cette sous-chaîne et de soustraire son équivalent numérique à celui de la sous-chaîne correspondante contenue dans l'enregistrement précédent pour obtenir la valeur recherchée. Une fonction SOUS-TEXTE permettra cela : STXT(chaîne, départ, longueur). Comme j'ai pris la précaution de faire écrire les données nécessaires dans chaque enregistrement en respectant le format d'écriture d'un « champ date » (mm jj aa hh:mm:ss), il est possible de demander une simple soustraction des valeurs récupérées sans avoir besoin de convertir les caractères alphabétiques en caractères numériques. Si la première donnée (Jun 04 1998 09:37:19) est contenue dans la cellule A16 et la suivante (Jun 04 1998 09:37:55) dans la cellule A17, il suffira d'entrer dans une cellule résultat (ici C17) la fonction suivante : =STXT(A17;13;8)-–STXT(A16;13;8). Le résultat obtenu : 0,00041667, correspond à une traduction numérique de l'écart de temps constaté.

51 Pour le rendre plus lisible, on pourra forcer un format d'affichage plus clair par l'ajout d'une fonction spéciale, de syntaxe : TEXTE (chaîne numérique ; format d'affichage). La fonction initiale sera donc transformée comme suit : =TEXTE(STXT(A17;13;8)-STXT(A16;13;8);"hh:mm:ss"), ce qui fournira le résultat suivant : 00:00:36, soit 36 secondes.

52 Un des avantages propres au tableur est qu'il suffit de calculer un résultat une fois pour que, si une série se présente organisée de la même façon sur une grande plage du fichier, une simple recopie de cette fonction dans une série de cellules correspondantes (ici seulement C18 et C19 mais cela pourrait se faire sur plusieurs milliers de cellules) donne automatiquement toute la série calculée, la fonction recopie automatique mettant à jour les références de cellules automatiquement.

Figure 4

	A	B	C
…	…	…	…
16	Jun 04 1998 09:37:19		00:00:00
17	Jun 04 1998 09:37:55		00:00:36
18	Jun 04 1998 09:38:02		00:00:07
19	Jun 04 1998 09:38:52		00:00:50

53 Les fonctions conditionnelles (par exemple : SOMME.SI (plage;critère;somme_plage)) permettront, quant à elles, de traiter dans l'ensemble d'un tableau uniquement les lignes de données répondant aux critères de choix qu'on aura déterminés. Elles compléteront, sur le plan calcul, ce que le filtrage automatique permettait sur le plan de l'affichage.

54 Il faut tenir compte du fait que, si on souhaite qu'une fonction de calcul n'intervienne que sur la partie filtrée d'une plage de données, il faut la faire passer comme argument dans la fonction SOUS.TOTAL(n;plage), le commutateur *n* permettant de choisir le type de fonction de calcul désiré. Si la fonction NBVAL (= nombre de valeurs) utilisée seule, inclut toutes les valeurs d'une plage, affichées ou non, lorsqu'elle sera passée comme argument dans la fonction SOUS.TOTAL (*n* prenant alors la valeur 3), elle n'affectera que les valeurs filtrées à l'affichage. La formule suivante, insérée dans une cellule résultat, permettra donc, après l'application d'un filtrage automatique, de récupérer le pourcentage des lignes affichées par rapport à toutes les lignes d'une plage de données (nommée ici : noms) :

=((SOUS.TOTAL(3;noms))/NBVAL(noms))*100

55 Si on ne cherche pas à faire apparaître des tendances ou des profils particuliers par des calculs précis, ou si on préfère dans un premier temps à faire parler les fichiers traces plus visuellement qu'arithmétiquement, il peut-être préférable d'avoir recours à des concordanciers. Leur facilité d'utilisation et leurs performances les rendent de plus en plus attrayants. Citons, entre autres, CONC, un gratuiciel fonctionnant sous Mac, récupérable très facilement sur l'Internet avec n'importe quel moteur de recherche. MONOCONC, conçu par Michael Barlowe, Rice University[27] fonctionne sous Windows et est d'un usage particulièrement aisé, ses fonctions étant volontairement simplifiées par rapport à ce que proposent d'autres produits plus ou moins réservés aux spécialistes. Qu'on ne s'y trompe pas, il possède déjà des atouts qui en font un outil d'exploration et d'enseignement très utile. MICROCONCORD, désormais retiré du catalogue OUP, était un compromis intéressant sous DOS. Quant à l'OCP, qui a marqué son époque et qui fonctionnait lui aussi sous DOS, il est désormais remplacé par un outil remarquable et déjà mentionné supra, le plus performant parmi ceux qui sont actuellement disponibles, tout en restant d'un accès aisé : 'Wordsmith Tools' (OUP[28]).

56 Si, à l'opposé, les possibilités de calcul d'un tableur sont jugées insuffisantes, un outil comme SPSS (*Statistical Programme for Social Sciences*) sera probablement un investissement intéressant, tant pour les possibilités de calculs statistiques qu'il offre, que pour la qualité et la maniabilité de ses graphiques. Son prix est un peu élevé (4.000F environ, en monoposte), et il ne faut pas sous-estimer le temps nécessaire pour le maîtriser.

Conclusion et perspectives proches

57 Si les possibilités augmentent, si les modes d'analyse deviennent de plus en plus performants et les outils pour le faire plus accessibles, beaucoup reste à faire pour systématiser l'approche des productions apprenantes. Un des aspects qui n'est certainement pas encore suffisamment pris en compte, qui pose des problèmes pratiques et surtout de fond apparemment loin d'être

résolus, est celui du traitement systématisé, modélisé, des réponses orales apprenantes. Les progrès de la technique et de la phonologie n'ont pas encore défini un cadre suffisamment clair pour qu'une évaluation dans ce sens des productions orales soit prochainement réalisable. Les avancées de la reconnaissance vocale pourront-elles contribuer à faire avancer au moins partiellement la question ? S'il devient véritablement possible de transformer de façon fiable des productions orales en traces écrites, dont une partie au moins de la validité sur le plan de la correction sémantique et syntaxique peut déjà aujourd'hui être modélisée et donc évaluée, alors certains espoirs sont permis.

58 Encore faudra-t-il que la fastidieuse épreuve, actuellement nécessaire, de l'apprentissage de la voix de l'utilisateur par le logiciel de reconnaissance vocale soit fortement améliorée et raccourcie. Il restera encore à traiter la forme sonore elle-même, la qualité, la cohérence et la diversité des sons, de leurs enchaînements, des accents, etc. Les moteurs de reconnaissance vocale sont aujourd'hui essentiellement préoccupés par la restitution d'une langue écrite oralisée (dictée de lettre, de compte-rendu d'analyse médicale, d'expertise juridique, etc.) ; ils devront aussi apprendre à traiter la nature particulière de la langue orale, ses hésitations, la part importante du sens qui dépend de ses fluctuations, sur le plan intonatoire, dynamique ou rythmique, avec les différentes valeurs d'expressivité que tout cela peut entraîner, ceci pour ne mentionner que quelques modes d'approche possible de la question.

59 Il semble pourtant qu'un tournant soit désormais pris et qu'un vaste travail d'analyse de productions apprenantes sur ordinateur, au moins pour ce qui est des données récupérées au clavier et par pointage à la souris, soit prêt à se développer rapidement. La récente sortie d'un ouvrage annoncé par ses éditeurs comme le premier dans son genre, *Learner English on Computers*, (1998, édité par Sylviane Granger, également co-auteur) témoigne de l'intérêt grandissant de la communauté scientifique pour cet aspect de la recherche. Il n'est pas certain pourtant qu'un travail équivalent soit à ce jour aussi avancé en ce qui concerne l'analyse plus spécifique des 'parcours apprenants' en environnement multimédia. L'étude des profils apprenants, qui existe depuis de longues années et intéresse de nombreux chercheurs, devrait pourtant connaître une forte accélération dans ce sens avec le développement actuel des centres de langues informatisés.

60 Les besoins en gestion de suivi apprenant par l'encadrement pédagogique et administratif de tels centres vont vraisemblablement progresser de façon exponentielle. Même si leur importance stratégique peut paraître dépasser largement les besoins de la recherche fondamentale —si l'on en croit, notamment, les projets de réformes actuels concernant les modes de validation des compétences en langues au niveau du DEUG—, les enjeux sont communs et les intérêts étroitement interdépendants. Pour assurer l'accompagnement sous toutes les formes possibles des apprenants, en tutorat local ou à distance, en individuel ou en groupe, en mode présentiel ou en suivi différé, avec les différents niveaux de proximité de l'institution que cela peut sous-entendre, il faudra bien récupérer une quantité importante de données apprenantes qui devront être traitées de façon systématisée. Les outils actuellement disponibles ne le permettent pas, ou pas suffisamment, tant s'en faut, mais les concepteurs, développeurs, utilisateurs et éditeurs de produits comme LAVAC (C^3), Cyber-Learning (*Learning Labs*), Speaker (*Neuroconcept*), ou encore de Tell Me More (*Auralog*), Logolab (*Logosapience/Tecnilab*), Divace (*Lab Média)*, pour n'en citer que quelques uns, en sont tous conscients. On sent bien que, pour garantir une pédagogie de qualité tout en maintenant un niveau de rentabilité suffisant à tous points de vue, il faudra que des outils de qualité soient développés en conséquence. La recherche dans ce domaine, à l'intersection de la technique, de la théorie didactique et d'une pratique pédagogique spécifique de l'apprentissage des langues, doit être capable de relever ce défi.

Bibliographie

Barlow, M. 1987. *Working with Computers. Computer Orientation for Foreign Students*. Stanford : Athelstan Publications.

Bickerton, D. 1994. « Organising information for Multimedia Learning ». *ASp* 4, 75-92.

Butler, C. S. (dir.). 1992. *Computers and Written Texts*. Oxford : Blackwell.

Cazade, A. 1998. « Souplesse et contrainte du tout numérique » (dans un laboratoire de langues multimédia). In *Dossiers de l'ingénierie éducative sur les Laboratoires multimédias*. Paris : CNDP.

Cazade, A. 1999. « L'électronique au service de la lecture ? ». In Taillefer, Gail & A. K. Pugh (dir.), *Lecture à l'université, langue maternelle, seconde et étrangère - Reading in the University, first, second and foreign languages*. Toulouse : Presses de l'Université des Sciences sociales de Toulouse.

Cazade, A. 1999. « Pour intégrer des outils de référence et de recherche dans l'enseignement des langues : Wordsmith Tools (OUP) ». *Les Cahiers de l'APLIUT* 19/1, 97-122.

Clear, J. 1993. « From Firth principles: Computational tools for the study of collocation ». In M. Baker, G. Francis & E. Tognini-Bonelli (dir.), *Text and Technology: in honour of John Sinclair*. Philadelphie, PA : John Benjamins, 271-92.

Ellis, R. 1994. *The Study of Second Language Acquisition*. Oxford : Oxford University Press.

Fillmore, C. 1978. « On the organization of semantic information in the lexicon ». In Farkas, D., W. Jacobsen & K. Todrys (dir.), *Papers from the Parasession on the Lexicon*. Chicago : Chicago Linguistic Society, 148-73.

Granger, S. 1998. *Learner English on Computer*. Londres : Longman.

Hardisty, D. & S. Windeatt. 1989. *CALL. Resource Books for Teachers*. Oxford : Oxford University Press.

Higgins, J. 1988. *Language, Learners and Computers*. Harlow : Longman.

Higgins, J. 1995. *Computers and English Language Learning*. Oxford : Intellect Books and Norwood NJ : Ablex Publications.

Higgins, J. (dir.). 1986. « Computer assisted language learning; a European view ». *System* (special issue) 14/2, 147-219.

Higgins, J. 1991. « The computer and grammar teaching ». In Leech, G., *Computers in English Language Teaching and Research*. Londres : Longman, 31-45.

Johns, T. & P. King (dir.). 1991. « Classroom concordancing ». *ELR Journal* (special issue) 4.

Kenning, M.J. & M.-M. Kenning. 1990. *Computers and Language Learning; Current theory and practice*. Chichester : Ellis Horwood.

Levy, M. 1997. *Computer-assisted Language Learning. Context and Conceptualisation*. Oxford : Clarendon Press.

McCain, C. 1996. *Plugged in and Turned on*. Thousand Oaks, CA : Corwin Press.

Papert, S. 1980. *Mindstorms: Children, computers and powerful ideas*. Brighton : The Harvester Press.

Pennington, M. C. (dir.). 1989. *Teaching Languages with Computers. The State of the Art*. La Jolla, CA : Athelstan Publications.

Pennington, M. C. (dir.). 1996. *The Power of CALL*. Hong Kong : Athelstan Publications.

Roach, P. (dir.). 1992. *Computing in Linguistics and Phonetics. Introductory Readings*. Londres : Academic Press Ltd.

Scott, M. 1997. « PC analysis of key words – and key key words ». *System* 25/2, 233-45.

Scrimshaw, P. (dir.). 1993. *Language, Classrooms and Computers*. Londres : Routledge.

Sinclair, J. M. 1991. *Corpus, Concordance, Collocation.* Oxford : Oxford University Press.

Adresses Internet utiles

Pour la recherche textuelle informatisée, visitées durant l'été 1999, en espérant qu'elles resteront valables jusqu'à la lecture de ces lignes

CLAWS : http://www.comp.lancs.ac.uk/ucrel/claws/

Marqueurs TOSCA/LOB : ftp://lands.let.kun.nl/pub/tosca/tlbtag/

Marqueur : http://www.lsi.upc.es/~padro/angles/recerca.html

Marqueur ENGCG : renseignements à demander à l'adresse : engcg-info@ling.helsinki.fi

QTAG: http://www-clg.bham.ac.uk/QTAG/ ou: http://www-clg.bham.ac.uk/oliver/java/qtag/

Xerox: ftp://ftp.parc.Xerox.com/pub/tagger/

Analyseur morphologique : ftp://scott.cogsci.ed.ac.uk/pub/phonology/tools/MAP/

Wordsmith Tools : http://www.oup.co.uk/elt/software/wsmith/

Logiciels référencés par Summer Institute of Linguistics (SIL) : http://www.sil.org

TACT : ftp.epas.utoronto.ca

MONOCONC : ftp.nol.net ou : www.nol.net

XTAG Project : http://www.cis.upenn.edu/~xtag/

arbres sémant. & synt. : http://www.ims.uni-stuttgart.de/Tools/DecisionTreeTagger.html

Winbrill : http://jupiter.inalf.cnrs.fr/WinBrill/

Brill : http://www.cs.jhu.edu/~brill/code.html (version DOS : http://www2n.meshnet.or.jp/~htakashi/dos/)

Autasys : http://www.phon.ucl.ac.uk/home/alex/project/tagging/tagging.htm

outils statistiques : http://www.ling.gu.se/~nivre/kurser/wwwstat/toolsTP.html

---

Notes

1 Geoffrey Leech et Sylviane Granger en font eux-mêmes le constat (Granger 1998 : xiv, xvi, xxi). La constitution des premiers corpus apprenants remonte au début des années 90. Sylviane Granger en mentionne trois principaux : le *International Corpus of Learner English*, le *Longman Learner's Corpus* et le *Hong Kong University of Science and Technology Learner Corpus.* Les lignes qui suivent font, en fait, tout autant référence à l'analyse de la production effective des apprenants qu'à celle de leurs parcours, choix de priorités, progressions vers une production, demandes d'aides, etc.

2 'Toolbook' (*Asymetrix*) est un peu cher pourtant – environ 11 000F pour la récente version 7. 'Director' (*Macromedia*) est certainement le plus performant de tous, et reprend désormais les atouts de son concurrent un peu moins sophistiqué et donc un peu plus accessible au profane averti : 'Authorware' (*Macromedia*).

3 Parmi d'autres procédures permettant de récupérer des *self-report data*, selon la terminologie d'Ellis (1994), qui les différencie des *language use data* et des *metalingual judgements*.

4 On peut donner cette propriété au fichier sous l''Explorateur' mais on peut aussi prévoir cette option sous 'Word' (*Microsoft*): menu Fichier/Propriétés/Général. Voir aussi l'utilisation des sauvegardes avec mots de passe sous 'Word'.

5 Les adeptes de la programmation pourront optimiser la récupération de réponse par l'utilisation de contrôles 'ActiveX' et le recours au langage 'Visual Basic' intégré à toute la suite 'Microsoft Office'. Voir aussi le logiciel : 'Exoset' (société *TNT*, Villeneuve d'Ascq) qui cherche à utiliser plus complètement de telles ressources de programmation.

6 Le seul fait d'inscrire désormais l'adresse complète d'un site dans le corps du texte écrit avec 'Word', depuis la version livrée avec 'Office 97', permet à l'étudiant dont l'appareil est déjà connecté au réseau Internet, en se contentant de cliquer sur cette adresse d'accéder immédiatement à ce site.

7 L'insertion se fait par copier/coller. La programmation 'objet' utilisée donne automatiquement au nouveau bouton de saisie de réponse ainsi créé une identité propre qui me permet d'identifier la réponse dans le fichier trace où sera écrit le texte de l'apprenant.

8 Le recours à une boîte 'combo', en programmation, entraîne l'affichage d'une fenêtre de réponse en milieu d'écran, avec quelques consignes éventuellement.

9 L'exemple est tiré d'un module travaillant sur les conséquences du défrichage actuel de la forêt amazonienne en vue de la construction d'autoroutes. Certains virus seraient ainsi mis (ou remis, suivant certaines données d'archives) en contact avec l'homme. Le texte du questionnement cité ici a été rédigé par ma collègue Evelyn Perry (MCF à Dauphine).

10 Sndrec32.exe, dans le répertoire C :\Windows. Comme le montrait la figure 1, un accès direct peut être proposé à cet outil ou à un autre plus performant depuis un logiciel de bureautique, en créant un lien OLE : menu (commun à de nombreuses applications) Insertion/Objet/Ensemble OLE. Attention aux limitations en temps d'enregistrement et aux questions d'identification des fichiers enregistrés. OLE = *Object Linking and Embedding*.

11 Les cartes SoundBlaster (*Creative Labs*) sont encore aujourd'hui un standard synonyme d'une réelle qualité. Elles sont accompagnés d'outils d'enregistrement, voire de synthèse vocale, intéressants. Une SB PCI 64 ne coûte que 150 F.

12 Logolab (*Logosapience-Tecnilab*), Divace (*Lab Média*), etc. Tony Toma a également prévu des fonctionnalités allant dans ce sens avec le LAVAC (C^3).

13 Cette procédure est notamment offerte avec la 'carte SB Live', dont le diffuseur peut lui aussi être lancé automatiquement en tant qu'objet OLE.

14 Voir un article à paraître dans une publication du LAIRDIL (LAboratoire Interuniversitaire de Recherche en DIdactique des Langues, Toulouse Paul Sabatier) spécialement dédiée au multimédia (septembre 1999).

15 Commission Nationale Informatique et Liberté.

16 Voir références supra.

17 Accessible en téléchargement à l'adresse : http://www.oup.co.uk/elt/software/wsmith/

18 Le terme nous vient du *Atlas Computer Laboratory*, et de *Didcot.* Il désignait l'acronyme : Count and Concordance on Atlas. Ce marquage était également beaucoup utilisé avec l'OCP (Oxford Concordance Program).

19 Voir Fillmore (1978).

20 Voir notamment, parmi d'autres programmes dont les références sont données en annexe : ENGCG, à Helsinki, CLAWS à Birmingham, ou encore TOSCA à Nijmegen.

21 J'aime cette phrase de Bergson, dans *L'évolution créatrice*,qui n'est pas sans nous mettre sur la voie de telle théorie du chaos développée plus récemment : « Le désordre est un ordre qu'on n'attendait pas. »

22 On pardonnera, je l'espère, l'utilisation de termes manifestement jargoniques, qui ne sont jamais très satisfaisants pour l'esprit. Pourtant, il n'est peut-être pas inutile d'en employer quelques uns, de temps en temps, s'ils permettent d'exprimer plus rapidement une référence sémantique trop peu économique à reformuler continuellement.

23 Pour certaines analyses de réaction à un stimulus donné, il faut descendre au centième de seconde pour pouvoir faire apparaître des différences parlantes entre apprenants.

24 L'avantage de ces fichiers est qu'ils sont très légers, quelques kilo-octets à peine pour plusieurs minutes parfois. L'Internet regorge d'exemples de ces thèmes.

25 Ceux que j'ai choisis sont : +, pour une réponse correcte ; # pour une réponse intéressante mais insuffisante ; - pour une réponse refusée, •- ou •• pour une tentative de remplissage partiel insatisfaisant ou couronnée de succès (cf. le jeu du pendu) – chaque type de réponse ayant un commentaire individuel approprié.

26 Un nom de fichier est composé de deux parties, le nom lui-même qui peut dépasser aujourd'hui 8 caractères sur PC –*Windows* traitant désormais les adresses de la même manière que sous Mac–, et une extension derrière un point, ex. : NOMDEFICHIER.EXT.

27 Voir quelques références Internet fournies en annexe.

28 Voir un article publié fin 1999 par les *Cahiers de l'APLIUT*, dans lequel j'analyse les possibilités de cet outil et des références déjà cités *supra*.

Pour citer cet article

Référence électronique

Alain Cazade, « De l'expérimentation multimédia à l'analyse », *ASp* [En ligne], 23-26 | 1999, mis en ligne le 18 janvier 2012, consulté le 06 mai 2013. URL : http://asp.revues.org/2620 ; DOI : 10.4000/asp.2620

Référence papier

Alain Cazade, « De l'expérimentation multimédia à l'analyse », *ASp*, 23-26 | 1999, 441-462.

À propos de l'auteur

Alain Cazade
Alain Cazade est maître de conférences HDR d'anglais à l'université Paris 9 Dauphine. Il est co-responsable de l'Atelier Nouvelles Technologies pour la Recherche et l'Enseignement (ANTRE) au sein de la SAES (Société des Anglicistes de l'Enseignement Supérieur) et du Centre de Ressources en Langues (CRL) de Dauphine. Il a conçu et réalisé un logiciel hypermédia sous Windows, *Help Yourself* (partagiciel) pour soutenir son travail de recherche (et celui d'autres collègues qui désireraient faire de même) sur le rôle des outils multimédias dans l'enseignement/apprentissage des langues.
cazade@dauphine.fr

Droits d'auteur

Tous droits réservés

Résumés

La recherche sur les « parcours apprenants » à l'aide d'outils électroniques est encore jeune. Pourtant, avec les développements actuels de l'enseignement des langues appuyé sur le multimédia, avec l'évolution des outils et des matériels dans ce secteur, de plus en plus accessibles et/ou adaptables aux besoins tant sur le plan de la recherche que sur celui de la pédagogie, elle ne peut que se révéler très prometteuse à maints égards. Malgré ce que d'aucuns seraient tentés de croire, la difficulté la plus sérieuse n'est pas liée à la maîtrise de ces outils, mais bien plutôt au choix des données que l'on souhaite analyser, à la mise au point de protocoles d'expérimentation productifs et fiables et donc à la réflexion sur ce que peut être un parcours apprenant.

Research on learning procedures based on electronic tools is in its infancy. With the current evolution in multimedia CALL environments and the development of increasingly specialised programs designed to analyse learners' trace files, however, this kind of research seems more promising today than ever before. Despite expectations to the contrary, the most serious difficulty we face does not lie in mastering the software or hardware, but in choosing the data to be analysed, in finalising experimentation protocols capable of yielding rich and reliable results and, most of all, in understanding what the progression of a given learner consists of or might consist of.

Entrées d'index

Mots-clés : analyse de fichiers traces, CALL, multimédia, profil apprenant
Keywords : CALL, learning style, multimedia, trace files analysis